M1 INTERMEDIATE ECONOMETRICS

CAUSALITY

Koen Jochmans

September 29, 2025

1. Potential outcomes

We may be interested in the causal effect of a treatment on an outcome of interest. To formalize this let $D \in \{0, 1\}$ be a binary indicator that captures whether treatment is assigned or not. The potential outcomes $Y(0)$ and $Y(1)$ are a pair of random variables that indicate the outcome of interest when the treatment was assigned or not. The causal effect of the treatment then is the difference

$$Y(1) - Y(0).$$

The econometric problem is that we do not observe both $Y(0)$ and $Y(1)$. We only observe

$$Y = Y(1)\, D + Y(0)\, (1 - D)$$

along with D. Moreover, when a unit is treated $D = 1$ and we observe $Y = Y(1)$. When, instead, a unit is untreated we observe $D = 0$ and $Y = Y(0)$.

An example of a causal parameter of interest is the average treatment effect on the treated,

$$\mathbb{E}(Y(1) - Y(0)|D = 1).$$

A naive comparison of conditional means, as in

$$\mathbb{E}(Y|D = 1) - \mathbb{E}(Y|D = 0)$$

does not usually yield this parameter. Indeed,

$$\begin{aligned}\mathbb{E}(Y|D=1) - \mathbb{E}(Y|D=0) &= \mathbb{E}(Y(1)|D=1) - \mathbb{E}(Y(0)|D=0) \\ &= \mathbb{E}(Y(1)|D=1) - \mathbb{E}(Y(0)|D=1) \\ &+ \mathbb{E}(Y(0)|D=1) - \mathbb{E}(Y(0)|D=0),\end{aligned}$$

and, in general,

$$\mathbb{E}(Y(0)|D=1) \neq \mathbb{E}(Y(0)|D=0).$$

To make progress on the identification of causal effects we will have to make assumptions about the stochastic relationship between the treatment and the potential outcomes.

2. Random assignment

The simplest case is when treatment is randomly assigned. Then

$$Y(0), Y(1) \perp\!\!\!\perp D,$$

potential outcomes are independent of treatment assignment.

To see how this is helpful suppose that we wish to learn the average treatment effect

$$\mathbb{E}(Y(1) - Y(0))$$

which, by independence, is equally equal to the average treatment effect on the treated here. We have

$$\begin{aligned}\mathbb{E}(Y|D=1) &= \mathbb{E}(Y(1)|D=1) = \mathbb{E}(Y(1)), \\ \mathbb{E}(Y|D=0) &= \mathbb{E}(Y(0)|D=0) = \mathbb{E}(Y(0)),\end{aligned}$$

and so the simple difference in means

$$\mathbb{E}(Y|D=1) - \mathbb{E}(Y|D=0)$$

identifies the average treatment effect.

To connect this to regression notice that we can write

$$Y = Y(0) + (Y(1) - Y(0))\, D.$$

Therefore, by independence,

$$\mathbb{E}(Y|D) = \mathbb{E}(Y(0)) + \mathbb{E}((Y(1) - Y(0)))\, D$$

and the slope coefficient in a simple regression of Y on D and a constant yields the average treatment effect.

Under the weaker assumption that only $Y(0) \perp\!\!\!\perp D$ we can no longer learn the average treatment effect but, as

$$\mathbb{E}(Y|D=1) = \mathbb{E}(Y(0)) + \mathbb{E}(Y(1) - Y(0)|D=1)$$

and $\mathbb{E}(Y|D=0) = \mathbb{E}(Y(0))$ we still obtain

$$\mathbb{E}(Y(1) - Y(0)|D=1) = \mathbb{E}(Y|D=1) - \mathbb{E}(Y|D=0)$$

to recover the average treatment effect on the treated by simple regression.

2.1. Selection on observables

Randomization is the experimental ideal for the recovery of treatment effects. This paradigm is suitable for well-designed lab experiments but less so for

observational data. There, it is possible for units to self-select into treatment.

To see why this can be a problem a Roy model makes for a natural example. Suppose that

$$D = \begin{cases} 1 & \text{if } Y(1) - Y(0) > 0 \\ 0 & \text{if } Y(1) - Y(0) \leq 0 \end{cases},$$

so that a unit selects into treatment when it pays off for him to do so. Then, clearly,

$$\mathbb{E}(Y|D=1) = \mathbb{E}(Y(1)|Y(1) > Y(0)) \neq \mathbb{E}(Y(1)),$$

so that the key independence assumption fails.

2.2. Conditional independence

A weaker condition than full independence is a conditional-independence requirement given a set of control variables, X. These variables are potential confounding factors. Moreover,

$$Y(0), Y(1) \perp\!\!\!\perp D \,|\, X.$$

In this case self-selection into treatment is allowed for, provided that it is based on the observable variables in X.

Then, in complete analogy to before,

$$\mathbb{E}(Y|D=1,X) = \mathbb{E}(Y(1)|D=1,X) = \mathbb{E}(Y(1)|X),$$
$$\mathbb{E}(Y|D=0,X) = \mathbb{E}(Y(0)|D=0,X) = \mathbb{E}(Y(0)|X).$$

So,

$$\mathbb{E}(Y(1) - Y(0)|X) = \mathbb{E}(Y|D=1,X) - \mathbb{E}(Y|D=0,X)$$

identifies the average treatment effect conditional on a given X. The marginal treatment effect, that is, unconditional on the control variables, is obtained by averaging out the control variables.

In a full-saturated design the conditional means can be estimated by linear regression techniques. Otherwise such regressions count as approximations. More flexible approaches are available.

2.3. Propensity score matching

The propensity score is

$$p(X) = \mathbb{P}(D = 1|X).$$

Observe that, if $Y(1), Y(0) \perp\!\!\!\perp D \,|\, X$, then also

$$Y(1), Y(0) \perp\!\!\!\perp D \,|\, p(X),$$

so that it suffices to compare treated and untreated based on the propensity score alone.

This results follows from the observation that

$$\begin{aligned}
\mathbb{P}(D = 1|Y(1), Y(0), p(X)) &= \mathbb{E}(\mathbb{E}(D|Y(1), Y(0), X, p(X))|Y(1), Y(0), p(X)) \\
&= \mathbb{E}(\mathbb{E}(D|Y(1), Y(0), X)|Y(1), Y(0), p(X)) \\
&= \mathbb{E}(\mathbb{E}(D|X)|Y(1), Y(0), p(X)) \\
&= \mathbb{E}(p(X)|Y(1), Y(0), p(X)) \\
&= p(X) \\
&= \mathbb{P}(D = 1|X),
\end{aligned}$$

so that indeed D is independent of $Y(1), Y(0)$ conditional on X.

A regression-based approach to estimation can then be based on a flexible estimator of $\mathbb{E}(Y|D, p(X))$, where the propensity score needs to be estimated in a first stage.

This characterisation suggests an alternative representation of the average treatment effect of interest. As

$$\mathbb{E}(YD|X) = \mathbb{E}(Y(1)|X)\,\mathbb{P}(D = 1|X),$$
$$\mathbb{E}(Y(1-D)|X) = \mathbb{E}(Y(0)|X)\,\mathbb{P}(D = 0|X),$$

we have

$$\mathbb{E}\left(\frac{YD}{p(X)} - \frac{Y(1-D)}{1-p(X)}\right) = \mathbb{E}(\mathbb{E}(Y(1)|X) - \mathbb{E}(Y(0)|X)) = \mathbb{E}(Y(1) - Y(0))$$

provided that the overlap condition $0 < p(X) < 1$ is satisfied for all values X.

One popular implementation via this route is to use a flexible nonlinear regression fit (such as logit or probit) for the propensity score, followed by a straightforward calculation of the weighted means for treated and control group.

3. Instrumental variables

The above approach fails when confounding arises from unobserved variables. In such cases instrumental variables may help. The interpretation of the resulting estimand is most clear under homogenous effects, i.e., when the difference $Y(1) - Y(0)$ is not random.

A simple example is non-compliance in a randomized control trail. Here, $Z \in \{0, 1\}$ represents treatment assignment, which is randomly assigned, but treated units can decide not to take up treatment and non-treated units can

pick up treatment. If the latter is ruled out we refer to the non-compliance as being one-sided. If the decision to deviate from assignment is related to the potential outcomes, $Y(0), Y(1) \perp\!\!\!\perp D$, will not hold.

3.1. Constant treatment effects

In the homogenous treatment effect setting, $Y(1) - Y(0) = \theta$. Therefore,

$$Y = Y(0) + (Y(1) - Y(0))\, D = Y(0) + \theta\, D.$$

Furthermore,

$$\mathbb{E}(Y|Z=1) = \mathbb{E}(Y(0)|Z=1) + \theta\, \mathbb{E}(D|Z=1) = \mathbb{E}(Y(0)) + \theta\, \mathbb{E}(D|Z=1)$$
$$\mathbb{E}(Y|Z=0) = \mathbb{E}(Y(0)|Z=0) + \theta\, \mathbb{E}(D|Z=0) = \mathbb{E}(Y(0)) + \theta\, \mathbb{E}(D|Z=0)$$

and so

$$\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0) = \theta\, (\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0)).$$

Therefore,

$$\theta = \frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0)}{\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0)} = \frac{\text{cov}(Y,Z)}{\text{cov}(D,Z)}$$

recovers the average treatment effect. This presumes that $\mathbb{E}(D|Z=1) \neq \mathbb{E}(D|Z=0)$. This requirement means that D and Z cannot be independent, so that there is some information on D contained in Z. For example, when non-compliance is one sided we know that $\mathbb{E}(D|Z=0) = 0$ and so as long as some units actually take up their assigned treatment, the relevance condition will be satisfied.

3.2. Heterogenous treatment effects

When the effect of the treatment is heterogeneous the interpretation of instrumental-variable estimands is more delicate. In the case of no control variables and both binary D and Z an instructive and clean interpretation is nonetheless available.

We consider two cases. The first is based on the eligibility rule that

$$\mathbb{P}(D = 1|Z = 0) = 0,$$

that is, non-compliance is one-sided. In this case,

$$\begin{aligned}\mathbb{E}(Y|Z = 1) &= \mathbb{E}(Y(0) + (Y(1) - Y(0))\, D|Z = 1) \\ &= \mathbb{E}(Y(0)|Z = 1) \\ &+ \mathbb{E}(Y(1) - Y(0)|D = 1, Z = 1)\, \mathbb{P}(D = 1|Z = 1) \\ &= \mathbb{E}(Y(0)) \\ &+ \mathbb{E}(Y(1) - Y(0)|D = 1)\, \mathbb{P}(D = 1|Z = 1)\end{aligned}$$

while

$$\mathbb{E}(Y|Z = 0) = \mathbb{E}(Y(0)) + \mathbb{E}(Y(1) - Y(0)|D = 1)\, \mathbb{P}(D = 1|Z = 0) = \mathbb{E}(Y(0)).$$

Therefore,

$$\mathbb{E}(Y|Z = 1) - \mathbb{E}(Y|Z = 0) = \mathbb{E}(Y(1) - Y(0)|D = 1)\, \mathbb{P}(D = 1|Z = 1),$$

and so, by re-arrangement

$$\mathbb{E}(Y(1) - Y(0)|D = 1) = \frac{\mathbb{E}(Y|Z = 1) - \mathbb{E}(Y|Z = 0)}{\mathbb{E}(D|Z = 1)}$$

recovers the average treatment effect on the treated.

The case with two-sided non-compliance is more delicate and requires an alternative restriction. To state it we define potential treatments $D(0)$ and $D(1)$ as the treatment as a function of the instrument. Then we observe actual treatment

$$D = D(1)\,Z + D(0)\,(1 - Z).$$

The potential treatment variables are useful as they define four types of units. Units with $D(0) = 0$ and $D(1) = 0$ or with $D(0) = 1$ and $D(1) = 1$ do not respond to variation in the instrument. Instead, they never/always participate. We call such units never takers and always takers, respectively. Units with $D(0) = 0$ and $D(1) = 1$ react to a shift in the instrument by participating when nudged to do so, and by not participation when not. As such, they are compliers. The final group show the opposite pattern, that is, they have $D(0) = 1$ and $D(1) = 0$. They participate when incentivized not to but decide to participate when not. As such, they defy the setup of the program.

We have

$$\begin{aligned}
\mathbb{E}(Y|Z=1) = {} & \mathbb{E}(Y|Z=1, D(0)=0, D(1)=0)\,\mathbb{P}(D(0)=0, D(1)=0|Z=1) \\
& + \mathbb{E}(Y|Z=1, D(0)=0, D(1)=1)\,\mathbb{P}(D(0)=0, D(1)=1|Z=1) \\
& + \mathbb{E}(Y|Z=1, D(0)=1, D(1)=0)\,\mathbb{P}(D(0)=1, D(1)=0|Z=1) \\
& + \mathbb{E}(Y|Z=1, D(0)=1, D(1)=1)\,\mathbb{P}(D(0)=1, D(1)=1|Z=1)
\end{aligned}$$

and, by random assignment of the instrument, this becomes

$$\begin{aligned}\mathbb{E}(Y|Z=1) = &\mathbb{E}(Y(0)|D(0)=0, D(1)=0)\,\mathbb{P}(D(0)=0, D(1)=0)\\ &+\mathbb{E}(Y(1)|D(0)=0, D(1)=1)\,\mathbb{P}(D(0)=0, D(1)=1)\\ &+\mathbb{E}(Y(0)|D(0)=1, D(1)=0)\,\mathbb{P}(D(0)=1, D(1)=0)\\ &+\mathbb{E}(Y(1)|D(0)=1, D(1)=1)\,\mathbb{P}(D(0)=1, D(1)=1).\end{aligned}$$

In the same way,

$$\begin{aligned}\mathbb{E}(Y|Z=0) = &\mathbb{E}(Y(0)|D(0)=0, D(1)=0)\,\mathbb{P}(D(0)=0, D(1)=0)\\ &+\mathbb{E}(Y(0)|D(0)=0, D(1)=1)\,\mathbb{P}(D(0)=0, D(1)=1)\\ &+\mathbb{E}(Y(1)|D(0)=1, D(1)=0)\,\mathbb{P}(D(0)=1, D(1)=0)\\ &+\mathbb{E}(Y(1)|D(0)=1, D(1)=1)\,\mathbb{P}(D(0)=1, D(1)=1).\end{aligned}$$

Therefore the difference $\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0)$ is equal to

$$\begin{aligned}&\mathbb{E}(Y(1)-Y(0)|D(0)=0, D(1)=1)\,\mathbb{P}(D(0)=0, D(1)=1)\\ +&\mathbb{E}(Y(0)-Y(1)|D(0)=1, D(1)=0)\,\mathbb{P}(D(0)=1, D(1)=0)\end{aligned}$$

and features only compliers and defiers. This is intuitive, as never-takers and always-takers do not react to a change in the instrument.

To make progress, the identifying condition we use as an alternative to one-sided compliance is the monotonicity condition

$$D(0) \leq D(1).$$

This rules out the existence of defiers in that

$$\mathbb{P}(D(0)=1, D(1)=0) = 0.$$

Then the instrumental-variable estimand is

$$\frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0)}{\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0)} = \mathbb{E}(Y(1) - Y(0)|D(0)=0, D(1)=1),$$

which equals the average treatment effect for the subpopulation of compliers.