Experimental methods

ADEC781001: Empirical Behavioral Economics

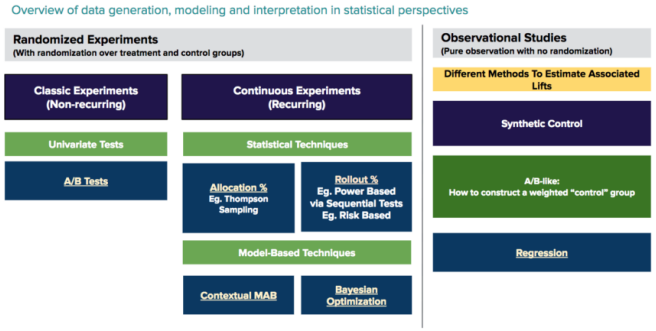
Lawrence De Geest (lrdegeest.github.io)

BOSTON COLLEGE
Woods College of Advancing Studies

Who runs experiments?

- Experiments used to be considered a "field" (like labor or environmental)
- Largely because not many economists used them
- Now "experimental economics" is considered by a method
- Experiments widely used
  - testing and building theory
    - e.g. labor: incentives and effort
    - e.g. environmental: valuation of non-market goods
    - e.g. micro: market and auction design (e.g. selling online ad space)
    - many other fields (macro, finance, ...)
  - testing policy interventions (e.g. RCTs and the 2019 Nobel Prize)

Uber's Experimentation Platform (XP)



https://eng.uber.com/xp/

Treatment effects

- Experiments are about treatments and what they do to behavior
- Let Y be some outcome of interest (e.g. effort at work, income, etc.) and i index an individual
- We want to know how Y_i changes due to some treatment T
- Let Y_{i1} be the outcome to i under the treated ($T = 1$) and Y_{i0} the outcome under the control ($T = 0$)
- The treatment effect for i is then $Y_{i1} - Y_{i0}$
- Of course this may not be consistent across individuals, so instead we calculate the average treatment effect (ATE)
  - $\mathbb{E}[Y_i|T = 1] - \mathbb{E}[Y_i|T = 0]$
  - also written as $\mathbb{E}[Y_1 - Y_0]$

Randomization

- But there is a bigger problem: no individual can receive both the treatment and the control *at the same time*
- That means we only observe $\mathbb{E}[Y_i|T=1]$ **or** $\mathbb{E}[Y_i|T=0]$
  - So we can't observe treatment effects at the individual level (i.e. we don't observe the counterfactual)
  - This is often referred to as the "Fundamental Problem of Causal Inference"
- One solution is to get a bunch of people who are similar in many characteristics and assign them to treatment or control
- Problem: if propensity to receive treatment is correlated with observed or unobserved subject characteristics, then the ATE will be biased
- Solution: randomize so that a coin flip determines whether i is assigned to treatment or control
  - This ensure $\mathbb{E}(Y_1) \perp \mathbb{E}(Y_0)$ (statistical independence)
  - Therefore $\mathbb{E}[Y_1 - Y_0] = \mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0]$
- You will of course never get completely identical control and treatment groups, but with large enough samples you can appeal to the asymptotic properties around your distributions and quantify your uncertainty of the ATE
- This is a Randomized Control Trial (Fisher 1925)

Calculating ATE

- What this means is that so long as randomization is properly done we can estimate the ATE by the difference-in-means estimator
- We can get this estimate using OLS:
  - $Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \varepsilon_i$ where $\hat{\beta}_1$ is the ATE
  - may need to correct the standard errors, add fixed/random effects, etc. depending on study design
    - assumption: fixed/random effects orthogonal to T_i
- And look at heterogenous treatment effects where the treatment effect varies on some condition (e.g. gender)
  - $Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 (T_i \times X_i) + \varepsilon_i$
  - Other approaches to heterogeneity: quantile treatment effects (QTEs)[1]
- Other commonly used methods possible
  - e.g. Wilcoxon Rank Sum tests on group means

[1] https://eng.uber.com/analyzing-experiment-outcomes/

Trusting ATE

- Randomization ensures no selection bias on T_i
- But there could be other problems
  - e.g. participation bias (subjects who participate in studies may be those with most to gain)
  - e.g. sample size (underpowered statistical tests are a problem)
  - e.g. stakes
- How does estimate of T_i generalize?
  - internal vs external validity
  - induced vs homegrown preferences

External validity

- Internal validity: do the data permit causal inference?
  - Driven by experiment design and analysis
- External validity: how well do the results generalize? (common criticism)
  - Response 1: the purpose of the experiment is to test theory
    - abstraction is a virtue *if* it clarifies and enhances understanding
  - Response 2: then do the experiment with "real people"

Induced preferences

Smith (1976)

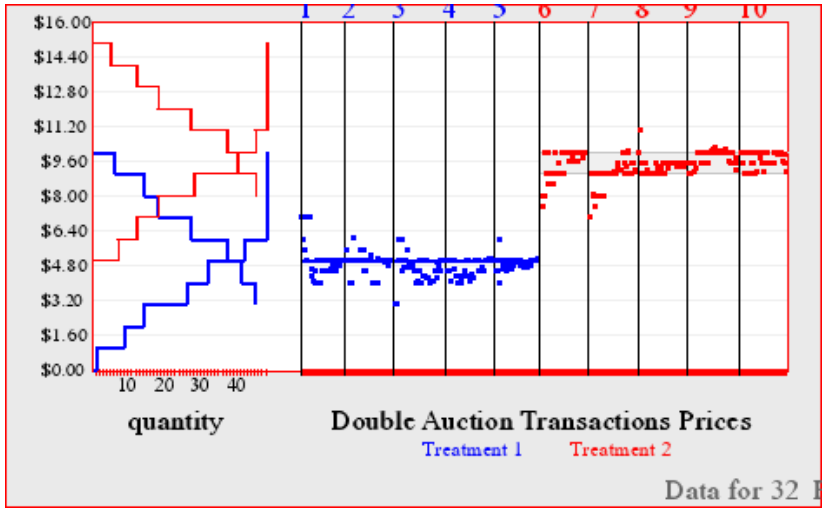
- How do we know subject behaves according to the rules and incentives in the experiment and not their "homegrown" preferences?
- Need to *induce* preferences and "neutralize" homegrown preferences
- Suppose subject has unobserved preferences $V(m_h + m_e, z)$
  - $V(\cdot)$ is the value function
  - m_h: homegrown money
  - m_e: experiment money
  - z: all other motives
- Assume $V_{m_e} > 0$
  - Then paying subject allows you to "control" subject's preferences
    - *marginal* incentives motivate subject to maximize utility independent of m_h and z
    - important: can't be a flat participation fee
- What if m_h is high?
  - make m_e sufficiently high
- What if z includes motivations like fairness and envy?
  - don't broadcast individual payoffs to other subjects
- What if z includes wanting to help/harm the experimenter? ("demand effects")
  - don't tip (drop hints to) the subject
  - use neutral language in instruction

Advantages of experiments

- Exogenous randomization
- Experimenter chooses exogenous variables and knows variables are endogenous
- Make endogenous real-world details exogenous in the lab
  - e.g. information, money supply shocks, stochastic processes
- Compare behavior to benchmarks
  - Game theory and experiments closely related
  - Experimenter can distinguish equilibrium and disequilibrium behavior
  - Is adjustment fast or slow?
- Substitute for complicated econometrics
- Replication
  - somebody else can run your experiment on a new sample

Example: markets

Confirming theories



Example: markets

Challenging theories

- Inefficiencies in real markets can be driven by behavioral anomalies
  - endowment effect
  - heuristics
- One heuristic: stereotypes
  - can lead to "statistical discrimination" (Arrow 1973, Phelps 1972)

Example: discrimination

- Consider a marketplace
- Prices p are determined through bargaining between buyers and sellers
- Suppose you observe prices p_i offered to individual i offered by a seller j
- Suppose you also observe i's race R_i and other buyer/seller variables $\mathbf{X}$
  - Assume one race for buyers $R_j = R = 0$
  - Assume there are only two races: $R_i = 1$ and $R_i = 0$
  - Interpretation: buyer's race is a "treatment" effect
- Suppose $w_i(R_i, \mathbf{X}) = \alpha + \beta R_i + \mathbf{X}'\gamma + \varepsilon$
- Do employers discriminate against members of $R_i = 1$?
  - Check sign and significance $\frac{\partial w_i}{\partial R_i} = \hat{\beta}$
- Lots of observational studies suggest $\hat{\beta} < 0$ (i.e. there is discrimination)
- But is j racist? Or is j exploiting bargaining power?
  - Source of discrimination is ambiguous

Example: discrimination

List (2004)[2]

- Sports cards market
  - people are both buyers and sellers of sports cards
  - initial observation "strong tendency for dealers to give nondealer minorities (women, nonwhite, and older agents) initial and final offers that are inferior to those received by their majority counterparts."
    - stronger "consumer-side discrimination" than "supply-side"
- Battery of experiments
  - "dictator games, which isolate "tastes" for discrimination"
  - "face-to-face continuous bilateral bargaining in a multilateral market"
  - "real auctions... estimate the underlying distribution of reservation values among the various consumer groups"
  - "dealer perception experiments...observe dealer beliefs about the reservation value distributions of the various groups"
- Findings support statistical discrimination over "animus" (prejudiced-based) discrimination
  - "—dealers use minority membership as a proxy for the distribution of reservation values"

[2] List, John A. "The nature and extent of discrimination in the marketplace: Evidence from the field." The Quarterly Journal of Economics 119, no. 1 (2004): 49-89.

Field vs Lab

- In research most methods are compliments rather than substitutes
  - e.g. experiments are not about *replacing* observational studies
  - the "vs" is a bit misleading
- But when running a *particular* study you often need to choose a single method
- Methods are driven by research questions
- If you run an experiment, assumption is you *need* an experiment
- Which type of experiment you choose matters
- Focus in this lecture: lab versus field experiments
  - many types of field experiments (Harrison and List (2004))

Aside: Mechanical Turk

- Online labor market where buyers contract with sellers to peform tasks[3]
  - Similar to Freelancer.com
- Used for experiments because of its friendly API
  - the experiment is a "job"
  - just like in a lab, "subjects" paid based on their actions
- Tends to produce qualitatively similar results as lab studies (Horton et al. 2011, Arechar et al. 2017)
  - However this has not been explored for all games/designs
  - Harder to implement complex games since you can't directly explain (e.g. read instructions) to subjects
  - Can't answer subject questions, can't enforce rules (e.g. no talking between subjects, no phones, etc.)
  - Tests for comprehension can alleviate this concern
- Advantage: lower marginal cost of data
  - cheaper to collect large samples
- Disadvantage: less control
  - is a worker one subject or multiple subjects or a bot? (workers can have multiple accounts or be robots)
    - generally dealt with by post-processing the data
  - possibly a narrower scope of feasible designs

[3] https://www.mturk.com/worker

Lab Experiments

What is a lab?

- Short answer: computer lab
- Long(er) answer: a computer lab that is "context-free"
- Experiments do not have to be run on computers
  - early experiments were done with paper and pencil
- But computers make it a lot easier to
  - collect data
  - manipulate variables of interest
  - scale
- Commonly used software
  - z-Tree (https://www.ztree.uzh.ch/en.html)
  - o-Tree (https://www.otree.org/)

Field Experiments

Artefactual field experiments

Harrison and List (2004)

- Also known as "lab in the field"
- Basically a lab study but with "non-standard populations" (read: students)
  - means you literally conduct study in the computer lab
  - subjects know they are in an experiment
  - Henrich et al. (2005) ultimatum games and dictator games in 15 different countries
- Argument that you often get qualitatively similar results with standard and non-standard population
  - e.g. Chavez et al. (2018) commons game with fishermen/students in Chile

Framed field experiment
Harrison and List (2004)

- Problem with any lab study is how much of the observed behavior is due to the lab setting
  - lab is a sterile setting, unlike real decision-making settings
  - sometimes you want this, sometimes you don't
- Framed field experiment is conducted in the non-standard subject's setting
  - go to the subject rather than bring the subject to you
  - subjects knows they are in an experiment

Natural field experiment
Harrison and List (2004)

- Randomization eliminates selection bias (whether subject is in treatment or control)
- But what if experiments participants are systematically different from non-participants?
- Or simply knowing they are in an experiment systematically affects their decisions?
  - ATE will be biased (despite randomization)
  - experimenter has less "control" than they think
- Natural field experiments eliminate participation decision
  - subjects are people already undertaking certain tasks
    - e.g. deciding whether to tip an Uber driver
  - experimenter manipulates something about the choice environment
    - people don't know they are in an experiment
- Under certain conditions produces best ATE estimate (Ubaydli and List, 2015)[4]

[4] https://www.nber.org/papers/w20877

RCTs in development economics

- 2019 Nobel Prize awarded for use of RCTs in development economics
  - recall RCT is a general method for causal inference
  - "RCT" now commonly associated with field experiments in developing countries
- Extension of field experiments where the researcher has a target population
  - e.g. recipients of some policy aimed at reducing poverty, increasing education, etc.
  - e.g. "which information intervention will lead to more contraception use in this particular population?"
  - e.g. "when is the optimal time to remind farmers to fertilize fields?"

So you want to run an experiment

First principles

- Your method should be driven by the research question
  - If you are running experiment it is assumed you can't answer the question with observational data
- So first things first, you need a research question
- Then you need to be sure that an experiment is the right way to address it
- Lab or field?
  - Does the subject pool matter? (or equivalently: does the *target population* matter?)
  - Does context matter?
  - How much money do you have?
    - field experiments tend to be more expensive than lab experiments (often stakes are a day's wage)
    - students are easier to recruit, sessions are easier to run

Theory

- Not all experiments are about testing theory
- But theory should motivate your research question
- It should also motivate your design
  - what is the subject's payoff function?
  - does the subject interact with other people?
  - if subjects are grouped, how big should the groups be?
  - what are the benchmarks produced by your payoff function/theoretical model?
  - how many rounds should the experiment be?
  - between or within designs?

(pre)Analysis

- Since you are designing the study, you are in effect "designing" the data
- This means you should already have an idea how to analyze it
  - e.g. average treatment effects
  - e.g. heterogeneous treatment effects
- Some journals use pre-registration to alleviate "publication bias"[5]

[5] https://www.nature.com/articles/s41562-016-0034

Implementation

- Test your design
  - pilot studies
  - cheaper to fix bugs than to throw out data and re-run sessions
  - this includes testing your instructions
  - are the parameter values and stakes appropriate?
- Get Internal Review Board (IRB)approval
  - time consuming
- Recruit subjects
  - lab experiment: usually straightforward (e.g. use software like ORSEE[6])
  - field experiment: challenging
    - how will you access your target population?
    - may need local government approval

[6] http://www.orsee.org/web/

Further reading

Further reading

- List, John A. “Why economists should conduct field experiments and 14 tips for pulling one off.” Journal of Economic Perspectives 25, no. 3 (2011): 3-16.
- Gneezy, Uri, and Alex Imas. “Lab in the field: Measuring preferences in the wild.” In Handbook of economic field experiments, vol. 1, pp. 439-464. North-Holland, 2017.
- Al-Ubaydli, Omar, and John A. List. “Field experiments in markets.” In Handbook of Economic Field Experiments, vol. 1, pp. 271-307. North-Holland, 2017.
- Kagel, John H., and Alvin E. Roth, eds. The handbook of experimental economics. Vol. 2. Princeton university press, 2016.
- Duflo, Esther, and Abhijit Banerjee, eds. Handbook of Field Experiments. Vol. 1. Elsevier, 2017.
- List, John A., Sally Sadoff, and Mathis Wagner. “So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design.” Experimental Economics 14, no. 4 (2011): 439.
- Levitt, Steven D., and John A. List. “Field experiments in economics: The past, the present, and the future.” European Economic Review 53, no. 1 (2009): 1-18.