Counterfactual approach 2

Prof. Roberto Gabriele

Methods for Empirical Economics

Department of Economics and Management, University of Trento,
Ph.D. programme in Economics and Finance

Outline

1 Key concepts

Let's recall some key concepts

1. **The treatment effect** is defined as the effect of a specific treatment variable on an **outcome (or target) variable**, once any **potential confounders** affecting the link between the cause and the effect are ruled out
2. **The treatment variable** may be, according to the disciplinary context, a new drug, a new type of physiotherapy method, as well as, in the economic context, a training program for unemployed workers, a subsidy to firms' capital investment, etc.
3. the terms ''**Treatment**'' and "**causal factor**" are exchangeable, thus meaning that the researcher is not looking for a mere association among phenomena, but rather a **precise causal link**

UNIVERSITÀ DI TRENTO

Some reflections

1. This places econometrics within the sphere of **nonexperimental statistical designs**, where the analyst cannot manipulate the design of the experiment.
2. In contrast,**experimental** and **quasi-experimental designs** are characterized by a scientist's capacity to control the experiment.
3. In the classical experimental setting, the scientist deliberately produces a **random assignment of the units** involved in the experiment.
4. In contrast, in quasi-experimental designs, although the assignment is nonrandom, the scientist can manage the form of this nonrandomness at least to some acceptable extent.

More about Causality

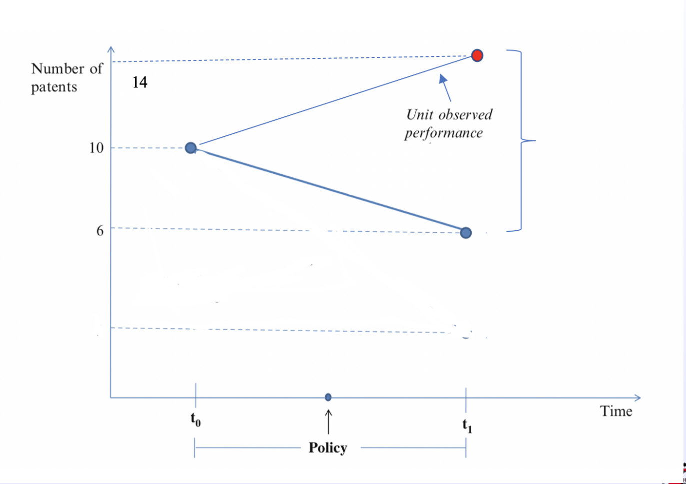
1. In experimental and quasi-experimental designs, the **treatment effect is generally estimated by the "counterfactual" approach**, so scientists in that field often refer to measuring "counterfactual causality" (Pearl 2000, 2009).
2. The concept of **counterfactual causality** draws upon the assumption that causality takes the form of a comparison between the outcome of a unit when this unit is treated in a certain way and the outcome of the same unit when it is not treated.
3. If one observes a unit only in its treated status, the untreated status is defined as the counterfactual status that is by definition not observable.

More about Causality II

1. Kant calls the 'law of causality' or the 'law of the connection of cause and effect'. It states that necessarily, in every event, there is something that is preceded and determined (according to a rule) by something else, i.e. that every event involves a cause
2. It presents the mind with puzzles. Hume's question, "Why a cause is always necessary", and the question of why the same cause should always have the same effect are examples of difficulties that have recurred throughout the history of thought
3. The debate

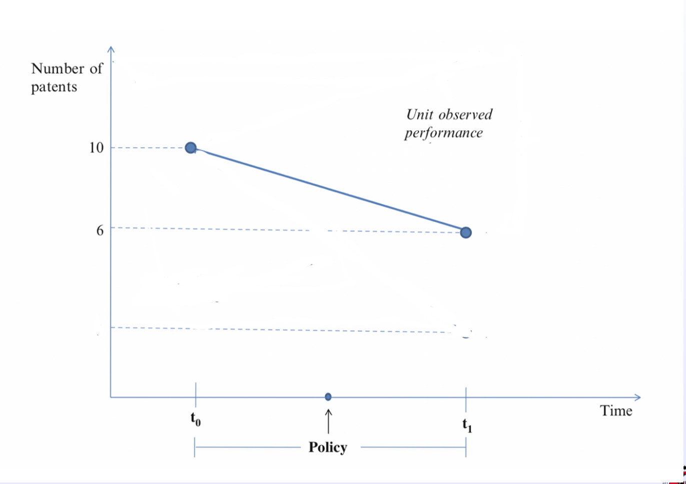
A representation of causal link

To better clarify the concept of counterfactual causality



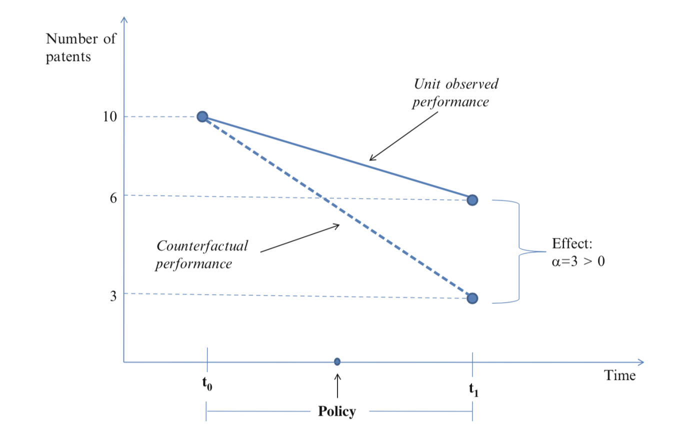
A representation of causal link

What we observe



A representation of causal link

Well: the policy has reduced the damage!



The policy maker problem of interpretation

The difference between monitoring and evaluating

1. If a public agency had only considered the **observed levels**, its conclusion about the usefulness of the policy would have been severely biased. This is the typical trap an agency can come across when considering just one side of the coin, the **observed side**
2. Only considering the observed performance is sometimes referred to as **policy monitoring**, whereas **policy evaluation** in the proper sense always needs to encompass the **counterfactual comparison**
3. ...in order to draw reliable conclusions about the **actual causal effect**.

Objective of the evaluation exercises

To **measure the "causal effects" of an intervention** on the part of an external authority (generally, a local or national government) on a set of subjects (people, companies, etc.) targeted by the program
Need of (see also later on...):

1. Data requirements
2. Institutional Setting requirements
3. Policy aim specification

Goals of program evaluation

Program evaluation may serve two related goals:

1. **Learning** aimed at providing improvements of various kinds for future policy programs, generally directed to managers and administrators,
2. **Legitimation** directed to higher political levels and to participants and other stakeholders involved in the program

UNIVERSITÀ DI TRENTO

Effective program evaluation

Preconditions for an econometric impact evaluation to be effective:

1. An appropriate evaluation design, based on the declared policy goals;
2. Detailed and well-documented data and information;
3. A broad and appropriate coverage of beneficiaries and non-beneficiaries;
4. A broad coverage of the spatial context when policies are geographically based

UNIVERSITÀ DI TRENTO

Where we can apply the "method" of causality

Some example to stimulate your interest:

- Is there an effect of a medical therapy?
- What is the effect of a R&D policy?
- what is the effect of subsidies to firms?
- What is the effect of a labour market program? (training...)
- what is the effect of social program (education to reduce discrimination or esclusion...)

Can you give me more examples? Other ideas?

More about causality in econometrics

1. Assume that causality assumes a linear form, where the analyst is interested in assessing the effect of a (usually) continuous variable (x) on a dependent variable (Y), by adding within the regression some control (or conditioning) factors
2. As we are embedded in a nonexperimental framework (a social experiment, as said above), at the heart of this causal framework there is the **exogeneity issue**:
   - the "true" causal effect of x on Y can be identified, as long as independent changes of x only produce a direct effect on Y, by ruling out any potential indirect effect of x on Y, via the relation of x with unobservable factors.
   - This is the condition under which x can be assumed to be exogenous; otherwise it is said to be "endogenously determined" and traditional estimation via ordinary least squares (OLS) produces biased estimates of the causal parameter.

UNIVERSITÀ DI TRENTO

Formally

1. $Y = \beta x + u$
2. β can be the causal effect of x on Y and u is an unobservable component
3. $\delta Y / \delta x = \beta$, only if $\delta u / \delta x = 0$

(necessary condition)

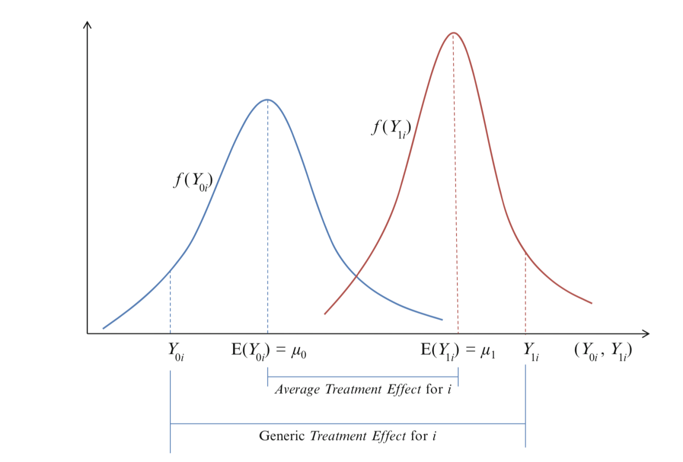
Statistical Setup

1. We are interested in estimating the so-called "treatment effect" of a policy program in a non-experimental setup, where a binary treatment variable D, taking value 1 for treated and 0 for untreated units
2. ...Is assumed to affect an outcome (or target) variable Y that can take a variety of forms: binary, count, continuous
3. We define the unit i treatment effect (TE) as: $TE_i = Y_{1i} - Y_{0i}$
4. is equal to the difference between the value of the target variable when the individual is treated (Y_1) and the value assumed by this variable when the same individual is untreated (Y_0)
5. **The analyst can observe just one of the two quantities, but not both!**

Potential Outcome Model

1. The analyst faces a fundamental **missing observation problem**...
2. that must be tackled econometrically in order to reliably **recover the causal effect**
3. What is observable to the analyst is the observable status of unit i:
4. $Y_i = Y_{1i}D_i - Y_{0i}(1 - D_i)$ (Potential Outcome Model)
5. D_i: dummy variable: 1 treated; 0 not treated
6. Estimating specific moments of distributions of Y_{0i}, Y_{1i} and in particular the mean, thus defining the so-called **population average treatment effect** (hereinafter ATE) of a policy intervention as:
7. $ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$

To fix ideas


$f(Y_{0i})$
$f(Y_{1i})$
Y_{0i}
$E(Y_{0i}) = \mu_0$
$E(Y_{1i}) = \mu_1$
Y_{1i}
(Y_{0i}, Y_{1i})
Average Treatment Effect for *i*
Generic *Treatment Effect* for *i*

ATEs definitions

1. By relying on the mean, however, the mainstream literature has emphasized two additional parameters as relevant to estimate. These are known as the average treatment effect on the treated (ATET) and the average treatment effect on the untreated (ATENT), defined respectively as:
2. $ATET = E[(Y_{1i} - Y_{0i})|D = 1]$
3. $ATENT = E[(Y_{1i} - Y_{0i})|D = 0]$
4. It holds that: $ATE = ATET \cdot p(D = 1) + ATENT \cdot p(D = 0)$

Law of iterated expectations

Example [edit]

Suppose that only two factories supply light bulbs to the market. Factory X's bulbs work for an average of 5000 hours, whereas factory Y's bulbs work for an average of 4000 hours. It is known that factory X supplies 60% of the total bulbs available. What is the expected length of time that a purchased bulb will work for?

Applying the law of total expectation, we have:

$$\begin{aligned} \mathrm{E}(L) &= \mathrm{E}(L \mid X)\,\mathrm{P}(X) + \mathrm{E}(L \mid Y)\,\mathrm{P}(Y) \\ &= 5000(0.6) + 4000(0.4) \\ &= 4600 \end{aligned}$$

where

- $\mathrm{E}(L)$ is the expected life of the bulb;
- $\mathrm{P}(X) = \frac{6}{10}$ is the probability that the purchased bulb was manufactured by factory X;
- $\mathrm{P}(Y) = \frac{4}{10}$ is the probability that the purchased bulb was manufactured by factory Y;
- $\mathrm{E}(L \mid X) = 5000$ is the expected lifetime of a bulb manufactured by X;
- $\mathrm{E}(L \mid Y) = 4000$ is the expected lifetime of a bulb manufactured by Y.

Thus each purchased light bulb has an expected lifetime of 4600 hours.

source: wikipedia

UNIVERSITÀ DI TRENTO

More on ATEs definitions

1. For each unit, beyond the values of Y and D, researchers (normally) have access also to a number of **observable covariates** which can be collected in a row vector $\mathbf{x}$:
2. $ATE(x) = E[Y_1 - Y_0|x]$
3. $ATET(x) = E[Y_1 - Y_0|D = 1, x]$
4. $ATENT(x) = E[Y_1 - Y_0|D = 0, x]$

Remember that: $ATE = E_x[ATE(x)]$

A simple example

1. **(Case 1)** first Suppose that, in evaluating a program through some econometric procedure, we find a value of ATET equal to 400 and a value of ATENT equal to 200
2. **Was this program successful?** The answer to case 1 seems to be positive: the group of treated individuals received, on average, a treatment effect of 400
3. With the knowledge of the value of ATENT things do not change: As the value of ATET is higher than that of ATENT, if the average untreated unit had been treated, then its outcome would have been lower by 200.
4. This is smaller than the increase in outcome obtained by the treated units when compared with their untreated status.

A simple example II

1. **(Case 2)** then suppose that in evaluating a program through some econometric procedure, we find a value of ATET equal to 100 and a value of ATENT equal to 200
2. Was this program successful? At first glance, the answer to case 2 seems to be positive: the group of treated individuals received, on average, a treatment effect of 100
3. Nevertheless, the knowledge of the value of ATENT might question this conclusion. As the value of ATENT is higher than that of ATET, if the average untreated unit had been treated, then its outcome would have been raised by 100.
4. This is higher than the increase in outcome obtained by the treated units when compared with their untreated status.
5. If the agency had been treating those who were not selected for treatment, the performance would had been better than in the opposite case. **In other words, one may conclude that the agency failed in selecting the right group to support, as they were not able to maximize the outcome.**

UNIVERSITÀ DI TRENTO

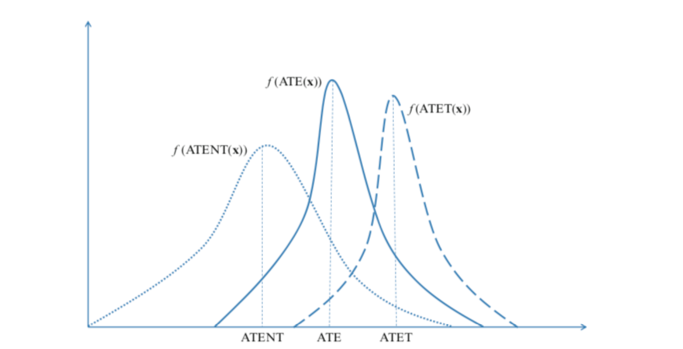
Although

Generally the agency is trying to maximize the outcome measure, in many cases this might not be the **prime objective** of an agency.

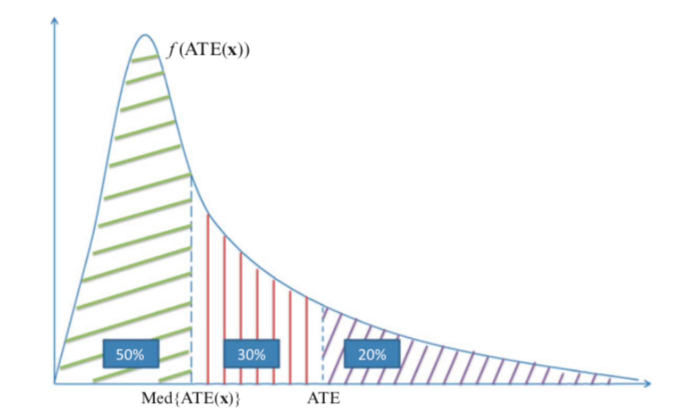
1. If welfare considerations are part of the policy's purposes, the agency might have been purposely aimed at **supporting lower performing units**
2. For instance, in a **microcredit program**, a public agency may find it consistent with its (social) objectives to support disadvantaged people living in depressed economic and social areas who are clearly in a position of weakness compared to those better off. It is not surprising that **these people will ultimately perform worse** than those better off

UNIVERSITÀ DI TRENTO

Distributions of ATEs



More on distribution of ATEs



The work of the econometrician

1. Recovering parameters from observational data
2. From an i.i.d. sample of observed variables for each individual i of this type:
3. $\{Y_i, D_i, x_i\}$
4. We exclude the possibility that the treatment of one unit affects the outcome of another unit (which is called SUTVA -or *stable unit treatment value assumption*)

UNIVERSITÀ DI TRENTO

Violations of SUTVA

1. In epidemiology, for instance, when treatments are vaccines for contagious diseases, it is quite intuitive that one unit treatment can influence the outcomes of others in their neighborhood
2. In economic applications such as the support to companies' research and development (R&D) activity, it might be hard to believe in the absence of "spillovers" from treated to untreated units activated by some form of subsidization

UNIVERSITÀ DI TRENTO

Identification under random assignment

The problem in estimating ATE (and thus ATET and ATENT) resides in the fact that for each observation we observe only one of the two states

1. If the sample was drawn at random (i.e., under random assignment), it would be possible to estimate the ATE as the difference between the sample mean of treated and the sample mean of untreated units
2. random assignment: $(Y_1;\ Y_0)$ are independent of D: $(Y_1; Y_0) \perp D$
3. independence assumption (IA):

$$E[Y|D=1]-E[Y|D=0]=E[Y_1|D=1]-E[Y_0|D=0]=$$
$$E[Y_1]-E[Y_0]=ATE=ATET=ATENT$$

DIM under randomization

Difference in Means

1. "Difference-in-means" (DIM) estimator
2. $\widehat{DIM} = \frac{1}{N_1} \sum_i Y_{i1} - \frac{1}{N_0} \sum_i Y_{i0}$
3. The knowledge of x is unnecessary for a correct estimation of this casual effect

Non-randomness

Policy programs hardly select individuals to treat (and, equivalently, to not treat) at random. This nonrandomness is inherent to a policy for two distinct reasons:

1. The self-selection into the program operated by individuals (cost and benefit of partecipating)
2. the selection mechanism of the agency managing the program (ex ante decision: picking the winner? aiding the poor?)

Thinking about non randomness

When the selection of treated and untreated units is done not randomly, depending on either individual "observable" or "unobservable" characteristics, the DIM estimator is no longer a correct estimation for ATE.

by using POM: $Y = Y_0 + D(Y_1 - Y_0)$ (baseline + causal effect)

1. $E[Y|D=1] - E[Y|D=0] = E[Y_1|D=1] - E[Y_0|D=0] + E[Y_0|D=1] - E[Y_0|D=1] =$
2. $= E[Y_0|D=1] - E[Y_0|D=0] + ATET*$
3. (2) contains *selection bias*

$*ATET = E[Y_1|D=1] - E[Y_0|D=1]$

Non-randomness II

In general $ATE \neq ATET \neq ATENT$

1. Define: $Y_1 = \mu_1 + U_1$ and $Y_0 = \mu_0 + U_0$
2. We can get: $Y_1 - Y_0 = \mu_1 - \mu_0 + U_1 - U_0 = ATE + U_1 - U_0$
3. By taking the expectation of this equation over $D = 1$
4. $E[Y_1 - Y_0|D = 1] = ATET = ATE + E[U_1 - U_0|D = 1]$*
5. $E[Y_1 - Y_0|D = 0] = ATENT = ATE + E[U_1 - U_0|D = 0]$
6. In general: $E[U_1 - U_0|D = 1] \neq E[U_1 - U_0|D = 0]$ so $ATET \neq ATENT$

*'participation gain" for those who actually participated in the program is present

Selection on observable

Observable and unobservable factors can affect the nonrandom assignment of beneficiaries:

In the case of Observables:

1. The analyst knows and can observe with precision which are the factors driving the self-selection of individuals and the selection of the agency
2. The knowledge of x, the structural variables that are supposed to drive the nonrandom assignment to treatment, are sufficient to identify the actual effect of the policy in question once adequately controlled for

UNIVERSITÀ DI TRENTO

Selection on observable II

In the case of Unobservables:

1. factors driving the nonrandom assignment are impossible or difficult to observe, then the only knowledge of the observable vector x is not sufficient to identify the effect of the policy

The nature of the unobservables:

1. Unobservable elements due to some lack of information in the available datasets (**contingent unobservables**)
2. **Genuine unobservables** would be fairly impossible to measure, even in the case of abundant information
3. Examples of this kind are represented by factors, such as entrepreneurial innate ability, propensity to bear risks, ethical attitudes...

Selection on observable III

Looking for an **additional assumptions** for estimating ATE, ATET, and ATENT under nonrandom selection

1. The analyst knows and can observe with precision which are the factors driving the self-selection of individuals and the selection of the agency
2. **Conditional independence assumption** (CIA), stating that conditional on the knowledge of x (sometimes called pretreatment covariates) Y_1 and Y_0 are probabilistically independent of D. Formally:
3. $(Y_1; Y_0) \perp D|x$
4. Once the knowledge of the factors affecting the sample selection is taken into account (or controlled for) by the analyst, then the condition of randomization is restored

Relaxing CIA

1. CIA assumption is too strong when we are interested, as we are, in average effects, so it is usual to rely on a weaker assumption, the so-called **conditional mean independence** (CMI):
2. $E[Y_1|D,x] = E[Y_1|x]$
3. $E[Y_0|D,x] = E[Y_0|x]$

under CMI:

1. $E[Y|D,x] = E[Y_0|D,x] + D(E[Y_1|D,x] - E[Y_0|D,x])$*$=$ $E[Y_0|x] + D(E[Y_1|x] - E[Y_0|x])$
2. Hence: if D=1: $E[Y_1|D,x] = E[Y_1|x]$
3. if D=0: $E[Y_0|D,x] = E[Y_0|x]$

* from POM: $Y_i = Y_{1i}D_i - Y_{0i}(1 - D_i)$

Relaxing CIA II

Subtracting the two expressions:

1. $E[Y|D=1,x] - E[Y|D=0,x] = E[Y_1|x] - E[Y_0|x] = ATE(x)$
2. $m_1(x) = E[Y_1|D=1,x]$, $m_0(x) = E[Y_0|D=0,x]$ are **observed quantities!**
3. $ATE(x) = m_1(x) - m_0(x) = m(x)$
4. $ATE = E_x[ATE(x)] = E_x[m(x)]$
5. Estimation of ATE can be obtained by the "sample equivalent":
6. $\widehat{ATE} = \frac{1}{N}\sum_i \hat{m}_i(x)$
7. The key is to find a consistent estimator of $m(x)$

UNIVERSITÀ DI TRENTO

Relaxing CIA III

Similarly for:
ATET:

1. $ATET = E_x[ATE(x)|D = 1]$
2. Looking at the sample equivalent: $\widehat{ATET} = \frac{1}{\sum_i D_i}[\sum_i D_i \hat{m}_i(x)]$

ATENT:

1. $ATENT = E_x[ATE(x)|D = 0]$
2. Looking at the sample equivalent: $\widehat{ATET} = \frac{1}{\sum_i (1-D_i)}[\sum_i (1 - D_i)\hat{m}_i(x)]$

UNIVERSITÀ DI TRENTO

Role of unbobservables

The CI (or CMI) assumption is not sufficient to identify program average effects when unobservable-to-analyst variables (either contingent or genuine) correlated with the potential outcomes, indeed:

1. $E[Y_1|D, x] \neq E[Y_1|x]$
2. $E[Y_0|D, x] \neq E[Y_0|x]$

Role of unbobservables II

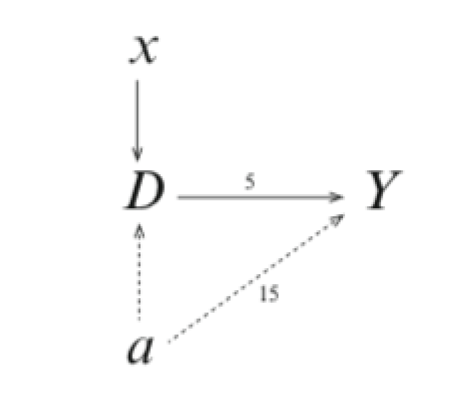
1. $E(Y|x, D = 1) - E(Y|x, D = 0) = E(Y_1|x, D = 1) - E(Y_0|x, D = 0) + E(Y_0|x, D = 1) - E(Y_0|x, D = 1) =$
2. $= E(Y_0|x, D = 1) - E(Y_0|x, D = 0) + ATET(x)$

...in this case the problem persists even if we control for x!
To correctly identify the direct effect of D on Y, more structural information needs to be added

UNIVERSITÀ DI TRENTO

A graphical representation

A path diagram:



The overlap assumption

Either in the case of selection on observables or selection on unobservables, the identification of ATEs requires a second fundamental assumption besides CMI, i.e., the so-called **overlap assumption**

1. Lets define a key notion of the econometrics of program evaluation, the propensity-score, defined as the "probability to get treated, given the knowledge of x":
2. $prob(D = 1, x)$
3. Overlap assumption states that: $0 < prob(D_i = 1, x_i) < 1$
4. For instance, if for $x = x_0$ the propensity-score assumes zero value, it means that there are no units in the treated group having that specific value of x, and this entails that ATEs cannot be calculated (i.e., identified).

An example

We have just two units with $x = 0$ (unit 1 and 2), both in the untreated group ($D = 0$), and no units in the treated group having such a value of x. In a situation like this, $p(D = 1 \mid x = 0) = 0$ and ATE cannot be identified:

Table 1.1 An example of unfeasible identification of ATE when the overlap assumption fails

	Treatment (D)	Covariate (x)	Outcome (Y)
1	0	0	5
2	0	0	8
3	0	1	6
4	0	1	4
5	1	1	10
6	1	1	20
7	1	1	80
8	1	1	70

UNIVERSITÀ DI TRENTO

An example II

Indeed:

$$\begin{aligned} \text{ATE} &= \text{E}_{\mathbf{x}}\{\text{ATE}(\mathbf{x})\} \\ &= p(x=1)\cdot\text{ATE}(x=1) + p(x=0)\cdot\text{ATE}(x=0) \end{aligned} \quad (1.44)$$

where according to Table 1.1, $p(x=1)=6/8$ and $p(x=0)=2/8$. Nevertheless, while when $x=1$ ATE can be identified (both treated and untreated present in fact this kind of attribute) so that:

$$\text{ATE}(x=1) = [(10+20+80+70)/4] - [(4+6)/2] = 45 - 5 = 40 \quad (1.45)$$

the same cannot be done for $\text{ATE}(x=0)$, as:

$$\text{ATE}(x=0) = [?] - [(5+8)/2] = ? \;=>\; \text{ATE} = ? \quad (1.46)$$

In order to identify all ATEs, each cell built by crossing the values taken by the various x-provided that they have finite discrete support- must have both treated and untreated units.