New Statistical Methods and Experimental Designs for the Identification of Causal Mechanisms

Kosuke Imai

Princeton University
Joint work with Luke Keele, Dutin Tingley, Teppei Yamamoto

October 9, 2009
Symposium on Experimental Social Science

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**
- Qualitative research uses process tracing

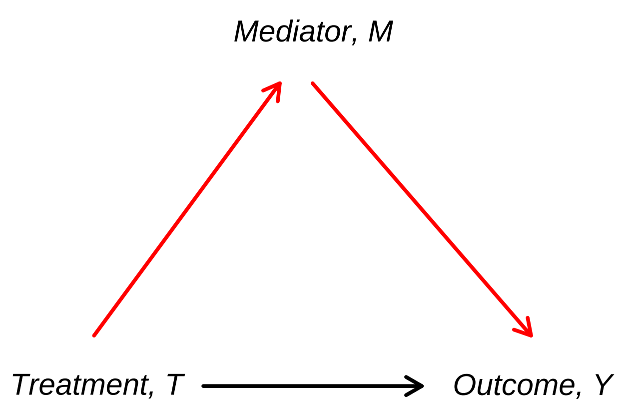
- Key Challenge: How can we design and analyze experiments to identify causal mechanisms?
- We propose new statistical methods and experimental designs for the identification of causal mechanisms

Overview of the Talk

- Identification of causal mechanisms in standard experiments
  1. Offer a general nonparametric identification and estimation strategy
  2. Modernize and extend causal mediation analysis
  3. Propose sensitivity analyses to assess the robustness
- New experimental designs for identification of causal mechanisms
  1. Derive the limitations of common approaches
  2. Propose alternative experimental designs
  3. Illustrate the ideas vis-à-vis a behavioral neuroscience experiment

Causal Mediation Analysis

- Graphical representation



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

Common Practice in the Discipline

- Regression
$$Y_i = \alpha + \beta T_i + \gamma M_i + \delta X_i + \epsilon_i$$
- Each coefficient is interpreted as a causal effect
- Sometimes, it's called marginal effect
- Idea: increase T_i by one unit while holding M_i and X_i constant

- Post-treatment bias: if you change T_i, that may also change M_i
- Usual advice: only include causally prior variables
- But, then you lose causal mechanisms!

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 outcome is observed

Defining and Interpreting Causal Mediation Effects

- Total causal effect:

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

- Indirect (causal mediation) effects:

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

- Causal effect of the change in M_i on Y_i that would be induced by treatment
- Change the mediator from $M_i(0)$ to $M_i(1)$ while holding the treatment constant at t
- Fundamental problem: For each unit i, $Y_i(t, M_i(t))$ is observable but one can *never* observe $Y_i(t, M_i(1-t))$

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 naturally occurring 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

Nonparametric Identification

- Quantity of Interest: Average Causal Mediation Effects

$$\bar{\delta}(t) \equiv \mathbb{E}(\delta_i(t)) = \mathbb{E}\{Y_i(t, M_i(1)) - Y_i(t, M_i(0))\}$$

- Problem: $Y_i(t, M_i(t))$ is observed but $Y_i(t, M_i(1-t))$ can *never* be observed

- Proposed identification assumption: Sequential Ignorability

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

Theorem 1 (Nonparametric Identification)

Under sequential ignorability,

$$\bar{\delta}(t) = \int\int \mathbb{E}(Y_i \mid M_i, T_i = t, X_i)\{dP(M_i \mid T_i = 1, X_i) - dP(M_i \mid T_i = 0, X_i)\}\, dP(X_i),$$
$$\bar{\zeta}(t) = \int\int \{\mathbb{E}(Y_i \mid M_i, T_i = 1, X_i) - \mathbb{E}(Y_i \mid M_i, T_i = 0, X_i)\}\, dP(M_i \mid T_i = t, X_i)\, dP(X_i).$$

Inference Under Sequential Ignorability

- Model outcome and mediator
- Outcome model: $p(Y_i \mid T_i, M_i, X_i)$
- Mediator model: $p(M_i \mid T_i, X_i)$
- A simplest setup: Linear Structural Equation Model (LSEM)

$$M_i = \alpha_2 + \beta_2 T_i + \epsilon_{i2},$$
$$Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{i3}.$$

Theorem 2 (Identification Under LSEM)

Under the LSEM and sequential ignorability, the average causal mediation effects are identified as $\bar{\delta}(0) = \bar{\delta}(1) = \beta_2\gamma$.

- Can include the interaction between T_i and M_i
- Can use parametric or nonparametric regressions; probit, logit, ordered mediator, GAM, quantile regression, etc.

Need for Sensitivity Analysis

- The sequential ignorability assumption is often too strong
- Need to assess the robustness of findings via sensitivity analysis
- Question: How large a departure from the key assumption must occur for the conclusions to no longer hold?
- Parametric sensitivity analysis by assuming

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

  but not

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

- Possible existence of unobserved *pre-treatment* confounder

Parametric Sensitivity Analysis

- Sensitivity parameter: $\rho \equiv \mathrm{Corr}(\epsilon_{i2}, \epsilon_{i3})$
- Sequential ignorability implies $\rho = 0$
- Set ρ to different values and see how mediation effects change

Theorem 3

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

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

- 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:

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

Interpreting Sensitivity Analysis with R squares

- Interpreting ρ: how small is too small?
- An unobserved (pre-treatment) confounder formulation:

$$\epsilon_{i2} = \lambda_2 U_i + \epsilon'_{i2} \quad \text{and} \quad \epsilon_{i3} = \lambda_3 U_i + \epsilon'_{i3}$$

- How much does U_i have to explain for our results to go away?
- Sensitivity parameters: R squares
  1. Proportion of previously unexplained variance explained by U_i

$$R_M^{2*} \equiv 1 - \frac{\text{var}(\epsilon'_{i2})}{\text{var}(\epsilon_{i2})} \quad \text{and} \quad R_Y^{2*} \equiv 1 - \frac{\text{var}(\epsilon'_{i3})}{\text{var}(\epsilon_{i3})}$$

  2. Proportion of original variance explained by U_i

$$\widetilde{R}_M^2 \equiv \frac{\text{var}(\epsilon_{i2}) - \text{var}(\epsilon'_{i2})}{\text{var}(M_i)} \quad \text{and} \quad \widetilde{R}_Y^2 \equiv \frac{\text{var}(\epsilon_{i3}) - \text{var}(\epsilon'_{i3})}{\text{var}(Y_i)}$$

- Then reparameterize ρ using (R_M^{2*}, R_Y^{2*}) (or $(\widetilde{R}_M^2, \widetilde{R}_Y^2)$):

$$\rho = \text{sgn}(\lambda_2\lambda_3) R_M^* R_Y^* = \frac{\text{sgn}(\lambda_2\lambda_3)\widetilde{R}_M \widetilde{R}_Y}{\sqrt{(1 - R_M^2)(1 - R_Y^2)}},$$

  where R_M^2 and R_Y^2 are from the original mediator and outcome models
- $\text{sgn}(\lambda_2\lambda_3)$ indicates the direction of the effects of U_i on Y_i and M_i
- Set (R_M^{2*}, R_Y^{2*}) (or $(\widetilde{R}_M^2, \widetilde{R}_Y^2)$) to different values and see how mediation effects change

Empirical Illustration: Nelson *et al.* (*APSR*)

- How does media framing affect citizens' political opinions?
- News stories about the Ku Klux Klan rally in Ohio

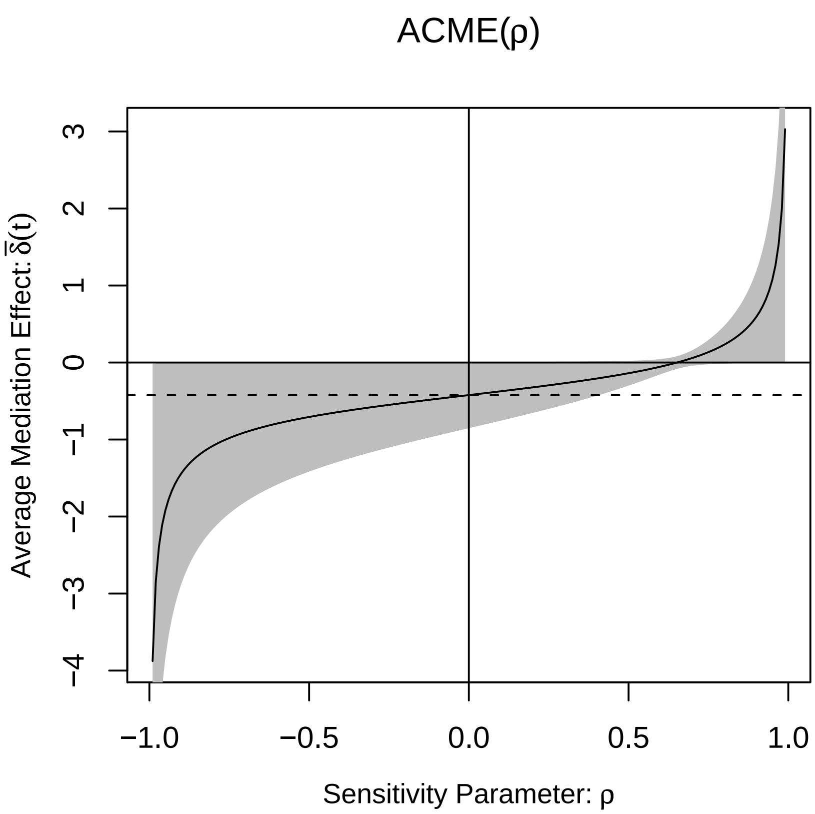
- Treatment: Free speech frame ($T_i = 0$) and public order frame ($T_i = 1$)
- Randomized experiment with sample size $= 136$

- Mediators: general attitudes about the importance of free speech and public order
- Outcome: tolerance for the Klan rally
- Expected findings: negative mediation effects

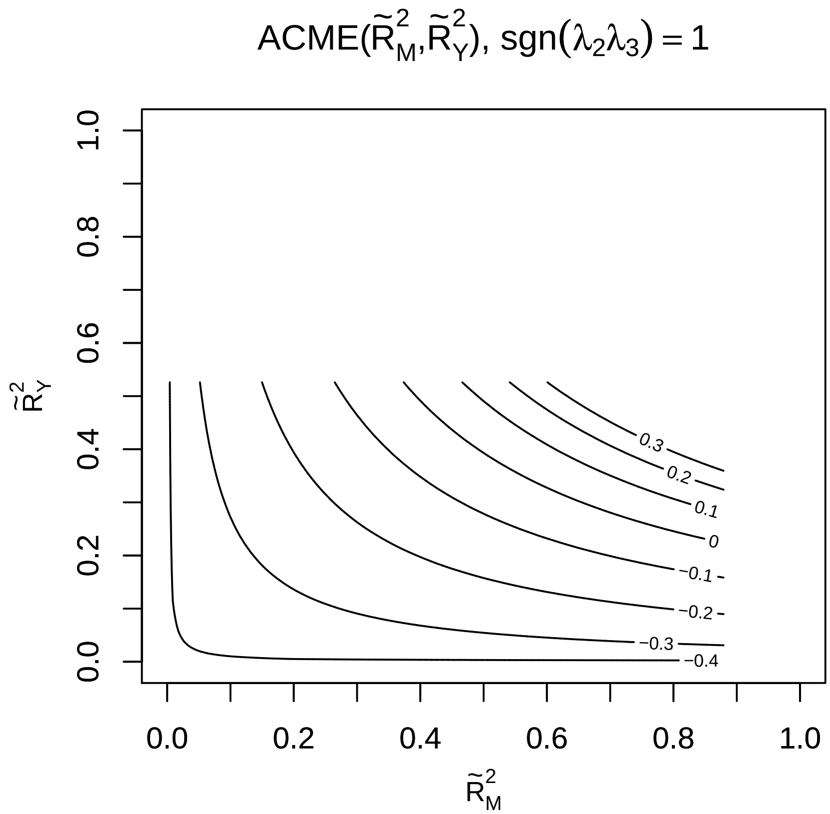
Analysis under Sequential Ignorability

Average Mediation Effects $\hat{\delta}(0) = \hat{\delta}(1)$	-0.44 $[-0.87, -0.01]$
Average Direct Effects $\hat{\zeta}(0) = \hat{\zeta}(1)$	-0.02 $[-0.49, \quad 0.47]$
Average Total Effect $\hat{\tau}$	-0.46 $[-1.11, \quad 0.23]$

Sensitivity Analysis with Respect to ρ



Sensitivity Analysis with Respect to $(\widetilde{R}^2_M, \widetilde{R}^2_Y)$



Experimental Designs and Causal Mechanisms

- *Statistical* vs. *Experimental* approach to the identification of causal mechanisms
- Can we design an experiment to facilitate the identification of causal mechanisms?
- Replace statistical assumptions with the assumptions about experimental design
- How do different experimental designs help or hinder the identification of causal mechanisms?

- Encourages experimentalists to be creative
- Technological developments facilitates the use of new designs

Single Experiment Approach

1) Randomize treatment

2) Measure mediator

3) Measure outcome

Key Identifying Assumptions

- Sequential Ignorability: conditional on treatment, mediator is random
- Violated if there are unobservables that affect mediator and outcome
- Not testable – sensitivity analysis at best

Identification Analysis

- Can never identify the sign of indirect effect

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

Key Identifying Assumptions

- Treatment in second experiment is random
- No Manipulation Effect: Manipulation of mediator has no direct effect on outcome
- No Interaction: Changing the mediator under the treatment produces same effect as changing mediator under the control

Identification Analysis

- More informative than single experiment
- In most cases, cannot identify the sign
- Statistical significant effects are neither necessary or sufficient

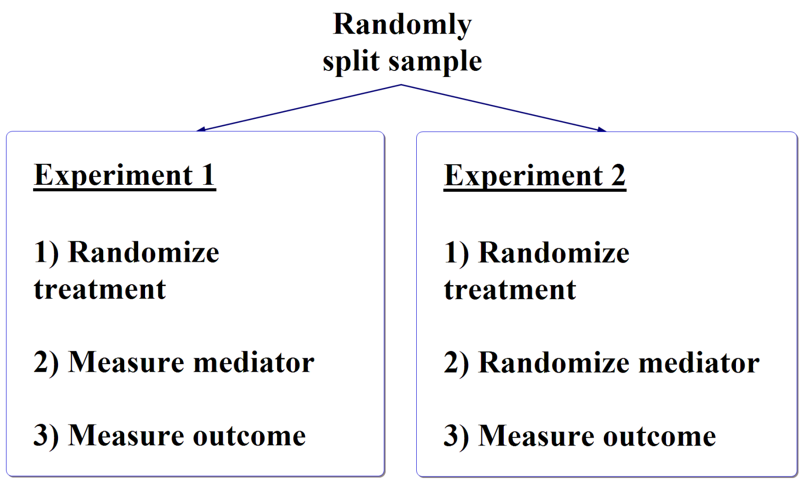
Comparison of Assumptions

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

Limitations of the existing approaches:

- Single experiment approach requires the SI assumption
- Causal chain approach replaces it with other untestable assumptions that are unrelated to experimental designs
- Can we come up with a better experimental design?

Parallel Design



Key Identifying Assumptions

- No Manipulation Effect
- No Interaction Effect

Identification Analysis

- Always more informative than causal chain

Comparison of Assumptions

Assumptions	Causal Chain	Parallel
Random Treatment	☹	☺
Sequential Ignorability		
Random Mediator	☺	☺
No Manipulation Effect	☹	☹
No Interaction Effect	☹	☹

- Difficult to justify the No Interaction Effect assumption
- Parallel design is more informative about causal mechanisms

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

Key Identifying Assumptions

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

Identification Analysis

- No information about carryover effect
- Use different crossover experiments

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

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 design is the most powerful, but requires the no carryover effect assumption
- Longer washout period
- Crossover encouragement design can be applied even if mediator is not directly manipulable
- Subtle 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

- Two proposed strategies:
  1. Sensitivity analysis to assess the robustness
  2. New experimental designs to improve the credibility
- Offer a comprehensive set of statistical methods
- Derive the identification power of different experimental designs

- Ongoing work:
  - Application to political psychology experiments
  - Experimental identification of causal effects of gene

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/