EEP/IAS 118 - Introductory Applied Econometrics, Lecture 10

Gregory Lane

July 2017

This Lecture

Topics

- Impact Evaluation
- RCTs

Assignments

- PS 4 due tomorrow
- Quiz 4 tomorrow

Office hours today 12:30 - 1:30

Impact Evaluation: Intro

Econometrics as a tool is often used to evaluate policies of some kind:

- Conditional cash transfers
- No Child Left Behind
- School voucher programs
- Infrastructure projects
- Air quality regulations
- Minimum wage

Want to assess changes in selected outcome / indicators that can be attributed to these interventions

- Generally, we only care about the effect of these programs not how other control variables influence the outcomes

Impact Evaluation: Intro

What econometricians care about is **causal relationships** - i.e. we want to know what would have happened to our y if the program had not occurred (the "counterfactual") and compare that to what did occur

- Omitted variable bias (and other problems) has prevented us from making any credible causal claims so far
- With purely observational data it is nearly impossible to credibly claim to have controlled for *all* x that may be causing bias
  - For example, how to control for intelligence, motivation, family connections, attractiveness, or friendliness?
  - All of these things matter for the result of many outcomes (e.g. earnings) and are likely correlated with many other x we care about (e.g. education)

Impact Evaluation: Intro

- To recover the true effect of the program, we would like to observe an individual's outcomes under a given policy and in a world where the policy did not occur. But this is impossible because we can only observe one of those outcomes!
  - Cannot just compare people affected by a policy because those affected may be different from those not affected $\rightarrow$ OVB!
  - In other words, we lack a credible comparison group for the treatment group
- We call this the "fundamental problem of causal inference"

Impact Evaluation: Intro

For example, let's say we want to examine the effect of a efficient appliance subsidy on energy consumption:

$$cons\_energy_i = \beta_0 + \beta_1 subsidy_i + \beta_2 Warm_i + \beta_3 Growth_i + \cdot + u_i$$

This seems straight forward - we know quite a bit about what drives energy consumption and can therefore control for many possible omitted varibles

- But we still need to assume that $E(u|X) = 0$.
- People who use the subsidy are likely to be more environmentally conscious, are more likely to have had old appliances, etc.

$\Rightarrow$ Therefore, can't interpret $\hat{\beta}_1$ as the *causal* effect of the subsidy on energy consumption

Impact Evaluation: Intro

- So does that mean econometric analysis is useless?
    - No, but it does mean that policy analysis is hard and needs to be done carefully
    - Think carefully when presented with claims that x causes y and use your judgement
    - This is especially true for things you *want* to be true
- We need more than a simple cross-section of data - we need a viable "identification" strategy in order to claim causality

Impact Evaluation: What is an RCT

- Randomized control trials are the closest we can get to solving this problem
- To ensure that the "control" individuals are a good counterfactual for "treated" individuals, we can randomly assign some individuals to receive treatment and others not
- If randomization is properly done, the two groups should not be statistically different
  - $\rightarrow$ Any difference can be attributed solely to the intervention/treatment
- In other words, the control group is a good counterfactual for the treatment group

Impact Evaluation: Measuring causal effect

When treatment is randomized and we have confirmed no statistical difference between treatment and control, we can estimate the causal effect of treatment as:

$$Impact = \bar{Y}_T - \bar{Y}_C$$

We can simply subtract the mean outcome in the control group from the mean outcome in the treatment group

Impact Evaluation: Measuring causal effect

In a regression framework, this is accomplished by:

$$Y_i = \beta_0 + \beta_1 T_i + u_i$$

Where T_i is an indicator for treatment and β_1 is the coefficient of interest.

- This provide us with an estimate as well as the standard error
- **Review Q:** Why is this identical to taking a difference in means?

Impact Evaluation: Key assumption

Key assumption!! If it were not for treatment, the control and treatment populations would be statistically identical, regardless of whether they are assigned to treatment/control:

$$E[Y_i | i \text{ in Treatment group}, T] = E[Y_j | j \text{ in Control group}, T]$$

or

$$E[u_i | T_i = 0] = E[u_i | T_i = 1] = 0$$

This key assumption cannot be empirically tested. Why?

Impact Evaluation: Testing key assumption

- The key assumption in the previous slide cannot be empirically tested because we never observe u_i
- Instead, we can provide evidence that the assumption is likely to hold by checking that the **observable** characteristics (e.g., age, income, education) between treatment and control are the same on average

$$E[x_i | i \text{ in Treatment group }] = E[x_i | i \text{ in Control group}]$$

Impact Evaluation: Testing key assumption

In Stata, this will run a t-test on the specified variables:

```
ttest age, by(treatment)
ttest education, by(treament)
OR
orth_out age education, by(treatment)
```

It is important that these x either be things we observed *before* treatment or be outcomes that should not be impacted by the intervention

Table: Information treatment balance: owners

Variable	Control	Treatment	Difference
Driver age	35.2	37.2	-2.04
			(0.86)**
Driver highest level of education	2.45	2.49	-0.040
			(0.080)
Driver experience	7.25	8.74	-1.49
			(0.68)*
Driver industry tenure	10.0	12.1	-2.09
			(0.75)
Weeks unemployed before job	3.13	2.19	0.94
			(0.70)
Number of vehicles driven before	5.68	5.40	0.28
			(0.55)
Number of conductors	1.23	1.16	0.064
			(0.053)
Number of past accidents	0.94	0.91	0.033
			(0.13)
Number of months employed	15.9	14.3	1.59

Impact Evaluation: Testing key assumption

Checking balance on observables does not ensure our key assumption is true, but it does make us feel better about it

- What if one of your observables is *not* balanced? Do you throw out the data?
- As you test more observed variables the probability that one of them is unbalanced is surprisingly high.
- E.g. if you test balance on 10 variables at the 5% level there is only a 60% chance that you won't reject at least once simply by chance:

$$1 - (.95)^{10} = .4$$

- If only one or two variables are unbalanced, the experiment is still probably okay
    - Should control for the unbalanced x

Impact Evaluation: RCT Steps

1. Decide (recruit) the universe of individuals that would be eligible for your study
2. Randomize that group into treatment or control
   - Need to decide on "unit" of randomization (e.g. individual, village, group)
3. Test that the randomization worked by checking balance on observables
4. Conduct the intervention with the treatment group
5. Estimate impact of the program via data collection

Impact Evaluation: Adding covariates

- Usually, we add covariates to a regression to prevent/reduce OVB. With proper randomization this is no longer necessary.
- However, adding covariates to the regression can serve two purposes:
  - Verify, as a robustness check, that $\hat{\beta}$ is invariant to the introduction of covariates in the
  - Add precision to the estimation regression
- We do not expect $\hat{\beta}$ to change because we do not expect the covariates to be correlated with treatment

Impact Evaluation: Adding covariates, Cont.

Adding covariates adds precision because it reduces our standard errors:

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{SST_x(1 - R_j^2)}}$$

$$\hat{\sigma}^2 = \frac{1}{n-k-1} \sum_i^n \hat{u}_i^2$$

If we include more covariates in our regression, we can reduce $\hat{u}_i^2$, i.e. the unexplained variation in Y goes down

Impact Evaluation: Heterogeneity

We may be interested in testing whether the treatment has a large impact on some groups than others $r \rightarrow$ treatment heterogeneity.

We can test this by interacting these characteristics (e.g., gender, age, socio-economic status, etc.) with the treatment variable.

$$Y_i = a + \beta_1 T_i + \beta_2 x_{2i} + \beta_3 T_i \times x_{2i} + u_i$$

If the variable x_2 represents a dummy for being female for example, then β_3 gives us the differential effect of the treatment for females relatives to males.

Impact Evaluation: Intention-to-treat

- Sometimes we do not have perfect compliance with a treatment
  - We can randomly assign whether a person receives a college scholarship, but we cannot force them to enroll accept the scholarship.
  - Maybe some entrepreneurial students in the control group find scholarships from other sources
- We cannot compare those that actually used a scholarship to those that didn't because these groups may be different in a way that is correlated with treatment (OVB!)
- We compare outcomes between the two groups to which people were originally assigned. This is called the Intention-to-treat (ITT) estimator

Impact Evaluation: Intention-to-treat

What can we do without perfect compliance:

- We compare outcomes between the two groups to which people were originally assigned. This is called the Intention-to-treat (ITT) estimator
- This effect is distinct from the "Average Treatment Effect" (ATE) which is what we estimate with perfect compliance
- However, the ITT may be more policy relevant in some cases
  - In most real-world settings there is not perfect compliance
  - E.g. Obamacare, vaccination, flossing
- If compliance is very bad, the ITT might be very different from the ATE

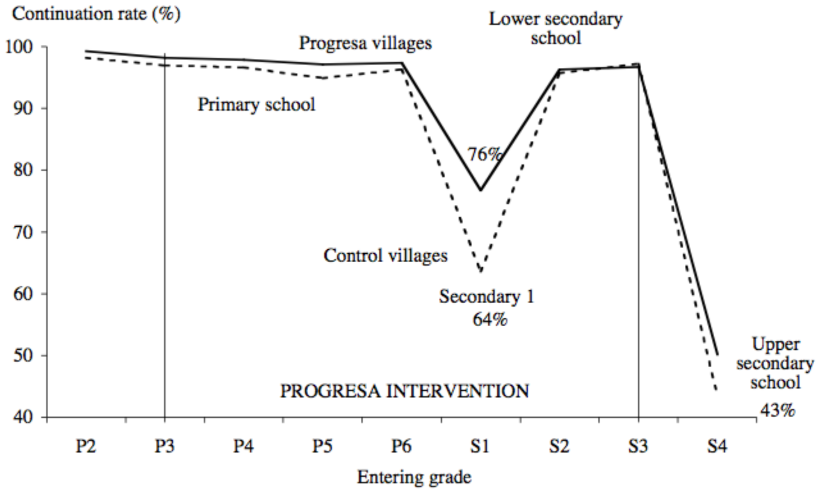
Progressa Example

Let's look at an example of an RCT to illustrate these concepts

Progressa was a Mexican program that offered cash transfers to poor families *conditional* on their children attending school

- Villages that initially received this program were chosen randomly
- A primary outcome of interest was school enrollment

Progressa Example



Progressa Example 1

```
reg enroll98 program

      Source |       SS       df       MS              Number of obs =     921
-------------+------------------------------           F(  1,   919) =    8.19
       Model |  1.21278015     1  1.21278015           Prob > F      =  0.0043
    Residual |  136.142269   919   .14814175           R-squared     =  0.0088
-------------+------------------------------           Adj R-squared =  0.0078
       Total |  137.355049   920  .149298966           Root MSE      =  .38489

------------------------------------------------------------------------------
    enroll98 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     program |   .0728048   .0254453     2.86   0.004     .0228671    .1227424
       _cons |   .7783019    .018692    41.64   0.000     .7416179    .8149858
------------------------------------------------------------------------------
```

Due to randomization, we this should be an unbiased estimate of the treatment effect

Note this was done with a sub-sample of the data

Progressa Example 2

```
orth_out male age97 h_edu hhsize exp98 distsec, by(program) pcompare

                                          control:    treatment:  (1) vs. (~e:
                                                 -             -             -
                        male:mean           0.512         0.499         0.699
                 age in 1997:mean          11.486        11.503         0.604
Education of the household head:mean        2.335         2.535         0.191
            (hhd) family size:mean          7.500         7.628         0.378
(hh) monthly per capita (nade~ :mean      104.744       109.967         0.244
(loc) Distance in km to the c~s:mean        2.680         2.575         0.507
```

Indirectly test the underlying assumption that

$$E[u_i|T_i = 0] = E[u_i|T_i = 1] = 0$$

by comparing balance on observables

Progressa Example 3

```
. reg enroll98 program distsec exp98 hhsize h_edu age97 male

      Source |       SS       df       MS              Number of obs =     921
-------------+------------------------------           F(  7,   913) =   14.94
       Model |  14.1151555     7  2.01645079           Prob > F      =  0.0000
    Residual |  123.239893   913  .134983454           R-squared     =  0.1028
-------------+------------------------------           Adj R-squared =  0.0959
       Total |  137.355049   920  .149298966           Root MSE      =   .3674

------------------------------------------------------------------------------
    enroll98 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     program |   .0740452   .0243661     3.04   0.002     .0262253    .1218652
     distsec |  -.0288689   .0051495    -5.61   0.000     -.038975   -.0187627
       exp98 |  -.0004827   .0001896    -2.55   0.011    -.0008549   -.0001106
      hhsize |  -.0049897   .0058742    -0.85   0.396    -.0165183    .0065389
       h_edu |    .012359   .0052988     2.33   0.020     .0019599    .0227582
       age97 |  -.1483166   .0242845    -6.11   0.000    -.1959766   -.1006566
        male |    .080076    .024256     3.30   0.001      .032472      .12768
       _cons |   2.577347   .2849907     9.04   0.000     2.018034     3.13666
------------------------------------------------------------------------------
```

Adding covariates to the regression

- Adds precision and doesn't change the coefficient

Progressa Example 4

```
. reg enroll98 male program male_program

      Source |       SS       df       MS              Number of obs =     921
-------------+------------------------------           F(  3,   917) =    5.61
       Model |  2.47455956     3  .824853187           Prob > F      =  0.0008
    Residual |  134.880489   917  .147088865           R-squared     =  0.0180
-------------+------------------------------           Adj R-squared =  0.0148
       Total |  137.355049   920  .149298966           Root MSE      =  .38352

------------------------------------------------------------------------------
    enroll98 |      Coef.   Std. Err.      t    P>|t|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .0765378   .0372613     2.05   0.040     .0034105    .1496651
     program |   .0761306   .0360734     2.11   0.035     .0053345    .1469267
male_program |   -.004702   .0507171    -0.09   0.926    -.1042371     .094833
       _cons |   .7391304   .0266566    27.73   0.000     .6868154    .7914455
------------------------------------------------------------------------------
```

Examine heterogenous effects across boys and girls

RCTs and You

RCTs are increasingly popular in academia and with governments, NGOs, and business (business speak calls them AB testing)

- My own work deals with several RCTs
  1. Randomize access to a new financial product in Bangladesh
  2. RCT on bus tracking systems in Kenya
  3. Contractor pay schedule in India
- Governments are increasingly willing to randomize policy
  1. Teacher payment systems in Indonesia
  2. School incentives for test scores
  3. Changing the "default" option for 401(k)

Understanding how RCTs work and how to conduct them will be useful