Experimental Identification of Causal Mechanisms

Kosuke Imai

Princeton University
Joint work with Dutin Tingley and Teppei Yamamoto

October 17, 2009
Columbia EGAP Conference

Experiments, Statistics, and Causal Mechanisms

- Causal inference is a central goal of social science
- Experiments as **gold standard** for estimating *causal effects*
- But, we really care about *causal mechanisms*

- A major criticism of experimentation (and statistics):

  it can only determine whether the treatment causes changes in the outcome, but not how and why

- Experiments are a **black box**

- Key Challenge: How can we design and analyze experiments to identify causal mechanisms?

Overview of the Talk

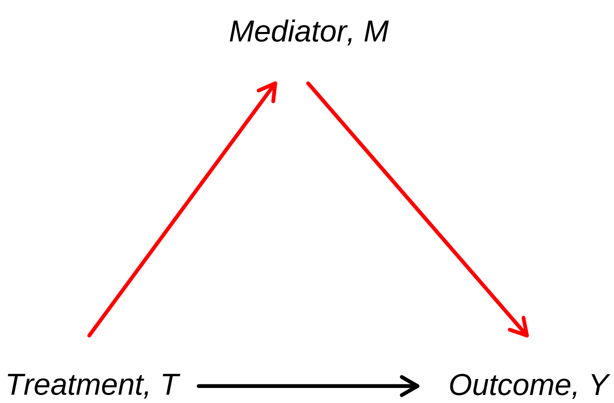
- Show the limitations of common approaches
- Propose alternative experimental designs

- What is a minimum set of assumptions required for identification under each design?
- How much can we learn without the key identification assumptions under each design?

- Identification of causal mechanisms is possible but difficult
- Replace statistical assumptions with design assumptions
- Roles of creativity and technological developments

Causal Mechanisms as Indirect Effects

- Causal mediation analysis



- Quantities of interest: Direct and indirect effects
- Fast growing methodological literature

Formal Statistical Framework of Causal Inference

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

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

- Fundamental problem of causal inference:

  Only one potential value is observed
- If $T_i = 1$, then $M_i(1)$ is observed but $M_i(0)$ is not
- If $T_i = 0$ and $M_i(0) = 0$, then $Y_i(0, 0)$ is observed but $Y_i(1, 0)$, $Y_i(0, m)$, and $Y_i(1, m)$ are not where $m \neq 0$

Defining and Interpreting Indirect Effects

- Total causal effect:

$$\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0))$$

- Indirect (causal mediation) effects:

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

- Change $M_i(0)$ to $M_i(1)$ while holding the treatment constant at t
- Effect of a change in M_i on Y_i that would be induced by treatment

- Fundamental problem of causal mechanisms:

  For each unit i, $Y_i(t, M_i(t))$ is observable but
  $Y_i(t, M_i(1-t))$ is not even observable

Defining and Interpreting Direct Effects

- Direct effects:

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

- Change T_i from 0 to 1 while holding the mediator constant at $M_i(t)$
- Causal effect of T_i on Y_i, holding mediator constant at its potential value that would be realized when $T_i = t$

- Total effect = indirect effect + direct effect:

$$\begin{aligned}\tau_i &= \frac{1}{2}\{\delta_i(0) + \delta_i(1) + \zeta_i(0) + \zeta_i(1)\} \\ &= \delta_i + \zeta_i \quad \text{if } \delta_i = \delta_i(0) = \delta_i(1) \text{ and } \zeta_i = \zeta_i(0) = \zeta_i(1)\end{aligned}$$

Mechanisms, Manipulations, and Interactions

Mechanisms

- Indirect effects:

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

- Counterfactuals about treatment-induced mediator values

Manipulations

- Controlled direct effects:

$$\xi_i(t, m, m') \equiv Y_i(t, m) - Y_i(t, m')$$

- Causal effect of directly manipulating the mediator under $T_i = t$

Interactions

- Interaction effects:

$$\xi(1, m, m') - \xi(0, m, m') \neq 0$$

- Doesn't imply the existence of a mechanism

Single Experiment Approach

1) Randomize treatment

2) Measure mediator

3) Measure outcome

Assumption Satisfied

- Randomization of treatment

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

Key Identifying Assumptions

- Sequential Ignorability:

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

- Selection on observables
- Violated if there are unobservables that affect mediator and outcome

Identification under Single Experiment Approach

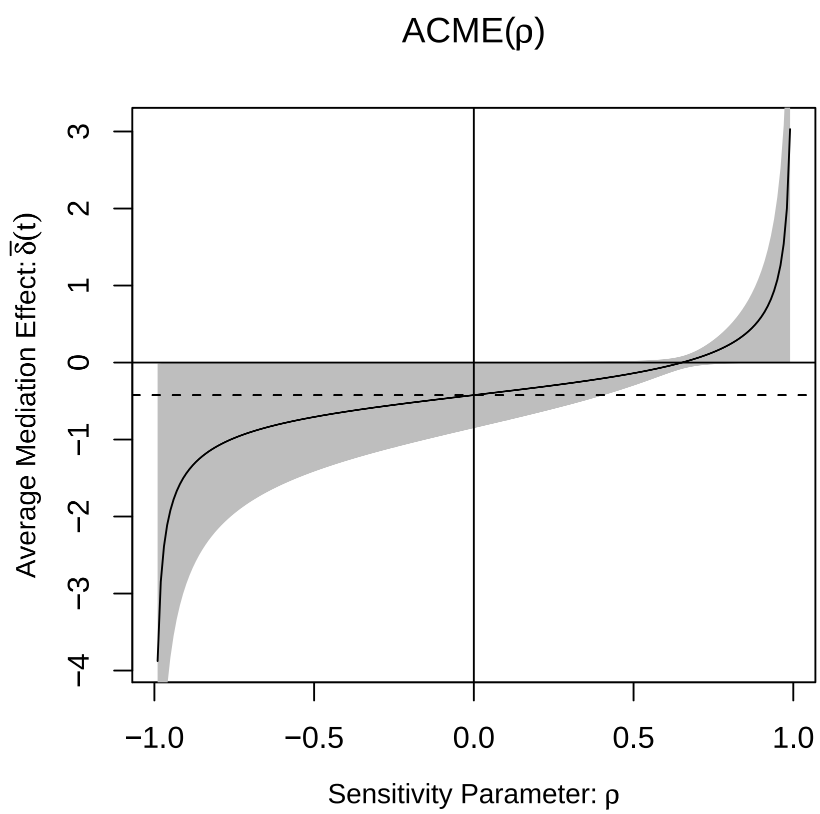
- Sequential ignorability yields nonparametric identification
- Linear regressions with no interaction: Baron-Kenny

- Untestable assumption
- How much can we learn without sequential ignorability?
- Sharp bounds on indirect effects: even sign cannot be identified

- Sensitivity analysis at best: How large a departure from sequential ignorability must occur for the conclusions to no longer hold?

- Can we replace the statistical assumption with the design assumption?
- Can we design experiments to help identify causal mechanisms?

Sensitivity Analysis



Causal Chain Approach

Experiment 1

1) Randomize treatment

2) Measure mediator

Different sample

Experiment 2

1) Measure treatment (though often not done)

2) Randomize mediator

3) Measure outcome

Assumption Satisfied

- Randomization of treatment in the first experiment but not the second
- Randomization of mediator given the treatment in the second experiment

Claim in the Literature

- Two statistically significant effects identify the causal mechanism without any additional assumption

Key Identification Assumptions

- Identification requires 3 more and untestable assumptions!

- No manipulation effect: Manipulation of mediator has no direct effect on outcome other than through the mediator value
- "Puppet" assumption: a good puppeteer can convince the audience that his puppets move under their own volition

- No interaction: For any $m \neq m'$,

$$Y_i(1, m) - Y_i(1, m') \; = \; Y_i(0, m) - Y_i(0, m')$$

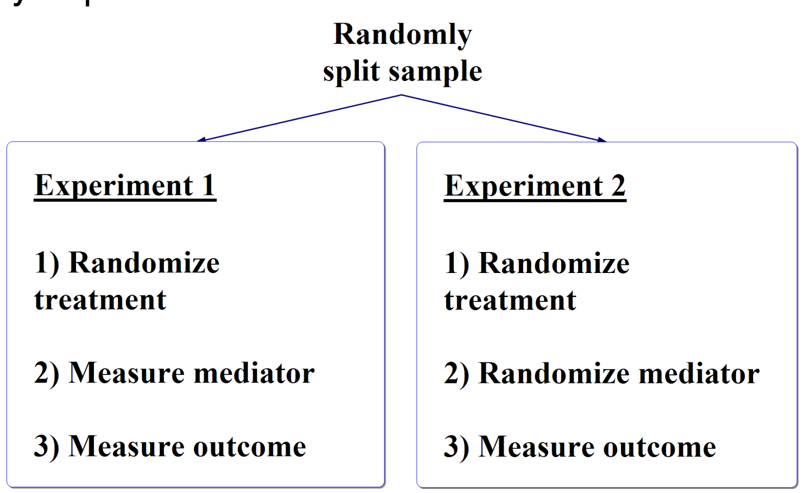
- Changing the mediator under the treatment produces same effect as changing mediator under the control

- No selection bias w.r.t. the treatment in second experiment

Identification Analysis and Parallel Design

- What happens if we do not make the no interaction assumption and the no selection bias assumption
- Bounds are narrower than those of single experiment approach
- The sign of indirect effect is not identified except rare cases

- Parallel Design: Bounds are always narrower and sometimes substantially improved



A Numerical Illustration

- Why aren't two statistically significant effects sufficient?
- Consider the following example:

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

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

Comparison of Assumptions

Assumptions	Single Experiment	Causal Chain	Parallel
Random Treatment	☺	☹	☺
Sequential Ignorability (SI)	☹		
Random Mediator		☺	☺
No Manipulation Effect		☹	☹
No Interaction Effect		☹	☹

- Single experiment approach requires the SI assumption
- Causal chain approach replaces it with other untestable assumptions that are unrelated to experimental designs
- Parallel design improves causal chain approach

- Can we come up with a more powerful design?

Crossover Design

Experiment 1

1) Randomize treatment

2) Measure mediator

3) Measure outcome

Same sample ↓

Experiment 2

1) Fix treatment opposite Experiment 1

2) Manipulate mediator to level observed in Experiment 1

3) Measure outcome

Basic Idea

- Want to observe $Y_i(1-t, M_i(t))$
- Figure out $M_i(t)$ and then switch T_i while holding the mediator at this value
- Subtract direct effect from total effect

Key Identifying Assumptions

- No Manipulation Effect
- No Carryover Effect: First experiment doesn't affect second experiment

Crossover Encouragement Design

Experiment 1

1) Randomize treatment

2) Measure mediator

3) Measure outcome (optional)

Same sample ↓

Experiment 2

1) Fix treatment opposite Experiment 1

2) Randomly encourage mediator to level observed in Experiment 1

3) Measure outcome

Motivation

- Imperfect and subtle manipulation

Key Identifying Assumptions

- No Defier: encouragement doesn't discourage anyone
- No Carryover Effect
- No Manipulation Effect

Identification Analysis

- Identify indirect effects for "pliable" units
- Can check carryover effect

Comparison of Assumptions

Assumptions	Crossover	Crossover Encouragement
Random Treatment	☺	☺
Sequential Ignorability		
Random Mediator		
Random Encouragement		☺
No Manipulation Effect	☹	☹
No Interaction Effect		
No Carryover Effect	☹	☹
No Defier		☹

- Crossover designs are most powerful
- No carryover effect: longer washout period
- Imperfect manipulation – indirect effect for pliable units
- Subtle and indirect encouragement – less manipulation effect

Example from Behavioral Neuroscience

Question: What mechanism links low offers in an ultimatum game with “irrational" rejections?

- Two brain regions more active when unfair offer received (single experiment design)

Design solution: manipulate mechanisms with TMS

- Knoch et al. use TMS to manipulate — turn off — one of these regions, and then observes choices (parallel design)

We discuss the applicability of each design and the credibility of its identification assumptions in this context

Concluding Remarks

- Identification of causal mechanisms is difficult but is possible
- Additional assumptions are required
- Three possible strategies:
  1. Single experiment design
  2. Parallel design
  3. Crossover (encouragement) design
- Statistical assumptions: sequential ignorability, no interaction
- Design assumptions: no manipulation, no carryover effect
- Experimenters' creativity and technological development to improve the validity of these design assumptions

Papers and Software

- "Experimental Identification of Causal Mechanisms"
- "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects."
- "A General Approach to Causal Mediation Analysis."
- "Causal Mediation Analysis in R."
- All available at
  http://imai.princeton.edu/projects/mechanisms.html
- **mediation**: R package for causal mediation analysis
- Available at
  http://cran.r-project.org/web/packages/mediation/