Randomized Controlled Trials

Sepehr Ekbatani
(TeIAS)

Spring 2022

Section 1

RCT Review

Why Experiments?

- Suppose the question of interest is a causal question: What is the effect of D on Y? consider a binary treatment D.
  - Would individual's health improve if we provide health insurance?
  - Do unions raise wages?
  - Does nutrition increase productivity?
  - Do training programs raise wages?
- Causal questions need a counterfactual
  - Y^1 outcome for an individual if treated,
  - Y^0 outcome for the same individual if not treated.
  - The treatment effect for that individual is $Y^1 - Y^0$.

Causality Is Defined In An Average Sense

- In general, individuals differ in how much they gain from treatment, so that we can imagine a distribution of gains over the population with mean:

$$ATE = E[Y^1 - Y^0]$$

- ATE is a standard measure of the causal effect of treatment 1 relative to treatment 0.

Can OLS Estimate The Causal Effect?

- Suppose we estimate the following regression:

$$Y = \alpha + \beta D + \varepsilon$$

- β is a measure of the association between Y and D:

$$\begin{aligned}
\beta &= E[Y|D=1] - E[Y|D=0] \\
&= E[Y^1|D=1] - E(Y^0|D=0) \\
&= \underbrace{E(Y^1 - Y^0|D=1)}_{\text{ATT}} + \underbrace{E(Y^0|D=1) - E(Y^0|D=0)}_{\text{Selection Effect}}
\end{aligned}$$

- β is biased because D is correlated with the error term

Why/When Does OLS Fail?

1. Selection/Omitted variable bias (OVB):
    - If treatment status is the result of individual decisions, and those with low Y^0 choose treatment more frequently than those with high Y^0.
    - Ex: who pays for training? Those with most to gain, lowest HK at start. If HK not observed, D is correlated with error term.
2. Reverse causality (*endogeneity*):
    - If (Y, D) are observed simultaneously: how do we know D causes Y and not the other way around?
    - Ex: training take-up is a function of low wages.
3. Measurement error:
    - If random: attenuation
    - If not: like OVB
4. functional form:
    - OLS best ~~linear~~ approximation. Sometime not good enough.

Randomized Experiments

- Treatment status is randomly by construction $(Y^0, Y^1) \perp D$:
  - We rule out selection, simultaneity, OVB.

1. No selection

$$E(Y^0|D=1) = E(Y^0|D=0)$$

2. Average potential outcomes are identical across treated and controls

$$E(Y^1 - Y^0|D=1) = E(Y^1|D=1) - E(Y^0|D=0) = E(Y^1 - Y^0)$$

  - The implication is $ATE = ATT = \beta$.

- An unbiased estimate of ATE is the difference between the average outcomes for treatments and controls, which you can estimate by OLS.

Testing For Random Assignment

- Compare mean differences in all observables. Test whether means are statistically different across treated and comparison.
  - If D is randomly assigned, only 5% of the variables should show significant differences across the two groups.
- Regress D on covariates. Test the null hypothesis that all the covariates are jointly insignificant using F-test.
  - If use linear probability model, need to adjust errors for heteroskedasticity, or can estimate logit/probit.

Should Covariates Be “Controlled For” In RCT ?

1. Adding covariates increases efficiency (lowers standard errors), especially if these covariates do actually affect Y.
2. Provides a check for randomization. If treatment is truly random then D is orthogonal to X and inclusion of X should not affect the estimates of β.
3. An argument against: the point of randomized trials is their transparency. When covariates are added, the investigator again has discretionary choices to make (which variables to add? Keep on adding until small enough se?).

Advantages of Experiments

- High internal consistency
  - Obtain causal estimate for population under study
- Easy estimation
- Causality can be inferred
- Transparency
  - little researcher discretion, easy to understand (e.g. to policy makers), and in principle easy to replicate in other settings

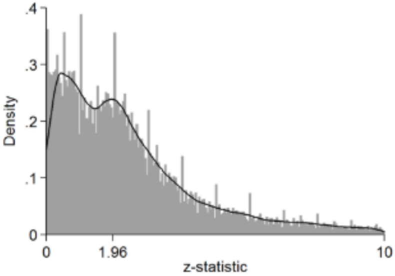
Why Are RCT's The "Gold Standard?"

- Other methods (propensity score matching, difference in differences, IV) require assumptions to identify causal effects that are not verifiable.
- Presumably, if a researcher undertakes the randomization, they know whether the treatment is randomized.
- Growing list of cases where randomized interventions and other methods produce different results.
    - Generally, presumed to mean that other methods produce biased estimates.
- Often compared to "scientific" random experiments in the medical literature.

Additional Benefits of RCTs

- Duflo , Glennerster , and Kremer argue that RCTs reduce "publication bias."
- What is publication bias?
  - 5% chance that a regression renders a significant result.
  - Researchers may search over specifications or identification strategies until they find significant results.
- Why would RCT's do better?
  - DGK argue that the "right" specification is less clear in non RCTs, so there is more room to specification search.
  - In RCTs, also possible to write a pre analysis plan specifying the right specification before researchers have contact with the data.
  - The cost of RCTs makes researchers more likely to disseminate null results.

Publication Bias: “Star Wars: The Empirics Strike Back”



(a) Raw distribution of z-statistics.

- Distribution of Z statistics in 50,000 tests published in the AER, JPE and QJE between 2005 and 2011 according to Brodeur et al.
- Argue that there is missing mass of p values to the left of 1.96 and extra mass at 1.96.

Some Important Issues With Experiments

- External validity
  - How representative are the results? Would they hold in other settings, or time periods? For other populations?
- General equilibrium effects
  - Partial equilibrium effects observed in study might not be the ones observed when program is implemented nation wide because at the national level there would be important GE effects (see externalities below).
  - Needs to have randomization at the “right” level.
- Hawthorne Effects
  - Does experiment itself affect outcomes?

Experimental Issues: Take-up

- We wish to estimate the effect of tutoring, x, on test scores, y.
  - Students are randomly offered tutoring sessions for free.
- Compliance: What if some individuals do not accept treatment?
  - Some of the students that are offered tutoring do not take it.
  - It is also possible that some students NOT offered tutoring get tutoring anyway. $\rightarrow$ The sample who receives tutoring is *self-selected*.
- A comparison of mean test scores, of those receiving tutoring, v, and those who didn't, provides a BIASED estimate of the ATE of tutoring, just like OLS.
- What can we learn from the effect of the *offer* (Intent to treat) on the outcome?

Experimental Issues: Attrition

- Not a problem if it is random.
- If treatment has different effects on participants, some treated individuals may be more likely to drop out in which case our naive mean comparisons will be biased.
  - Eg. In drug trials those with bad side effects may stop.
  - In this case we would over estimate benefits of drugs.
- Two solutions:
  - Estimate bounds
  - Model attrition
  - In either case we (generally) need to make additional assumptions to get meaningful estimates of the treatment effects.

Other Important Limitations

- Externalities in treatment
  - At extreme can find 0 effect if full externality. Need to think carefully about level of randomization.
- Causal estimates provide no evidence on mechanisms.
  - Why does aspirin lower probability of heart attacks? Need model.
  - Multiple mechanisms might be at play, including behavioral responses. Eg. Aspirin's physical effect but person might change other behaviors as result of aspirin consumption.
  - However experiments can be combined with theory.

Critiques of RCTs (Deaton)

- Not truly gold standards
  - Often implementation issues that mean randomization or program implementation are imperfect.
- May not identify parameters of interest
  - External validity and GE effects
- Detachment from Theory Decreases Generalizability
  - Without broader theory about human behavior, difficult to know how RCT results from one context or for one population apply to another.

Section 2

Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility

Petra E. Todd, Kenneth I. Wolpin (2006)

Policy Evaluation

- Choice of subsidy scheme is a critical design element of the program.
- Comparison of the treatment and control groups assesses the impact only of the particular subsidy scheme that was implemented.
- Ex-ante evaluation of alternative policies requires a theoretical framework.

Paper Overview

- Uses data from a randomized social experiment in Mexico called PROGRESA.
- Estimate and validate a dynamic behavioral model of parental decisions about *fertility* and *child schooling*.
- Evaluate the impact of different policy interventions to increase schooling.

Basic Model of School Attendance With Child Wages

- Household maximization problem:

$$\max_{s} U(c, s, \mu)$$
$$\text{s.t.} \;\; c = y + w(1 - s)$$

- In this example, the optimal choice $s^* = \phi(y, w, \mu)$, where μ denotes unobservable heterogeneity.

Introducing School Attendence Subsidy

- Consider the effects of a policy that provides a subsidy in the amount τ for school attendance, so the problem becomes:

$$\max_{s} U(c, s, \mu)$$
$$\text{s.t.} \;\; c = y + w(1-s) + \tau s$$

- The constraint of the model can be rewritten as:

$$c = (y+\tau) + (w-\tau)(1-s)$$

- Which shows that the optimal choice of s in the presence of the subsidy is $s^{**} = \phi(\tilde{y}, \tilde{w}, \mu)$ where $\tilde{y} = y + \tau$ and $\tilde{w} = w - \tau$.

Implication and Required Assumption

- Schooling choice of a family with income y and child wage w receiving the subsidy is the same as the schooling choice of a family with income $\tilde{y}$ and child wage $\tilde{w}$.
- Assumption
$$f(\mu|y, w) = f(\mu|\tilde{y}, \tilde{w}).$$
- Or
$$f(\mu|y, w, x) = f(\mu|\tilde{y}, \tilde{w}, x).$$

Intent-To-Treat Estimator

- ITT

$$\frac{1}{n}\sum_{j=1}^{n}\{E(s_i|w_i = w_j - \tau, y_i = y_j + \tau) - s_j(w_j, y_j)\}$$

- Where $s_j(w_j, y_j)$ denotes the school attendance decision for a child of family j with characteristics (w_j, y_j).

Coverage Rate and Treatment-on-The-Treated Estimator

- Coverage rates:

$$Pr(s(w-\tau, y+\tau) = 1) = E(s(w-\tau, y+\tau)).$$

- The relationship between ITT and TT:

$$\begin{aligned} ITT(w,y) = Pr&(\text{participates}|w,y)TT(w,y) \\ &+ Pr(\text{does not participate}|w,y)0. \end{aligned}$$

- Thus,

$$TT(w,y) = \frac{ITT(w,y)}{E(s(w-\tau, y+\tau))}$$

- Estimated by:

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\{E(s_i|w_i = w_j - \tau, y_i = y_j + \tau) - s_j(w_j, y_j)\}}{E(s_i|w_i = w_j - \tau, y_i = y_j + \tau)}.$$

Extension to Multiple Children, Exogenenous Fertility

- Suppose there are two children in the family who are eligible for subsidies τ^1 and τ^2, have wage offers w^1 and w^2

$$\max_{(s^1,s^2)} U(c, s^1, s^2)$$

$$\text{s.t. } c = (y + \tau^1 + \tau^2) + (w^1 - \tau^1)(1 - s^1) + (w^2 - \tau^2)(1 - s^2)$$

- Match families with income level y and child wages w^1 and w^2 to other families with income level $\tilde{y} = (y + \tau^1 + \tau^2)$ and child wage offers $\tilde{w^1} = w^1 - \tau^1$ and $\tilde{w^2} = w^2 - \tau^2$.

Multiple Children, Endogenous Fertility

- n denotes the number of children.
- s^i the schooling decision for child i.
- Assume the potential earnings and subsidy level for each child is the same.
- Parents get utility over the number of children and their children's schooling levels.
- Parents decide on the number of children and on schooling decisions, and both decisions are potentially affected by the subsidy level.

$$\max_{(n, s^1 \ldots s^n)} U(c, n, s^1, \ldots, s^n)$$

$$\text{s.t. } \; c = (y + n\tau) + (w - \tau) \sum_{i=1}^{n} (1 - s^i),$$

where w is the per child potential wage and τ is the subsidy.

Model Prediction - No Subsidy

- In the absence of subsidy, expected fertility for a family with income y facing wage w is:

$$\sum_{j=1}^{J} jPr(n = j|\tilde{y} = y, \tilde{w} = w)$$

where j indexes the range of potential number of children.

- The expected schooling level can be written as:

$$\sum_{j=1}^{J} Pr(s = 1|\tilde{y} = y, \tilde{w} = w, \tilde{n} = j)Pr(\tilde{n} = j|\tilde{y} = y, \tilde{w} = w)$$

Model Prediction - With Subsidy τ

- With the subsidy, the expected number of children is:

$$\sum_{j=1}^{J} jPr(n = j | \tilde{y} = y + j\tau, \tilde{w} = w - \tau)$$

- The expected schooling level is:

$$\sum_{j=1}^{J} Pr(s = 1 | \tilde{y} = y + j\tau, \tilde{w} = w - \tau, \tilde{n} = j) Pr(\tilde{n} = j | \tilde{y} = y + j\tau, \tilde{w} = w - \tau)$$

Model Prediction - Program Effect

- Noting that:

$$Pr(s = 1|\tilde{y} = y + j\tau, \tilde{w} = w - \tau, \tilde{n} = j) = E(I(\tilde{s} = 1)|\tilde{y} = y + j\tau, \tilde{w} = w - \tau, \tilde{n} = j)$$

$$Pr(n = j|\tilde{y} = y + j\tau, \tilde{w} = w - \tau) = E(I(\tilde{n} = j)|\tilde{y} = y + j\tau, \tilde{w} = w - \tau)$$

- The program effect can be calculated by taking the difference, replacing the probabilities with their corresponding estimators.

Extension to Two Period Model

$$\max_{(c_1,c_2,s_1,s_2)} U(c_1, c_2, s_1, s_2)$$

$$\text{s.t. } \; c_1 + c_2 \leq y_1 + y_2 + w_1(1 - s_1) + w_2(1 - s_2).$$

- The schooling choices in each period can be written as functions

$$s_1 = \phi_1(\tilde{y}, w_1, w_2)$$

$$s_2 = \phi_2(\tilde{y}, w_1, w_2)$$

where $\tilde{y} = y_1 + y_2$.

Extension to Two Period Model With Subsidy

- With the subsidy, the constraint becomes

$$\begin{aligned} c_1 + c_2 &= y_1 + y_2 + w_1(1 - s_1) + w_2(1 - s_2) + \tau_1 s_1 + \tau_2 s_2 \\ &= (y_1 + \tau_1 + y_2 + \tau_2) + (w_1 - \tau_1)(1 - s_1) + (w_2 - \tau_2)(1 - s_2). \end{aligned}$$

- So that the optimal schooling choices are

$$s_1^* = \phi_1(\tilde{y}, \tilde{w_1}, \tilde{w_2})$$

$$s_2^* = \phi_2(\tilde{y}, \tilde{w_1}, \tilde{w_2})$$

Description of PROGRESA

- Large scale anti-poverty program
  - Begun in 1997.
  - Originally provided aid to 10 million poor families (40% of all rural households).
  - Operates in 31 states with a budget $\sim$ 1 billion USD.
  - Recent expansion into urban areas.
- Provides educational grants to parents (mothers) to encourage children's school attendance.
  - Must attend 85% of days.
- Subsidies amounted to about 25 percent of average annual income over all children that actually attended in the first year of the program.

Benefit Levels

- Benefit levels increase with grade level, higher for girls.

School level	Grade	Monthly payment in pesos	
		Females	Males
Primary	3	70	70
	4	80	80
	5	105	105
	6	135	135
Secondary	1	210	200
	2	235	210
	3	255	225

Figure 1: Monthly Transfers Under PROGRESA

Experimental Design and Data

- Program implemented as a randomized social experiment
- 506 villages randomly selected from 7 states in Mexico
    - 320 randomly assigned to the treatment group and 186 to the control group
- Controls incorporated after third year of the program, but not told about the program until incorporated
- Use Oct. 1997 baseline and Oct. 1998 follow-up surveys.
- Data elements:
    - School attendance and grade attainment, information on employment and wages (to construct total family income net of child income)
    - Village level data in the minimum wage paid to daily laborers
- Subsample:
    - Children from the program eligible families, age 12 to 15 in 1998, who are the son or daughter of the household head, and for whom information is available in the 1997 and 1998 surveys.

Model Validation

- Estimate the model using households that **did not** receive the subsidy.
- Evaluate the performance of the model in forecasting the behavior of the **treated** households.

- Out-of-sample validation
  - Compares the actual post-program school attendance rates in treated households to the rates predicted by the model based on simulating the introduction of the subsidy schedule.
  - Predict what the treatment effect would have been for the nontreated households and compare it to the experimental estimate based on the treated households.

In-Sample Validity

Table 9—Actual and Predicted Choice Distribution by Child Age and Sex
(*Pooled* 1997 *and* 1998)

	Actual			Predicted			
Age	School	Work	Home	School	Work	Home	χ^2
Boys							
6	0.933	—	0.066	0.923	—	0.077	0.58
7	0.981	—	0.019	0.980	—	0.020	0.02
8	0.987	—	0.013	0.980	—	0.020	0.99
9	0.994	—	0.006	0.979	—	0.021	3.49
10	0.982	—	0.018	0.974	—	0.026	0.86
11	0.977	—	0.023	0.964	—	0.036	1.45
12	0.885	0.021	0.094	0.846	0.039	0.115	3.99
13	0.780	0.084	0.136	0.736	0.078	0.186	4.51
14	0.677	0.157	0.166	0.619	0.191	0.190	3.41
15	0.490	0.276	0.235	0.520	0.251	0.229	0.88
Girls							
6	0.965	—	0.035	0.942	—	0.058	3.84
7	0.976	—	0.024	0.968	—	0.032	0.77
8	0.989	—	0.011	0.976	—	0.024	1.96
9	0.991	—	0.009	0.975	—	0.025	3.26
10	0.979	—	0.021	0.970	—	0.030	0.93
11	0.969	—	0.031	0.948	—	0.052	2.97
12	0.896	0.007	0.097	0.854	0.020	0.126	4.61
13	0.726	0.028	0.245	0.676	0.025	0.299	2.85
14	0.582	0.089	0.329	0.566	0.092	0.342	0.22
15	0.419	0.123	0.458	0.402	0.157	0.442	1.68

Note: χ^2 (0.05, 1) = 3.84, χ^2 (0.05, 2) = 5.99.

- As children finish primary school, the model captures and *overstates* the drop.

Out-of-Sample Validity

Table 12—Actual and Predicted School Attendance Rates by Child Age, Sex, and School Attainment: Control and Treatment Groups by Year[a]

	Girls				Boys			
	Control group		Treatment group		Control group		Treatment group	
	1997	1998	1997	1998	1997	1998	1997	1998
Age 6–11								
Actual	96.9	96.5	97.6	98.5	96.6	96.7	97.6	98.7
Predicted	96.1	96.2	96.4	97.1	96.4	96.4	96.3	97.1
No. obs.	449	451	632	600	471	460	671	678
Age 12–15								
Actual	65.3	66.5	62.9	74.4	68.8	72.5	69.5	76.3
Predicted	61.6	61.8	61.8	74.9	68.8	68.8	68.0	77.1
No. obs.	190	176	[illegible]	[illegible]	189	182	279	262
Age 12–15 behind in school								
Actual	58.3	58.7	56.9	71.4	64.0	67.4	64.2	71.6
Predicted	54.2	55.5	55.6	72.3	63.9	65.3	62.7	72.9
No. obs.	127	121	144	161	139	135	204	190
Age 13–15 HGC ≥ 6 behind in school								
Actual	40.9	44.4	30.3	51.5	59.0	57.1	52.6	58.3
Predicted	40.2	45.3	37.3	58.7	55.0	53.0	51.7	66.7
No. obs.	66	72	66	66	61	56	95	96

Within-Sample

Out-of-Sample

[a] Based on 200 simulation draws per family.

Long-Term Impact on Fertility

- Fertility outcomes are essentially invariant to the subsidies.

TABLE 18—PREDICTED EFFECT OF SUBSIDY ON COMPLETED FERTILITY: ALL CHILDREN EVER BORN

	No subsidy	Subsidy
Mean number of children ever born	4.24	4.28
Percent of families with		
Zero children	0.05	0.04
One child	1.16	1.12
Two children	9.23	8.75
Three children	22.97	22.49
Four children	24.43	24.64
Five children	21.54	21.55
Six children	14.78	15.23
Seven children	5.05	5.32

Figure 2: Predicted Effect of Subsidy on Fertility

Counterfactual Policy Experiments

TABLE 19—THE EFFECTIVENESS AND COST OF ALTERNATIVE PROGRAMS

	Baseline[a]	Compulsory school attendance through age 15	Original subsidy	2× subsidy	0.5× subsidy	Restricted subsidy[b]	1.43× restricted subsidy
Mean completed schooling							
Girls	6.29	8.37	6.83	7.30	6.56	6.67	6.97
Boys	6.42	8.29	6.96	7.44	6.68	6.79	7.07
Percent completed grade 6 or more							
Girls	75.8	95.1	82.3	86.9	79.3	77.4	82.0
Boys	78.8	93.7	83.3	86.7	81.1	79.6	82.8
Percent completed grade 9 or more							
Girls	19.8	55.5	25.9	31.6	23.1	26.2	29.3
Boys	22.8	54.7	28.0	34.6	25.5	29.2	31.8
Cost per family	0	—	26,096	59,935	11,989	15,755	25,193
Mean number of children	4.24	4.21	4.28	4.32	4.27	4.25	4.27

	Bonus for completing 9th grade[c]	Junior secondary school in each village	Unconditional income transfer 5,000 pesos/yr	No child labor through age 15	Original subsidy and 25% wage increase
Mean completed schooling					
Girls	6.50	6.39	6.41	6.30	6.75
Boys	6.58	6.55	6.53	6.52	6.79
Percent completed grade 6 or more					
Girls	74.9	76.0	77.6	76.1	81.5
Boys	76.9	79.0	80.0	79.9	81.8
Percent completed grade 9 or more					
Girls	28.8	21.2	20.8	19.7	25.2
Boys	32.7	24.1	23.7	23.5	26.5
Cost per family	36,976	—	237,000	—	25,250
Mean number of children	4.20	4.24	4.24	4.25	4.29

[a] Predicted: control and treatment families.
[b] Subsidy for attending school in grades 6–9 only.
[c] The bonus is set at 30,000 pesos for girls and boys.

Caveat?

- Their analysis is partial equilibrium.
- As school attendance rates rise due to the program,
  $\rightarrow$ children withdraw from the child labor market,
  $\rightarrow$ child wage rates to rise,
  $\rightarrow$ increase in school attendance rates to be somewhat mitigated.
- They performed a counterfactual experiment which combines the original subsidy program with a concomitant increase in child wage rates of 25 percent.