Field Experiments in Economics

Mini course on Causal inference: Lecture 3

Dr. Miri Stryjan

Department of Economics
Aalto University
2021-22

Plan for lecture

1. Introduction to Randomized experiments and why we use them
   - Some terminology
   - Recap on our goal: introduce exogenous variation in X
   - How can Randomized experiments help us reach this goal?
2. Discuss Randomized Field Experiments, or RCTs, in more detail:
   - When and why a can RCTs be used/considered?
   - What can be studied in RCTs?
   - What are the criticisms against RCTs
3. Discuss very briefly other types of experiments: Lab experiments, Artefactual field experiments and and Lab in the field experiments.

Empirical analysis in Economics

- In empirical analysis in economics, we are often interested in testing if the relationship between 2 variables is causal. In general terms: we look at the relationship between:
  - the dependent variable (the outcome), Y and
  - the independent variable (that supposedly affects the outcome), X.
- We write this as: Y=f(X) meaning that Y is a function of X.

The basic challenge of causal inference

We would like to understand the effect of X on Y.

- Ideally: we would like to watch the same individual, in the same time period, being exposed to X **and** not being exposed to X, and compare the outcomes of these two scenarios.
- In reality, we never observe both scenarios:
  - We observe one (the factual scenario, that actually happened)
  - but we do not observe the other (the counterfactual scenario).
- In empirical work in economics that focuses on causal inference, we try to convincingly describe the "counterfactual scenario".

Terms: Endogeneity vs. Exogeneity

- Endogeneity: A variable is endogenous in an empirical model if it is (at least partly) a function of other variables in that model.
- The name comes from Endo- : a prefix from Greek *endon* meaning "within, inner, absorbing, or containing"
- Exogeneity: something from outside of the system. A variable is exogenous if it is not a function of other variables in that model.
- The name comes from Exo- (or ex-): a prefix from Greek *ekso* meaning "outer, external"

Empirics - Endogeneity

Endogeneity: A variable is endogenous in an empirical model if it is (at least partly) a function of other variables in that model

- The outcome variable Y is always explained inside the model and is therefore always endogenous.
- But the "explanatory" or independent variable X should be exogenous, since The model is supposed to explain Y, not X.

$\Rightarrow$ **Concerns about endogeneity in empirical analysis are always about X, the independent variable.**

Empirics - Endogeneity

- If X is correlated with other variables that also affect Y, we say that X is endogenous, and this is a problem because it makes us unable to measure the true effect of X on Y.
- Many of these "other variables" are hard to measure, and will remain omitted from our model and we are likely to have a biased estimate due to omitted variable bias/selection bias.
- We then need to either

  (a) control for these omitted variables as well as we can, or
  (b) replace X with some *exogenous* independent variable, or
  (c) find or create a situation with *exogenous variation* in X.

Exogenous variation in X through Randomization

$\Rightarrow$ Goal: introduce exogenous variation in X.

Exogenous variation in X through Randomization

$\Rightarrow$ Goal: introduce exogenous variation in X.

...that is, variation in X that is not correlated with any omitted variables that affect both X and Y.

Exogenous variation in X through Randomization

$\Rightarrow$ Goal: introduce exogenous variation in X.

...that is, variation in X that is not correlated with any omitted variables that affect both X and Y.

- To introduce exogenous variation in the X variable, some argue that the most convincing way is to conduct a **Randomized experiment**, where X is manipulated by the researchers. This manipulation is called "Treatment".

Randomized experiments

Also called Randomized Control/Controlled Trials (RCTs)

- Some individuals/units are randomly placed in the "treatment group" (T=1 for them)
- Others are randomly placed in a "control group" (T=0 for them)
- Whether an individual gets the treatment or not depends only on the randomization, and not on any factors that could also affect the outcome (preferences, ability).

- Common in natural sciences, harder to do in social sciences.
- BUT: Experiments are becoming increasingly common in economics: especially in the fields of micro-development and labor.

When do people randomize?

Some things can not be randomized (X variables in red):

- Suppose we want to study the effect of political propaganda on Political views or beliefs
- Suppose we want to study the effect of fertility on women's wages.
- Suppose we want to study effect of exposure to violence, war or pollution on human capital development or risk preferences.

When do people randomize?

Some things can not be randomized (X variables in red):

- Suppose we want to study the effect of political propaganda on Political views or beliefs
- Suppose we want to study the effect of fertility on women's wages.
- Suppose we want to study effect of exposure to violence, war or pollution on human capital development or risk preferences.

  - not ethical or possible to attempt to introduce random variation in these X variables.

$\Rightarrow$ In such cases: Find some exogenous factor that introduces random variation in the participation in / exposure to war or violence or pollution. IV or Diff in Diff approaches (later lectures).

When do people randomize?

So when is it possible and appropriate to randomize?

1. When randomization is not too ethically problematic.
2. When randomization is not so costly (in terms of money and time of both researchers, involved organizations, and study subjects)

AND

3. When randomization can teach us something we would not otherwise be able to measure empirically.

When do people randomize?

Some things can quite easily be randomized (I):

- If too many individuals apply for a course and fulfill the requirement, the course intake can be decided by a lottery among those who are eligible. (oversubscription design)
- This creates a situation where those who fulfill the requirement and are accepted to the course are not systematically different than those who fulfill the requirement but did not get in (for example in motivation or skill before the course). The lottery eliminates selection bias.

When do people randomize?

Some things can quite easily be randomized (I):

- If too many individuals apply for a course and fulfill the requirement, the course intake can be decided by a lottery among those who are eligible. (oversubscription design)
- This creates a situation where those who fulfill the requirement and are accepted to the course are not systematically different than those who fulfill the requirement but did not get in (for example in motivation or skill before the course). The lottery eliminates selection bias.

$\Rightarrow$ The effect of the course can be evaluated by comparing those who won the lottery and got accepted (the "treated" group) to those who lost the lottery and were not accepted (the "control group")

When do people randomize?

Some things can quite easily be randomized (II):

- A new program for education or health can be introduced to randomly selected areas or districts, while other districts that randomly did not get the program can serve as a control group.

When do people randomize?

Some things can quite easily be randomized (II):

- A new program for education or health can be introduced to randomly selected areas or districts, while other districts that randomly did not get the program can serve as a control group.

$\Rightarrow$ If the decision of where to start the program is decided by randomization, and not by what areas need the program most (e.g. the poorest areas) the effect of the program can be evaluated by comparing the "treated" districts to those in the "control" districts (who did not get the program).

When do people randomize?

Some things can quite easily be randomized (III):

- When there is no consensus about what procedure will work better in a specific context. Randomize between 2 procedures.
- Example: what rule for decision making in groups works best for electing group leaders. Two system are popular and commonly used: (a) secret vote and (b) public group discussion. In one research project, I randomized instructions: 46 groups were instructed to use secret ballot vote and 46 groups public discussion.

When do people randomize?

Some things can quite easily be randomized (III):

- When there is no consensus about what procedure will work better in a specific context. Randomize between 2 procedures.
- Example: what rule for decision making in groups works best for electing group leaders. Two system are popular and commonly used: (a) secret vote and (b) public group discussion. In one research project, I randomized instructions: 46 groups were instructed to use secret ballot vote and 46 groups public discussion.

$\Rightarrow$ Since groups did not choose what method to use, the randomization eliminates omitted variable bias related to group preferences. We can compare groups to learn about the effect of the rule used for decision making.

`https://voxdev.org/topic/institutions-political-economy/selecting-better-community-leaders-experimental-evidence-uganda`

Example from inequality literature

- Observation: African Americans have a higher unemployment rate than whites
- Hypothesis: African Americans are discriminated against by employers, and that is one reason that they have higher unemployment (they get less chances and job offers than whites).
- **Challenge**: African Americans also *on average* are poorer and have lower education, which could affect employment prospects. There are also many unobservable factor in the interaction between employers and candidate employees that can affect the chance that a job application $\rightarrow$ employment.

Example from inequality literature

- How can we show whether discrimination by employers is an explanation for higher unemployment, and not all these other factors?
- We can control for education and poverty level of applicants, but we can not control for all the unobservables.
- Solution: find a situation in which we can compare white and black individuals with identical levels of education and other relevant variables.
- **Example** experiment with job applications where the only thing that was randomly changed between applicants was their name (which signaled if the applicant was black or white): Marianne Bertrand and Sendhil Mullainathan (2004)

Field experiments -Intro

Field experiments (sometimes called natural field experiments - not to be confused with natural experiments) take place in the real world.

- Participants are a representative, randomly chosen, non-self-selected subset of the treatment population of interest.
- Participants don't always know that they are part of an experiment.
  - If data is collected on the participants, such as survey data, they are informed that they are part of a research study but often not that it is in the form of an experiment, and they are not informed if they are in the treated or the control group.

Field experiments - What do people randomize?

1. Interventions that introduce exogenous (random) variation in access to a program or product

- Program access (who gets access to a specific health or education policy, or a program)

  Examples: Miguel & Kremer 2004: Randomized access to a health program between schools. Budget to treat 25 schools, but collected data on an additional 50 control group schools. Banerjee and Duflo (2015): Randomized access to a loan product (microcredit) at the neighbourhood level. Control group areas got access to the product later.

- requires a collaboration with a program implementor (Government, firm or NGO).
- Long term outcomes (at least 1 year, typically more)
- evaluations that measure the entire effect of a program are called "impact evaluations".

Field experiments - What do people randomize?

2. Interventions that manipulate a part of a program
   - Prices
     - Example: Cohen & Dupas (2010) randomize subsidies (discounts) on bednets (against Malaria)
   - Contract type/rules
     - Examples: Burchardi et al. (2019): randomize contract types for small scale farmers. Deserranno, Stryjan et al. (2019): randomize instructions to community groups about what rule to use for a leadership election.
   - Also requires collaboration with a program implementor.
   - Long term outcomes (at least 1 year, typically more)
   - experiments that focus on **one** aspect of a program often have a clearer contribution to our understanding of an underlying theory as compared to "impact evaluations" that measure overall effects. So these papers are great references!

Field experiments - What do people randomize?

3. Smaller interventions (in terms of budget, scope or time)
   - information/framing of a message. Examples:
     - Deserranno (2019) Randomizes if a job ad emphasizes the salary or the "mission" of the job, measures how this affects who applies for the job. Bertrand & Mullainathan (2004), Bartos et al. (2016) randomize information given to employers about job applicants and measure discriminatory behavior.
   - If the outcome is immediate (e.g. how many and who applies?) this type of experiment can be quick.
   - experiments focusing on **one** aspect of a program often have a clearer contribution to our understanding of an underlying theory as compared to "impact evaluations" that measure overall effects. So these papers are great references!
   - Small interventions of this type could be made in an MA thesis.

Estimating equations

The estimating equations in RCT papers are typically simple: Treatment (T) is random and by definition exogenous, so we can in principle just measure the effect of T without worrying about omitted variable bias:

$$Y_i = \alpha + \beta T_i + \epsilon_i$$

But often, we run an Analysis of covariance (ANCOVA) specification, where we also control for covariates, including the Y value at baseline (before experiment):

$$Y_{it+1} = \alpha + \beta T_i + \gamma Y_{it} + \sigma X_{it} + \epsilon_{it+1}$$

Where $Y_i t$ and $X_i t$ are characteristics of indiv. i at time t (baseline). (To control for baseline differences, it is also possible to run a Difference in Differences specification where the baseline value is averaged out, but this often reduces statistical power compared to ANCOVA.)

Criticism towards randomized experiment approach

- Low external validity (too much focus on internal validity: estimating and isolating causal effects in a specific context).
- Cannot address "general equilibrium effects"?
- Too simplistic to break down reality into isolated causal relationships?
- Focus on things that are easy to evaluate in this way, rather than the important questions?
- Ethically problematic?

Ethics and randomized field experiments

Are these experiments ethically problematic?

- Fairness concerns (control group do not get the good)
- Information (affected people not always informed)
- There is an interesting recent discussion on ethics in field experiments, after the 2019 Economics Nobel Prize was awarded to three development economists that pioneered such experiments:

`https://www.project-syndicate.org/commentary/ethics-of-random-controlled-trials-to-fight-poverty-by-peter-singer-et`

Lab experiments

- Test economic behavior in a controlled environment: typically a computer lab where subjects are faced with scenarios and make decision about how to allocate resources
- **Purpose 1**: to test theory (Smith 1982):
  - Create a microeconomic environment in the lab,
  - with players, institutions, prices, producers, consumers...
  - the researcher controls all ingredients of the environment except players' preferences.
  - Payment to participants is done according to their performance - in a way that gives incentives to participants to behave according to the theoretical model.
- **Purpose 2**: to understand mechanisms as part of a broader study.
- A number of standard "work horses" (games) have been developed by researchers.

Commonly used "work horses" in lab experiments

"work horses" (games) that have been developed by researchers to measure specific types of behavior and preferences.

- dictator game - altruism vs. self interest
- ultimatum game - preferences for fairness, honor and more
- trust game - trust, discrimination
- public goods game - cooperation and efficiency
- prisoner's dilemma games -coordination, cooperation and more

Moving toward a more "natural" setting

To increase external validity, or learn about the choices of a specific population (not the WEIRD group), sometimes lab experiments are taken to the field.

- "Artefactual field experiments" - inviting a "non standard" population to lab experiments - example poor, less educated, non western... Busara center for behavioral economics in Nairobi, Kenya specialize in this. `https://www.busaracenter.org/`
- Lab in the field or "Framed field experiments": here, the lab is moved to the relevant field setting. Example: groups of villagers who had first done a course on cooperation. The public good game provides outcomes of the program: did it increase cooperation and decrease free riding?

References

- Bertrand, M. and Mullainathan, S., 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American economic review, 94(4), pp.991-1013.
- Bartos, V., Bauer, M., Chytilová, J. and Matejka, F., 2016. Attention discrimination: Theory and field experiments with monitoring information acquisition. American Economic Review, 106(6), pp.1437-75.
- Miguel, E. and Kremer, M., 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica, 72(1), pp.159-217.
- Banerjee, A., Duflo, E., Glennerster, R. and Kinnan, C., 2015. The miracle of microfinance? Evidence from a randomized evaluation. American Economic Journal: Applied Economics, 7(1), pp.22-53.
- Burchardi, K.B., Gulesci, S., Lerva, B. and Sulaiman, M., 2019. Moral hazard: Experimental evidence from tenancy contracts. The Quarterly Journal of Economics, 134(1), pp.281-347.

References

- Cohen J. and P. Dupas. 2010. Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment, Quarterly Journal of Economics 125 (1), pp.1-45.
- Deserranno, E., 2019. Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda. American Economic Journal: Applied Economics, 11(1), pp.277-317.
- Deserranno, E., M. Stryjan, and M. Sulaiman. 2019. "Leader Selection and Service Delivery in Community Groups: Experimental Evidence from Uganda." American Economic Journal: Applied Economics, 11 (4): 240-67.
- Smith, V.L., 1982. Microeconomic systems as an experimental science. The American Economic Review, 72(5), pp.923-955.

Field experiments - references*

For more discussion of Field Experiments in Economics, see

- https://www.nobelprize.org/uploads/2019/10/popular-economicsciencesprize2019-2.pdf
- https://www.nobelprize.org/uploads/2019/10/advanced-economicsciencesprize2019.pdf