Identification and Causal Inference (Part I)

Kosuke Imai

Princeton University

June 2011
Empirical Implications of Theoretical Models (EITM)
Summer Institute

What is "Identification"?

- Inference: Learn about what you do not observe (*parameters*) from what you do observe (*data*)
- Identification: How much can we learn about parameters from infinite amount of data?

- Ambiguity vs. Uncertainty
- Identification assumptions vs. Statistical assumptions
- Point identification vs. Partial identification

- FURTHER READING: C. F. Manski. (2007). *Identification for Prediction and Decision*. Harvard University Press.

What is "Causal Inference"?

- Learning about *counterfactuals* from *factuals*
- Potential outcomes framework (Neyman-Holland-Rubin)
  - Units: $i = 1, \ldots, n$
  - Data: Y_i (outcome), T_i (treatment), X_i (pre-treatment covariates)
  - Potential outcomes: $Y_i(t)$ where $Y_i = Y_i(T_i)$

Voters	Contact	Turnout		Age	Gender
i	T_i	$Y_i(1)$	$Y_i(0)$	X_{1i}	X_{2i}
1	1	1	?	20	M
2	0	?	0	55	F
3	0	?	1	40	M
⋮	⋮	⋮	⋮	⋮	⋮
n	1	0	?	62	F

  - (Unit-level) Causal effect: $\tau_i = Y_i(1) - Y_i(0)$
  - Average causal effects: $\frac{1}{n}\sum_{i=1}^{n} \tau_i$ and $\mathbb{E}(\tau_i)$
- Causal inference as a missing data problem

The Key Assumption

- No interference between units:

$$Y_i(T_1, T_2, \ldots, T_n) = Y_i(T_i)$$

- Stable Unit Treatment Value Assumption (SUTVA)
- Potential violations: spill-over effects, carry-over effects, contagion

- Potential outcomes are thought to be fixed for each individual
- J-valued treatment $\longrightarrow$ J potential outcomes for each unit

Causal Effects of Immutable Characteristics

- "No causation without manipulation" (Holland, 1986 *JASA*)
- Immutable characteristics; gender, race, age, etc.
- What does the causal effect of gender mean?

- Causal effect of a female politician on policy outcomes (Chattopadhyay and Duflo, 2004 *QJE*)
- Causal effect of a discussion leader with certain preferences on deliberation outcomes (Humphreyes *et al*. 2006 *WP*)
- Causal effect of a job applicant's gender/race on call-back rates (Bertrand and Mullainathan, 2004 *AER*)

Classical Randomized Experiments

- Units: $i = 1, \ldots, n$
- Treatment: $T_i \in \{0, 1\}$
- Outcome: $Y_i = Y_i(T_i)$
- Complete randomization of the treatment assignment
- Exactly n_1 units receive the treatment
- $n_0 = n - n_1$ units are assigned to the control group
- Assumption: for all $i = 1, \ldots, n$, $\sum_{i=1}^{n} T_i = n_1$ and

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i, \quad \Pr(T_i = 1) = \frac{n_1}{n}$$

Estimation of Average Treatment Effects

- Key idea (Neyman 1923): Randomness comes from treatment assignment (plus sampling for PATE) alone
- Design-based (randomization-based) rather than model-based
- Statistical properties of $\hat{\tau}$ based on design features
- Another important idea (skipped): Fisher's permutation inference

- Estimand $=$ SATE or PATE
- Estimator $=$ Difference-in-means:

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

Sample Inference

- Define $\mathcal{O} \equiv \{Y_i(0), Y_i(1)\}_{i=1}^n$
- Unbiasedness (over repeated treatment assignments):

$$\mathbb{E}(\hat{\tau} \mid \mathcal{O}) \quad = \quad \text{SATE}$$

- Exact variance of $\hat{\tau}$:

$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) \quad = \quad \frac{1}{n}\left(\frac{n_0}{n_1}S_1^2 + \frac{n_1}{n_0}S_0^2 + 2S_{01}\right),$$

where for $t = 0, 1$,

$$S_t^2 \quad = \quad \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(t) - \overline{Y(t)})^2 \quad \text{sample variance of } Y_i(t)$$

$$S_{01} \quad = \quad \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(0) - \overline{Y(0)})(Y_i(1) - \overline{Y(1)}) \quad \text{sample covariance}$$

- The variance is NOT *identifiable*

Population Inference

- Now assume that units are randomly sampled from a population
- Unbiasedness (over repeated sampling):

$$\mathbb{E}\{\mathbb{E}(\hat{\tau} \mid \mathcal{O})\} \quad = \quad \mathbb{E}(\text{SATE}) \quad = \quad \text{PATE}$$

- Exact variance

$$\begin{aligned} \mathbb{V}(\hat{\tau}) &= \mathbb{V}(\mathbb{E}(\hat{\tau} \mid \mathcal{O})) + \mathbb{E}(\mathbb{V}(\hat{\tau} \mid \mathcal{O})) \\ &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0} \end{aligned}$$

where σ_t^2 is the population variance of $Y_i(t)$ for $t = 0, 1$

Relationships with Regression

- Simple regression: $Y_i = \alpha + \beta T_i + \epsilon_i$
- Potential outcomes: $Y_i(T_i) = \alpha + \beta T_i + \epsilon_i(T_i)$
- Causal effects: $\tau_i = \beta + (\epsilon_i(1) - \epsilon_i(0))$ and $\tau = \beta$
- Algebraic equivalence: $\hat{\beta}_{LS} = \hat{\tau}$
- Bias of usual standard error:

$$\frac{(n_1 - n_0)(n-1)}{n_1 n_0 (n-2)}(\sigma_1^2 - \sigma_0^2)$$

- Bias of "robust" standard error:

$$-\left(\frac{\sigma_1^2}{n_1^2} + \frac{\sigma_0^2}{n_0^2}\right)$$

- Heteroskedasticity: $\sigma_1 \neq \sigma_0$

Cluster Randomized Experiments

- Units: $i = 1, 2, \ldots, n_j$
- Clusters of units: $j = 1, 2, \ldots, m$
- Treatment at cluster level: $T_j \in \{0, 1\}$
- Outcome: $Y_{ij} = Y_{ij}(T_j)$
- Random assignment: $(Y_{ij}(1), Y_{ij}(0)) \perp\!\!\!\perp T_j$
- Estimands at unit level:

$$\begin{aligned} \text{SATE} &\equiv \frac{1}{\sum_{j=1}^{m} n_j} \sum_{j=1}^{m} \sum_{i=1}^{n_j} (Y_{ij}(1) - Y_{ij}(0)) \\ \text{PATE} &\equiv \mathbb{E}(Y_{ij}(1) - Y_{ij}(0)) \end{aligned}$$

- Random sampling of clusters and units

Merits and Limitations of CREs

- Interference between units within a cluster is allowed
- Assumption: No interference between units of different clusters
- Often easy to implement: Mexican health insurance experiment

- Opportunity to estimate the spill-over effects
- D. W. Nickerson. Spill-over effect of get-out-the-vote canvassing within household (*APSR*, 2008)

- Limitations:
  1. A large number of possible treatment assignments
  2. Loss of statistical power

Design-Based Inference

- For simplicity, assume equal cluster size, i.e., $n_j = n$ for all j
- The difference-in-means estimator:

$$\hat{\tau} \equiv \frac{1}{m_1}\sum_{j=1}^{m} T_j \overline{Y}_j - \frac{1}{m_0}\sum_{j=1}^{m}(1-T_j)\overline{Y}_j$$

  where $\overline{Y}_j \equiv \sum_{i=1}^{n_j} Y_{ij}/n_j$
- Easy to show $\mathbb{E}(\hat{\tau} \mid \mathcal{O}) = \text{SATE}$ and thus $\mathbb{E}(\hat{\tau}) = \text{PATE}$
- Exact population variance:

$$\text{Var}(\hat{\tau}) = \frac{\text{Var}(\overline{Y_j(1)})}{m_1} + \frac{\text{Var}(\overline{Y_j(0)})}{m_0}$$

- Intracluster correlation coefficient ρ_t:

$$\text{Var}(\overline{Y_j(t)}) = \frac{\sigma_t^2}{n}\{1 + (n-1)\rho_t\} \leq \sigma_t^2$$

Relationship with Cluster Standard Error in Regression

- Cluster-adjusted robust variance estimator:

$$\widehat{\mathbb{V}((\hat{\alpha},\hat{\beta}) \mid T)} = \left(\sum_{j=1}^{m} X_j^\top X_j\right)^{-1} \left(\sum_{j=1}^{m} X_j^\top \hat{\epsilon}_j \hat{\epsilon}_j^\top X_j\right) \left(\sum_{j=1}^{m} X_j^\top X_j\right)^{-1}$$

  where in this case $X_j = [1\, T_j]$ is an $n_j \times 2$ matrix and $\hat{\epsilon}_j = (\hat{\epsilon}_{1j}, \ldots, \hat{\epsilon}_{n_j j})$ is a column vector of length n_j

- Design-based evaluation (assume $n_j = n$ for all j):

$$\text{Finite Sample Bias} = -\left(\frac{\mathbb{V}(\overline{Y_j(1)})}{m_1^2} + \frac{\mathbb{V}(\overline{Y_j(0)})}{m_0^2}\right)$$

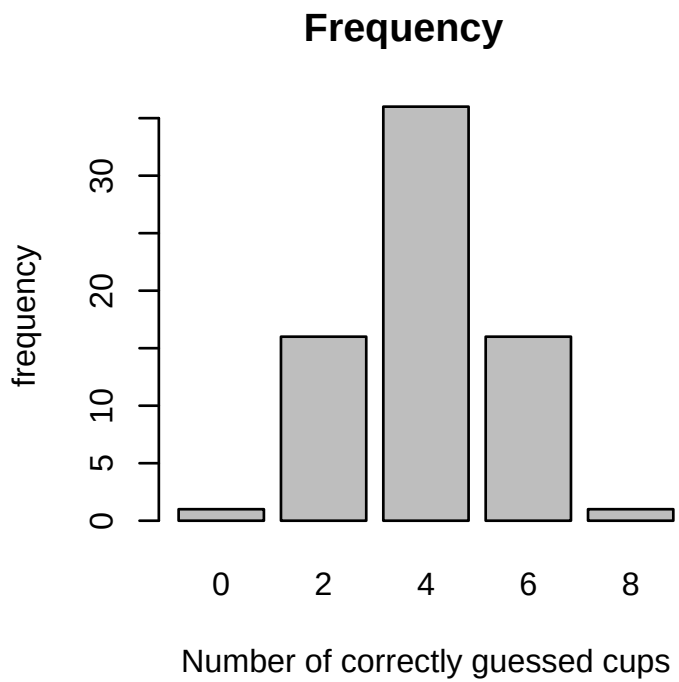
- Bias vanishes asymptotically as $m \to \infty$ with n fixed
- Clustering should be done at the level of treatment assignment

Fisher's Lady Tasting Tea

- Does tea taste different depending on whether the tea was poured into the milk or whether the milk was poured into the tea?
- 8 cups; $n = 8$
- Randomly choose 4 cups into which pour the tea first ($T_i = 1$)
- Null hypothesis: the lady cannot tell the difference
- $H_0 : Y_i(1) = Y_i(0)$ for all $i = 1, \ldots, 8$
- Statistic: the number of correctly classified cups
- The lady classified all 8 cups correctly!
- Did this happen by chance?

Permutation Inference

cups	guess	actual	scenarios	...
1	M	M	T	T
2	T	T	T	T
3	T	T	T	T
4	M	M	T	M
5	M	M	M	M
6	T	T	M	M
7	T	T	M	T
8	M	M	M	M
correctly guessed		8	4	6



- ${}_8C_4 = 70$ ways to do this and each arrangement is equally likely
- What is the p-value?
- No assumption, but the sharp null may be of little interest

Formalization of the "Lady Tasting Tea"

- Sharp null hypothesis, $H_0 : Y_i(1) - Y_i(0) = \tau_0$ for all i
- Test statistic: $f(\mathbf{Y}, \mathbf{T}, \tau_0)$ for some function $f(\cdot, \cdot, \cdot)$
- Exact p-value: $p_{\text{exact}} \equiv \Pr(f(\mathbf{Y}, \mathbf{t}^{obs}, \tau_0) \leq f(\mathbf{Y}, \mathbf{T}, \tau_0))$ under H_0
- Nonparametric, exact, computationally intensive
- Commonly used test statistics: sum of successes, sum of ranks
- Exact population inference without the constant additive treatment effect assumption (Wilcoxon's rank-sum statistic):

$$F_{Y(1)}(y) \quad = \quad F_{Y(0)}(y + \tau)$$

- Example: California Alphabet Lottery (Ho and Imai *JASA*, 2006)
- Inference with complex treatment assignment mechanisms

Exact Confidence Sets and Population Inference

- Invert the exact test
- Collect null values that cannot be rejected by α-level test
- Yields $(1-\alpha) \times 100\%$ confidence set
- Restrictive assumption: Constant additive treatment effect

$$A_\alpha \quad = \quad \{\tau_0 : \Pr(f(\mathbf{Y}, \mathbf{t}^{obs}, \tau_0) \leq f(\mathbf{Y}, \mathbf{T}, \tau_0)) \geq \alpha\}.$$

- Coverage probability equals exactly $(1-\alpha)$ over *repeated (hypothetical) experiments*

- Confidence intervals for the causal effect estimate of one observation

List Experiment: An Example

- The 1991 National Race and Politics Survey (Sniderman et al.)
- Randomize the sample into the treatment and control groups
- The script for the control group

Now I'm going to read you three things that sometimes make people angry or upset. After I read all three, just tell me HOW MANY of them upset you. (I don't want to know which ones, just how many.)

(1) the federal government increasing the tax on gasoline;
(2) professional athletes getting million-dollar-plus salaries;
(3) large corporations polluting the environment.

List Experiment: An Example

- The 1991 National Race and Politics Survey (Sniderman et al.)
- Randomize the sample into the treatment and control groups
- The script for the treatment group

Now I'm going to read you four things that sometimes make people angry or upset. After I read all four, just tell me HOW MANY of them upset you. (I don't want to know which ones, just how many.)

(1) the federal government increasing the tax on gasoline;
(2) professional athletes getting million-dollar-plus salaries;
(3) large corporations polluting the environment;
(4) a black family moving next door to you.

Identification Assumptions

1. No Design Effect: The inclusion of the sensitive item does not affect answers to control items
2. No Liars: Answers about the sensitive item are truthful

Under these assumptions, difference-in-means estimator is unbiased

Potential Outcomes Framework

- Define a type of each respondent by
  - total number of yes for control items $Y_i(0)$
  - truthful answer to the sensitive item Z_i^*
- Under the above assumptions, $Y_i(1) = Y_i(0) + Z_i^*$
- A total of $(2 \times (J + 1))$ types
- Example: three control items $(J = 3)$

Y_i	Treatment group	Control group
4	(3,1)	
3	(2,1) (3,0)	(3,1) (3,0)
2	(1,1) (2,0)	(2,1) (2,0)
1	(0,1) (1,0)	(1,1) (1,0)
0	(0,0)	(0,1) (0,0)

Potential Outcomes Framework

- Define a type of each respondent by
  - total number of yes for control items $Y_i(0)$
  - truthful answer to the sensitive item Z_i^*
- Under the above assumptions, $Y_i(1) = Y_i(0) + Z_i^*$
- A total of $(2 \times (J + 1))$ types
- Example: three control items ($J = 3$)

Y_i	Treatment group	Control group
4	(3,1)	
3	(2,1) (3,0)	(3,1) (3,0)
2	(1,1) (2,0)	(2,1) (2,0)
1	~~(0,1)~~ ~~(1,0)~~	(1,1) ~~(1,0)~~
0	~~(0,0)~~	~~(0,1)~~ ~~(0,0)~~

- *Joint distribution* of $(Y_i(0), Z_i^*)$ is identified

Likelihood Inference

- $g(x, \delta) = \Pr(Z^*_{i,J+1} = 1 \mid X_i = x)$
- $h_z(y; x, \psi_z) = \Pr(Y_i(0) = y \mid X_i = x, Z_{i,J+1}(0) = z)$
- The "Mixture Model" likelihood function:

$$\prod_{i \in \mathcal{J}(1,0)} (1 - g(X_i, \delta)) h_0(0; X_i, \psi_0) \prod_{i \in \mathcal{J}(1,J+1)} g(X_i, \delta) h_1(J; X_i, \psi_1)$$

$$\times \prod_{y=1}^{J} \prod_{i \in \mathcal{J}(1,y)} \{g(X_i, \delta) h_1(y-1; X_i, \psi_1) + (1 - g(X_i, \delta)) h_0(y; X_i, \psi_0)\}$$

$$\times \prod_{y=0}^{J} \prod_{i \in \mathcal{J}(0,y)} \{g(X_i, \delta) h_1(y; X_i, \psi_1) + (1 - g(X_i, \delta)) h_0(y; X_i, \psi_0)\}$$

where $\mathcal{J}(t, y)$ represents a set of respondents with $(T_i, Y_i) = (t, y)$

- Maximizing this function is difficult

Missing Data Formulation

- Consider $Z^*_{i,J+1}$ as (partially) missing data
- The complete-data likelihood has a much simpler form:

$$\prod_{i=1}^{N} \left\{ g(X_i, \delta) h_1(Y_i - 1; X_i, \psi_1)^{T_i} h_1(Y_i; X_i, \psi_1)^{1-T_i} \right\}^{Z_{i,J+1}(0)}$$

$$\times \{(1 - g(X_i, \delta)) h_0(Y_i; X_i, \psi_0)\}^{1-Z_{i,J+1}(0)}$$

- The EM algorithm (Dempster, Laird, and Rubin):
  1. E-step: Compute

$$Q(\theta \mid \theta^{(t)}) \equiv \mathbb{E}\{l_n(\theta \mid Y_{obs}, Y_{mis}) \mid Y_{obs}, \theta^{(t)}\}$$

     where $l_n(\theta \mid Y_{obs}, Y_{mis})$ is the complete-data log-likelihood
  2. M-step: Find

$$\theta^{(t+1)} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, Q(\theta \mid \theta^{(t)})$$

- Monotone convergence: $l_n(\theta^{(t+1)} \mid Y_{obs}) \geq l_n(\theta^{(t)} \mid Y_{obs})$
- Stable, no derivative required

EM Algorithm for List Experiments

1. E-step:

$$
\begin{aligned}
w_i &= \mathbb{E}(Z^*_{i,J+1} \mid Y_i = y, T_i = t, X_i = x) \\
&= \begin{cases} 0 & \text{if } (t,y) = (1,0) \\ 1 & \text{if } (t,y) = (1, J+1) \\ \dfrac{g(x,\delta) h_1(y-t;x,\psi_1)}{g(x,\delta) h_1(y-t;x,\psi_1) + (1-g(x,\delta)) h_0(y;x,\psi_0)} & \text{otherwise} \end{cases}
\end{aligned}
$$

2. M-step:
   - weighted regression for $g(x, \delta)$
   - weighted regression for $h_z(y; x, \psi_z)$
   - weights are w_i and $1 - w_i$

3. Connection to data augmentation in Bayesian MCMC
   - Sample Z^* given $(\delta, \psi_z, Y, T, X)$
   - Sample (δ, ψ_z) given (Y, T, X, Z^*)

Hypothesis Test for List Experiment Failures

- Under the null hypothesis of no design effect and no liar, we

$$\pi_1 = \Pr(\text{type} = (y, 1)) = \Pr(Y_i \leq y \mid T_i = 0) - \Pr(Y_i \leq y \mid T_i = 1) \geq 0$$
$$\pi_0 = \Pr(\text{type} = (y, 0)) = \Pr(Y_i \leq y \mid T_i = 1) - \Pr(Y_i < y \mid T_i = 0) \geq 0$$

  for each y
- Alternative hypothesis: At least one is negative
- Test of two stochastic dominance relationships

- Watch out for multiple testing
- Failure to reject the null may arise from the lack of power

Modeling Ceiling and Floor Effects

- Potential liars:

Y_i	Treatment group	Control group
4	(3,1)	
3	(2,1) (3,0) (3,1)*	(3,1) (3,0)
2	(1,1) (2,0)	(2,1) (2,0)
1	(0,1) (1,0)	(1,1) (1,0)
0	(0,0) (0,1)*	(0,1) (0,0)

- Proposed strategy: model ceiling and/or floor effects under an additional assumption
- Identification assumption: conditional independence between items given covariates
- ML estimation can be extended to this situation
- More on list experiments: Imai (2011, *JASA*), Blair and Imai (2011)

Key Points

- Identification and inference
- Potential outcomes framework of causal inference
- Design-based inference
- Connections to regression models
- Causal inference as a missing data problem

Important Topics in the Methodological Literature

- Identification of heterogenous treatment effect
- Derivation of individualized treatment rules
- Extrapolation from an experimental sample
- Identification of spill-over effects
- Identification of causal mechanisms

Identification of the Average Treatment Effect

- Assumption 1: Overlap (i.e., no extrapolation)

$$0 < \Pr(T_i = 1 \mid X_i = x) < 1 \text{ for any } x \in \mathcal{X}$$

- Assumption 2: Ignorability (exogeneity, unconfoundedness, no omitted variable, selection on observables, etc.)

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid X_i = x \text{ for any } x \in \mathcal{X}$$

- Conditional expectation function: $\mu(t, x) = \mathbb{E}(Y_i(t) \mid T_i = t, X_i = x)$
- Regression-based Estimator:

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \{\hat{\mu}(1, X_i) - \hat{\mu}(0, X_i)\}$$

- Delta method is pain, but simulation is easy (Zelig)
- FURTHER READING: Imbens (2004, *Rev. Econ. Stat.*)

Partial Identification

- A special case with binary outcome

$$[-\Pr(Y_i = 0 \mid T_i = 1, X_i = x)\pi(x) - \Pr(Y_i = 1 \mid T_i = 0, X_i = x)\{1 - \pi(x)\},$$
$$\Pr(Y_i = 1 \mid T_i = 1, X_i = x)\pi(x) + \Pr(Y_i = 0 \mid T_i = 0, X_i = x)\{1 - \pi(x)\}]$$

  where $\pi(x) = \Pr(T_i = 1 \mid X_i = x)$ is called propensity score

- The width of the bounds: 1 "A glass is half empty/full"

- Monotone treatment selection (Manski):

$$[\Pr(Y_i = 1 \mid T_i = 1, X_i = x)\pi(x) - \Pr(Y_i = 1 \mid X_i = x),$$
$$\Pr(Y_i = 1 \mid X_i = x) - \Pr(Y_i = 1 \mid T_i = 0, X_i = x)(1 - \pi(x))].$$

- The width of the bounds: $\Pr(Y_i \mid X_i = x)$

- FURTHER READING: Manski (2007, *Harvard UP*)

Sensitivity Analysis

- Consider a simple pair-matching of treated and control units
- Assumption: treatment assignment is "random"
- Difference-in-means estimator
- Question: How large a departure from the key (untestable) assumption must occur for the conclusions to no longer hold?
- Rosenbaum's sensitivity analysis: for any pair j,

$$\frac{1}{\Gamma} \leq \frac{\Pr(T_{1j} = 1)/\Pr(T_{1j} = 0)}{\Pr(T_{2j} = 1)/\Pr(T_{2j} = 0)} \leq \Gamma$$

- Under ignorability, $\Gamma = 1$ for all j
- How do the results change as you increase Γ?
- Limitations of sensitivity analysis
- FURTHER READING: P. Rosenbaum. *Observational Studies*.

Covariate Adjustments in Experiments

- Adjusting for covariates may lead to efficiency gain
- Dangers of post-randomization covariate adjustment
  - Bias due to statistical methods
  - Bias due to post-hoc analysis

- Make adjustments *before* the randomization of treatment
- Employ design-based inference rather than model-based

Randomized-block Design

- Form a group of units based on the pre-treatment covariates so that the observations within each block are similar
- Complete randomization of the treatment within each block
- Inference based on the weighted average of within-block estimates

- Blocking can never hurt; unbiased and no less efficient
- Difference in asymptotic variance:

$$\mathbb{V}(\overline{Y(1)}_b + \overline{Y(0)}_b) \geq 0$$

where $\overline{Y(t)}_b$ is the within-block mean of $Y_i(t)$

- Efficiency gain is greater if across-block heterogeneity is greater

Matched-Pair Design

- Blocking where the size of all blocks is 2
- Create pairs of units based on the pre-treatment covariates so that within the units within a pair are similar to each other
- Randomly assign the treatment within each matched-pair
- Inference based on the average of within-pair differences

- Difference in variances:

$$\frac{1}{n/2}\text{Cov}(Y_{ij}(1), Y_{i'j}(0))$$

- Greater within-pari similarity leads to greater efficiency
- Multivariate blocking/matching methods

Matching as Nonparametric Preprocessing

- Assume exogeneity holds: matching does not solve endogeneity
- Need to model $\mathbb{E}(Y_i \mid T_i, X_i)$
- Non-parametric regression – curse of dimensionality
- Parametric regression – functional-form/distributional assumptions
- Preprocess the data so that treatment and control groups are similar to each other w.r.t. the observed pre-treatment covariates

- Goal of matching: achieve balance

$$\widetilde{F}(X \mid T = 1) \quad = \quad \widetilde{F}(X \mid T = 0)$$

  where $\widetilde{F}(\cdot)$ is the *empirical* distribution
- Reduced model dependence: minimal role of statistical modeling

The Role of Propensity Score

- The probability of receiving the treatment:

$$\pi(X_i) \equiv \Pr(T_i = 1 \mid X_i)$$

- The balancing property under exogeneity:

$$T_i \perp\!\!\!\perp X_i \mid \pi(X_i)$$

- Exogeneity given the propensity score:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i \mid \pi(X_i)$$

- Dimension reduction
- But, true propensity score is unknown: propensity score tautology

Classical Matching Techniques

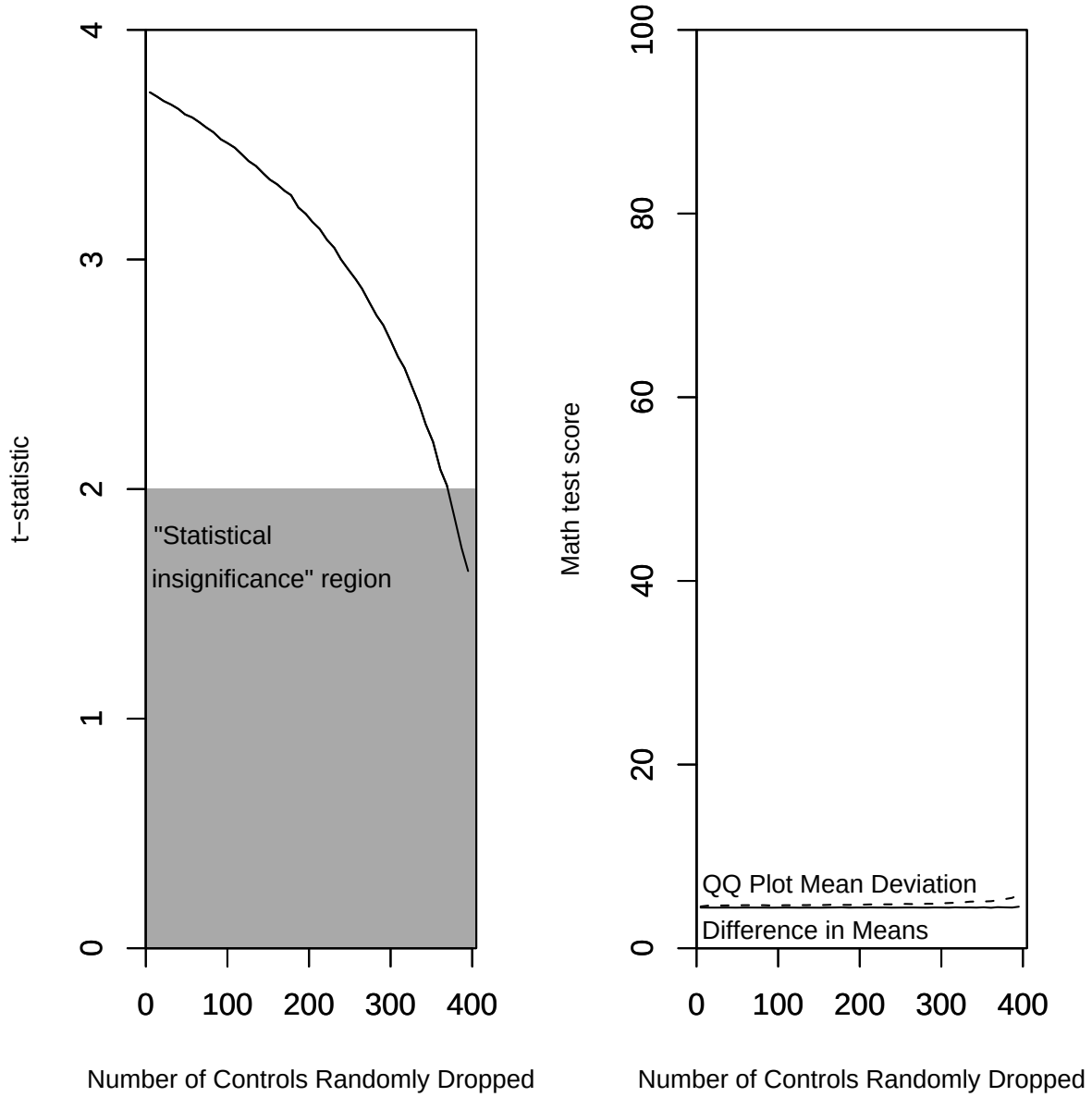
- Exact matching
- Mahalanobis distance matching: $\sqrt{(X_i - X_j)^\top \widetilde{\Sigma}^{-1} (X_i - X_j)}$
- Propensity score matching
- One-to-one, one-to-many, and subclassification
- Matching with caliper
- Which matching method to choose?
- Whatever gives you the "best" balance!
- Importance of substantive knowledge: propensity score matching with exact matching on key confounders
- FURTHER READING: Rubin (2006). *Matched Sampling for Causal Effects* (Cambridge UP)

How to Check Balance

- Success of matching method depends on the resulting balance
- How should one assess the balance of matched data?
- Ideally, compare the joint distribution of all covariates for the matched treatment and control groups
- In practice, this is impossible when X is high-dimensional
- Check various lower-dimensional summaries; (standardized) mean difference, variance ratio, empirical CDF, etc.

- Frequent use of balance test
  - t test for difference in means for each variable of X
  - other test statistics; e.g., χ^2, F, Kolmogorov-Smirnov tests
  - statistically insignificant test statistics as a justification for the adequacy of the chosen matching method and/or a stopping rule for maximizing balance

An Illustration of Balance Test Fallacy



Problems with Hypothesis Tests as Stopping Rules

- Balance test is a function of both balance and statistical power
- The more observations dropped, the less power the tests have
- t-test is affected by factors other than balance,

$$\frac{\sqrt{n_m}(\overline{X}_{mt} - \overline{X}_{mc})}{\sqrt{\frac{s^2_{mt}}{r_m} + \frac{s^2_{mc}}{1-r_m}}}$$

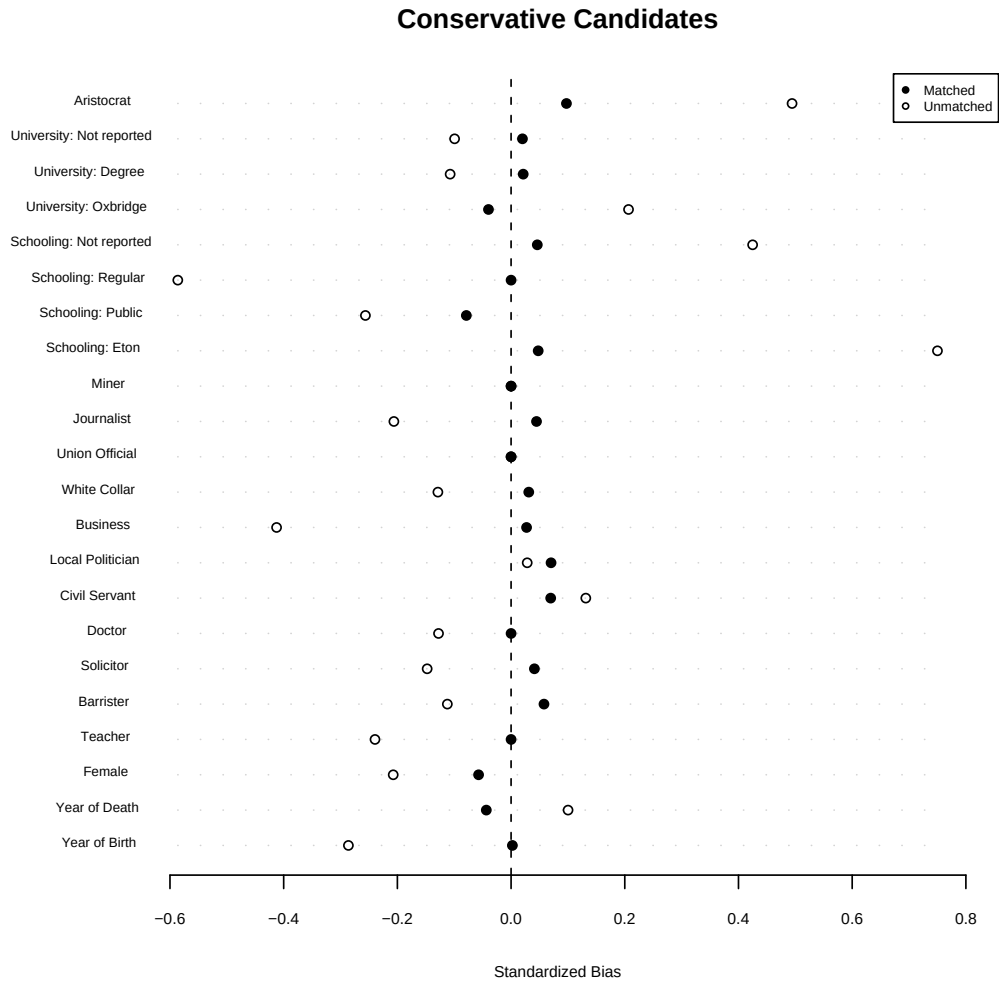
  - $\overline{X}_{mt}$ and $\overline{X}_{mc}$ are the sample means
  - s^2_{mt} and s^2_{mc} are the sample variances
  - n_m is the total number of remaining observations
  - r_m is the ratio of remaining treated units to the total number of remaining observations

- Balance is a characteristic of sample rather than population

An Empirical Example

- "Value of political power" by Eggers and Hainmueller (APSR)

Figure 3: Covariate Balance Before and After Matching



Inverse Propensity Score Weighting

- Matching is inefficient because it throws away data
- Weighting by inverse propensity score

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\hat{\pi}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{\pi}(X_i)}\right)$$

- An improved weighting scheme:

$$\frac{\sum_{i=1}^{n}\{T_i Y_i/\hat{\pi}(X_i)\}}{\sum_{i=1}^{n}\{T_i/\hat{\pi}(X_i)\}} - \frac{\sum_{i=1}^{n}\{(1-T_i)Y_i/(1-\hat{\pi}(X_i))\}}{\sum_{i=1}^{n}\{(1-T_i)/(1-\hat{\pi}(X_i))\}}$$

- Unstable when some weights are extremely small

Efficient Doubly-Robust Estimators

- The estimator by Robins *et al.* :

$$\hat{\tau}_{DR} \equiv \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(1, X_i) + \frac{1}{n} \sum_{i=1}^{n} \frac{T_i(Y_i - \hat{\mu}(1, X_i))}{\hat{\pi}(X_i)} \right\} - \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(0, X_i) + \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - T_i)(Y_i - \hat{\mu}(0, X_i))}{1 - \hat{\pi}(X_i)} \right\}$$

- Consistent if either the propensity score model or the outcome model is correct
- (Semiparametrically) Efficient
- FURTHER READING: Lunceford and Davidian (2004, *Stat. in Med.*)

- Estimator can behave poorly when both models are incorrect (especially if weights are highly variable)
- Recent work on stabilized weights

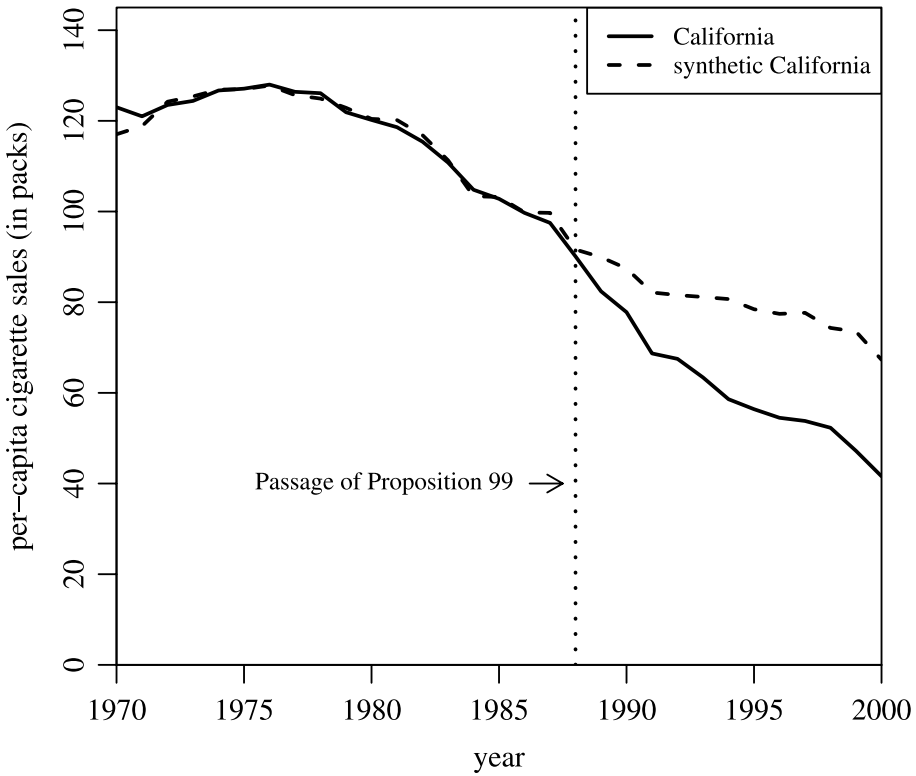
Weighting for Panel Data

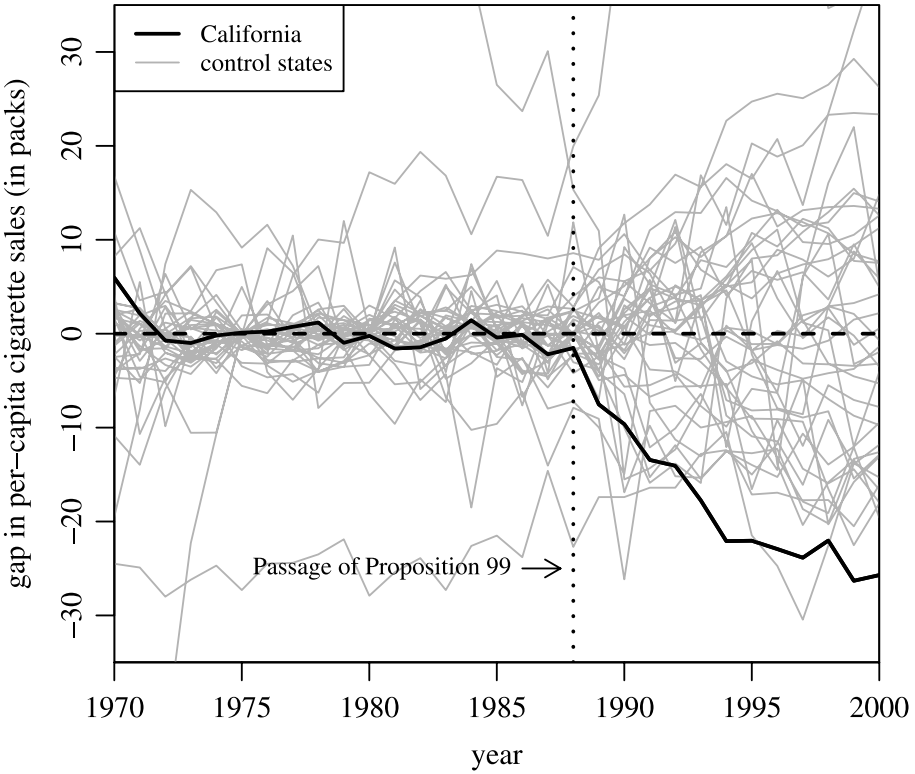
- Synthetic Control Method: Abadie *et al.* (2010, *JASA*)
- Setting: one treated unit, observations before and after the treatment assignment
- Idea: Use the weighted average of control units to estimate the counterfactual for the treated unit

$$\widehat{Y_{1t}(0)} = \sum_{i=2}^{N} w_i Y_{it} \quad \text{for any } t$$

- Key assumptions:
  1. No interference between units
  2. Weights exist
  3. Extrapolation of weights is valid

Permutation Inference for Synthetic Control Method





Matching vs. Regression

- They are based on the same qualitative assumption: ignorability
- Neither method solves endogeneity
- Matching is nonparametric: more flexible
- It also forces researchers to look at covariate balance
- Importance of the overlap assumption
- Matching and regression can be used together: matching as nonparametric preprocessing for reducing model dependence

- Matching (and the potential outcomes framework in general) clarifies what information is used to "impute" counterfactual outcomes

Matching Representation of Simple Regression

- Simple regresison: $Y_i = \alpha + \beta X_i + \epsilon_i$
- Binary treatment: $X_i \in \{0, 1\}$
- Matching representation:

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{Y_i(1)} - \widehat{Y_i(0)} \right)$$

where

$$\widehat{Y_i(1)} = \begin{cases} Y_i & \text{if } X_i = 1 \\ \frac{1}{N_1} \sum_{i'=1}^{N} X_{i'} Y_{i'} & \text{if } X_i = 0 \end{cases}$$

$$\widehat{Y_i(0)} = \begin{cases} \frac{1}{N_0} \sum_{i'=1}^{N} (1 - X_{i'}) Y_{i'} & \text{if } X_i = 1 \\ Y_i & \text{if } X_i = 0 \end{cases}$$

Matching Representation of Fixed Effects Regression

- Simple fixed effects regression: $Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$
- Binary treatment: $X_i \in \{0, 1\}$
- Matching representation: Prop. 1 of Imai and Kim (2011)

$$\hat{\beta}^{FE} = \frac{1}{K}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)}\right)\right\}$$

where

$$\widehat{Y_{it}(x)} = \begin{cases} Y_{it} & \text{if } X_{it} = x \\ \frac{1}{T-1}\sum_{t' \neq t} Y_{it'} & \text{if } X_{it} = 1 - x \end{cases} \quad \text{for } x = 0, 1,$$

$$K = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{X_{it} \cdot \frac{1}{T-1}\sum_{t' \neq t}(1 - X_{it'}) + (1 - X_{it}) \cdot \frac{1}{T-1}\sum_{t' \neq t} X_{it'}\right\}.$$

Matching and Weighted Fixed Effects Estimator

- A more natural *unadjusted* matching estimator (Prop. 2):

$$\hat{\beta}^M = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)} \right)$$

where

$$\widehat{Y_{it}(1)} = \begin{cases} Y_i & \text{if } X_{it} = 1 \\ \frac{\sum_{t'=1}^{T} X_{it'} Y_{it'}}{\sum_{t'=1}^{T} X_{it'}} & \text{if } X_{it} = 0 \end{cases}$$

$$\widehat{Y_{it}(0)} = \begin{cases} \frac{\sum_{t'=1}^{T} (1 - X_{it'}) Y_{it'}}{\sum_{t'=1}^{T} (1 - X_{it'})} & \text{if } X_{it} = 1 \\ Y_{it} & \text{if } X_{it} = 0 \end{cases}$$

- Equivalent to the weighted fixed effects regression where the weights are the inverse of the estimated propensity score:

$$W_{it} \equiv \begin{cases} \frac{T}{\sum_{t'=1}^{T} X_{it'}} & \text{if } X_{it} = 1, \\ \frac{T}{\sum_{t'=1}^{T} (1 - X_{it'})} & \text{if } X_{it} = 0. \end{cases}$$

General Equivalence Result (Theorem 1)

- Consider a general class of unadjusted matching:

$$\widehat{Y_{it}(1)} = \begin{cases} Y_{it} & \text{if } X_{it} = 1 \\ \sum_{t'=1}^{T} v_{it}^{it'} X_{it'} Y_{it'} & \text{if } X_{it} = 0 \end{cases}$$

$$\widehat{Y_{it}(0)} = \begin{cases} \sum_{t'=1}^{T} v_{it}^{it'} (1 - X_{it'}) Y_{it'} & \text{if } X_{it} = 1 \\ Y_{it} & \text{if } X_{it} = 0 \end{cases}$$

  where $\sum_{t'=1}^{T} v_{it}^{it'} X_{it'} = \sum_{t'=1}^{T} v_{it}^{it'} (1 - X_{it'}) = 1$.

- Example: estimated inverse-propensity score weighting

$$v_{it}^{it'} = \begin{cases} \dfrac{(1-\hat{\pi}(Z_{it'}))^{-1}}{\sum_{t^*=1}^{T} (1-\hat{\pi}(Z_{it^*}))^{-1}(1-X_{it^*})} & \text{if } X_{it} = 1 \\ \dfrac{\hat{\pi}(Z_{it'})^{-1}}{\sum_{t^*=1}^{T} \hat{\pi}(Z_{it^*})^{-1} X_{it^*}} & \text{if } X_{it} = 0 \end{cases}$$

- The one-way fixed effects regression weights can be derived from any non-negative (normalized) weight $v_{it}^{it'}$.

What about the Two-Way Fixed Effects Estimator?

- The Model:

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \epsilon_{it}$$

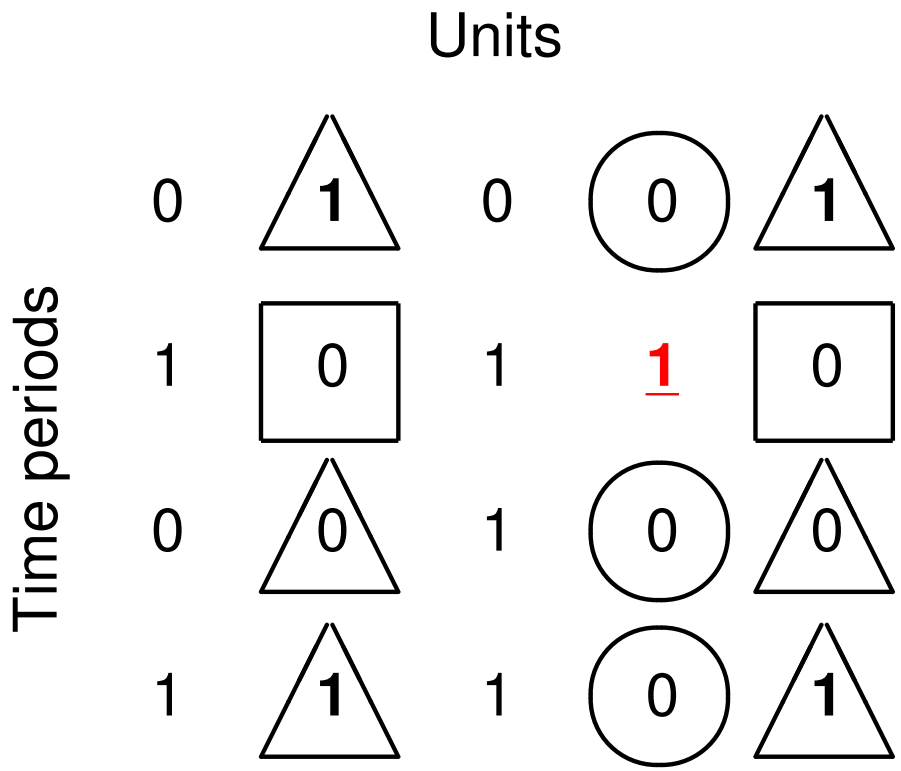
  where a restriction such as $\sum_{t=1}^{T} \gamma_t = 0$ is needed

- The matching representation (Prop. 3):

$$\widehat{Y_{it}(x)} = \begin{cases} Y_{it} & \text{if } X_{it} = x \\ \frac{\sum_{t' \neq t} Y_{it'}}{T-1} + \frac{\sum_{i' \neq i} Y_{i't}}{N-1} - \frac{\sum_{i' \neq i} \sum_{t' \neq t} Y_{i't'}}{(T-1)(N-1)} & \text{if } X_{it} = 1 - x \end{cases}$$

$$K = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ X_{it} \left(\frac{\sum_{t'=1}^{T} (1 - X_{it'})}{T-1} + \frac{\sum_{i'=1}^{N} (1 - X_{i't})}{N-1} - \frac{\sum_{i' \neq i} \sum_{t' \neq t} (1 - X_{i't'})}{(T-1)(N-1)} \right) + (1 - X_{it}) \left(\frac{\sum_{t'=1}^{T} X_{it'}}{T-1} + \frac{\sum_{i'=1}^{N} X_{i't}}{N-1} - \frac{\sum_{i' \neq i} \sum_{t' \neq t} X_{i't'}}{(T-1)(N-1)} \right) \right\}.$$

Can We Improve It? (Prop. 4)



- Implications for applied data analysis:
  1. Two-way fixed effects estimator is difficult to justify from a causal inference perspective
  2. One-way fixed effects can be improved by the weighted one-way fixed effects based on propensity scores

Recent Developments in the Methodological Literature

- The main problem of matching/weighting: balance checking
- Skip balance checking all together
- Specify a balance metric and optimize it

- Optimal matching
- Genetic matching
- Fine matching
- Coarsened exact matching
- Entropy balancing
- SVM matching

- Matching and weighting in panel data settings
- Dynamic treatment regimes via inverse propensity score weighting
- Synthetic control method

Coping with Endogeneity in Observational Studies

- Selection bias in observational studies
- Two research design strategies:
  1. Find a plausibly exogenous treatment
  2. Find a plausibly exogenous instrument
- A valid instrument satisfies the following conditions
  1. Exogenously assigned – no confounding
  2. It monotonically affects treatment
  3. It affects outcome only through treatment – no direct effect
- Challenge: plausibly exogenous instruments with no direct effect tends to be weakly

Identifying Causal Mechanisms

- Randomized experiments as **gold standard** for causal inference
- But, experiments are a **black box**
- Can only tell *whether* the treatment causally affects the outcome
- Not *how* and *why* the treatment affects the outcome
- Qualitative research uses process tracing

- How can quantitative research be used to identify causal mechanisms?
- Causal mediation analysis: direct vs. indirect effects
- Identification of causal mechanisms is more difficult than that of causal effects
- "Causal chain approach" does not work

Partial Compliance in Randomized Experiments

- Unable to force all experimental subjects to take the (randomly) assigned treatment/control
- Intention-to-Treat (ITT) effect $\neq$ treatment effect
- Selection bias: self-selection into the treatment/control groups

- Political information bias: effects of campaign on voting behavior
- Ability bias: effects of education on wages
- Healthy-user bias: effects of exercises on blood pressure

- Encouragement design: randomize the encouragement to receive the treatment rather than the receipt of the treatment itself

Potential Outcomes Notation

- Randomized encouragement: $Z_i \in \{0, 1\}$
- Potential treatment variables: $(T_i(1), T_i(0))$
  1. $T_i(z) = 1$: would receive the treatment if $Z_i = z$
  2. $T_i(z) = 0$: would not receive the treatment if $Z_i = z$
- Observed treatment receipt indicator: $T_i = T_i(Z_i)$
- Observed and potential outcomes: $Y_i = Y_i(Z_i, T_i(Z_i))$
- Can be written as $Y_i = Y_i(Z_i)$

- No interference assumption for $T_i(Z_i)$ and $Y_i(Z_i, T_i)$
- Randomization of encouragement:

$$(Y_i(1), Y_i(0), T_i(1), T_i(0)) \perp\!\!\!\perp Z_i$$

- But $(Y_i(1), Y_i(0)) \not\perp\!\!\!\perp T_i \mid Z_i = z$, i.e., selection bias

Principal Stratification Framework

- Imbens and Angrist (1994, *Econometrica*); Angrist, Imbens, and Rubin (1996, *JASA*)
- Four principal strata (latent types):
  - compliers $(T_i(1), T_i(0)) = (1, 0)$,
  - non-compliers $\begin{cases} always-takers & (T_i(1), T_i(0)) = (1, 1), \\ never-takers & (T_i(1), T_i(0)) = (0, 0), \\ defiers & (T_i(1), T_i(0)) = (0, 1) \end{cases}$
- Observed and principal strata:

	$Z_i = 1$	$Z_i = 0$
$T_i = 1$	Complier/Always-taker	Defier/Always-taker
$T_i = 0$	Defier/Never-taker	Complier/Never-taker

Instrumental Variables and Causality

- Randomized encouragement as an instrument for the treatment
- Two additional assumptions
  1. Monotonicity: No defiers

$$T_i(1) \geq T_i(0) \quad \text{for all } i.$$

  2. Exclusion restriction: Instrument (encouragement) affects outcome only through treatment

$$Y_i(1, t) = Y_i(0, t) \quad \text{for } t = 0, 1$$

  Zero ITT effect for always-takers and never-takers
- ITT effect decomposition:

$$\begin{aligned} \text{ITT} &= \text{ITT}_c \times \Pr(\text{compliers}) + \text{ITT}_a \times \Pr(\text{always} - \text{takers}) \\ &\quad + \text{ITT}_n \times \Pr(\text{never} - \text{takers}) \\ &= \text{ITT}_c \Pr(\text{compliers}) \end{aligned}$$

IV Estimand and Interpretation

- IV estimand:

$$\begin{aligned} \text{ITT}_c &= \frac{\text{ITT}}{\Pr(\text{compliers})} \\ &= \frac{\mathbb{E}(Y_i \mid Z_i = 1) - \mathbb{E}(Y_i \mid Z_i = 0)}{\mathbb{E}(T_i \mid Z_i = 1) - \mathbb{E}(T_i \mid Z_i = 0)} \\ &= \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(T_i, Z_i)} \end{aligned}$$

- ITT_c = Complier Average Treatment Effect (CATE)
- Local Average Treatment Effect (LATE)
- CATE $\neq$ ATE unless ATE for noncompliers equals CATE
- Different encouragement (instrument) yields different compliers
- Debate among Deaton, Heckman, and Imbens in *J. of Econ. Lit.*

An Example: Testing Habitual Voting

- Gerber *et al.* (2003) *AJPS*
- Randomized encouragement to vote in an election
- Treatment: turnout in the election
- Outcome: turnout in the next election

- Monotonicity: Being contacted by a canvasser would *never* discourage anyone from voting
- Exclusion restriction: being contacted by a canvasser in this election has no effect on turnout in the next election other than through turnout in this election
- CATE: Habitual voting for those who would vote if and only if they are contacted by a canvasser in this election

Multi-valued Treatment

- Angrist and Imbens (1995, *JASA*)
- Two stage least squares regression:

$$\begin{aligned} T_i &= \alpha_2 + \beta_2 Z_i + \eta_i, \\ Y_i &= \alpha_3 + \gamma T_i + \epsilon_i. \end{aligned}$$

- Binary encouragement and binary treatment,
  - $\hat{\gamma} = \widehat{\text{CATE}}$ (no covariate)
  - $\hat{\gamma} \xrightarrow{P} \text{CATE}$ (with covariates)
- Binary encouragement multi-valued treatment
- Monotonicity: $T_i(1) \geq T_i(0)$
- Exclusion restriction: $Y_i(1, t) = Y_i(0, t)$ for each $t = 0, 1, \ldots, K$

- Estimator

$$\hat{\gamma}_{TSLS} \xrightarrow{P} \frac{\mathrm{Cov}(Y_i, Z_i)}{\mathrm{Cov}(T_i, Z_i)} = \frac{\mathbb{E}(Y_i(1) - Y_i(0))}{\mathbb{E}(T_i(1) - T_i(0))}$$

$$= \sum_{k=0}^{K} \sum_{j=k+1}^{K} w_{jk} \mathbb{E}\left(\frac{Y_i(1) - Y_i(0)}{j - k} \,\middle|\, T_i(1) = j, T_i(0) = k \right)$$

where w_{jk} is the weight, which sums up to one, defined as,

$$w_{jk} = \frac{(j-k)\Pr(T_i(1) = j, T_i(0) = k)}{\sum_{k'=0}^{K} \sum_{j'=k'+1}^{K} (j'-k')\Pr(T_i(1) = j', T_i(0) = k')}.$$

- Easy interpretation under the constant additive effect assumption for every complier type
- Assume encouragement induces at most only one additional dose
- Then, $w_k = \Pr(T_i(1) = k, T_i(0) = k-1)$

Partial Identification of the ATE

- Balke and Pearl (1997, *JASA*)
- Randomized binary encouragement, Z_i
- Binary treatment, $T_i = T_i(Z_i)$
- Suppose exclusion restriction holds
- Binary outcome, $Y_i = Y_i(T_i, Z_i) = Y_i^*(T_i)$
- 16 Latent types defined by $(Y_i(1), Y_i(0), T_i(1), T_i(0))$

$$q(y_1, y_0, t_1, t_0) \equiv \Pr(Y_i^*(1) = y_1, Y_i^*(0) = y_0, T_i(1) = t_1, T_i(0) = t_0)$$

- ATE

$$\begin{aligned} & \mathbb{E}(Y_i^*(1) - Y_i^*(0)) \\ = & \sum_{y_0} \sum_{t_1} \sum_{t_0} q(1, y_0, t_1, t_0) - \sum_{y_1} \sum_{t_1} \sum_{t_0} q(y_1, 1, t_1, t_0) \end{aligned}$$

Derivation of Sharp Bounds

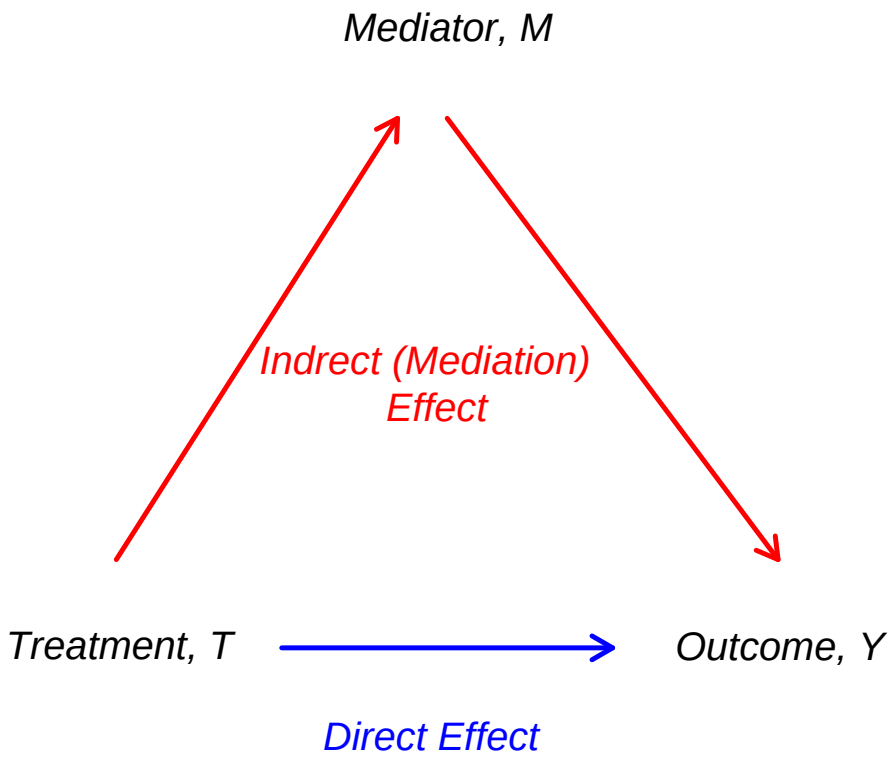
- Data generating mechanism implies

$$\begin{aligned}
\Pr(Y_i = y, T_i = 1 \mid Z_i = 1) &= \sum_{y_0} \sum_{t_0} q(y, y_0, 1, t_0) \\
\Pr(Y_i = y, T_i = 0 \mid Z_i = 1) &= \sum_{y_1} \sum_{t_0} q(y_1, y, 0, t_0) \\
\Pr(Y_i = y, T_i = 1 \mid Z_i = 0) &= \sum_{y_0} \sum_{t_1} q(y, y_0, t_1, 1) \\
\Pr(Y_i = y, T_i = 0 \mid Z_i = 0) &= \sum_{y_1} \sum_{t_1} q(y_1, y, t_1, 0).
\end{aligned}$$

- Monotonicity (optional): $q(y_1, y_0, 0, 1) = 0$
- Obtain sharp bounds via linear programming algorithms
- Bounds are sometimes informative

What is Causal Mechanism?

- Causal mediation analysis:



- Direct and indirect effects; intermediate and intervening variables

Examples

1. Incumbency effect (Cox and Katz *AJPS*)
   - Treatment: incumbency status
   - Mediator: challenger's quality
   - Outcome: reelection
   - Mechanism: incumbents deter high-quality challengers
2. Vietnam draft lottery (Erikson and Stoker *APSR*)
   - Treatment: Vietnam draft lottery
   - Mediator: military service
   - Outcome: Political attitudes
   - Mechanism: the expectation rather than the actuality of military service influences political attitudes

Potential Outcomes Framework

- Binary treatment: $T_i \in \{0, 1\}$
- Mediator: $M_i \in \mathcal{M}$
- Outcome: $Y_i \in \mathcal{Y}$
- Observed pre-treatment covariates: $X_i \in \mathcal{X}$

- Potential mediators: $M_i(t)$, where $M_i = M_i(T_i)$ observed
- Potential outcomes: $Y_i(t, m)$, where $Y_i = Y_i(T_i, M_i(T_i))$ observed

- Causal mediation (Indirect) effects:

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0))$$

- Causal effect of the change in M_i on Y_i that would be induced by treatment

Total Effect = Indirect Effect + Direct Effect

- Direct effects:

$$\zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t))$$

- Causal effect of T_i on Y_i, keepting mediator constant at its potential value that would realize when $T_i = t$

- Total effect = mediation (indirect) effect + direct effect:

$$\tau_i = \delta_i(t) + \zeta_i(1-t) = \frac{1}{2}\{\delta_i(0) + \delta_i(1) + \zeta_i(0) + \zeta_i(1)\}$$

- Quantities of interest: average direct/indirect effects
- Identification problem: $Y_i(t, M_i(t))$ is observed but $Y_i(t, M_i(t'))$ can never be observed

Traditional Estimation Method

- Linear structural equation model (LSEM):

$$\begin{aligned} M_i &= \alpha_2 + \beta_2 T_i + \xi_2^\top X_i + \epsilon_{i2}, \\ Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3^\top X_i + \epsilon_{i3}. \end{aligned}$$

  together implying

$$Y_i = \alpha_1 + \beta_1 T_i + \epsilon_{i1}$$

- Fit two least squares regressions separately
- Product of coefficients ($\hat{\beta}_2\hat{\gamma}$) or Difference of coefficients ($\hat{\beta}_1 - \hat{\beta}_3$)
- Asymptotic test of significance (Sobel test)
- What's the identification assumption?

Identification under Sequential Ignorability

- Identification assumption: Sequential Ignorability

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i \mid X_i = x \qquad (1)$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) \mid T_i = t, X_i = x \qquad (2)$$

- (1) is guaranteed to hold in a standard experiment
- (2) does **not** hold unless X_i includes all confounders

Theorem: Under sequential ignorability, ACME and average direct effects are nonparametrically identified
(= consistently estimated from observed data)

Exogeneity Is Insufficient

- Difference between manipulation and mechanism

Prop.	$M_i(1)$	$M_i(0)$	$Y_i(t,1)$	$Y_i(t,0)$	$\delta_i(t)$
0.3	1	0	0	1	-1
0.3	0	0	1	0	0
0.1	0	1	0	1	1
0.3	1	1	1	0	0

- Here, $\mathbb{E}(M_i(1) - M_i(0)) = \mathbb{E}(Y_i(t,1) - Y_i(t,0)) = 0.2$, but $\bar{\delta}(t) = -0.2$

- Commonly used causal chain approach is invalid

Need for Sensitivity Analysis

- Standard experiments require sequential ignorability to identify mechanisms
- The sequential ignorability assumption is often too strong
- Parametric sensitivity analysis by assuming

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i \mid X_i = x$$

but not

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) \mid T_i = t, X_i = x$$

- Possible existence of unobserved *pre-treatment* confounder

Sensitivity Analysis for LSEM

- Sensitivity parameter: $\rho \equiv \text{Corr}(\epsilon_{i2}, \epsilon_{i3})$
- Sequential ignorability implies $\rho = 0$
- Set ρ to different values and see how ACME changes
- **Result:**

$$\bar{\delta}(0) = \bar{\delta}(1) = \frac{\beta_2 \sigma_1}{\sigma_2} \left\{ \tilde{\rho} - \rho \sqrt{(1-\tilde{\rho}^2)/(1-\rho^2)} \right\},$$

  where $\sigma_j^2 \equiv \text{var}(\epsilon_{ij})$ for $j = 1, 2$ and $\tilde{\rho} \equiv \text{Corr}(\epsilon_{i1}, \epsilon_{i2})$.

- When do my results go away completely?
- $\bar{\delta}(t) = 0$ if and only if $\rho = \tilde{\rho}$
- Easy to estimate from the regression of Y_i on T_i:

$$Y_i = \alpha_1 + \beta_1 T_i + \epsilon_{i1}$$

- Alternative parameterizations via R^2
- Extensions to nonlinear models

Crossover Design

- Need for alternative research designs
- Recall ACME can be identified if we observe $Y_i(t', M_i(t))$
- Get $M_i(t)$, then switch T_i to t' while holding $M_i = M_i(t)$

- Crossover design:
  1. Round 1: Conduct a standard experiment
  2. Round 2: Change the treatment to the opposite status but fix the mediator to the value observed in the first round
- Very powerful – identifies mediation effects for each subject
- Must assume no carryover effect: Round 1 cannot affect Round 2
- Can be made plausible by design

Labor Economics Experiment

Bertrand & Mullainathan (2004, AER)

- Treatment: Black vs. White names on CVs
- Mediator: Perceived qualifications of applicants
- Outcome: Callback from employers
- Quantity of interest: Direct effects of (perceived) race
- Would Jamal get a callback if his name were Greg but his qualifications stayed the same?
- Round 1: Send Jamal's actual CV and record the outcome
- Round 2: Send his CV as Greg and record the outcome
- Assumptions are plausible

Observational Studies Example

- Estimation of incumbency advantages goes back to 1960s
- Why incumbency advantage? Scaring off quality challenger
- Use of cross-over design (Levitt and Wolfram)
  1. 1st Round: two non-incumbents in an open seat
  2. 2nd Round: same candidates with one being an incumbent
- Assume challenger quality (mediator) stays the same
- Estimation of direct effect is possible
- Redistricting as natural experiments (Ansolabehere et al.)
  1. 1st Round: incumbent in the old part of the district
  2. 2nd Round: incumbent in the new part of the district
- Challenger quality is the same but treatment is different
- Estimation of direct effect is possible

Recent Developments in the Methodological Literature

- Alternative research designs
  - Use of randomized encouragement (i.e., instruments)
- Alternative definitions and approaches of causal mechanisms
  - Principal strata direct effects
  - Causal components
- Statistical methods for multiple mediators
  - Identification assumptions
  - Sensitivity analysis