Econometrics 1
Lecture 11: Treatment effects and instrumental variables II

ENSAE 2014/2015

Michael Visser (CREST-ENSAE)

Evaluation of program effects

Today we discuss program evaluation methods.

Examples of programs: university grants for well-performing students, social programs for low-income families, class-size reductions in poor urban zones, job training programs for low-skilled workers, new medical drugs, etc...

It is important to have adequate econometric tools to correctly evaluate such programs.

Based on econometric evaluations, policy makers can then decide to implement the program or not.

We present the evaluation methods within Rubin's counter-factual framework (see Lectures 1 and 2).

Let us first recall the notations.

The binary variable D_i is the treatment indicator, equal to 1 if individual i participates in the program, and 0 otherwise.

Notations

Y_{i1} is the outcome with treatment (salary of person i who followed the job training program, exam result of a child i in a small class, etc.), and Y_{i0} the outcome without treatment (salary without job training program, exam result in the large class).

Since a person cannot be in both states, we cannot simultaneously observe Y_{i0} and Y_{i1}. That is why these two variables are called *potential outcomes.* Although they are not both observed, one can imagine that they both exist.

Finally, the observed outcome is denoted Y_i ($Y_i = Y_{i0}$ if $D_i = 0$, and $Y_i = Y_{i1}$ if $D_i = 1$):

$$Y_i = Y_{i0} + D_i(Y_{i1} - Y_{i0}). \tag{1}$$

The effect of the program for i, or treatment effect, is $Y_{i1} - Y_{i0}$, i.e., the difference of the outcome with and without treatment.

ATE and ATT

Throughout we allow for treatment heterogeneity, i.e., the treatment effect is assumed to be individual-specific: $Y_{i1} - Y_{i0}$ may be different for each i.

Since the two potential outcomes are not both observed for a given individual, we cannot calculate the individual-level treatment effect.

Given a sample of data, (Y_i, D_i), $i = 1, ..., n$, we can only hope to be able to estimate certain average treatment effects.

One quantity of interest if the *average treatment effect* (ATE):

$$E[Y_1 - Y_0] \tag{2}$$

ATE is the expected effect of a treatment on a randomly drawn individual from some population.

Another quantity of interest is the *average treatment effect on the treated (ATT):*

$$E[Y_1 - Y_0 | D = 1] \tag{3}$$

ATT is the mean treatment effect for those who actually participated in the program.

ATE and ATT: Experimental data

How can we estimate (2) or (3) using the data?

Suppose first that D is statistically independent of (Y_0, Y_1), as would be the case in a controlled experiment where treatment assignment is random.

Estimation of ATE is then simple. Using (1) we have

$$E[Y|D=1] = E[Y_1|D=1] = E[Y_1]$$

Similarly:

$$E[Y|D=0] = E[Y_0|D=0] = E[Y_0]$$

So the ATE is

$$ATE = E[Y|D=1] - E[Y|D=0]$$

The right-hand side can be consistently estimated by a difference in sample means (the average of the outcome variable among the treated minus the average among the non-treated).

ATE and ATT: Experimental data (cntd)

Note that under independence between D and (Y_0, Y_1), ATT and ATE coincide:

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

Note also that ATT can also be estimated consistently as a difference in appropriate means under the weaker assumption that D is independent of Y_0 (and without any restriction on the relationship between Y_1 and D).

Indeed we have:

$$\begin{aligned} &E[Y|D = 1] - E[Y|D = 0] \\ &\quad = E[Y_0|D = 1] - E[Y_0|D = 0] + E[Y_1 - Y_0|D = 1] = ATT \end{aligned}$$

Unfortunately, experiments are rare in econometrics. In the large majority of data, treatment is not randomy assigned. Instead, assignment is typically decided by the program organizers, or the participants of the program themselves.

ATE and ATT: observational data

In a job training program, for instance, individuals themselves typically decide whether they receive treatment (get the training) or not, and their decision may be related to the wage gain $Y_1 - Y_0$ (wage after training minus wage without training).

Similarly, in class-size reduction programs, the schoolheads often decide which children to put in large classes, and which children in the smaller ones. Again their decisions may be based on the gain in test results $Y_1 - Y_0$ (highly performing and mature pupils are typically assigned to large classes, the others to small classes).

With non-experimental data, treatment assignment is therefore typically non-random, and it is then likely that the treatment indicator is not independent of the potential outcomes.

ATE and ATT: observational data (cntd)

We then have

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

With non-experimental data the comparison between the mean outcome of the treated and the mean outcome of the non-treated does not identify the ATT.

Instead, what can be identified (and estimated) is the ATT plus a selection effect which reflects the fact that in observational data individuals are not randomly assigned into treatment groups.

ATE and ATT: Ignorability of treatment

The question arises how we can identify treatments effects with non-experimental (observational) data.

Fortunately we can estimate treatment effects under assumptions that are weaker than the independence assumption (between D and (Y_0, Y_1)).

One method consists in assuming that the outcome variables are independent of the treatment indicator conditional on certain other variables.

In the job training example, for instance, it may be that the treatment decision is partly determined by the education level and gender of individuals (women and more educated persons are more likely to particpate in the program), but that, conditional on education and gender, treatment assignment is random.

Such an assumption is called *ignorability of treatment.*

ATE and ATT: Ignorability of treatment (cntd)

Supose that we also observe for each individual a vector of characteristics X. So the sample of data we have is of the form (Y_i, D_i, X_i), $i = 1, ..., n$.

Suppose also that the the following *ignorability of treatment* assumption holds:

$$(Y_0, Y_1) \perp D \mid X \tag{4}$$

Condition (4) states that, conditionally on X, (Y_0, Y_1) and D are independent.

Under (4), conditionally on X, the average treatment effect and the average treatment effect on the treated coincide:

$$ATT(X) \equiv E[Y_1 - Y_0 | X, D = 1] = E[Y_1 - Y_0 | X] \equiv ATE(X)$$

However, the unconditional versions of these treatment effects are generally not identical.

ATE and ATT: Ignorability of treatment (cntd)

Indeed, denoting $r(X) = E[Y_1 - Y_0|X]$, we have:

$$ATE = E[r(X)]; \; ATT = E[r(X)|D = 1]$$

where the first expectation is with respect to X, and the second with respect to X given $D = 1$.

To obtain estimators of $ATE(X)$ and $ATT(X)$ (and the unconditional versions) we proceed as follows. Using (1) and the ignorability assumption (4) we get

$$\begin{aligned} E[Y|X,D] &= E[Y_0|X,D] + D(E[Y_1|X,D] - E[Y_0|X,D]) \\ &= E[Y_0|X] + D(E[Y_1|X] - E[Y_0|X]) \end{aligned}$$

Therefore

$$E[Y|X,D=1] - E[Y|X,D=0] = E[Y_1 - Y_0|X] = ATE(X) = ATT(X)$$

ATE and ATT: Ignorability of treatment (cntd)

$ATE(X)$ and $ATT(X)$ can be estimated using non-parametric estimations of $E[Y|X, D = 1]$ and $E[Y|X, D = 0]$. In a second step one can then obtain non-parametric estimates of ATE and ATT.

This non-parametric estimation approach (i.e., without imposing any parametric structure on the relationship between Y on the one hand, and X and D on the other) is beyond the scope of this course (see, Wooldridge, Econometric Analysis of Cross Section and Panel Data, chapter 18).

Instead we turn to estimation of the treatment effects within the familiar regression framework.

We need to put more structure on the relationship between Y, X, and D.

ATE and ATT: Ignorability of treatment (cntd)

Suppose that

$$
\begin{aligned}
Y_0 &= E[Y_0] + \nu_0 = \mu_0 + \nu_0 \\
Y_1 &= E[Y_1] + \nu_1 = \mu_1 + \nu_1
\end{aligned}
$$

where μ_0 and μ_1 are scalar parameters and ν_0 and ν_1 are error terms with expectation equal to zero: $E[\nu_0] = E[\nu_1] = 0$.

Note that $ATE = E[Y_1 - Y_0] = \mu_1 - \mu_0$.
Inserting the specifications of the potential outcomes in (1) gives

$$
Y = \mu_0 + (\mu_1 - \mu_0)D + \nu_0 + D(\nu_1 - \nu_0) \tag{5}
$$

Assume finally that

$$
E[\nu_0|X] = g_0(X); \;\; E[\nu_1|X] = g_1(X) \tag{6}
$$

where g_0 and g_1 are scalar functions of X.

ATE and ATT: Ignorability of treatment (cntd)

Taking the expectation of equation (5) given (X, D), and using (4) and (6), gives:

$$E[Y|X, D] = \mu_0 + (\mu_1 - \mu_0)D + g_0(X) + D(g_1(X) - g_0(X))$$

so we get the regression model

$$Y_i = \mu_0 + (\mu_1 - \mu_0)D_i + g_0(X_i) + D_i(g_1(X_i) - g_0(X_i)) + u_i, \; i = 1, ..., n \tag{7}$$

where $E[u_i|X_i, D_i] = 0$ by construction.

Choosing parametric functions for g_0 and g_1 (linear in parameters), we can estimate model (7) by OLS. The ATE can be estimated by

$$\widehat{ATE} = \widehat{\mu_1 - \mu_0}$$

ATE and ATT: Ignorability of treatment (cntd)

How to estimate the ATT?

Use that

$$Y_1 - Y_0 = (\mu_1 - \mu_0) + (\nu_1 - v_0) = ATE + (\nu_1 - v_0)$$

we get

$$ATT = ATE + E[\nu_1 - v_0 | D = 1] = ATE + E[E[\nu_1 - v_0 | D = 1, X] | D = 1]$$

Using that $E[\nu_1 - v_0 | D = 1, X] = E[\nu_1 - v_0 | X] = g_1(X) - g_0(X)$, ATT can be estimated by

$$\widehat{ATT} = \widehat{ATE} + (\sum_{i=1}^{n} D_i)^{-1} \left[\sum_{i=1}^{n} D_i(\hat{g}_1(X_i) - \hat{g}_0(X_i)) \right]$$

where $\hat{g}_0$ and $\hat{g}_1$ are the estimates of g_0 and g_1.

LATE

The validity of the previous estimators of the treatment effects hinges crucially on the validity of the ignorability of treatment assumption (4).

In many practical settings this assumption is unlikely to hold (i.e., we do not have a set of characteristics such that treatment assignment is independent of the potential outcomes given these characteristics), and then the treatment effects cannot be estimated consistently.

If we have an instrumental variable, we can, however, identify the so-called Local Average Treatment effect (LATE).

We consider the simplest case where the instrument, denoted z, is binary.

For each i, z_i is either 1 or 0.

LATE: Counterfactual treatments

Associated with the two possible outcomes on z_i, are counterfactual treatment indicators D_{i0} and D_{i1}:

- D_{i1} is person i's treatment status if z_i would equal 1
- D_{i0} is person i's treatment status if z_i would equal 0

For each i, we only observe one of the counterfactual treatment variables (but we can imagine that they both exist).

The observed treatment indicator is

$$D_i = D_{i0} + z_i(D_{i1} - D_{i0})$$

Using this formulation of the treatment indicator, we can rewrite the outcome equation (1) as

$$Y_i = Y_{i0} + D_{i0}(Y_{i1} - Y_{i0}) + z_i(D_{i1} - D_{i0})(Y_{i1} - Y_{i0})$$

LATE: Independence

A key assumption is the independence between the instrument on the one hand, and the potential outcomes and potential treatments on the other:

$$z_i \perp (Y_{i0}, Y_{i1}, D_{i0}, D_{i1}) \tag{8}$$

Under assumption (8) we have

$$\begin{aligned} & E[Y_i|z_i = 1] - E[Y_i|z_i = 0] \\ &= E[Y_{i0} + (Y_{i1} - Y_{i0})D_{i1}|z_i = 1] - E[Y_{i0} + (Y_{i1} - Y_{i0})D_{i0}|z_i = 0] \\ &= E[(Y_{i1} - Y_{i0})(D_{i1} - D_{i0})] \end{aligned}$$

Furthermore

$$\begin{aligned} & E[(Y_{i1} - Y_{i0})(D_{i1} - D_{i0})] \\ &\quad = E[Y_{i1} - Y_{i0}|D_{i1} - D_{i0} = 1]Pr(D_{i1} - D_{i0} = 1) \\ &\quad - E[Y_{i1} - Y_{i0}|D_{i1} - D_{i0} = -1]Pr(D_{i1} - D_{i0} = -1) \end{aligned} \tag{9}$$

LATE: Monotonicity

We can distinguish 4 different types of individuals

1. Always takers: $D_{i1} = D_{i0} = 1$
2. Never takers: $D_{i1} = D_{i0} = 0$
3. Compliers: $D_{i1} = 1,\ D_{i0} = 0$
4. Defiers: $D_{i1} = 0,\ D_{i0} = 1$

The monotonicity assumption states that there are no defiers:

$$D_{i1} - D_{i0} \geq 0\ \forall i = 1, ..., n$$

Under monotonicity, $Pr(D_{i1} - D_{i0} = -1) = 0$, and (9) becomes

$$E[(Y_{i1} - Y_{i0})(D_{i1} - D_{i0})] = E[Y_{i1} - Y_{i0} | D_{i1} - D_{i0} = 1] Pr(D_{i1} - D_{i0} = 1)$$

LATE: Definition

We therefore have

$$
\begin{aligned}
E[Y_i|z_i = 1] - E[Y_i|z_i = 0] &= E[(Y_{i1} - Y_{i0})(D_{i1} - D_{i0})] \\
&= E[Y_{i1} - Y_{i0}|D_{i1} - D_{i0} = 1] Pr(D_{i1} - D_{i0} = 1)
\end{aligned}
$$

The LATE is now defined as

$$
LATE \equiv E[Y_{i1} - Y_{i0}|D_{i1} - D_{i0} = 1] = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{Pr(D_{i1} - D_{i0} = 1)} \tag{10}
$$

LATE: Definition (cntd)

LATE can be interpreted as the average treatment effect for those whose treatment status changes because of the variation in the instrument.

Note that LATE is defined for individuals in a sub-population that is not identified (we cannot identify individuals for whom $D_1 - D_0 = 1$ since we do not observe both counterfactual treatments).

Note also that LATE depends on the instrument z. If we choose another instrument then LATE is modified. In contrast, the entities ATE and ATT are defined without reference to an IV.

LATE: Definition (cntd)

We can rewrite the denominator in (10) as follows

$$Pr(D_{i1} - D_{i0} = 1) = E[D_{i1} - D_{i0}] = E[D_i|z_i = 1] - E[D_i|z_i = 0]$$

So the LATE can be rewritten as

$$LATE = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[D_i|z_i = 1] - E[D_i|z_i = 0]} \tag{11}$$

Provided that

$$E[D_i|z_i = 1] \neq E[D_i|z_i = 0]$$

the LATE exists.

LATE: estimation

The LATE can be estimated by contrasting, both the numerator and denominator appearing in (11), by appropriate sample differences:

$$\widehat{LATE} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{D}_1 - \bar{D}_0} \tag{12}$$

where $\bar{Y}_1$ is the average of Y among individuals for whom $z = 1$, $\bar{Y}_0$ the average of Y among individuals for whom $z = 0$, and analogous definitions for $\bar{D}_0$ and $\bar{D}_1$.

Note that (12) is the Wald estimator of the slope coefficient in a regression model of Y on a constant and D (with z as an instrument for D))!

Put differently, in the simple case of a binary instrument for the binary treatment, the usual IV estimator consistently estimates LATE.

I understand/can apply...

- Definition of ATE and ATT
- Estimation of ATE and ATT (with experimental data, with observational data under ignorability of treatment)
- Definition of LATE
- Identification of LATE
  - monotonicity assumption
  - independence assumption
  - always takers, never takers, compliers, defiers
- Estimation of LATE