ECON 21110
Applied Microeconometrics
Winter 2022
Lecture 3
Experiments in Economics

Eyðfríð Juanna Schrøter Joensen

University of Chicago

January 27, 2022

Randomized Experiments

- How can we seek to estimate a causal effect of X on Y?
- Consider the point of randomized experiments by use of the example of drug trials
- Suppose we want to evaluate the causal effect of a new medical drug, X, on some health outcome, Y
- We proceed in six steps

Randomized Experiments

1. We recruit a sample of 100 patients who may benefit from the drug
2. We then **randomly** select 50 patients into a **treatment** group and 50 patients into a **control** group
3. The treatment group is given the drug, the control group a placebo (i.e. a sugar pill)
4. We then collect data on the outcomes of interest (e.g. health) of all patients
5. Finally, we compare the average outcomes in the treatment and control groups:
$$\overline{Y}_T - \overline{Y}_C$$
6. If outcomes are significantly better in the treatment group (e.g. better health), then this suggests that drug X has a positive causal effect on Y for this group of patients

Randomized Experiments

- The key thing in the example on the previous slide is the **randomization** in step 2
  - The fact that we randomly selected patients into treatment and control implies that the treatment and control groups will not differ systematically with respect to other characteristics that affect outcomes
  - This does not imply that the difference in outcomes between the treatment and control groups are exactly equal to the true causal effect. The reason is that pure chance may imply that some patients with better or worse outcomes happen to end up in the treatment or control group, but the procedure itself does not contribute to this
  - As the size of the treatment and control groups increases, the influence of pure chance on the results will fall and eventually disappear

Randomized Experiments

- What if we let patients decide themselves whether to get the drug or the placebo?
  - In this case, it is possible (or even likely) that the treatment and control groups differ systematically with respect to other characteristics that affect outcomes
  - For example: Patients who worry that their health (Y) is declining may be more prone to opt for the drug, while patients who don't worry about declining health may prefer the placebo. As a result, we can't tell whether differences in outcomes between the treatment and control groups reflect the effect of the drug (X), differences in characteristics ($X_1, X_2, ..., X_k$) between the treatment and control groups, or a combination of both
  - Self-selection into treatment and control groups create a systematic difference in outcomes that does not go away as we increase sample size

Randomized Experiments

- The underlying logic in the medical drug case pertains to many other circumstances
- For example: A colleague found a positive correlation between eating a healthy breakfast and college performance (GPA)
- However, the study did not randomize different types of breakfast to students, it simply compared the performance among students who ate a healthy or an unhealthy breakfast. Hence, the *students* rather than the *researchers* selected the type of treatment
- It is possible that the difference in performance indeed reflects a causal effect of breakfast, but an alternative explanation that cannot be ruled out is that the students who ate a healthy breakfast would have performed better anyway; e.g. because they are more forward-looking and live healthier lives in general

Experiments in Economics

- Randomized experiments are used to an increasing extent in economics, in particular within development economics and behavioral economics
  - Laboratory experiments
    - ⋆ Typically faces people with new situations in the “lab” (in practice often a classroom)
  - Field experiments
    - ⋆ Experiments in peoples’ natural environment; e.g. students and teachers in schools
  - Natural experiments (also called “quasi-experiments”)
    - ⋆ “As if” randomized experiments may be introduced due to institutional rules, borders, weather shocks, etc.
    - ⋆ Example: North and South Korea may be considered an example of a “natural” experiment on the effect of economic policy on economic performance
    - ⋆ We will return to natural experiments when discussing Instrumental Variables (IV) in Lecture 4, Regression Discontinuity (RD) in Lecture 5, and Difference-in-Differences (DiD) in Lecture 6

The Selection Problem

- For simplicity of illustration, assume we want to estimate the causal effect of a binary treatment D on the outcome Y
- Assume that for any individual i there are two potential outcomes:

$$Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

  - Y_{1i} could be the earnings if the individual attained a college degree
  - Y_{0i} could be the earnings if the individual did NOT attain a college degree
  - $D_i = 1$ if the individual was treated; i.e. attained a college degree
  - $D_i = 0$ if the individual was NOT treated

- The causal effect of treatment is: $Y_{1i} - Y_{0i}$

The Selection Problem

- The observed outcome Y_i can be written terms of potential outcomes as:

$$Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

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

- The last term is the causal effect of going to university on earnings
- Impossible to estimate in reality since we do not see the same individual attaining a university degree AND not attaining a university degree
- A fundamental problem is that we cannot observe both Y_{1i} and Y_{0i} for each individual. We can therefore not directly observe: $\mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 1]$

The Selection Problem

- Assume constant effect of treatment
- What are we measuring if we compare the average earnings between the two groups?

$$
\begin{aligned}
& \underbrace{\mathbb{E}\left[Y_i \mid D_i=1\right]-\mathbb{E}\left[Y_i \mid D_i=0\right]}_{\text{Observed difference in average earnings}} \\
= & \underbrace{\mathbb{E}\left[Y_{1i} \mid D_i=1\right]-\mathbb{E}\left[Y_{0i} \mid D_i=0\right]}_{\text{Observed difference in average earnings IF full compliance}} \\
= & \mathbb{E}\left[Y_{1i} \mid D_i=1\right]-\mathbb{E}\left[Y_{0i} \mid D_i=1\right]+\mathbb{E}\left[Y_{0i} \mid D_i=1\right]-\mathbb{E}\left[Y_{0i} \mid D_i=0\right] \\
= & \underbrace{\mathbb{E}\left[Y_{1i}-Y_{0i} \mid D_i=1\right]}_{\text{Average Treatment Effect on the Treated (TT)}} \\
& +\underbrace{\mathbb{E}\left[Y_{0i} \mid D_i=1\right]-\mathbb{E}\left[Y_{0i} \mid D_i=0\right]}_{\text{Selection Bias}}
\end{aligned}
$$

The Selection Problem

- Two effects
  - Average treatment effect on the treated (TT): $\mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1]$
  - The selection bias: $\mathbb{E}[Y_{0i} \mid D_i = 1] - \mathbb{E}[Y_{0i} \mid D_i = 0]$
- Examples where we may have a selection problem when comparing treated and non-treated
  - Effect of education on earnings
  - Effect of special teaching on test scores
  - Effect of hospitalization on health
- The selection problem arises because at a give point in time an individual can only be treated *or* not treated

Random Assignment

- In order to consistently estimate the average causal effect we must rely on different identifying assumptions
- Random assignment is one (the "best") approach which solves the selection problem under the independence assumption: $D_i \perp\!\!\!\perp Y_{0i}, Y_{1i}$
- Random assignment means that there are no observed or unobserved differences between the treated and the untreated:

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

- This means that the observed $\mathbb{E}[Y_{0i} \mid D_i = 0]$ is a good counterfactual for the unobserved $\mathbb{E}[Y_{0i} \mid D_i = 1]$ and that the selection bias term is zero with **random assignment** and **full compliance**

Random Assignment

- Since treatment is randomly assigned:

$$
\begin{aligned}
& \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] \\
= \; & \mathbb{E}[Y_{1i} \mid D_i = 1] - \mathbb{E}[Y_{0i} \mid D_i = 1] + \underbrace{\mathbb{E}[Y_{0i} \mid D_i = 1] - \mathbb{E}[Y_{0i} \mid D_i = 0]}_{=0} \\
= \; & \mathbb{E}[Y_{1i} \mid D_i = 1] - \mathbb{E}[Y_{0i} \mid D_i = 1] \\
= \; & \mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1] \\
= \; & \mathbb{E}[Y_{1i} - Y_{0i}]
\end{aligned}
$$

- Thus, the observed differences between treated and untreated is the Average Treatment Effect (ATE)

Random Assignment

- For good decisions based on empirical evidence you need to understand the problems with different methods and to be able to assess the quality of different studies
- For a decision maker it is also important to know what to think about before implementation in order to make it possible to evaluate
- Typically, we use regression to study the effect from an experiment
- A random sample of individuals i is draw from a population:

$$Y_i = \alpha + \beta D_i + U_i$$

- $D_i = 1$ if individual i is treated and $D_i = 0$ if individual i is not treated
  - U_i is the error term which includes all other determinants of Y_i

Random Assignment

- Evaluate the conditional expectations for treated and untreated, the difference:

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = \beta + \underbrace{\mathbb{E}[U_i \mid D_i = 1] - \mathbb{E}[U_i \mid D_i = 0]}_{\textit{Selection bias}}$$

- The OLS estimate, $\widehat{\beta}$, of the parameter β from the regression of Y on D is called the *difference estimator*
- Selection bias if correlation between the error term and treatment status $\mathbb{E}[U_i \mid D_i] \neq 0$
- Informal test of whether there is a problem with the randomization:
  - Run a regression using data before the treatment started using the treatment status $\Rightarrow$ should not have any effect!
  - Add control variables. If the treatment is random the estimated coefficient shall not change much

Random Assignment: Randomized Controlled Trial (RCT)

- Random assignment is what we would like to have in order to estimate causal effects
- Always useful to think about "ideal" randomized experiment
- Can randomize at the individual, village, district, school level. For the three latter, all individuals in a village, district, school are treated with an intervention
- Randomized Control Trial (RCT)
  - Very popular approach in development economics (e.g. Poverty Action Lab)
  - Increasingly popular in other subfields (e.g. economics of education, labor economics, environmental economics, and IO)

Random Assignment: Randomized Controlled Trial (RCT)

- Randomized experiments were first conducted in the sciences (commonly traced back to Galileo Galilei who used experiments to test his theories of falling bodies)
- Randomized experiments in the social sciences in particular suffer from a major problem: the missing counterfactual
  $\rightarrow$ Individuals or firms can usually not be observed with and without treatment at the same time

Experiments and Regression Analysis

- Assume treatment is the same for everyone

$$Y_{1i} - Y_{0i} = \beta$$

- With constant treatment effects, we can write the regression equation as

$$Y_i = \underbrace{\alpha}_{\mathbb{E}[Y_{0i}]} + \underbrace{\beta}_{Y_{1i}-Y_{0i}} D_i + \underbrace{U_i}_{Y_{0i}-\mathbb{E}[Y_{0i}]}$$

- where U_i is the random part of Y_{0i}

Experiments and Regression Analysis

- Evaluating the conditional expectations of this equation gives us:

$$\begin{aligned} \mathbb{E}[Y_i \mid D_i = 1] &= \alpha + \beta + \mathbb{E}[U_i \mid D_i = 1] \\ \mathbb{E}[Y_i \mid D_i = 0] &= \alpha + \mathbb{E}[U_i \mid D_i = 0] \end{aligned}$$

so that:

$$\begin{aligned} \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] &= \underbrace{\beta}_{\textit{Treatment effect}} + \\ &+ \underbrace{\mathbb{E}[U_i \mid D_i = 1] - \mathbb{E}[U_i \mid D_i = 0]}_{\textit{Selection bias}} \end{aligned}$$

- Selection bias $\Rightarrow$ correlation between the error term U_i and the treatment status D_i

Experiments and Regression Analysis

$$\mathbb{E}\left[U_i \mid D_i = 1\right] - \mathbb{E}\left[U_i \mid D_i = 0\right] = \mathbb{E}\left[Y_{0i} \mid D_i = 1\right] - \mathbb{E}\left[Y_{0i} \mid D_i = 0\right]$$

- is there difference in (no-treatment) potential outcomes between those who get treated and those who don't?
  - This can be indirectly tested by running a regression using pre-treatment data; i.e. test "balance" in pre-treatment variables

Experiments and Regression Analysis: Inclusion of Control Variables

- If control variables, X_i, are uncorrelated with the treatment, D_i, they will not affect the estimated treatment effect β
- Multiple ("long") regression model:

$$Y_i = \alpha + \beta D_i + X_i \gamma + U_i$$

- Estimates of β in the multiple ("long") regression model, will be close to the estimates of β in the simpler ("short") regression
- Inclusion of variables X_i may generate more precise estimates of the causal effect
  - If the control varibles X_i (uncorrelated with treatment, D_i) have substantial explanatory power for $Y_{i,}$ the standard error of the treatment effect will be smaller in the "long" regression model
  - Including control variables with explanatory power reduces the residual variance, which in turn lowers the standard error of the regression estimates

Threats to Validity

- Internal validity: if the estimated causal effects are valid for the studied population. Thus, no correlation between the error term and treatment
- Threats to internal validity, $\mathbb{E}[U_i \mid D_i] \neq 0$:
  - **Selection bias:** Treatment is in fact not randomized and depends on characteristics of i that also affect the outcome
  - **Partial compliance:** Not everyone who is treated actually takes part of the treatment
  - **Attrition:** Units with certain characteristics leave the experiment
  - **Hawthorne effect:** The experiment changes the individuals' behavior

Threats to Validity

- External validity: the extent to which causal effects of a particular program in a particular situation (or environment) at a particular time can be generalized to other situations (or environments) and time periods
  - Is the population representative?
  - Is the situation (or environment) representative?
  - Is the program (treatment) representative
  - Can we expect general equilibrium effects if we expand the program?

Threats to Validity

Other potential concerns for validity:

- Treatment affects also the untreated by spillovers or changes in market prices
  - Important to think about what happens to market prices or if there are other spillovers through, for example, peer effects

Non-compliers:

- A situation when treatment is offered randomly, but some participants do not participate
  - Causes a selection into treatment
  - When units drop out of treatment, the difference in outcomes between the treatment and the control groups now estimates the average impact of offering the treatment, usually called intention-to-treat (ITT)

Threats to Validity

- Assume we have an experiment that is designed such that the option to treatment is randomly assigned:
  - $Z_{it} = 0$: No treatment is offered
  - $Z_{it} = 1$: Treatment offered
  - If we have perfect compliance $Z_{it} = D_{it}$
  - If not:
    - $\mathbb{E}[Y_i \mid Z_i = 1] = \alpha + \beta\,\mathbb{E}[D_i \mid Z_i = 1] + \mathbb{E}[U_i \mid Z_i = 1]$
    - $\mathbb{E}[Y_i \mid Z_i = 0] = \alpha + \beta\,\mathbb{E}[D_i \mid Z_i = 0] + \mathbb{E}[U_i \mid Z_i = 0]$

Threats to Validity

- Difference by treatment status:

$$
\begin{aligned}
& \mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right] \\
= \; & \beta\left(\mathbb{E}\left[D_i \mid Z_i = 1\right] - \mathbb{E}\left[D_i \mid Z_i = 0\right]\right) \\
& + \left(\mathbb{E}\left[U_i \mid Z_i = 1\right] - \mathbb{E}\left[U_i \mid Z_i = 0\right]\right)
\end{aligned}
$$

- This estimates the *intent-to-treat (ITT)* parameter
- The ITT parameter divided by the the difference in compliance rate between treatment and control groups is the effect of treatment on the treated (TT):

$$
\begin{aligned}
\frac{ITT}{compliance\ rate} &= \frac{\mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right]}{\mathbb{E}\left[D_i \mid Z_i = 1\right] - \mathbb{E}\left[D_i \mid Z_i = 0\right]} \\
&= \mathbb{E}\left[Y_{1i} - Y_{0i} \mid D_i = 1\right] \\
&= TT
\end{aligned}
$$

Example of Large Randomized Experiment: Tennessee Project STAR

- Krueger (1999) econometrically re-analyses a randomized experiment of the effect of class size on student achievement
- The project is known as Tennessee Student/Teacher Achievement Ratio (STAR) and was run in the 1980s
- 11,600 students and their teachers were randomly assigned to one of three groups
  - 1) Small classes (13-17) students
  - 2) Regular classes (22-25) students
  - 3) Regular classes (22-25) students with a full time teacher's aide
- After the assignment, the design called for students to remain in the same class type for four years
- Randomization occurred within schools

Example of Large Randomized Experiment: Tennessee Project STAR

- Krueger (1999) estimates the following econometric model:
- $Y_{ics} = \beta_0 + \beta_1 SMALL_{cs} + \beta_2 \text{RegAide}_{cs} + \beta_3 X_{ics} + \alpha_s + \varepsilon_{isc}$
  - Y_{ics} = percentile score
  - $SMALL_{cs}$=indicator whether student was assigned to a small class
  - RegAid_{cs}=indicator whether student was assigned to regular class with aide
  - α_s= school fixed effects. Because random assignment occurred within schools
  - X_{ics}=control variables

Regression Results: Kindergarten

Explanatory variable	OLS: actual class size (1)	(2)	(3)	(4)
			A. Kindergarten	
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian (1 = yes	—	—	8.35	8.44
			(1.35)	(1.36)
Girl (1 = yes)	—	—	4.48	4.39
			(.63)	(.63)
Free lunch (1 = yes)	—	—	−13.15	−13.07
			(.77)	(.77)
White teacher	—	—	—	−.57
				(2.10)
Teacher experience	—	—	—	.26
				(.10)
Master's degree	—	—	—	−.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

Regression Results: 1st Grade

Explanatory variable	OLS: actual class size			
	(1)	(2)	(3)	(4)
			B. First grade	
Small class	8.57	8.43	7.91	7.40
	(1.97)	(1.21)	(1.17)	(1.18)
Regular/aide class	3.44	2.22	2.23	1.78
	(2.05)	(1.00)	(0.98)	(0.98)
White/Asian (1 = yes)	—	—	6.97	6.97
			(1.18)	(1.19)
Girl (1 = yes)	—	—	3.80	3.85
			(.56)	(.56)
Free lunch (1 = yes)	—	—	−13.49	−13.61
			(.87)	(.87)
White teacher	—	—	—	−4.28
				(1.96)
Male teacher	—	—	—	11.82
				(3.33)
Teacher experience	—	—	—	.05
				(0.06)
Master's degree	—	—	—	.48
				(1.07)
School fixed effects	No	Yes	Yes	Yes
R^2	.02	.24	.30	.30

Problem 1: Attrition

- If attrition is random and affects the treatment and control groups in the same way the estimates would remain unbiased
- Here the attrition is likely to be non-random: especially good students from large classes may have enrolled in private schools creating a selection bias problem
- Krueger (1999) addresses this concern by imputing test scores (from their earlier test scores) for all children who leave the sample and then reestimates the model including students with imputed test scores

Regression Results Imputing Test Scores to Address Attrition

	Actual test data		Actual and imputed test data	
Grade	Coefficient on small class dum.	Sample size	Coefficient on small class dum.	Sample size
K	5.32 (.76)	5900	5.32 (.76)	5900
1	6.95 (.74)	6632	6.30 (.68)	8328
2	5.59 (.76)	6282	5.64 (.65)	9773
3	5.58 (.79)	6339	5.49 (.63)	10919

- Non-random attrition does not seem to bias estimates

Problem 2:
Students changed Classes After Random Assignment

Example: Transitions between Grade 1 and Grade 2

	Second grade			
First grade	Small	Regular	Reg/aide	All
Small	1435	23	24	1482
Regular	152	1498	202	1852
Aide	40	115	1560	1715
All	1627	1636	1786	5049

- Ideally *all* observations on diagonal, but actually only 89%

Problem 2:
Students changed Classes After Random Assignment

- Students moved between treatment and control groups
- A common solution to this problem is to use initial assignment (here initial assignment to small or regular classes) as an instrument for actual assignment
- Krueger reports reduced form results where he uses initial assignment and not current status as explanatory variable
  $\rightarrow$ recovers ITT
- Kindergarten OLS and reduced form are the same because students remained in their initial class for at least one year
- From grade 1 onwards OLS and reduced form results are different

Problem 2:
Students changed Classes After Random Assignment

Explanatory variable	OLS: actual class size				Reduced form: initial class size			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	B. First grade							
Small class	8.57	8.43	7.91	7.40	7.54	7.17	6.79	6.37
	(1.97)	(1.21)	(1.17)	(1.18)	(1.76)	(1.14)	(1.10)	(1.11)
Regular/aide class	3.44	2.22	2.23	1.78	1.92	1.69	1.64	1.48
	(2.05)	(1.00)	(0.98)	(0.98)	(1.12)	(0.80)	(0.76)	(0.76)
White/Asian (1 = yes)	—	—	6.97	6.97	—	—	6.86	6.85
			(1.18)	(1.19)			(1.18)	(1.18)
Girl (1 = yes)	—	—	3.80	3.85	—	—	3.76	3.82
			(.56)	(.56)			(.56)	(.56)
Free lunch (1 = yes)	—	—	−13.49	−13.61	—	—	−13.65	−13.77
			(.87)	(.87)			(.88)	(.87)
White teacher	—	—	—	−4.28	—	—	—	−4.40
				(1.96)				(1.97)
Male teacher	—	—	—	11.82	—	—	—	13.06
				(3.33)				(3.38)
Teacher experience	—	—	—	.05	—	—	—	.06
				(0.06)				(.06)
Master's degree	—	—	—	.48	—	—	—	.63
				(1.07)				(1.09)
School fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
R^2	.02	.24	.30	.30	.01	.23	.29	.30

Potential Problems when Running Experiments

(1) **Randomization Bias**

- The experimental sample may be different from the population of interest because of randomization IF
  - Experiment changes who selects into receiving treatment
  - Experiment changes nature of treatment
- Can occur if treatment effects are heterogeneous $\rightarrow$
  People selecting to take part in the randomized trial may have different returns compared to the population average

Potential Problems when Running Experiments

Potential Problems when Running Experiments

(2) **Supply Side Changes**

- If programmes are scaled up the supply side implementing the treatment may be different
- In the trial phase the supply side may be more motivated than during the large scale roll-out of a programme

(3) **Attrition Bias**

- Attrition rates (i.e. leaving the sample between the baseline and the follow-up surveys) may be different in treatment and control groups
- The estimated treatment effect may therefore be biased

Potential Problems when Running Experiments

(4) **"Hawthorne" Effects**

- People behave differently because they are part of an experiment
- If they operate differently on treatment and control groups they may introduce biases
- If people from the control group behave differently these effects are sometimes called "John Henry" effects

(5) **Substitution Bias**

- Control group members may seek substitutes for treatment
- This would bias estimated treatment effects downwards
- Can also occur if the experiment frees up resources that can now be concentrated on the control group

A Good Experimental Paper?

- A promising avenue for experimental papers seem to be the ones that combine experimental data with economic theory:
    1. Discriminating between important theories
    2. First obtain "reduced form" results of a causal effect and then use structural econometric methods to disentangle economic mechanisms
    3. Use an experiment to estimate externalities or other market failures
- Two experimental papers with particularly close links between empirics and theory:
    - Miguel and Kremer (2004) - Worms
    - DellaVigna, List, and Malmendier (2010) - Charitable Giving

Other Econometric Methods

- IF we do not have random assignment into treatment; i.e. a randomized controlled experiment, then we can use other econometric methods for evaluation:

- **Selection Observables:** Regression Analysis (Done!)
- **Selection Unobservables:** Fixed Effects (FE) or First Differences (FD) (Lecture 7), Difference-in-Differences (DiD) (Lecture 6)

- **(Quasi-)Experiments:** Random Assignment (Done!), Instrumental Variables (IV) (Next: Lecture 4), Regression-Discontinuity (RD) (Lecture 5)