Workshop on Experiments in Political Economy 18-20 May 2011

Columbia Center for the Study of Development Strategies & the Harriman Institute

Contents

Preface

In May 2011 the Center for the Study of Development Strategies and the Harriman Institute at Columbia University hosted a short training workshop on political economy experiments. The goal was to provide full immersion training on causal inference, and the design, implementation, and analysis of experiments. This booklet contains compiled notes from the workshop lectures.

The lectures were put together by a talented group of researchers working right at the cutting edge of this literature: Bernd Beber, Don Green, Guy Grossman, Becky Morton, Eric Mvukieye, Laura Paler, Cyrus Samii, Alex Scacco, Rocio Titiunik. Our thanks to them. Thanks too to Kate Baldwin, Chris Blattman, Raul Sanchez de la Sierra, and Peter van der Windt, Maria Amelina, Scott Gelbach, Omar Ponce, Marion Dumas, Aleksei Belyanin, Yanilda Gonzalez, Ana Bracic, Noah Buckley, Jorge Gallego, Pavithra Suryanarayan, and David Szakonyi, for leading other parts of our discussions during the workshop, to David Szakonyi for putting this handbook together, to Caroline Peters for running the whole show, to Grant Gordon and Lauren Young for preparing background material, and to Raul Sanchez de la Sierra for the cover photograph. More information can be found at: `http://cu-csds.org/events/2011-experiments-conference`. Special thanks to the group from the Center for the Study of Institutions and Development in the Higher School of Economics in Moscow who supported the workshop and provided spirited discussion throughout: best of luck with your experiments ahead!

Macartan Humphreys and Timothy Frye

May 2011

1 Causal Inference I: The Fundamental Problem of Causal Inference

Bernd Beber

1.1 Key Ideas

- Potential outcomes framework (Rubin model):
  - Denote units as $i = 1, 2, 3, \ldots, n$
  - Observed outcome Y_i
  - Pretreatment covariates X_i
  - Treatment for unit i, D_i
  - Focus on binary case, where $D_i = 1$ if treated and $D_i = 0$ otherwise
  - Potential outcomes $Y_i(D_i = 1) = Y_i(1)$ and $Y_i(D_i = 0) = Y_i(0)$
  - Causal effect $Y_i(1) - Y_i(0)$
- Note, implicitly in our notation we have made the "Stable Unit Treatment Value Assumption" (SUTVA), that is, potential outcomes of unit i depend only on the treatment of unit i:

  $$Y_i(d_1, d_2, d_3, \ldots, d_n) = Y_i(d_i)$$

  where d_i is a realization of D_i

1.2 The Problem

The fundamental problem of causal inference is that we just observe one of the two potential outcomes for each individual. We observe either $Y(0)$ or $Y(1)$.

1.3 The Solution

If potential outcomes are independent of treatment assignment, then:

$$E(Y_i(1)) = E(Y_i(1)|D_i = 1), \text{ and}$$
$$E(Y_i(0)) = E(Y_i(0)|D_i = 0)$$

Then we can identify *average* causal effects

$$\begin{aligned} E(Y_i(1) - Y_i(0)) &= E(Y_i(1)) - E(Y_i(0)) \\ &= E(Y_i(1)|D_i = 1) - E(Y_i(0)|D_i = 0) \end{aligned}$$

This means that even if units differ in any number of ways, with independence, the average treatment effect is, in expectation, just the difference between the (expected) average outcome among treated units and control units.

2 Causal Inference II: Randomization and Alternatives

Bernd Beber

2.1 Recap

- Basic setup of randomized experiments:
  - A sample of units $i = 1, 2, \ldots, n$
  - Could be a random or a convenience sample
  - Number of units assigned to treatment is n_1, remainder $n_0 = n - n_1$ assigned to control
  - Completely randomized treatment assignment $D_i \in \{0, 1\}$, so $\Pr(D_i = 1) = \frac{n_1}{n}$
  - Outcome $Y_i(D_i)$
  - Treatment randomization ensures the following assumption is plausible:

$$(Y_i(1), Y_i(0)) \perp D_i$$

- Estimand τ is the average treatment effect (ATE)
- Random treatment assignment ensures that difference in means is unbiased estimator $\hat{\tau}$ of τ:

$$\hat{\tau} = \frac{\sum_{i=1}^{n} D_i Y_i}{n_1} - \frac{\sum_{i=1}^{n} (1 - D_i) Y_i}{n_0}$$

2.2 Variance of Estimator

- Variance of estimator in sample:

$$var(\hat{\tau}_{SATE}) \leq \frac{s^2_{Y_i(0)}}{n_0} + \frac{s^2_{Y_i(1)}}{n_1}$$

where s^2 denotes the sample variance of its subscript

- Variance of estimator in population:

$$var(\hat{\tau}_{PATE}) = \frac{\sigma^2_{Y_i(0)}}{n_0} + \frac{\sigma^2_{Y_i(1)}}{n_1}$$

where σ^2 is the relevant population variance, with the sample variance as unbiased estimator

- Confidence intervals: Typically given by $\hat{\tau} \pm \sqrt{var(\hat{\tau})} \cdot Q(1 - \frac{1-\alpha}{2}, \nu)$, where Q is the quantile function for the t-distribution with ν degrees of freedom and significance level α. For large ν and $\alpha = .95$, the confidence interval is approximately

$$\hat{\tau} \pm 1.96 \cdot \sqrt{var(\hat{\tau})}$$

- Optimal treatment allocation: $n_0^* = \frac{n}{1+\frac{\sigma_{Y_i(1)}}{\sigma_{Y_i(0)}}}$ and hence $n_1^* = \frac{n}{1+\frac{\sigma_{Y_i(0)}}{\sigma_{Y_i(1)}}}$

2.3 Principles

- Ensuring independence in experiments:
  - Randomization
  - Well-specified control condition(s)
  - Double-blind procedures
  - Parallelism in procedures, staff, and timing
  - Block to improve efficiency

2.4 Regression discontinuity (RD)

- Locate arbitrary cut point c for treatment assignment such that $D_i = 1$ for $X_i \geq c$ and $D_i = 0$ otherwise (for sharp design)
- Assume $E(Y_i(D_i)|X_i)$ continuous around c for $D_i = 1, 0$

- Estimand is

$$\begin{aligned} &E(Y_i(1) - Y_i(0)|X_i = c) \\ =&E(Y_i(1)|X_i = c) - E(Y_i(0)|X_i = c) \end{aligned}$$

- Causal effect in RD design at discontinuity:

$$\begin{aligned} &\lim_{X_i \downarrow c} E(Y_i(1)|X_i = c) - \lim_{X_i \uparrow c} E(Y_i(0)|X_i = c) \\ =&\lim_{X_i \downarrow c} E(Y_i|X_i = c) - \lim_{X_i \uparrow c} E(Y_i|X_i = c) \end{aligned}$$

2.5 Instrumental variables:

- Suppose we estimate $Y = \alpha + X\beta + \varepsilon$, but suspect $cov(X, \varepsilon) \neq 0$
- For some other variable Z, we have $cov(Y, Z) = cov(X, Z)\beta + cov(\varepsilon, Z)$
- If $cov(\varepsilon, Z) = 0$ and $cov(X, Z) \neq 0$, we have $\beta = \frac{cov(Y,Z)}{cov(X,Z)}$
- Two conditions: Instrument has to meet **exlusion restriction** and has to be **strong**

- Encouragement in the potential outcomes framework:
  - Randomized encouragement $Z_i \in \{0, 1\}$
  - Potential treatment indicators $(D_i(Z_i = 1), D_i(Z_i = 0))$
  - Observed and potential outcomes $Y_i = Y_i(Z_i, D_i(Z_i)) = Y_i(Z_i)$
  - Since encouragement is randomized, we have

$$(Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp Z_i$$

  - Distinguish four latent types (principal strata):

    1. $(D_i(1), D_i(0)) = (1, 0)$ (Complier)
    2. $(D_i(1), D_i(0)) = (1, 1)$ (Non-complier, always-taker)
    3. $(D_i(1), D_i(0)) = (0, 0)$ (Non-complier, never-taker)

4. $(D_i(1), D_i(0)) = (0, 1)$ (Non-complier, defier)

- Observed strata:

	$D_i = 1$	$D_i = 0$
$Z_i = 1$	Complier or always-taker	Defier or always-taker
$Z_i = 0$	Defier or never-taker	Complier or never-taker

- Required assumptions:
  1. **Monotonicity:** $D_i(1) \geq D_i(0)$, i.e. no defiers
  2. **Exclusion restriction:** $Y_i(1, t) = Y_i(0, t)$ for $t = 0, 1$, i.e. encouragement affects outcome only through treatment
- We can then write

$$\begin{aligned} & E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0) \\ & = E(Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = 1) \Pr(D_i(1) - D_i(0) = 1), \text{ and} \\ & E(D_i|Z_i = 1) - E(D_i|Z_i = 0) \\ & = E(D_i(1)) - E(D_i(0)) = \Pr(D_i(1) - D_i(0) = 1) \end{aligned}$$

- It follows that

$$\begin{aligned} E(Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = 1) &= \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)} \\ &= \frac{cov(Y_i, Z_i)}{cov(D_i, Z_i)} \end{aligned}$$

- Exclusion restriction implies no ITT effect for always-takers or never-takers!
- Instrumental variable approach yields *complier* average treatment effect
- Different instrument (encouragement) will yield different compliers and estimate, so this is a *local* effect (LATE)

3 Analysis of Experimental Data I: Estimands and Estimators

Cyrus Samii

3.1 Potential outcomes in large population

We imagine a large population with potential outcomes under treatment and control. We want to estimate the average treatment effect. We suppose as well that there is a pre-treatment covariate. Some hallmarks of the setting are (i) heterogenous treatment effects, (ii) irregular functional relationships between covariates and potential outcomes, (iii) heteroskedasticity. Finite means and variances for the potential outcomes are presumed. The PATE is defined as the difference in mean potential outcomes for this population. Refer to Fig. 1.

3.2 Simple random sample

We do not have access to the full population, and so some kind of sampling is necessary. A simple random sample is the most basic. The SATE is defined as the difference in mean potential outcomes for the sample. The SATE equals the PATE in expectation if all units in the population have equal probability of selection. However, for any given sample, the SATE will not equal the PATE exactly. Refer to Fig. 2.

3.3 Simple randomized experiment

We cannot observe the full schedule of potential outcomes. Rather, the potential outcome corresponding to treatment assignment is revealed in an experiment. A simple randomized experiment assigns a fixed number of units to treatment without replacement, with the remaining sample members assigned to control. So long as assignment probabilities are uniform, we can estimate the SATE without bias via the simple difference in mean observed outcomes conditional on treatment. However, for any given experiment this estimate will not equal the SATE exactly. The presumption here is that at minimum, we want to estimate SATE, and presuming a sample from a well-defined population, we want to get back to PATE. Refer to Fig. 3.

3.4 Inference: $S\hat{AT}E$

Let's formalize the setting. Suppose a simple randomized experiment with M of N sampled units assigned to treatment, $D_i \in \{0,1\}$.. Index such that $i = 1, ..., M$ are treated, and $M+1, ...N$ are control. For each unit the experiment yields, $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$. Then,

$$S\hat{AT}E = \frac{1}{M}\sum_{i=1}^{M} Y_i - \frac{1}{N-M}\sum_{i=M+1}^{N} Y_i,$$

is unbiased for SATE, with variance defined by randomization distribution.

Conditioning on the sample, $\mathcal{S}$, by elementary sampling theory,

$$\begin{aligned}\mathrm{V}\left(S\hat{AT}E|\mathcal{S}\right) = & \frac{N}{N-1}\left[\frac{\mathrm{V}\left(Y(1)|\mathcal{S}\right)}{M} + \frac{\mathrm{V}\left(Y(0)|\mathcal{S}\right)}{N-M}\right] \\ & + \frac{1}{N-1}\left[2\mathrm{Cov}\left(Y(1), Y(0)|\mathcal{S}\right)\right. \\ & \left. -\mathrm{V}\left(Y(1)|\mathcal{S}\right) - \mathrm{V}\left(Y(0)|\mathcal{S}\right)\right].\end{aligned}$$

- Use $s^2(\{Y_i\}_{i=1}^{M})$ and $s^2(\{Y_i\}_{i=M+1}^{N})$ for the first part.
- $\mathrm{Cov}\left(Y(1), Y(0)\right)$ cannot be estimated without auxiliary info.
- Conservative ("Neyman") estimator ignores second part.
- Conservative estimator equals the HC2 heteroskedasticity robust estimator.
- In large fixed samples, $S\hat{AT}E$ is approximately normal with variance typically a bit smaller than conservative estimator.

Refer to Fig. 4.

3.5 Inference: from $S\hat{AT}E$ to PATE

Presuming an equal-probability-selection-mechanism (epsem) sample from the population, $S\hat{AT}E$ is unbiased for PATE, and,

$$\mathrm{V}\,(S\hat{AT}E) = \frac{\mathrm{V}\,(Y(1))}{M} + \frac{\mathrm{V}\,(Y(0))}{N-M}.$$

- A consistent variance estimator is equivalent to the conservative estimator given above (without the finite sample correction).
- For large samples, $S\hat{AT}E$ is approximately normal with variance equal to the above.
- Thus, regression on constant & treatment dummy, with "robust" s.e.'s, provides either slightly conservative or unbiased inference (confidence intervals) for simple experiments on moderately-sized or large samples.

3.6 Covariance adjustment

Covariance adjustment estimates, $\bar{Y}(1)_{adj} - \bar{Y}(0)_{adj}$, where

$$\begin{aligned}\bar{Y}(1)_{adj} &= \bar{Y}(1)_{treated} + \beta^1_{adj}(\bar{X} - \bar{X}_{treated})\\ \bar{Y}(0)_{adj} &= \bar{Y}(0)_{control} + \beta^0_{adj}(\bar{X} - \bar{X}_{control}),\end{aligned}$$

β^1_{adj} and β^0_{adj} are OLS coefficients, and $\beta^1_{adj} = \beta^0_{adj}$ with simple adjustment.

- No presumption that the regression is "correct."
- Biased but consistent for SATE.
- Interacted reg. improves asymptotic precision (AP) relative to difference-in-means for PATE. Simple adjustment improves (AP) if experiment is not strongly imbalanced ($min[M/N, (N-M)/N] > .25$) and $\mathrm{Cov}\,(D_i, Y_i)$ is sufficiently large relative to $\mathrm{Cov}\,(D_i, Y_i(1) - Y_i(0))$.
- Robust se's are consistent, as above.
- Results carry through to multiple regression.

Refer to Fig. 5.

3.7 Block randomization

- Regression in previous slides motivated by efficiency. It may be used to address "incidental confounds," with consistency following from usual regression assumptions.
- A design-based approach to addressing covariate imbalance is block randomization.
- Observations are divided into blocks, $b = 1, .., B$, typically by prognostic covariates.
- By principles of stratified sampling, an unbiased estimator for SATE is, $S\hat{A}TE_{block} = \sum_b (N_b/N) S\hat{A}TE_b$.
- $S\hat{A}TE_{block}$ has lower variance if outcome variation is reduced within strata.
- $S\hat{A}TE_{block}$ is algebraically equivalent to coefficient from regression with block FEs and inverse propensity score weights. Robust se's are consistent, as above.

Refer to Fig. 6.

3.8 Are permutation tests an alternative?

- The mode of randomization inference presented here is known as the "Neyman" approach.
- An alternative, associated with Fisher, examines the distribution of test statistics under the randomization-based permutation distribution of the treatment variable.
- The Fisher approach is very robust for testing the "sharp null" hypothesis, $H_0 : Y_i(1) = Y_i(0)$, that implies no causal effect.
- However, using this method to produce confidence intervals can mislead. E.g., if the test statistic is $S\hat{A}TE$ under the null hypothesis of constant effects, then the resulting confidence interval is exactly $(N - 1)/N$ times that which would be produced from a regression assuming homoskedastic variance. We know that is wrong.
- The take-away is that with moderate to large samples, the Neyman approach is robust and the Fisher approach provides not apparent benefits. For small samples (less than 30 total units?), we may have to content ourselves with Fisher style sharp null hypothesis tests.

3.9 Wrinkles

Binary outcomes All above results apply. Logit odds ratios are biased, though differences in predicted probabilities are consistent.

Non-compliance Randomization also identifies unbiased estimates for "treatment received" on "complier" subpopulation – a type of LATE. Estimation via ITT/(Compliance Rate), or, equivalently, TSLS.

Cluster randomization Finite sample bias if (and only if) cluster sizes are variable. Also, inferences must attend to within-group dependence (e.g. with RE-GLS, or OLS with cluster-robust se's). Area of active research.

Missing data Undermines randomization. More later.

Interference & spill-over Designs and estimates require model of "indirect exposure." More later.

Figures

Figure 1:

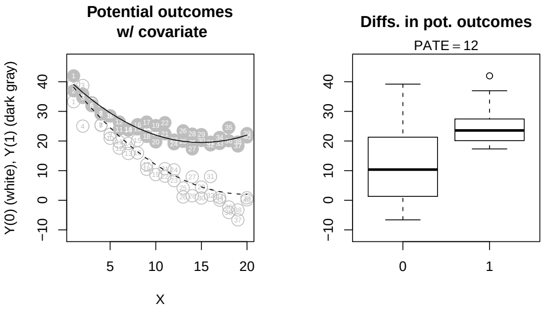


Figure 2:

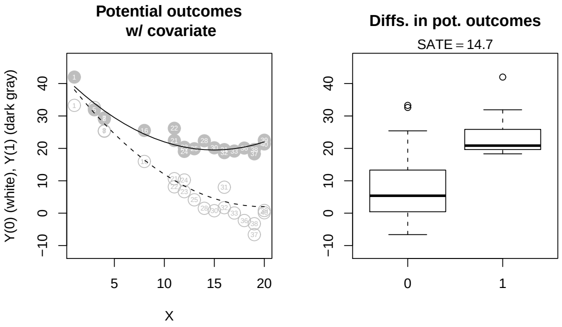


Figure 3:

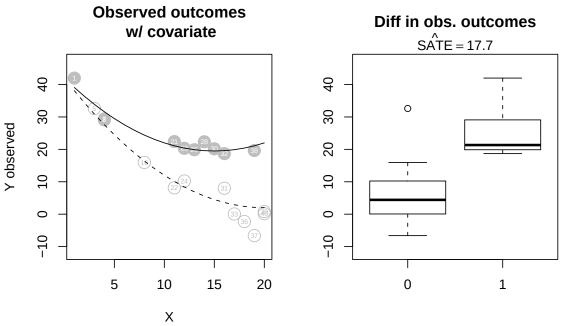


Figure 4:

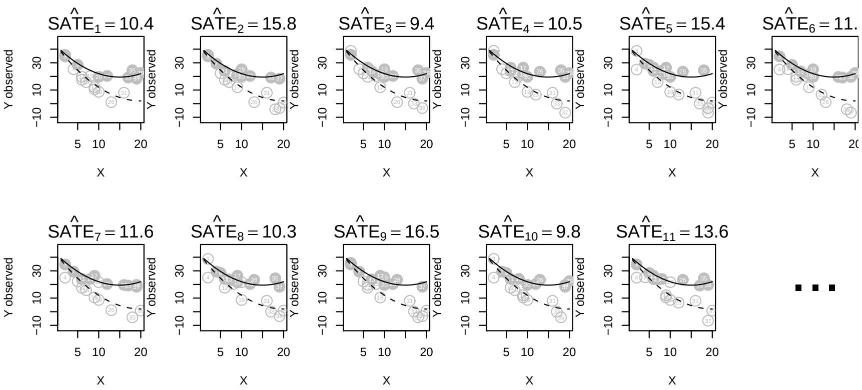


Figure 5:

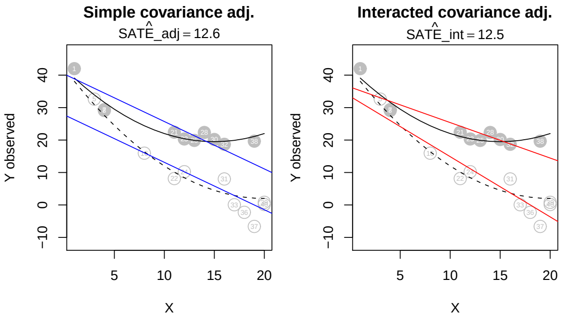
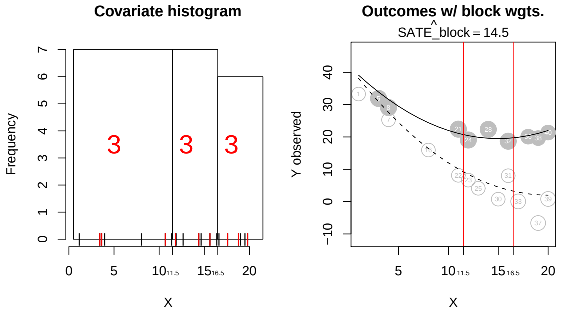


Figure 6:



4 Analysis of Experimental Data II: Randomization Inference

Rocío Titiunik

Here we describe how you can use the randomization procedure to generate hypotheses tests without having to appeal to any assumptions whatsoever.

4.1 Potential Outcomes Framework

We can summarize the potential outcomes framework as follows:

- $D_i = 1$ or $D_i = 0$: binary treatment assignment
- $Y_i(1)$: potential outcome under treatment
- $Y_i(0)$: potential outcome under control
- $Y_i = Y_i(0)(1 - D_i) + Y_i(1)D_i$: observed outcome
- Fundamental Problem of Causal Inference: we see either $Y_i(0)$ or $Y_i(1)$, but never both simultaneously for the same i

- The effect of the treatment on unit i is

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

We know that we cannot know the treatment effect for any particular individual, τ_i, but random assignment of treatment offers a way to learn about, for example, the average treatment effect. This is because, under randomization, *potential outcomes and treatment assignment are independent.*

It turns out that when potential outcomes and treatment assignment are independent, statistical inferences can be done with very little (and sometimes no) assumptions. This is, randomization does not only give us identification, it also gives us the basis to make statistical inferences that do not require assuming that a particular statistical model is true and they don't even require assuming that large-sample approximations hold. This was first shown by Fisher.

Let $\mathbf{D}$ be an n-dimensional column vector whose elements are the D_i for all units. When we randomize a treatment, we determine the value of $\mathbf{D}$ using a random mechanism that is by definition known because it is the one *we* used to randomize.

This takes us to a few distinctions. In the model we are discussing, $\mathbf{D}$ is the *only* random variable. On the contrary, (Y_0, Y_1, X) are fixed (where X is used to refer to pre-treatment covariates). The observed outcome $Y = DY_1 + (1 - D)Y_0$ is also a random variable, but it is only a random variable *because* D is a random variable. The only randomness in the model is coming from $\mathbf{D}$ and *nowhere* else. This observation is crucial, and it is the basis of randomization inference.

4.2 The general approach to testing the sharp null

Imagine that we want to test the hypothesis that the treatment is without effect. In statistical inference, this is referred to as a *null* hypothesis. Note that the null hypothesis is *not* an assumption. Rather, we know that when the null hypothesis is true, certain conditions follow. If those conditions are not plausible in our data, then we have evidence against the null hypothesis.

The so-called "sharp null" is the hypothesis that the treatment has no effect for all units:

$$H_0 : Y_i(1) = Y_i(0) \text{ for } i = 1, 2, \ldots n$$

Under the sharp null H_0, the observed outcome $\mathbf{Y}$ is no longer a random variable, it is fixed, because $Y_i = Y_i(1) = Y_i(0)$ for every i. Now we have to define a test-statistic in order to test our null hypothesis. A test-statistic is a function of the data (in our case, the observed outcomes $\mathbf{Y}$) and the treatment assignment ($\mathbf{D}$). We write it as $t(\mathbf{D}, \mathbf{Y})$.

When we have an experiment, someone assigns $\mathbf{D}$ randomly. And as a consequence, the randomization mechanism that assigned units to treatment and control is *known*. The insight of Fisher was to realize that under the sharp null, the only source of randomness is the assignment of $\mathbf{D}$, and because this assignment is entirely known, the distribution of the random variable $\mathbf{D}$ is known, and we can use it in turn to derive the distribution *under the null* of *any* test-statistic $t(\mathbf{D}, \mathbf{Y})$. This is exactly what we did in the examples in class.

Now, to get a little more formal, we imagine that all possible realizations of $\mathbf{D}$ are collected in the set Ω (the only restriction that we place is that the assignment mechanism must give a positive probability of receiving treatment and control to unit every unit, this is, it must be the case that $0 < \Pr(D_i = 1) < 1$ for all i).

There a lot of different random mechanisms that can be used to assign $\mathbf{D}$. One mechanism that *does not* fix the number of treatments and controls is a mechanism that assigns treatment according to $\Pr(D_i = 1) = p$ where p is between 0 and 1. Note that in this case the event where *all units* are assigned to the same group has a positive probability, although small if n is large. When $p = 1/2$, we have $\Pr(\mathbf{D} = \mathbf{d}) = (1/2)^n$ for all $\mathbf{d} \in \Omega$, and $|\Omega| = 2^n$ (where we use $|A|$ to denote the number of elements in the set A).

In the exercise we saw in class, the random mechanism that determined $\mathbf{D}$ was different, because in that mechanism the number of treatments and the number of controls was*fixed*. This is usually referred to as an experiment with "fixed margins" or, as we will see in my second class on Thursday, as a *random allocation rule*. In this case, we have $|\Omega| = \binom{n}{n_t} = \frac{n!}{n_t!(n-n_t)!}$. When all elements of Ω that assign n_t units to treatment and $n - n_c$ units to control are equally likely, we have $\Pr(\mathbf{D} = \mathbf{d}) = \frac{1}{\binom{n}{n_t}}$ for all $\mathbf{d} \in \Omega$.

So, given all this, how do we calculate the p-value or significance level of our test of the sharp null hypothesis? Well, since under H_0 the only randomness is coming from $\mathbf{D}$, and we know its distribution, we only need to calculate the observed value of our test-statistic of choice, which we call T, and then we calculate the probability of observing a value of $t(\mathbf{D}, \mathbf{Y})$ equal to or greater than T. Remember, T is just the value of the test-statistic in our data, using the values $\mathbf{d}$ and $\mathbf{y}$ that were realized. This p-value is therefore calculated as follows:

$$\Pr(t(\mathbf{D}, \mathbf{Y}) \geq T) = \sum_{\mathbf{d} \in \Omega} \mathbf{1}\left\{t(\mathbf{d}, \mathbf{Y}) \geq T\right\} \cdot \Pr(\mathbf{D} = \mathbf{d}) \tag{1}$$

4.3 Extending the framework to interval and point estimation

As we saw, under the sharp null,the observed outcomes are not random variables, because they are fixed for all values of $\mathbf{D}$, which is the only source of randomness in our model. As we also saw, we can always test the sharp null hypothesis, and to do so we need no assumptions of any kind. But many times we will want to do more than just testing this null hypothesis, we will want to construct confidence intervals and estimate the treatment effect. To do so, even if we have random assignment, we need extra assumptions, namely, we need to assume a certain model about how the treatment affects the outcomes.

We assume SUTVA (the observed outcome of every unit depends solely on the treatment status of that unit, and not on the treatment status of the rest of the units) and that the treatment effect is constant and additive:

$$Y_i(1) = Y_i(0) + \tau \quad \text{for every } i = 1, 2, \cdots, n \tag{2}$$

Why do we need these extra assumptions? Well, if the treatment has an effect, then the vector of observed responses $\mathbf{Y}$ is no longer a constant for every value of $\mathbf{d} \in \Omega$. On the contrary, in this case $\mathbf{Y}(\mathbf{d})$ varies with every $\mathbf{d} \in \Omega$. Since $\mathbf{D}$ is a random variable, the observed outcomes are also random variables because they depend on $\mathbf{D}$. In principle, there could be a different pattern of outcomes $\mathbf{Y}(\mathbf{d})$ for every value of $\mathbf{d} \in \Omega$. But it is hard for us to think about a treatment that has so many different possible effects, that is why we write down a simplified model.

As we saw in class, under this model, we can achieve interval and point estimation using the idea of adjusted responses.

5 Troubleshooting I: Causal inference with missing data: Nonparametric and semi-parametric approaches

Cyrus Samii

Missing data can undermine the gains from randomization. Here we describe some solutions.

5.1 Missing data undermines randomization

We suppose the following setting:

- Sample of N units from a large population.
- $1 < M < N-1$ assigned to treatment ($D_i = 1$), remaining to control ($D_i = 0$).
- Potential outcomes, (Y_{1j}, Y_{0j}).
- Response potential, (R_{0j}, R_{1j}) where $R_{0j}, R_{1j} = 1$ if outcome observed, 0 otherwise.
- We observe $R_i = D_i R_{1i} + (1 - D_i) R_{0i}$ for everyone, but only observe Y_i for units with $R_i = 1$

By total probability and randomization, the PATE can be decomposed as,

$$\begin{aligned} PATE =& [\overbrace{\mathrm{E}\,(Y_i|D_i=1,R_{1i}=1)}^{A}\,\Pr(R_{1i}=1) - \overbrace{\mathrm{E}\,(Y_i|D_i=0,R_{0i}=1)}^{B}\,\Pr(R_{0i}=1)] \\ &+ [\underbrace{\mathrm{E}\,(Y_i|D_i=1,R_{1i}=0)}_{C}\,\Pr(R_{1i}=0) - \underbrace{\mathrm{E}\,(Y_i|D_i=0,R_{0i}=0)}_{D}\,\Pr(R_{0i}=0)] \end{aligned}$$

- C and D are unidentified in observed data.

- If $A \neq C$ or $B \neq D$, analysis of the complete data will be biased. Equal under missingness completely at random (MCAR).
- The size of the bias depends on the degree of inequality and the missingness rates.
- One idea is to attempt a second round of data collection to fill in at least part of C and D ("double sampling").

Refer to Figures 7 and 8.

5.2 Non-parametric approaches: Worst case bounds

Suppose that the potential outcomes are bounded such that $\Pr(y_t^L \leq Y_{ti} \leq y_t^H) = 1$ for $t = 0, 1$. Define,

$$\beta^L = [\mu_{1,obs} \Pr(R_{1i} = 1) - \mu_{0,obs} \Pr(R_{0i} = 1)] + [y_1^L \Pr(R_{1i} = 0) - y_0^H \Pr(R_{0i} = 0)]$$
$$\beta^H = [\mu_{1,obs} \Pr(R_{1i} = 1) - \mu_{0,obs} \Pr(R_{0i} = 1)] + [y_1^H \Pr(R_{1i} = 0) - y_0^L \Pr(R_{0i} = 0)]$$

where $\mu_{t,obs} = \mathrm{E}\left(Y_i | D_i = t, R_{it} = 1\right)$.

- It must be that $\beta^L \leq PATE \leq \beta^H$. Thus, $[\beta^L, \beta^H]$ are "worst case" bounds on the treatment effect.
- Width depends on depends on $|y_1^L - y_0^H|$ and missingness rates.
- Provide a "zone of consensus."
- Additional assumptions can narrow the bounds, e.g., if $X_i \perp\!\!\!\perp (Y_{1i}, Y_{0i}) | R_i$, then bounds can be estimated at values of X_i with lowest missingness rate.

Refer to Figure 9.

5.3 Non-parametric approaches: Trimming bounds

Suppose "monotonicity", $\Pr(R_{1i} = 0, R_{0i} = 1) = 0$. Treatment never induces missingness.

- Observed control units are representative sample of units with $(R_{1i} = 1, R_{0i} = 1)$.

- Observed treated units are a mixture of units with $(R_{1i} = 1, R_{0i} = 1)$ and $(R_{1i} = 1, R_{0i} = 0)$.

- $\mathrm{E}\,(Y_{1i}|R_{1i} = 1, R_{0i} = 1)$ must be less than the mean of the Y_{1i} distribution that excludes the lowest $100 * \Pr(R_{1i} = 1, R_{0i} = 0|D_i = 1, R_i = 1)\%$ of values for the $D_i = 1, R_i = 1$ units. Symmetric logic gives lower bound.

- By monotonicity and random assignment, $\Pr(R_{1i} = 1, R_{0i} = 0|D_i = 1, R_i = 1)$ is identified.

- By monotonicity, equal missingness rates implies that observed data unbiased for treatment effect among $(R_{1i} = 1, R_{0i} = 1)$ units.

Refer to Figure 10.

5.4 Statistical inference with bounds

- Common approach is to presume that sample is drawn from a large population. Then, the usual frequentist inference is valid (e.g., via bootstrap).

- Bayesian constructions allow for inferences conditional on the sample. E.g., if one assumes missing data are parameters, and then assigns a prior with mass one that they equal the boundary values, then the variance of the posterior for the difference-in-means is given by the randomization distribution.

5.5 Semi-parametric approaches: IPW & imputation

- Key assumption is that dataset contains all info necessary to account for dependence between missingness and potential outcomes. Formally, $Y_i(t) \perp\!\!\!\perp R_i(t)|(D_i, W_i)$ for $t = 0, 1$, where W_i refers to other variables in dataset. Known as "missing at random" (MAR).

- The conditioning data, W_i may include pre-treatment covariates and, under the assumption of no confounding due to unobserved potential post-treatment outcomes, post-treatment covariates.

- Violations of MAR can be explored through sensitivity analysis, although judging the results of sensitivity analysis is quite subjective.

Refer to Figure 11

- If one is willing to invoke MAR, then primary concern becomes, how to use data most efficiently with fewest assumptions? Semi-parametric estimators do this.

- The most efficient methods combine IPW and imputation – known as "augmented IPW" estimators. Suppose $Z_i = (1 \quad D_i)'$ and $\hat{\beta} = (Y\hat{\hat{(}}0) \quad \hat{\tau})'$. Then an efficient AIPW estimator solves the following with respect to $\hat{\beta}$,

  $$\sum_{i=1}^{N} \frac{R_i}{\Pr(R_i = 1|D_i, W_i)} Z_i(Y_i - \hat{\beta}' Z_i)$$
  $$+ \left[1 - \frac{R_i}{\Pr(R_i = 1|D_i, W_i)}\right] Z_i(Y_i^* - \hat{\beta}' Z_i) = 0,$$

  where Y_i^* are imputed Y_i values.

- The methods are agnostic about how one obtains the estimates of the missingness rates and imputations. You can use your favorite method (from OLS to Bayesian Additive Regression Trees).

- As above, for inference the presumption that the sample is drawn from a large population justifies frequentist inference, e.g. via the bootstrap or sandwich estimators.

- Bayesian alternatives are conceivable, although they have not been developed as far as I know.

5.6 Conclusion

- All of the methods shown above can also be adapted to missing treatments or covariates.

- Nonetheless, our tour of methods for missing data should provoke some worry. The worst case bounds were obscenely wide, and the alternative methods relied on strong untestable assumptions.

- This should inspire effort to minimize missingness and to consider “double sample” methods when possible.
- Nonetheless, you need a plan for missingness. An ex ante analysis plan should specify what are the primary analyses, and what supplemental analyses will be done to examine implications of missingness.

Figures

Figure 7: No missing data

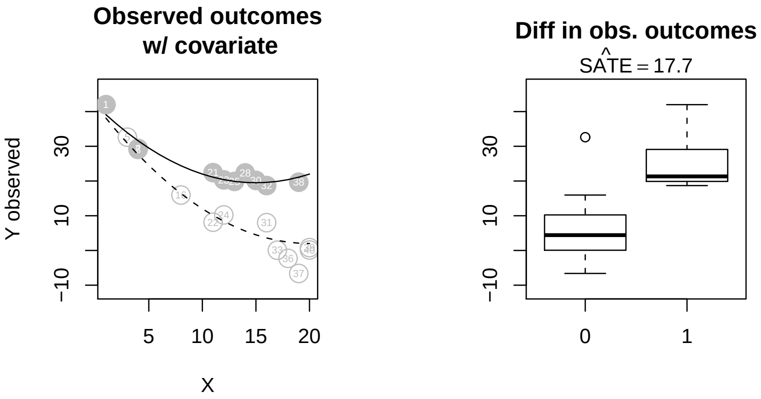


Figure 8: With missing data

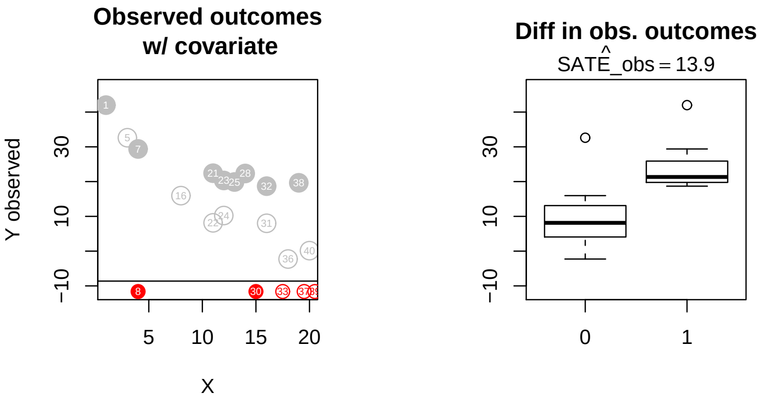


Figure 9: Worst case bounds

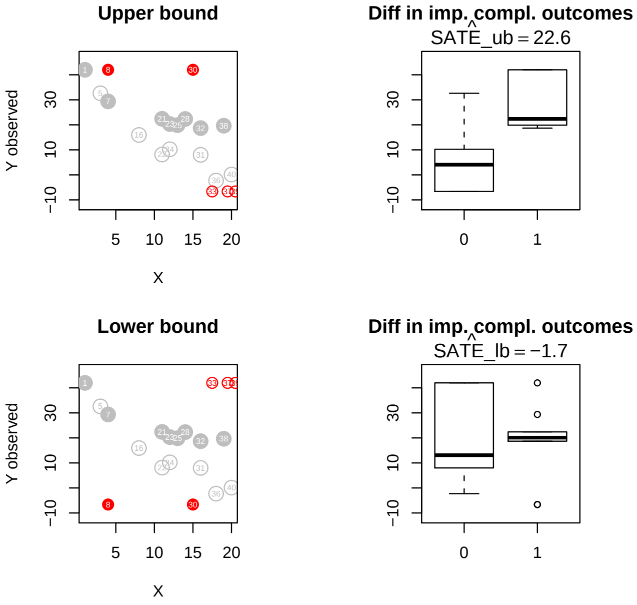


Figure 10: Trimming bounds

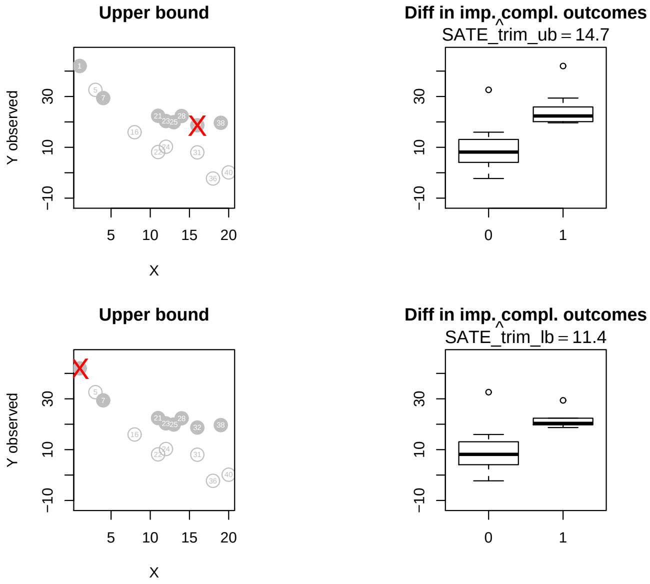
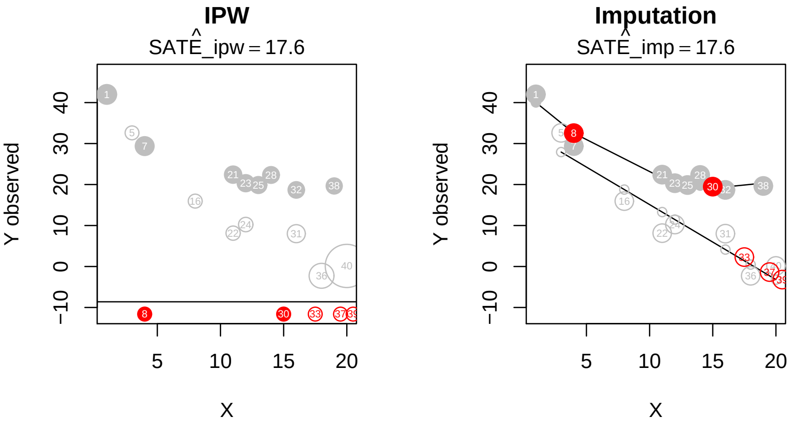


Figure 11: IPW and imputation



6 Troubleshooting II: Interference between Experimental Units

Donald Green

Note: Summary prepared by Lauren Young and based on forthcoming book by Donald Green and Alan Gerber

Subject: SUTVA: Stable Unit Treatment Value Assumption (Rubin 1990), or Individualistic Treatment Response (Manski 2008). Here we describe the spillover problem and how to address it.

6.1 Introduction

The non-interference (SUTVA) assumption requires that potential outcomes for subject i respond only to the subject's own treatment status, D_i, such that $Y_i(D_i) = Y_i(D)$. Or, in the case of noncompliance with the vector of assignment Z_i and actual treatment D_i, the assumption is $Y_i(Z_i, D_i) = Y_i(Z, D)$.

Examples of violations in social sciences: contagion, displacement, communication, social comparison, signaling, and memory.

When designing experiments, think about potential outcomes models that go beyond $Y_i(1)$ if i is treated and $Y_i(0)$ if not. Spillovers require more complex potential outcomes models.

6.2 Identifying Causal Effects in the Presence of Spillover

Example: multi-level design for turnout encouragement intervention. Four potential outcomes for each voter: neither voter nor housemate Y_{00}, only housemate Y_{01}, only voter Y_{10}, and both Y_{11}. Interesting causal effects: $Y_{10} - Y_{00}$ is effect of personal target conditional on housemate receiving nothing, $Y_{11} - Y_{01}$ is effect of personal target conditional on housemate also targeted. Spillover effects: $Y_{01} - Y_{00}$ is spillover effect on those who receive no mail, and $Y_{11} - Y_{10}$ is spillover effect on those who receive mail.

Still requires non-interference assumption that potential outcomes are unaffected by treatments administered to those outside their own household! No unmodeled spillovers.

Error terms are still correlated across observations, which violates the assumption of independent errors. However this does not bias estimators and can be corrected.

Direction of bias: incorrectly ignoring spillover effects would lead one to this equation: $\frac{\overline{Y}_{11} - \overline{Y}_{10}}{2} - \frac{\overline{Y}_{01} - \overline{Y}_{00}}{2} = \hat{Y}_{10} - \hat{Y}_{00}$. When there are no spillovers (plus randomization), this is an unbiased ATE. But if there are spillovers, the ATE will be biased upward if $\overline{Y}_{11} - \overline{Y}_{10} > \overline{Y}_{01} - \overline{Y}_{00}$, or downward if the inequality is flipped.[1]
Multilevel designs shed light on spillovers by varying the degree of first- and second-hand exposure to the treatment.

6.3 Mixing Non-Interference with Non-Compliance

Non-compliance forces us to grapple with the problem of heterogeneous treatment effects among different latent types of people: Compliers, Never-Takers, Always-Takers, and Defiers. In the presence of spillover effects, there is an even larger schedule of potential outcomes to allow for direct and spillover effects among each latent type. This requires additional assumptions to identify causal effects.

Now the expected voting rate in each group (meaning households with no treatment $\hat{Y}_{00}$, subject treated $\hat{Y}_{10}$, housemate treated $\hat{Y}_{01}$, and both treated $\hat{Y}_{11}$) is the

[1] I believe that the hats and overlines are slightly off on page 8 of the chapter draft, and have changed this inequality from $\overline{Y}_{11} - \overline{Y}_{10} > \hat{Y}_{01} - \hat{Y}_{00}$, as it is in the draft.

average of the four compliance groups' base rates weighted by the share of all voters belonging to that group. For subject compliers we add the term t_D and t'_D if they are in a two-complier household. For housemate compliers, we add in a term t_S and for two compliers, t'_S. Then an interaction term for direct and spillover effects $\triangle$. This leaves us with five treatment parameters and four equations, so we need additional assumptions. If we assume that $t_S = t'_S$ and $t_D = t'_D$, then we are able to pin down the direct and indirect treatment effects, as well as the interaction term. This assumption suggests that individuals in two-complier households are not very different from individuals in one-complier households.

Placebo design enable direct comparison of voters in two-complier households where both receive a placebo call and voters where both receive a treatment call. With high non-compliance, the placebo design may generate more precise estimates.

6.4 Spatial Spillover

Spillovers are not always confined to neat units: sometimes you're interested in proximity to treated locations. You must develop a metric for proximity, and regress outcomes on treatment and the proximity metric. This is more difficult than it seems!

- Developing a measure for proximity presupposes a model of how spillovers are transmitted. Euclidean distance to treatment, density of treatments in an area, distance in travel time, etc.
- Flexible models (where the data determine the rate of decay) generate imprecise estimates. Getting the right model is critical.

6.5 Beware of Naive Spatial Regression

Proximity to treated locations is NOT random just because treatment assignment is random. Proximity to treated locations is random only in the conditional sense: for those subjects that share the same spatial orientation to other subjects, proximity is random. Thus, if you estimate spillover effects without fully accounting for blocking on spatial orientation you will be prone to bias.

Two methods to deal with spatial orientation:

- Match observations according to their proximity to all potentially treated units and estimate treatment effects within matching strata.
- Measure the spatial arrangement of observations using metrics such as average distance to all other observations or the number of other observations within a certain radius. Then these measures are included as covariates in the regression. This is riskier because it introduces parametric assumptions (dummies do not).

Both methods depend heavily on assumed models of proximity.

6.6 Making Use of Nonexperimental Units

"Failure to investigate the effects of spillovers on non-experimental units is like leaving money on the table." Discovery of spillovers (or no spillovers) inspires follow-up substantive research and justifies modeling assumptions for future studies.
Must still block on spatial orientation..

6.7 Within-Subjects Design and Time-Series Experiments

Within-subjects experimentation: studies where a single person or entity is tracked over time and random assignment determines when a treatment is administered.

- Pluses: subjects are compared with themselves to hold background characteristics constant, and conditions are typically controlled.

- Minuses: depend on substantive assumptions that are difficult to justify in social science. Thus, they are rarely used outside of psychology.

Parallel in social sciences are intervention studies, or interrupted time-series. But these do not even randomize the time of intervention and thus might just pick up regression to the mean, or the intervention's onset being driven by some third factor.

In the causal framework, we would like to estimate $Y_{\underline{1}0} - Y_{\underline{0}0}$, where the first number in the subscript is the treatment assignment in the first period and the second subscript is treatment assignment in the second period, and the underline is the period of observations. Yet what we really observe is $Y_{\underline{1}0} - Y_{1\underline{0}}$. The same logic applies to the second period. Thus, we need additional assumptions to equate these two quantities.

- No anticipation. A potential outcome in a given period is unaffected by treatments administered in subsequent periods. $Y_{\underline{0}0} = Y_{\underline{0}1}$.
- No period effects. The response to a series of untreated states is constant over time. $Y_{\underline{0}0} = Y_{0\underline{0}}$.
- No persistence. Potential outcomes in one period are unaffected by treatments in prior periods. $Y_{1\underline{0}} = Y_{0\underline{0}}$.

Bottom line: despite random assignment, within-subjects designs depend on supplementary non-interference-type assumptions that researchers must justify.

6.8 Summary

The non-interference assumption requires researchers to specify how potential outcomes respond to all possible random assignments and how treatment effects will be designed. This chapter relaxed the assumption that subject i is unaffected when others are treated by randomly assigning varying degrees of second-hand exposure. These designs invoke modeling assumptions about the way that indirect effects travel.

As we increasing complexity of models, we risk introducing bias. Studies of spillover effects are often de facto blocked experiments, and failure to control for the (correct!) strata may result in severely biased estimates. But there are huge substantive and methodological gains in estimating spillovers!

7 Design I: Ways of Randomizing

Rocío Titiunik

7.1 Overview

We can summarize the potential outcomes framework as follows:

- $D_i = 1$ or $D_i = 0$: binary treatment assignment
- $Y_i(1)$: potential outcome under treatment
- $Y_i(0)$: potential outcome under control
- $Y_i = Y_i(0)(1 - D_i) + Y_i(1)D_i$: observed outcome
- Fundamental Problem of Causal Inference: we see either $Y_i(0)$ or $Y_i(1)$, but never both simultaneously for the same i
- The effect of the treatment on unit i is

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

As we know, randomization provides a solution to this fundamental problem of causal inference, or missing data problem. In this lecture, we will explore different ways in which we can assign treatments randomly or, in other words, *how* to randomize.

7.2 Restricted versus Unrestricted Randomization

We begin by distinguishing two broad types of randomization procedures:

- Simple or unrestricted randomization
- Restricted randomization

7.3 Unrestricted randomization

Unrestricted or simple randomization has two main cases. In the first case, referred to as *Random Allocation Rule*, the total sample size n and the sample sizes in each group (treatment, control) are fixed and under the control of the researcher. A subset of $n/2$ out of n is randomly chosen and assigned to treatment, the remainder to control. Thus, this randomization is appropriate when the researcher knows the final sample size before the experiment begins.

In the second case, referred to as *Complete Randomization*, the final sample size is not known with certainty, although a target sample size is usually established. In this case, the randomization procedure is analogous to tossing of fair coin multiple times, one per each subject in the study. The sample sizes in each group are binomially distributed random variables. This is the most commonly used procedure in clinical trials.

How do these two cases compare to each other? In both cases, the marginal probability of assignment is 1/2 for all assignments, although in the random allocation rule the conditional probability of assignment at the j^{th} step given prior assignments is not always 1/2. In complete randomization, there is a chance of having imbalanced sample sizes in both groups, but this becomes extremely unlikely for $n \geq 200$. Since statistical power is maximized for equal sample sizes in both groups, random allocation rule might be preferable for experiments with small sample size.

There is a greater likelihood of covariate imbalance with the random allocation rule than with complete randomization, but the difference is trivial for large sample sizes. Furthermore, in an unmasked study with staggered patient entry (common in clinical trials), there is substantial potential for selection bias with random allocation rule, not with complete randomization, because which treatment condition will be assigned to the next patient may be predictable.

In both cases, the conditional permutational variance for common tests is usually asymptotically equivalent to the population model variance, so that special analyses are not required to consider the permutational distribution of statistical tests.

7.4 Restricted randomization

With unrestricted randomization, during the recruitment process there is a chance of undesirable differences in the number of subjects assigned to each group. If baseline characteristics of subjects change over time, this periodic imbalance in sample sizes may result in significant differences between treatment and control groups in the distribution of pre-treatment characteristics. These are important concerns in small trials.

In contrast, restricted randomization procedures control the probability of obtaining an allocation sequence with severely imbalanced sample sizes in the treatment and control groups. This is, restricted randomization ensures that the sample sizes are fixed (indeed, some people call the random allocation rule procedure a restricted randomization with a single block). The most common form of restricted randomization is *blocked randomization* or simply *blocking.*

Blocked randomization refers to a randomization procedure that forces (periodic) balance in the number of subjects assigned to each treatment group, and its most common form is the *permuted-block design*, in which the size of each block is decided, and the treatment is randomly assigned within each block. Blocked randomization is usually combined with *stratification* to increase statistical power of treatment-control comparisons. A stratified blocked randomization divides experimental units into homogeneous strata (blocks), and then randomly assigns the treatment within blocks.

Performing a stratified blocked randomization involves incorporating *pre-treatment covariates* in the design of the experiment. These are covariates that both predict potential outcomes (ideally) and are determined *before* treatment is assigned (always). In randomized experiments, covariates are not really needed to identify treatment effects, but by blocking on important covariates in the design stage we can improve power in the analysis stage.

Stratified blocked randomization can provide considerable benefits when blocks are defined on the basis of a pre-treatment covariate that is strongly correlated with the outcome of interest. The general procedure is as follows:

- Divide total observations into homogeneous blocks
- Each block has observations with same or similar values of pre-treatment covariates
- Randomly assign treatment to observations *within* blocks
- If we have J blocks, we run J "mini-experiments"

Blocking on the basis of an important pre-treatment covariate has several advantages. It eliminates the randomness in outcomes that occurs when covariates are imbalanced, since imbalance in covariates used for blocking are ruled out by design. It also eliminates correlation between blocked covariates and treatment assignment, which introduces a collinearity penalty when using regression if these covariates are correlated with the treatment. It increases the efficiency of estimation and the power of hypothesis tests, and *forces the researcher to think about which covariates should be included in the analysis before randomization occurs.* Moreover, unless the sample size is very small, blocking on a covariate that fails to predict experimental outcomes does not invalidate inferences.

7.5 Factorial Designs

In this lecture, we also saw the basic definition of factorial experimental designs. This type of designs is appropriate when we want to learn about the effect of two or more treatments on some response. In this case, we can either run a separate experiment for each treatment, or we can use a factorial design in which both treatments are studied in the same experiment. The different treatments that we vary randomly are referred to as "factors" and the possible values a treatment may take are referred to as "levels" of the factor.

A complete factorial design refers to an experiment in which two or more treatments are under study, each treatment takes two or more levels, and all combinations of treatments and levels are represented.

The main advantages of using factorial designs over one-factor-at-a-time designs are that with factorial designs we have greater precision for estimating overall or main effects of factors (because all groups contribute to the analysis of all of the factors), and we also have the possibility of exploring interaction effects between the factors (which is by construction unavailable in the one-factor-at-a-time approach).

Although factorial designs offer advantages, we must be careful not to use too complicated of a design. In particular, note that for n factors that have two levels each, the number of treatment groups to be compared is 2^n, a number that gets too large too quickly. When there are too many groups, we will need a very large sample size to have a reasonable sample size in each and every cell. To deal with this issue, fractional factorial designs are available, which omit one ore more possible combinations to make the design feasible.

8 Design II: Formal Theory and Experimental Design

Lecture by Rebecca Morton

Note: Summary prepared by Grant Gordon and based on Chapter 6 of Rebecca Morton and Kenneth Williams, *Experimental Political Science and the Study of Causality, From Nature to the Lab*

8.1 Introduction

In this chapter, Morton and Williams discuss how the Formal Theory Approach (FTA) can be used to guide study design, data collection and analysis. According to the authors, a formal model is:

- Formal Model: A set of precise abstract assumptions or axioms about the Data Generating Process (DGP) presented in symbolic terms that are solved to derive predictions about the DGP.

This synopsis will discuss the role of formal models in experiments, the key elements of a successful formal model, as well as the choices in experimental design that reflect the formal model and the analysis stage of empirical work.

8.2 The Formal Theory Approach and Formal Modeling

The Formal Theory Approach embraces formal models as a means of deriving hypotheses that are motivated from explicit and theoretically consistent assumptions. For comparison, a nonformal model is:

- Nonformal Model: A set of verbal statements or predictions about the DGP

> that involve idealization, identification, and approximation but are given in terms of real observables or unobservable rather than symbols or abstracts.

A formal model, however, allows the researcher to directly derive predictions from explicit assumptions or axioms about the data generating process rather than idealization, identification, and approximation. A formal model makes explicit the assumptions about the axioms guiding individual or collective behavior as well as the DGP that underlines the phenomenon explored in the research.

8.3 The Formal Theory Approach vs. the Rubin Causal Model

Comparing the FTA to the Rubin Causal Model (RCM) illuminates the way in which assumptions are made theoretically consistent and explicit. The RCM, which is often used to establish causality in the social science, makes assumptions about the form of the data specifically, Ignorability and Stable Unit Treatment Value Assumption which often go untested in experimental work. Formal models are therefore more likely to meet the Theoretical Consistency Assumption, which is defined as:

- Theoretical Consistency Assumption: Assumptions made by researchers working with an RCM approach to causality that the assumptions underlying the causal predictions evaluated are consistent with the assumptions that underlie the methods used by the researcher to infer causal relationships.

Theoretical consistency underlines the connection between the empirical analysis and the formal model.

8.4 The 5 Components of a Formal Model

There are five primary components that must be specified to craft a fully inclusive and successful formal model. These components and assumptions motivate experimental design and empirical tests.

1. The political environment must be defined. Political environments include the institutions, political actors and information available to the actors.

2. Primitives, or assumptions must be described. Preferences of actors and institutional characteristics are examples of primitives.

3. Variables exogenous to actors and the political environment must be modeled. These include the constraints on actors' behavior outside the environment specified as well as other variables that alter behavior.

4. Decision variables, time horizons, and the means by which actors make their choices must be specified.

5. An equilibrium concept must be explicitly stated. These often take the form of game-theoretic solution concepts.

8.5 Predictions and Tests Derived from a Formal Model

From the formal model derived, predictions should be made, and theory tests, stress tests and comparative static studied. A formal model provides transparent derivations of:

- Point predictions: Predictions from a formal model about the values of the variables in the model when in equilibrium
- Relationship predictions: Predictions from a formal model about how two variables in the model are related.

Both point and relationship predictions are quantities of interest that map from the formal model to the data collected in order to test the hypotheses. Theoretical tests harness these predictions can be defined as:

- Theoretical Tests: When a researcher investigates the predictions of a formal model while attempting to make all the assumptions underlying the empirical study as close as possible to the theoretical assumptions.

This includes taking steps to ensure that subjects are 'primed' into specified preferences, that the experimental environment mirrors the theoretical environment, as well as any other assumption explicitly made. Stress tests are also of use and are defined as:

- Stress Tests: When a researcher investigates the predictions of a formal model while explicitly allowing for one or more assumptions underlying the empirical study to be at variance with the theoretical assumptions.

By relaxing or examining a certain assumption, the variable associated with that assumption can be tested. Additionally, comparative statics facilitate further analysis of how modeling assumptions and variables impact outcomes and can be defined as follows:

- Comparative Statics: Causal relationship predictions from formal model in which researchers compare how one variable takes divergent value given changes in another variable, holding time constant.

8.6 Designing an Experiment to Reflect Modeling Assumptions

Designing an experiment to recreate the components of the formal model often requires unique strategies in order to recreate the political environment, identify and recruit appropriate subjects, and frame the experiment. In recreating the political environment, either the lab or field can be used. The lab can be easier because the researcher has more control over possible confounding variables, however, a number of FTA experiments have been conducted in the field by economists (not by political scientists) and in internet and lab in the field experiments.

In these experiments, recruited subjects should be assigned roles that reflect the actors in the models. Subjects are generally motivated into certain preference profiles as assumed by the theory using financial incentives or other motivation techniques. However, subjects' choices are generally unconstrained and their behavior is the focus of the experimental study.

There are many complicated issues to address in such experiments. For example, how can one capture a game that has an indefinite end? Research has shown that one way to capture such a situation is to use a randomization process (such as tossing a coin or die) that determines when a game will end.

8.7 Experiment Design Continued, Framing

In designing an experiment, researchers should take care to ensure that exogenous variables can be manipulated, that random assignment of subjects to roles occurs, and that within- and between-subject comparisons can be made.

- In between-subject design: subjects make choices in only one state of the world
- In within-subject design: subjects in an experiment make choices in multiple states of the world.

In both approaches, it is critical to make sure that all subjects fully recognize the form of the game or experiment being tested when the goal is to evaluate the theory. An assumption of the equilibrium a formal model tests are that individuals in the game recognize and fully understand the game. As such, fully informing subjects of the game and making it recognizable is crucial to a successful experiment. This can be done using rigorous protocols and framing techniques. In an experiment run by Chou et al (2009) testing a game-theoretic equilibrium, subjects responded differently to various frames and hints offered by the authors. Priming a subject doesn't hardwire the responses that researchers are searching for; a wealth of experiments that document subject behavior varies from predictions even when fully prompted.

8.8 Experiment Design Continued, Repetition

In addition to using framing to communicate the game-theoretic nature of the model to subjects, repetition can be used.

- Repetition facilitates learning on behalf of the subject that therefore allows the subject to update their strategies to an equilibrium that the researcher can then observe.

Various randomization procedures can be used to avoid learning outside of the context of the game, however care should be taken as the number of independent observations decreases in this format.

However, sometimes we are interested in subjects' behavior in a new situation, in which repetition does not occur.

8.9 Empirical Analysis from the FTA

Once data from the experiment is collected, researchers should analyze the data not assuming theoretical consistency, but examining explicitly the degree to which the theoretical assumptions are consistent with statistical analysis.

In doing so, attention might be paid to strategic errors' or the non-equilibria choices strategically played by a subject given his or her assumptions about the other subject's likelihood of erring. This type of analysis follows the ways in which the Formal Theory Approach offers researchers a means of deriving, testing, and explicitly confronting the theoretical assumptions analyzed in experimental work.

9 Design III: Validity

Lecture by Rebecca Morton

Note: Summary prepared by Grant Gordon and based on Chapters 7 and 8, Rebecca Morton and Kenneth Williams, *Experimental Political Science and the Study of Causality, From Nature to the Lab*

Validity and Experimental Manipulations (Chapter 7)

9.1 Introduction to Validity

Ideal research design provides internally valid results, or results that are true for the population being studied, and externally valid results, or results that can be generalized to populations beyond those in the study. Specifically, internal and external validity can be defined as follow:

- **Internal Validity**: The approximate truth of the inference of knowledge claim within a target population studied.
- **External Validity**. The approximate truth of the inference or knowledge claim for observations beyond the target population studied.

While the dichotomy of internal and external validity is often adopted in the social sciences and is helpful in understanding experimental research, this chapter explores more precise types of validity that touch on construct, causal, statistical, and ecological validity.

9.2 Types of Internal Validity in Experiments

Internal validity, which facilitates inference on the target population studied, can be divided into construct, causal, and statistical validity. Construct validity is defined as:

- **Construct Validity**: Whether the inferences from the data are valid for the theory (or constructs) the researcher is evaluating in a theory testing experiment.

In other words, do the experiment and the measures accurately and precisely correspond to the theory being tested with the data? To ensure construct validity, steps should be taken to recreate the political environment, preferences, and assumptions embedded within the theory driving the experiment. Causal validity is defined as:

- **Causal Validity**: Whether the relationships the research finds within the target population analyzed are causal.

Experiments facilitate the identification of causal relationships when the appropriate design is implemented and the relevant counterfactual identified. Statistical validity is defined as:

- **Statistical Validity**: Whether there is a statistically significant covariance between the variables the researcher is interested in and whether the relationship is sizeable.

To ensure statistical validity, estimates must be efficient, accurate, significant, and sizable. This touches on both the need to produce accurate estimates and estimates that are clinically meaningful. Statistical replication, or replications of the estimation procedures on a different sample from the same target population, can be used to statistically validate research.

9.3 Types of External Validity in Experiments

External validity, which allows a researcher to generalize the results of a study to populations beyond those examined in a study, is often confused with Ecological Validity. However, external validity is NOT the same. Ecological validity is defined as:

- **Ecological Validity**: Whether the methods, materials, and setting of the research are similar to a given target environment.

Important: Whether an experiment is ecologically valid does not tell us anything about the external validity of the experiment. We can have an experiment that has a high degree of ecological validity, similarity to one target environment, but has absolutely no similarity to another target environment. Since external validity is about generalization beyond the target environment, ecological validity tells us nothing about external validity.

The ONLY way to establish external validity is empirically. How can we do that? We can use scientific replication.

Scientific replication is defined as:

- **Scientific Replication**: When a researcher uses a different sample from the same population to evaluate the same theoretical implications as in the previous study with equivalent construct validity or uses the same sample from the same population but comparing statistical techniques to evaluate the same theoretical implications as in the previous study, again with equivalent construct validity.

Replication is often used to verify that results found in one population can be found in another population. Stress tests, as discussed in chapter 6, can be used to validate results, as can sampling a nonrandom holdout, or a population that differs significantly than that used for initial estimation, over which results can be compared. In evaluating the validity of a study, internal validity must be established for external validity to be applicable. However attention should be paid to both forms of validity during research design.

Location, Artificiality, and Related Design Issues (Chapter 8)

9.4 Introduction to a Series of Design Choices

This chapter discusses the choices a researcher faces when determining the level of analysis, location of experimentation, baseline, and artificiality of a study. The researcher must determine all of these elements as a function of the question at hand, aware of the tradeoffs within each of the decisions.

9.5 Determining the Level of Analysis in Experiments

Level of analysis refers to the unit analyzed in an experiment. These can take the form of individuals or groups, the latter of which can be conceptualized in various ways. An individual decision-making experiment and group decision-making experiment are defined as:

- Individual Decision-Making Experiment: The subjects' choices are not interactive and the experimenter only observes individual-level behavior,
- Group Decision-Making Experiment: The subjects' choices are interactive and the experiment observes both individual and group choices.

While individuals are often studied because group level studies require a larger and more costly sample, the formal model and theory determine which unit is relevant for the study.

9.6 Determining the Location of Experiments

The location of an experiment is often considered the most salient dimension of an experiment. Experiments can take place:

1. Over the internet
2. In Labs
3. In Labs-in-the-field
4. In the Field

Each location has advantages and disadvantages that must be considered in the context of the experiment implemented.

1. The internet

The internet can be a useful and efficient way of recruiting subjects. Concerns arise over whether subjects truly believe that they are interacting with other subjects or change their behavior because they think they are interacting with virtual subjects. The internet is the ideal venue in which to test new institutions or mechanisms that don't exist in real life; for example, new voting procedures.

2. Labs-in-the-Field

In lab-in-the-field experiments:

- Lab-in-the-Field: Subjects participate in a common physical location, but the experiment, to some degree, brings the laboratory to the subjects' natural environment more than the subjects come to the laboratory.

Lab-in-the-field experiments offer a balance between the control a researcher has in the laboratory and the benefits of the natural setting of the field. Often, this approach is taken when the researcher is interested in a particular element of the field, for example, ensuring a randomized draw from a population or examining a particular post-disaster environment.

3. The Field A field experiment is when a researcher's intervention takes place in subjects' natural environments and the researcher has only limited only beyond the intervention conducted. Usually the relationship between the researcher and the subject is conducted through variables outside the researcher's control.

9.7 Choosing a Baseline Measurement for Experiments

To measure the impact of a treatment or intervention, a researcher must make a comparison against a baseline. The baseline group is most often constructed as a group that does not receive the treatment or manipulation. Often times though, absence of treatment might not make substantial sense and a treatment is compared against a different baseline. For example, when comparing voting rules, a control group will have one voting rule that is compared against the others rather than no rule. In cases where multiple comparisons are made between various treatments, researchers should adjust for a higher probability of false significance. In some cases, the baseline might actually be a theoretical, rather than empirical baseline.

9.8 Concerns with Artificiality in Experiments

Artificiality refers to the effect that the experiment itself has on the subject. Often, subjects will alter their behavior because they are aware they are participating in an experiment. This is called the experimental effect and is defined as follows:

- Experimental Effect: When subjects' choices in an experiment are influenced by the fact that they are participating in an experiment and are different from what they would be if the manipulations and control exercised in the experiment took place via the DGP without experimental intervention.

Making the experimental procedures unobtrusive can ameliorate these effects. Importantly, for subjects that participate in multiple studies, there might exist learning that changes behavior as well. These experimental cross-effects should also be minimized, possibly by recruiting from subject pools that haven't participated in experiments previously.

However, experimental effects are by definition what an experiment is designed to produce. As Vernon Smith points out, everything about an experiment is inducing experimental effects and a huge advantage of experimental design is our ability to both vary and manipulate the various aspects of the experimental environment in order to determine when experimental effects exist. Experimental effects are best embraced and recognized, rather than minimized and ignored.

10 Design IV: Handling Sensitive Questions

Alex Scacco

10.1 Motivation

- The study of sensitive attitudes and behavior is now widespread in survey research in political science.
- Researchers often need to ask questions about private information that respondents would prefer not to answer in public.
- The presentation highlights three sets of innovations in methods for asking sensitive survey questions: (1) randomized response designs, (2) list experiments, and (3) enhanced anonymity or survey mechanics procedures.
- How is this relevant for experimental research? Survey questions are often used as outcome measures to study the impact of randomized interventions.
- Example: Christia et al. (2011) study the effects of community?driven development projects on government popularity in rural Afghanistan. To evaluate a policy intervention, they use surveys to measure attitudes toward the government.
- Example: Green and Wong (2008) ask whether increased interracial contact reduces prejudice. The intervention randomly assigns participants to racially heterogeneous and homogeneous Outward Bound wilderness courses. Follow?up surveys measure racial prejudice.

10.2 Sensitive Questions Research Examples:

Randomized response:

- Corrupt behavior (e.g. bribe-taking) by bureaucrats in Latin America (Gingerich, 2009)

List experiments:

- Racial prejudice and attitudes toward affirmative action in the United States (Kuklinksi et al., 1997)
- Vote-buying in Lebanon (Corstange, 2009) Enhanced anonymity/ survey mechanics:
- Sexual behavior and sexuality in the United States (Laumann et al., 1994) 1
- Participation in religious riots in Nigeria (Scacco, 2010)
- Attitudes toward partition in Sudan (Beber, Roessler and Scacco, 2011)

10.3 The Problem:

When asked sensitive questions, respondents may have incentives not to respond truthfully (response bias), or not to respond at all (non?response). Possible sources of response bias:

- Interviewers seem to be looking for a particular answer (example: leading questions).
- The respondent wants to please the questioner by answering what appears to be the morally correct answer (example: female interviewers asking male respondents about domestic abuse).
- Truthful answers carry the risk of punishment, embarrassment or other harmful outcomes for respondents (example: questions about participation in illegal activities).

10.4 Three Innovations in Survey Design

I. Randomized Response (RR)

- The intuition: deliberately introduce noise into responses. Individual responses are obscured such that the researcher does not know which of two questions any given respondent has answered. The researcher can recover population estimates for sensitive questions because the noise probability is known in advance.
- An example: Survey on corruption in Bolivia, Brazil and Chile (Gingerich 2010) includes questions about bribe?taking.
- Design: 80% of respondents were asked to respond to the following yes/no question: Have you ever taken a bribe? and the remaining 20% are asked the opposite question: Have you never taken a bribe?
- Suppose you have 100 responses and 60 yes answers and you want to know how many people are corrupt. The observed yes?rate R is a function of the true corruption incidence C and the probability p of being asked Have you ever taken a bribe, as opposed to Have you never taken a bribe. That is, R = pC + (1?p) (1?C), which we can solve for C = (R + p 1) / (2p 1). In the example, this gives us an estimated corruption incidence of (.6 + .8 1)/ (1.6 1) =2/3.
- Implementation: researchers need some sort of a randomization device. A popular choice is a spinner (others have flipped coins, rolled dice).
- A nice feature for respondents: it gives them plausible deniability.

Drawbacks:

- Deliberate introduction of noise into responses is inefficient (if you could ask the question directly, you'd be more certain of the proportion of the population that is corrupt).
- The novelty or elaborateness of devices used in RR questions (e.g. spinners) may draw attention to the measurement and may make respondents more uncomfortable.

II. List Experiments

The intuition: use a survey experiment to elicit responses to lists of both sensitive and non?sensitive items in order to obscure and protect sensitive responses.

- An example: How prevalent is racial prejudice in the United States? (Kuklinski et al., 1997)
- Design: Randomly assign respondents to treatment and control groups. Ask respondents in the control group to answer how many items out of the following three make them angry:

1. Higher tax on gasoline
2. Million-dollar salaries for athletes
3. Large corporations polluting the environment"

- For the treatment group, add "a black family moving in next door." ? Compare mean number of items from treatment to control. The average treatment effect gives us the estimated share of the sample that is prejudiced.
- Suppose control group respondents answered yes to 2 questions on average and treatment group answered yes to 2.3 questions on average, we infer that .3 or 30% of respondents were upset by the sensitive item.
- Some constraint on plausible deniability (constrained by ceiling and floor effects).

Drawbacks:

- As with randomized response: by deliberately obscuring what you're trying to measure, any response will be noisier, less efficient.
- Most list experiment designs don't allow the researcher to run individual?level models (though Corstange 2009 and Blair and Imai 2011 have begun moving in this direction).

III. Direct Questions with Enhanced Anonymity The intuition: Use the mechanics of survey administration to preserve anonymity while asking sensitive questions directly to respondents.

1. Make it impossible for enumerators, local authorities, or researchers to learn or guess respondent answers during survey administration.
2. Make it impossible to link answers to sensitive questions to respondent profiles or other sets of responses.
3. Examples: Riot Participation in Nigeria (Scacco 2010), Attitudes toward partition in Sudan (Beber et al. 2011)

- Design: Respondents answer sensitive questions themselves, without observation by interviewers. Sensitive and non?sensitive questions are then physically separated in a way that is transparent to respondents.
- In the Sudan partition study, three documents were required for each interview:
- Main questionnaire
- Sensitive questions sheet (includes any questions for which respondents would have incentives to misreport)
- Sensitive answers sheet (example: bubble sheet if low levels of literacy in target population)

After sensitive answer sheets filled out, respondent placed them in an envelope (containing other sheets, some of which may be decoys) and in a ballot box. These answers could only be linked to main questionnaires with numerical code key left in New York.

- Smaller logistical tips: Avoid skip patterns. Limit circulation of sensitive questions sheet.
- Gives respondents plausible deniability as long as either: (1) no names or contact information collected or (2) answer key is fully protected.

- Drawbacks: Relatively time- consuming administration; Measurement somewhat obtrusive (envelopes, ballot boxes, etc.)

10.5 References:

Beber, Bernd, Philip Roessler and Alexandra Scacco, 2011. Who Supports Partition and Why? New Survey Evidence from Sudan, working paper, New York University.

Corstange, Daniel, 2009. Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT, Political Analysis 17: 1, 45?63.

Gingerich, Daniel W., 2010. Understanding Off?the?Books Politics: Conducting Inference on the Determinants of Sensitive Behavior with Randomized Response Surveys, Political Analysis 18:3, 349? 380.

Green, Donald P. and Janelle S.Wong. 2009. Tolerance and the Contact Hypothesis: A Field Experiment, in The Political Psychology of Democratic Citizenship, ed. Eugene Borgida, Christopher M. Federico and John L. Sullivan. Oxford, UK: Oxford University Press, 228?244.

Kuklinksi, James H., Paul M. Sniderman, Kathleen Knight, Thomas Piazza, Philip E.Tetlock, Gordon R. Lawrence, and Barbara Mellers, 1997. American Journal of Political Science 41: 2, 402?419.

Laumann, Edward O., John H. Gagnon, Robert T. Michael, and Stuart Michaels, 1994. The Social Organization of Sexuality: Sexual Practices in the United States. Chicago: University of Chicago Press.

Scacco, Alexandra, 2010. Who Riots? Explaining Participation in Ethnic Violence, manuscript, Columbia University.

11 Applications I: Types of Experimentation in Political Economy

Laura Paler

11.1 Types of Experiments

Previous sessions have emphasized that experiments strive for validity, robustness and generalizability beyond the target population. This session presents different types of experiments and reviews the strengths and weaknesses of these different approaches, with reference to political economy research on taxation and public goods provision. The goal is to highlight that different types of experiments have different advantages and disadvantages, and that researchers often must make tradeoffs when choosing an experiment type. Being aware of this not only helps a researcher decide what type of experiment to use but also can help produce a good design.

This session centers around five of the main types of experiments:

1. Lab: When subjects are recruited to a common physical location called a laboratory and the subjects engage in behavior under a researcher's direction at that location (Morton and Williams 2010, p. 42).

2. Lab-in-the-field: Experiment where the subjects participate in a common physical location (called the lab in the field) but the experimenter, to some degree, brings the laboratory to the subjects' natural environment more than the subjects come to the laboratory (Morton and Williams 2010, p. 296).

3. Survey: An individual decision-making experiment embedded in a survey (Morton and Williams 2010, p. 279). Can be about priming, framing, or actually changing subjects' priors.

4. Field: When a researcher's intervention takes place in subjects' natural environments and the researcher has only limited control beyond the interven-

tion conducted. Usually the relationship between the researcher and subject is conducted through variables outside of the researcher's control (Morton and Williams 2010, p. 46).

5. Policy: A field experiment in which a government agency or other institution chooses to intervene and act like an experimentalist (Morton and Williams 2010, p. 54)

11.2 Differences among Experiment Types: Five Key Dimensions

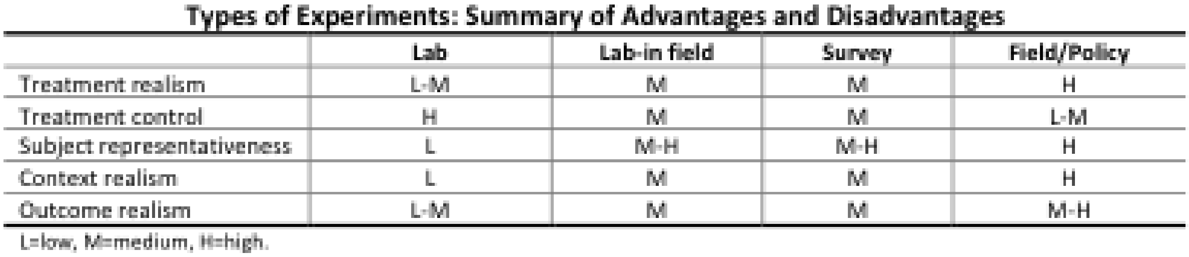
While there are numerous ways in which experiments of different types differ, I focus here on five key dimensions to help organize the discussion. Keep in mind the following when we turn to the examples.

1. Treatment control: How do types of experiments differ in the type of control over treatments afforded to researchers? Typically lab experiments give researchers more control over the treatments, make possible fine-grained variations in treatments, and let researchers control which subjects are exposed to the treatment and how. The treatments in field experiments are often complex and more constrained by feasibility, resources, and interests of different actors.

2. Treatment realism: How do types of experiments differ in the extent to which the treatments resemble real-world policies or interventions? What are the implications of this for validity? An advantage of field experiments is that (compared to lab experiments) they are realistic, naturalistic and less obtrusive.

3. Subject representativeness: How do different types of experiments vary in the extent to which subjects resemble the target population of interest? How does this affect the inferences that can be made? Lab experiments often rely on student subjects, while lab-in-field, survey and field experiments have the advantage of using the target population.

4. Context realism: How do different types of experiments vary in the realism or artificiality of the context in which they are implemented? For instance, experimenter effects (whether subjects change their behavior if they know they



Types of Experiments: Summary of Advantages and Disadvantages

	Lab	Lab-in field	Survey	Field/Policy
Treatment realism	L-M	M	M	H
Treatment control	H	M	M	L-M
Subject representativeness	L	M-H	M-H	H
Context realism	L	M	M	H
Outcome realism	L-M	M	M	M-H

L=low, M=medium, H=high.

are being studied) are potentially a bigger concern in lab experiments than in field experiments. Lab experiments also often rely on monetary rewards or punishments into the experiment to make the stakes feel real, whereas these incentives arise naturally in field experiments.

5. Outcome realism: How do different types of experiments vary in how closely the experimental outcomes reflect the real-world outcomes of interest?

The table below summarizes the general strengths and weaknesses of the different experiment types, although as we will see in the examples there is often substantial variation.

11.3 Example 1: Taxation

2.1 Lab Experiment: Tax Compliance (Alm, Jackson and McKee 1992)

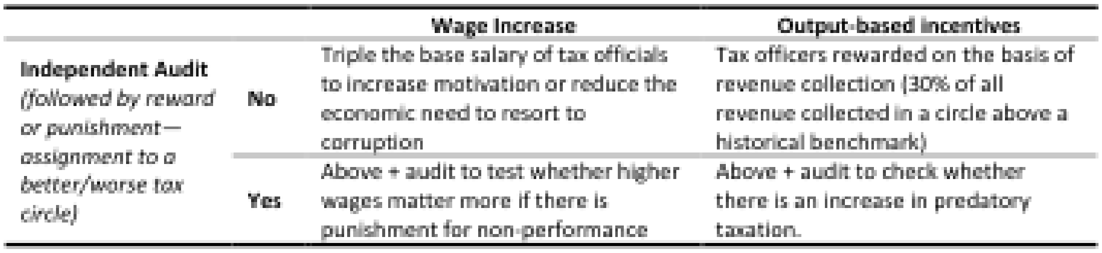
- Study goal: Identify how different government policies affect taxpayer compliance.
- Challenge: Tax evasion is illegal and difficult to obtain reliable data on individual compliance choices. Also difficult to test how individuals respond to different government policy choices.
- Experiment: In a lab setting, varied the tax rate, audit probability, penalty level and public good provision to study impacts on the amount of total income

reported. Values of policy parameters chosen to approximate real world values.

- Subjects: Student volunteers from undergraduate classes.
- Implementation: In each session, subjects are organized into three groups of five and play 25 rounds. In each round, they first receive income, then pay taxes on income voluntarily reported, then face an audit probability, then pay a penalty if found cheating. In one session, all tax payments of group members are paid into a group fund, which is multiplied by two and then redistributed equally.
- Outcomes: How much income subjects voluntarily report.

2.2 Survey Experiment: Taxation and Political Engagement (Paler 2011)

- Study goal: Does paying taxes make citizens more politically engaged and willing to hold government accountable?
- Location: Bloraa resource rich district in Indonesia.
- Experiment: Embedded a revenue experiment in a public awareness campaign and survey. The treatment group paid a simulated tax to the district government from income they earned as part of the experiment and the share of taxes in the district budget was emphasized. The control group captured a situation where government depends on revenue windfalls (natural resources, aid): Subjects paid no simulated tax and the share of windfall revenue in the district budget was primed.
- Subjects: 1863 citizens from 93 villages in Blora. Subjects were randomly sampled from adult voting-age population.
- Randomization: At the individual-level (blocked within villages)
- Implementation: Trained canvassers implemented the campaign and survey one-on-one with respondents in their homes.
- Outcomes: Actual participation in a postcard campaign + survey.



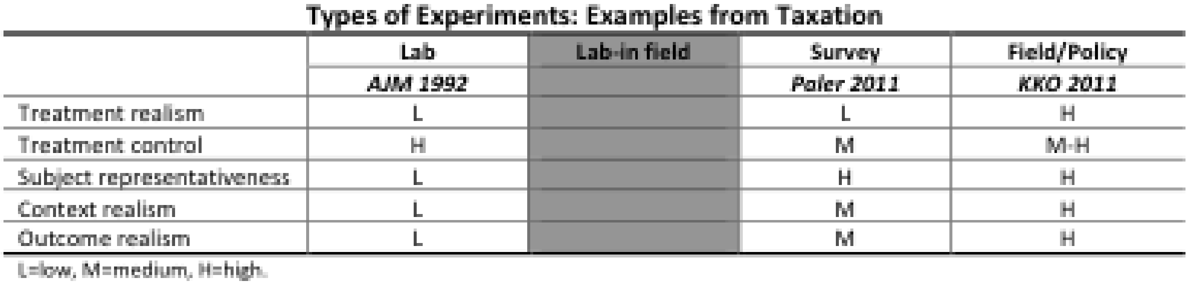
		Wage Increase	Output-based incentives
Independent Audit *(followed by reward or punishment—assignment to a better/worse tax circle)*	**No**	Triple the base salary of tax officials to increase motivation or reduce the economic need to resort to corruption	Tax officers rewarded on the basis of revenue collection (30% of all revenue collected in a circle above a historical benchmark)
	Yes	Above + audit to test whether higher wages matter more if there is punishment for non-performance	Above + audit to check whether there is an increase in predatory taxation.

2.3 Field Experiment: Property Taxes (Khan, Khwaja, Olken, ongoing)

- Study goal: Given corruption and low public sector efficiency in tax collection, what is the optimal way to improve the behavior of tax collectors, minimize tax evasion, and improve overall tax performance?
- Location: Punjab, Pakistan
- Experiment: Four treatment conditions to test different mechanisms (plus a pure-control):

Wage Increase Output-based incentives Independent Audit (followed by reward or punishmentassignment to a better/worse tax circle) No Triple the base salary of tax officials to increase motivation or reduce the economic need to resort to corruption Tax officers rewarded on the basis of revenue collection (30% of all revenue collected in a circle above a historical benchmark) Yes Above + audit to test whether higher wages matter more if there is punishment for non-performance Above + audit to check whether there is an increase in predatory taxation.

- Subjects: Property tax collectors within about 300 tax circles (geographic areas with approximately an equal number of properties).
- Randomization: First identified 25 treatment and 15 control zones of about 12 tax circles each. Consenting tax circles in treatment zones randomly assigned to one of the four treatments (Design intended to minimize resentment within



Types of Experiments: Examples from Taxation

	Lab	Lab-in field	Survey	Field/Policy
	AJM 1992		*Paler 2011*	*KKO 2011*
Treatment realism	L		L	H
Treatment control	H		M	M-H
Subject representativeness	L		H	H
Context realism	L		M	H
Outcome realism	L		M	H

L=low, M=medium, H=high.

zones and to minimize spillovers with pure control since they are removed from the experiment.)

- Implementation: Collaboration with the Excise and Taxation Department, who will implement all schemes. All funding for incentive payments from the Punjab government.
- Outcomes: Real measures of revenue performance + surveys.

The following table summarizes how the different experiments just reviewed perform according to the five key considerations. Prepare to discuss whether you think this is an accurate characterization. What changes would you make to the table below?

11.4 Example 2: Public Goods Provision and Social Cohesion

The production of public goods (security, healthcare, education, sanitation) is widely considered important to development and social welfare. There is a wide degree of variation in the extent to which communities contribute to public goods production in reality. Lab and lab-in-the-field studies have explored the determinants of public goods provision. More recently, field experiments have been used to approach the question from the reverse: Do new opportunities/institutions for public goods production increase social cohesion among community members?

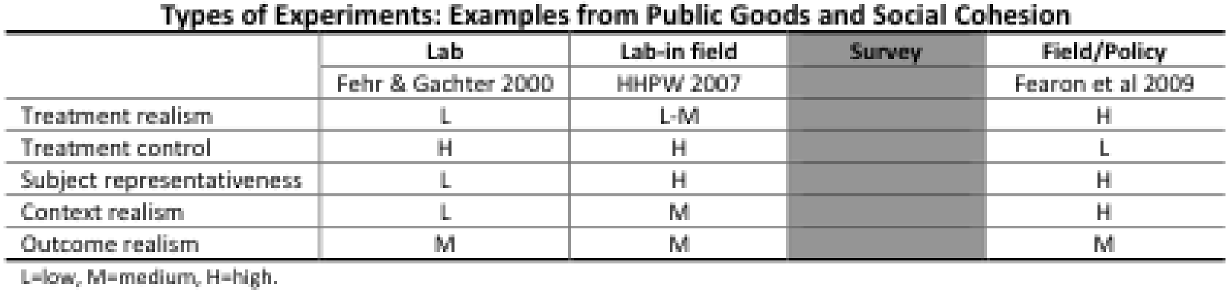
3.1 Lab: Cooperation in Public Goods Games (Fehr and Gachter, 2000)

- Study goal: Show that people are willing to engage in costly punishment of free-riding, that punishment opportunities reduce free riding.
- Experiment: Played a public goods game with two over-lapping experiments: (1) Stranger-partner experiment; and (2) Punishment and no punishment experiment. The theoretical prediction is that all subjects should contribute nothing to the public goods in all periods.
- Subjects: Students from different fields (except economics) at the University of Zurich (Switzerland).
- Outcomes: Game play.

3.2 Lab-in-the-Field: Ethnic Diversity and PG Provision (Habyarimana et al 2007)

- Study goal: Why do some communities generate a high level of public goods where others do not? How does ethnic diversity affect the willingness of community members to contribute to the public good?
- Location: Kampala, Uganda
- Experiment: Designed different games to test three different mechanisms: (1) preferences, for instance different ethnic groups prefer different goods, so hard to agree; (2) technology, for instance simply easier for members of the same ethnic group to communicate and work together; or (3) strategy selection, for instance, if there are social norms of cooperation within the ethnic group but not across ethnic groups.
- Subjects: Random sample of 300 subjects recruited from a high diversity, low public goods area in Kampala.
- Implementation: Common location where subjects participated in different games where the treatments varied the ethnic identity and anonymity of fellow-players to test specific mechanisms.
- Outcomes: Game play.

3.3 Field: Public Goods and Social Cohesion (Fearon et al 2009)



Types of Experiments: Examples from Public Goods and Social Cohesion

	Lab	Lab-in field	Survey	Field/Policy
	Fehr & Gachter 2000	HHPW 2007		Fearon et al 2009
Treatment realism	L	L-M		H
Treatment control	H	H		L
Subject representativeness	L	H		H
Context realism	L	M		H
Outcome realism	M	M		M

L=low, M=medium, H=high.

- Study goal: Can efforts to build local institutions for public goods provision build community social cohesion?
- Location: Northern Liberia
- Experiment: A randomized community-driven reconstruction (CDR) program that involves the organization of community structures for making and implementing decisions about local public goods provision in a transparent and accountable way. Treatment complex in that communities received both funds and new decision-making institutions.
- Subjects: Individuals in 42 treatment (and 41 control) communities.
- Implementation: International Rescue Committee (IRC).
- Outcomes: Use a public goods experiment (behavioral measure). In both treatment and control villages, 24 households randomly sampled to participate in a PG game to win up to an additional $420 for their villages. There was also a cross-cutting treatment where the gender composition of the players was varied.

The following table summarizes how the different experiments just reviewed perform according to the five key considerations. Prepare to discuss whether you think this is an accurate characterization. What changes would you make to the table below?

11.5 Some Concluding Points

- Often the type of experiment selected is primarily a function of feasibility and appropriateness.

  While it might seem desirable to do a field experiment, it is not always possible.

- Different types of experiments present tradeoffs. For instance, lab experiments often give researchers the most control over designing nuanced treatments but at the expense of realism. Being aware of the advantages and disadvantages of different types of treatments can help create a stronger design.

- Sometimes what might seem like a disadvantage on the surface is actually a factor of interest. For instance, while it might be harder to control spillover or non-compliance in a field experiment, if the goal is to estimate the impact of a real policy intervention then these should be taken into account rather than avoided.

11.6 References

Alm, James, Betty Jackson and Michael McKee, Estimating the Determinants of Taxpayer Compliance with Experimental Data, National Tax Journal 45(1), 1992.

Alm, James, Betty Jackson and Michael McKee, Institutional Uncertainty and Taxpayer Compliance American Economic Review 82(4), 1992

Fearon, Jame, Macartan Humphreys and Jeremy Weinstein, Development Assistance, Institution-Building and Social Cohesion after Civil War: Evidence from a Field Experiment in Liberia American Economic Review Papers and Proceedings

Fehr, Ernst and Gachter, Cooperation and Punishment in Public Goods Experiments American Economic Review 90(4), 2000.

Habyarimana, James, Macartan Humphreys, Daniel Posner, Jeremy Weinstein, Why Does Ethnic Diversity Undermine Public Goods Provision, American Political Science Review 101(4), 2007.

Khan, Adnan, Asim Khwaja and Bejamin Olken, Property Tax Experiment in

Punjab, Pakistan: Testing the role of wages, incentives and audit on tax inspectors' behavior (working document).

Morton, Rebecca and Kenneth Williams, Experimental Political Science and the Study of Causality: From Nature to the Lab, Cambridge University Press: 2010.

Paler, Laura Keeping the Public Purse: An Experiment in Windfalls, Taxes and Transparency working paper, last revised May 2011.

12 Applications II: Political Economy Field Experiments

Guy Grossman

12.1 Overview of applications in political economy:

1. What are the outcomes that we may care about?
2. What can be randomized?
3. Notable examples (worth thinking about:)

- At what level (or site) does the randomization take place?
- What are the different sorts of partnerships researchers can form?
- What are the different types of data sources?
- What are the different research designs?

12.2 Selected Outcomes:

1. Provision of Key Social Services
2. Corruption (electoral fraud, mismanagement of public resources)
3. Electoral Accountability (vote choice, turnout, other forms of participation)
4. Violence and order (crime, security)
5. Post-conflict reconstruction (collective action, trust, cohesion)

12.3 What can be randomized (manipulated)?

Political institutions

- rules for selecting projects
- rules for selecting representatives
- rules for protecting property rights (land tenure)

Monitoring institutions

- top-down vs. bottom-up
- internal vs. external / domestic vs. international
- political vs. apolitical
- technology vs. human interaction
- election observers (presence, intensity, type)

Information (content)

- on quality of service providers (schools, clinics, roads)
- on politician's performance (effort, spending, leakages)
- on rights and responsibilities of communities (service standards)
- on inter-ethnic relations

Messenger (delivery method)

- traditional vs. political authority
- media (radio, newspapers) vs. community meetings

Incentives Schemes

- rewards, hiring and contractual arrangements
- cash transfers
- source of funding

12.4 Notable Examples:

1. Improve the Provision of Key Social Services

1.1 Interventions designed to alter the incentives of social service providers

- Test the relative efficiency of additional resources versus changing hiring policies in the public education sector (Duflo, Dupas and Kremer, 2009)
- Test different monitoring schemes and technologies designed to reduce teachers' absenteeism (Duflo and Hanna, 2007) and nurses' absenteeism (Banerjee, Deaton and Duflo, 2004)

1.2 Interventions designed to mobilize social services consumers

Several studies have tested different ways of increasing the social pressure on service providers by increasing citizens' direct involvement. The idea is to measure the causal effect of

- Providing citizens with information about schools' performance (Andrabi, Das and Khwaja, 2009)
- Encouraging communities to form PTAs: Banerjee et al., (2008): minor effect (India) ; Duflo, Dupas and Kremer (2009): positive effect (Kenya)
- Encouraging communities to form community health committees that monitor the local health clinic (Bjorkman and Svensson, 2009)
- Different reward schemes for students excelling in school: Sorting by performance (Duflo, Dupas and Kremer, 2010) ; Blimpo, (2010)

1.3 Altering Rules for Selecting Development Projects: An experiment in 49 Indonesian villages that were randomly assigned to choose development projects through either representatives-based meetings or direct elections (Olken 2010)

- A 2X2 factorial design field experiment implemented in 250 villages in Afghanistan, testing the impact of two methods for electing local development council and two methods for selecting development projects (Beath, Christia, Enikolopov, 2010)

2. Corruption (leakage and electoral fraud/violence) Several studies have tested the relative effectiveness of different ways to reduce corruption.

- Monitoring institutions: Randomized field experiment that tests the relative efficiency of increasing government audits versus increasing grassroots participation in monitoring on corruption in over 600 Indonesian village road projects (Olken 2007)
- Information campaign: Study that uses distance to newspaper outlet to test the causal relation between information on central government capitation grants to schools on reduction in leakage of funds (Reinikka and Svensson, 2011)
- Electoral violence: a nationwide field experiment based on randomized anti-violence grassroots campaigning during the 2007 Nigerian elections (Collier and Vicente, 2010)
- Information campaign: study that exploits random selection of Brazilian municipalities to receive an audit (policy experiment). The study tests the effect of disseminating information on corruption practices of the randomly selected Brazilian municipalities on the electoral outcomes of incumbents (Ferraz and Finan, 2008)

3. Electoral Accountability

- Information campaign: Banerjee et al., (2010) provide slum Indian dwellers with newspapers containing report cards giving information on candidate qualifications and legislator performance (state level).

- Information campaign: Chong et al., (2010) examine the effects of an information campaign on electoral participation and incumbent parties' vote share in the 2009 municipal elections in Mexico. The information that was distributed was taken from reports produced by the Mexican Federal Auditor's Office.

- MPs Scorecard: study that uses the dissemination of performance scorecards to a random subset of the constituencies of Members of Parliament in Uganda to distinguish between selection-based and incentive-based accounts of legislator responsiveness (Humphreys and Weinstein, 2007)

13 Implementation: Forging Partnerships and Mitigating Threats to Validity in Field Experiments

Eric Mvukiyehe

In order to study the effects of real-word intervention, the researcher must understand what the treatment means in practical termswhen it is delivered, by whom, in what form... (Green, forthcoming).

This memo discusses requirements to successfully implement field experiments as well as some practical implementation issues, namely: (i) how to forge partnerships and sustain cooperation with a partner throughout the project's life; and (ii) how to mitigate pitfalls that can potentially threaten validity of the study.

13.1 I. Forging partnerships and sustaining cooperation

Perhaps one of the most important requirements to conduct a successful field experiment is to find partner willing and able to implement the study according to specific research protocols, especially those pertaining to the assignment of treatment to different units/subjects and to the administration of treatment regimes. Below I discuss possible outlets through which a field experiment can be organized; the types of partnerships researchers can negotiate; and the conditions make collaboration with an implementing partner more or less likely.

What are possible outlets for potential partnerships?

There are several possible outlets through which researchers can carryout randomized studies. These are typically individuals, agencies or organizations that have some control over (or working with) the population you wish to study and include government agencies, political parties, community organizations, non-government organizations, international organizations, to name a few. Different partners have different degrees of flexibility, which should be borne in mind when shopping around for a suitable partner. Government entities are typically expected to serve the entire populations and may perhaps be least flexible. NGOs, on the other hand, are not expected to serve entire populations and can buy into randomization approaches relatively easily (Duflo and Kremer 2003). The main issues with international institutions are bureaucratic red tapes, relatively high turnovers in personnel and sensitivity to extraneous events. Political entities such as political parties are the least amenable to randomized interventions, in part because of the sensitive and consequential nature of their business (e.g., election), but even then it has been shown that researchers can conduct experimental manipulations without tempering the ultimate outcome.[2] In short, it is important to investigate the types of constraints potential partners may face before entering into a partnership.

13.2 What types of partnerships to negotiate?

It depends. In some cases, partnership is initiated by an NGO or agency seeking credible evaluations of its programs and researchers come in as consultants to provide technical assistance (e.g., to do the randomization and data collection/analysis). In these types of partnerships, researchers don't usually have a hand in the development of the intervention being randomized and so negotiation tends to be over the types of design that would be feasible. Examples of successful partnerships of this sort include: Miguel and Kremer's work with Internationaal Christelijk Steunfonds Africa (ICS) on school programs in Kenya; Blattman's work with Land Mines Action and UNHCR in Liberia; and Fearon et.al's work with the International Rescue Committee in Liberia (as well as Humphreys' work with the same NGO in the DRC).

In other cases, field experiments tend to be initiated by a researcher who might have explicit theories or hypotheses she wishes to test experimentally and tries to find a partner organization that can provide a testing ground. In these types of

[2] See, for example, Wantchekon's (2003) study of different campaign messages in a presidential election in Benin. Randomization occurred in the first round of the elections (where the stakes are often fairly low given the large number of candidates). He also carefully screened villages and only selected those where the votes were not close in the previous election to help ensure that the experiment would not influence the result. See, Browning (2002) for a discussion on the merits and ethnics of this experiment.

partnerships, researchers often have a hand in the design and implementation of the intervention itself, but collaboration depends on many of the conditions I discuss below. Examples of successful partnerships of this sort include: Collier and Vicente work with the Nigeria chapter of Action Aid International to study how violence is used as an electoral strategy; Loewen and Rubenson's work with a campaign for the leadership for the Liberal Party of Canada; and Mvukiyehe and Samii's work with the United Nations Mission in Liberia (UNMIL).
Whatever the case, certain conditions must be met in order to forge a successful partnership and maintain cooperation at least until the end of the experiment.

What are the conditions for collaboration? *Potential for mutual interests:* Field experiments are not a charity. Researchers come to potential partners because they want to gain something (e.g., improve knowledge about the real world; advance their career, etc.) Potential partners need to something for them too before the can agree to the experiment proposal. So, unless a potential partner already understands the value of randomized control trials or they are required to integrate these approaches in their programs (which is usually the case with many NGO that receive funding from external donors), the burden is on the researcher to convince prospective implementing partners that they stand to gain something from the partnership.

Feasibility of the experiment: Feasibility here refers both to the moral/ethics of the study (e.g., studies that may be harmful to the subjects and/or carry little benefits to society) as well as to its costs and logistics (e.g., studies interfere with a partner's operations or divert resources away from programs). Either or both of these two problems would discourage potential partners from agreeing to a field experiment or stop cooperating if one has already started. Green (forthcoming) and Loewen et. al (2011) suggest specific conditions that should make field experiments more likely:

- Uncertainty about the outcome of an intervention; Intervention is known to work, but mechanisms of an effect are not known;
- Field experiment carries low likelihood of harm (physical, mental and emo-

tional) to the subjects and project staff; and

- Implementation can be done in a flexible, scalable and non-intrusive manner (both practical and financial.)

How to negotiate partnerships?

Knowing when the conditions are right for a field experiment is not enough, however. In addition, researchers must also have the skills necessary to secure partnership and maintain cooperation with the partner to keep the project on tract, at least until the conclusion of the experiment. To this end, Green (forthcoming) suggests that the researcher be able to play multiple roles at different stages, including that of a diplomat, an ethnographer and a business consultant. Furthermore, partnerships should be formalized in a Memorandum of Understanding (MOU) that defines respective roles, responsibilities, obligations and expectations. The MOU should be especially explicit about the use of randomization in the project, staffing and financing of research activities and about data ownership and usage terms.

13.3 Mitigating pitfalls that might threaten validity of a field experiment

In implementing field experiments, things don't always go according to plans. There is always a chance that something might go wrong or be done incorrectly. Such missteps could pose serious threats to the validity of the study (e.g., undo random assignment) and lead to what Barrett and Carter (2010) call 'faux exogeneity.' Below I discuss potential sources of these threats and how they can be mitigated.

Potential pitfalls and their sources The literature on field experiments suggests a number of problems that may threaten validity of experimental studies. These include:

- Compliance problems (i.e. subjects don't take treatment assigned to them);
- Attrition (i.e. subjects or units drop out of the study);
- Interference between units or spillover (i.e. subjects on the control group get second hand treatment).

There are several sources of these problems (Barrett and Carter 2010; Loewen et.al 2011; Green, forthcoming):

- Randomization protocols may be compromised or impractical;
- Treatment regimes might be administered poorly (e.g., not administered or administered to everyone, including in the control group;)
- Communities may be inaccessible (especially in unstable countries);
- Collaboration between researchers and the implementing partner might strain; and
- Resources might be insufficient or come in late.

Researchers must have the ability to foresee these problems and take preventive (or mitigating) measures. I discuss possible mitigating measures below. **How to mitigate threats to validity' problems?**

- Coordinate closely with the implementing partner (better yet, the point of contact), but don't expect them to do the research for you. At the minimum, hire your own research manager who would work alongside the partner to ensure the strict respect of research protocols, especially those pertaining to the assignment of randomization of the intervention and administration of treatment regimes.
- Limit knowledge of the experiments. Everyone among your partner's staff does not need to know that you are conducting experiments. This minimizes the chances of treatment distortion (e.g., eager staff may feel the need to compensate non-treated units/subjects in some other ways) and of spillovers.
- Learn as much as possible about the research environment: What are the key features of the research setting? Who are the stakeholders or key players? How do they perceive the project? Are people willing participate in the study/to talk about issues frankly and openly or are certain topics off-limit?) What are the security, social, and political dynamics that may compromise implementation?

In short, is this project appropriate for this setting at this particular time? This is where you may need to play the role of an ethnographer.

- Be attuned to logistics (i.e. are there any particular logistical challenges that might require to make adjustments in the original design? If so, can you make such adjustments without undermining the robustness of the design?)
- Documentation, Documentation, Documentation: Corrective measures are often necessary either during the project or in the analysis. However, As Loewen et. al (2011) pointed out, mistakes can only be corrected if they are discovered and documented. Thus, it is imperative for researchers and their partners to establish a 'paper trail' and carefully record all instructions, decisions and actions during the process of randomization and treatment administration. Moreover, careful documentation can also provide data that may be use gauge the movement on key indicators of interests during the rollout of the program.

13.4 Background reading:

- Green, Donald and Alan Gerber (forthcoming.) "Experimental Challenges and Opportunities." Chap 13.
- Loewen, John, Daniel Rubenson and Leonard Wantchekon. 2011. "Conducting Field Experiments with Political Elites"
- Barrett, Christopher and Michael Carter. 2010. "The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections." (esp, pp 5-24.)
- Duflo, Ester and Michael Kremer (2003). "Use of Randomization in the Evaluation of Development Effectiveness," (from pp 17).

Lecturer Biographies

Bernd Beber

Assistant Professor of Politics, New York University
bhb2102@columbia.edu

Dr. Beber’s research has focused in particular on the causes and consequences of international mediation in wars. He has addressed the strategic selection of mediators both theoretically and empirically and have argued that such an analysis produces substantively different results from an analysis that treats mediation as a non-strategic, exogenous intervention, as much of the relevant quantitative literature does. He has also co-authored papers on the economics of rebel recruitment and on how to detect election fraud in data-poor environments.

Selected Papers

Beber, B. C. Blattman. 2010. “The Industrial Organization of Rebellion: The Logic of Forced Labor and Child Soldiering”

Beber, B. 2009.”The Effect of International Mediation on War Settlement: An Instrumental Variables Approach.”

Beber, B., A. Scacco. 2008. Ẅhat the Numbers Say: A Digit-Based Test for Election Fraud Using New Data from Nigeria.“

Timothy Frye

Professor, Department of Political Science, Columbia University
tmf2@columbia.edu

Tim is the director of the Harriman Institute and professor of Political Science at Columbia University . His research and teaching interests are in comparative politics and political economy with a focus on the former Soviet Union and Eastern

Europe. He is the author of Brokers and Bureaucrats: Building Markets in Russia, and has published articles on property rights, protection rackets, economic reform, presidential power, and trade liberalization in a wide range of academic journals. Current projects include a book manuscript on the politics of economic reform in 25 postcommunist countries from 1990-2004 and articles on property rights and the rule of law drawing on surveys of business elites and the mass public in Russia.

Selected Papers

Building States and Markets After Communism: The Perils of Polarized Democracy. 2010. Cambridge University Press

Donald Green

A. Whitney Griswold Professor of Political Science, Yale University
donald.green@yale.edu

Donald Green is the Director of the Institution for Social and Policy Studies and the A. Whitney Griswold Professor of Political Science at Yale University. He hold a Ph.D. in Political Science from UC Berkeley. Dr. Green studies American Politics with a focus on campaigns and elections. His interests in the field of political behavior include voter turnout, partisanship, and prejudice.

Selected Papers

Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2010. A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Experimental Benchmark. Sociological Methods and Research 39: 256-282.

Gerber, Alan S., Donald P. Green, Edward H. Kaplan, and Holger L. Kern. 2010. Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments. Political Analysis 18:297-315.

Green, Donald P., and Alan S. Gerber. 2010. Introduction to Social Pressure and Voting: New Experimental Evidence. Political Behavior 32(3): 331-336.

Guy Grossman

PhD, Department of Political Science, Columbia University

gsg2102@columbia.edu

Dr. Grossman received his PhD from the Department of Political Science at Columbia University in May 2011 and was a Graduate fellow at Columbia's Center for the Study of Development Strategies and at the Mellon Interdisciplinary Graduate Fellows Program. He studies comparative politics, with a regional focus on Sub-Saharan Africa. His research spans a wide range of issues in the political economy of development, such as political selection, legitimacy and leadership, cooperation and collective action in voluntary associations, accountability and public goods provision, and forms of political participation in developing countries.
In his research Dr. Grossman applies a variety of methods, including Control Randomized Trials (CRT), Social Network Analysis, Behavioral Experiments, and surveys.

Selected Papers

Baldassarri, D., and G. Grossman. 2011. The Impact of Elections on Public Goods Production: Evidence from a Lab-in-the-Field Experiment in Uganda [under review]

Grossman, G, R. Kaplan. 2006. Courage to Refuse, Peace Review: A Journal

Macartan Humphreys

Associate Professor, Department of Political Science, Columbia University

mh2245@columbia.edu

Macartan works on the political economy of development and formal political theory. Ongoing research focuses on civil wars, post conflict development, ethnic politics, natural resource management, political authority and leadership and democratic development. He uses a variety of methods including survey work, lab experimentation, field experimentation, econometric analysis, game theoretic analysis and classical qualitative methods. He has conducted field research in Chad, the Democratic Republic of Congo, Ghana, Haiti, Indonesia, Liberia, Mali, Sao Tome and Principe, Sierra Leone, Senegal, Uganda and elsewhere. He has new series of projects underway examines democratic decision making in post conflict and developing areas.

Selected Papers

2009 "Can Development Aid Contribute to Social Cohesion After Civil War? Evidence from a Field Experiment in Post-Conflict Liberia" (with J Fearon and J Weinstein) American Economic Review

2009 "Field Experiments and the Political Economy of Development" (with J Weinstein) Annual Review of Political Science

Rebecca Morton

Professor, Department of Politics, New York University rbm85@nyu.edu

Dr. Morton's research focuses on voting processes as well as experimental methods. She is the author or co-author of four books and numerous journal articles, which have appeared in noted outlets such as the American Economic Review, American Journal of Political Science, American Political Science Review, Journal of Law and Economics, Journal of Politics, and Review of Economic Studies.

Selected Papers

Morton, R, and K. Williams. Experimental Political Science and the Study of Causality: From Nature to the Lab. Cambridge University Press, 2010.

Morton, R. Methods and models: A guide to the empirical analysis of formal models in political science. Cambridge University Press, 1999.

Cyrus Samii

Assistant Professor of Politics, New York University
cds82@columbia.edu

Dr. Samii's research addresses substantive questions about civil conflict and political development andmethodological questions about study design, measurement, and causal inference. His substantive work tests behavioral theories that inform policies for preventing civil war and promoting durable peace. His current methodological research focuses on semi-parametric methods for analyzing broken experiments and observational studies. His current methodological research focuses on semi-parametric methods for analyzing broken experiments and observational studies.

Selected Papers

Samii, C. 2010. "Do Quotas Exacerbate or Reduce Ethnic Conflict? Evidence from Burundi's Military." Presented at MIT and UCSD.

Gilligan, M., B. Pasquale, and C. Samii. 2010. Civil War and Social Capital: Behavioral Game Evidence from Nepal. Presented at NYU, Stony Brook, and Yale.

King, E., C. Samii, and B. Snilstveit. 2010. Interventions to Promote Social Cohesion in Sub-Saharan Africa. Journal of Development Effectiveness. 2(3):336-370.

Weighting and Augmented Weighting for Causal Inference with Missing Data: New Directions. Presented at Polmeth 2010, Yale, and NYU (updated October 2010).

Laura Paler

PhD Candidate, Columbia University
lbp2106@columbia.edu

Laura Paler researches the political economy of development, with a focus on the resource curse; transparency, information and accountability; and conflict and post-conflict reintegration. To investigate causal relationships at the micro-level, she primarily uses experiments (natural, lab and field) and original survey and behavioral data. Regionally, she focuses on Asia–Indonesia and China in particular–where she conducted 20 months of field research for my dissertation.

Selected Papers

Barron, P, M. Humphreys, L. Paler, and J. Weinstein. 2009. Community-Based Reintegration in Aceh: Assessing the Impacts of BRA-KDP Indonesian Social Development Paper #12.

Paler, L. 2005. China's Legislation Law and the Making of a More Orderly Legislative System. The China Quarterly. vol 182: 301-318.

Alex Scacco

Assistant Professor of Politics, New York University
alex.scacco@nyu.edu

Alexandra Scacco is Assistant Professor in the Department of Politics at New York University. She studies comparative and ethnic politics, with a regional focus on Sub-Saharan Africa. Her book project asks why ordinary people participate in communal violence in contemporary Nigeria. Other ongoing research projects focus on post-conflict peace-building, the causes and consequences of partition, electoral manipulation, effective methods for asking sensitive questions, and the costs and benefits of respondent-driven sampling.

Selected Papers

Beber, B. and A. Scacco. 2008. ”What the Numbers Say: A Digit-Based Test for Election Fraud Using New Data from Nigeria.”

Scacco, A. 2007. Individual Participation in Violent Demonstrations in Nigeria.

Rocio Titinuk

Assistant Professor of Political Science, University of Michigan
titiunik@michigan.edu

Rocio Titiunik works on political methodology and American politics. Her methodological interests center on the validity and limitations of employing experimental and non-experimental research designs to the study of politics. She is particularly interested in causal inference in the study of political institutions. Her current projects focus on incumbency advantage, minority representation and turnout, legislative behavior, and party identification.
Rocio was born and raised in Buenos Aires, Argentina, where she completed her undergraduate education at the Universidad de Buenos Aires. She received her Ph.D. from UC-Berkeley in 2008. She joined the Michigan faculty in September 2010, after spending one year as a postdoctoral fellow.

Selected Papers

Sekhon, J., R. Titiunik. When Natural Experiments Are Neither Natural Nor Experiments: Lessons from the Use of Redistricting to Estimate the Personal Vote." Under review.

Cattaneo, M., S. Galiani, P. Gertler, S. Martinez, R. Titiunik. 2009. Housing, Health and Happiness" American Economic Journal: Economic Policy. 1(1): 75-105.