Unpacking the Black-Box of Causality:
Learning about Causal Mechanisms from Experimental and Observational Studies

Kosuke Imai

Princeton University

November 2, 2011

Joint work with

L. Keele (Penn State)   D. Tingley (Harvard)   T. Yamamoto (MIT)

Quantitative Research and Causal Mechanisms

- Causal inference is a central goal of scientific research
- Scientists care about *causal mechanisms*, not just *causal effects*

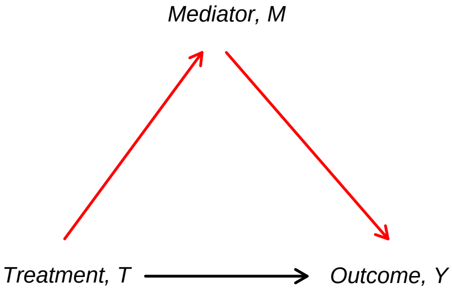
- Randomized experiments often only determine whether the treatment causes changes in the outcome
- Not how and why the treatment affects the outcome
- Common criticism of experiments and statistics:

black box view of causality

- Qualitative research uses process tracing
- **Question:** How can quantitative research be used to identify causal mechanisms?

Overview of the Talk

- **Goal:** Convince you that statistics *can* be useful for learning about causal mechanisms
- **Method:** Causal Mediation Analysis

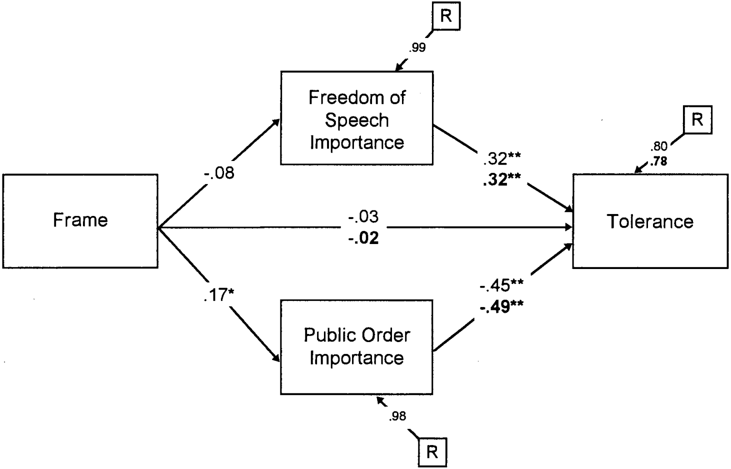


Direct and indirect effects; intermediate and intervening variables

- **New tools:** framework, estimation algorithm, sensitivity analysis, research designs, easy-to-use software

Causal Mediation Analysis in **American Politics**

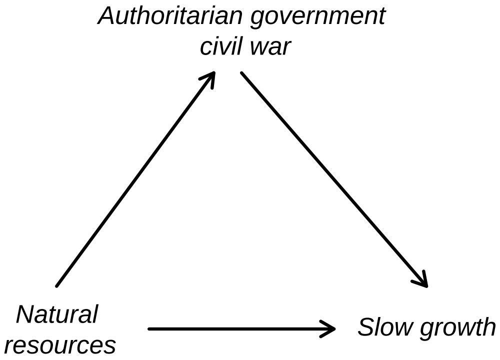
- The political psychology literature on media framing
- Nelson *et al*. (*APSR*, 1998)



- Popular in social psychology

Causal Mediation Analysis in **Comparative Politics**

- Resource curse thesis



- Causes of civil war: Fearon and Laitin (*APSR*, 2003)

Causal Mediation Analysis in **International Relations**

- The literature on international regimes and institutions
- Krasner (*International Organization*, 1982)

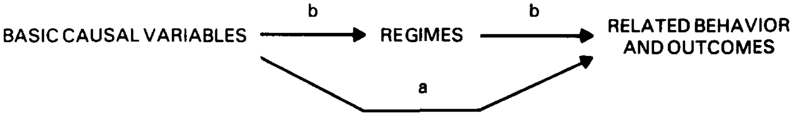


Figure 2

- Power and interests are mediated by regimes

Current Practice in **Political Science**

- 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

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

Formal Statistical Framework of Causal Inference

- Units: $i = 1, \ldots, n$
- "Treatment": $T_i = 1$ if treated, $T_i = 0$ otherwise
- Pre-treatment covariates: X_i
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$
- Observed outcome: $Y_i = Y_i(T_i)$

Voters	Contact	Turnout		Age	Party ID
i	T_i	$Y_i(1)$	$Y_i(0)$	X_i	X_i
1	1	1	?	20	D
2	0	?	0	55	R
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	1	0	?	62	D

- Causal effect: $Y_i(1) - Y_i(0)$
- Problem: **only one potential outcome can be observed per unit**

Potential Outcomes Framework for Mediation

- Binary treatment: T_i
- Pre-treatment covariates: X_i

- Potential mediators: $M_i(t)$
- Observed mediator: $M_i = M_i(T_i)$

- Potential outcomes: $Y_i(t, m)$
- Observed outcome: $Y_i = Y_i(T_i, M_i(T_i))$

- Again, **only one potential outcome can be observed per unit**

Causal Mediation Effects

- Total causal effect:

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

- Causal mediation (Indirect) effects:

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

- Causal effect of the treatment-induced change in M_i on Y_i
- Change the mediator from $M_i(0)$ to $M_i(1)$ while holding the treatment constant at t
- Represents the mechanism through M_i

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, holding mediator constant at its potential value that would be realized when $T_i = t$
- Change the treatment from 0 to 1 while holding the mediator constant at $M_i(t)$
- Represents all mechanisms other than through M_i

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

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

What Does the Observed Data Tell Us?

- Quantity of Interest: Average causal mediation effects (ACME)

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

- Average direct effects ($\bar{\zeta}(t)$) are defined similarly

- $Y_i(t, M_i(t))$ is observed but $Y_i(t, M_i(t'))$ can never be observed
- We have an identification problem

$\Longrightarrow$ Need additional assumptions to make progress

Identification under Sequential Ignorability

- Proposed identification assumption: Sequential Ignorability (SI)

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

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

- (1) is guaranteed to hold in a standard experiment
- (2) does **not** hold unless X_i includes all confounders
- Limitation: X_i cannot include post-treatment confounders

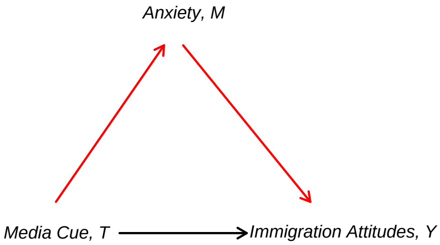
Under SI, ACME is nonparametrically identified:

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

Example: Anxiety, Group Cues and Immigration

Brader, Valentino & Suhat (2008, AJPS)

- How and why do ethnic cues affect immigration attitudes?
- Theory: Anxiety transmits the effect of cues on attitudes



- ACME = Average difference in immigration attitudes due to the change in anxiety induced by the media cue treatment
- Sequential ignorability = No unobserved covariate affecting both anxiety and immigration attitudes

Traditional Estimation Method

- Linear structural equation model (LSEM):

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

- Fit two least squares regressions separately
- Use product of coefficients ($\hat{\beta}_2 \hat{\gamma}$) to estimate ACME

- The method is valid under SI
- Can be extended to LSEM with interaction terms

- Problem: Only valid for the simplest LSEMs

Proposed General Estimation Algorithm

1. Model outcome and mediator
   - Outcome model: $p(Y_i \mid T_i, M_i, X_i)$
   - Mediator model: $p(M_i \mid T_i, X_i)$
   - These models can be of any form (linear or nonlinear, semi- or nonparametric, with or without interactions)
2. Predict mediator for both treatment values ($M_i(1)$, $M_i(0)$)
3. Predict outcome by first setting $T_i = 1$ and $M_i = M_i(0)$, and then $T_i = 1$ and $M_i = M_i(1)$
4. Compute the average difference between two outcomes to obtain a consistent estimate of ACME
5. Monte Carlo or bootstrap to estimate uncertainty

Example: Estimation under Sequential Ignorability

- Original method: Product of coefficients with the Sobel test
  — Valid only when both models are linear w/o T–M interaction (which they are not)
- Our method: Calculate ACME using our general algorithm

Outcome variables	Product of Coefficients	Average Causal Mediation Effect (δ)
Decrease Immigration	.347	.105
$\bar{\delta}(1)$	[0.146, 0.548]	[0.048, 0.170]
Support English Only Laws	.204	.074
$\bar{\delta}(1)$	[0.069, 0.339]	[0.027, 0.132]
Request Anti-Immigration Information	.277	.029
$\bar{\delta}(1)$	[0.084, 0.469]	[0.007, 0.063]
Send Anti-Immigration Message	.276	.086
$\bar{\delta}(1)$	[0.102, 0.450]	[0.035, 0.144]

Need for Sensitivity Analysis

- Even in experiments, SI is required to identify mechanisms
- SI is often too strong and yet not testable
- 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?
- 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

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 ACME changes

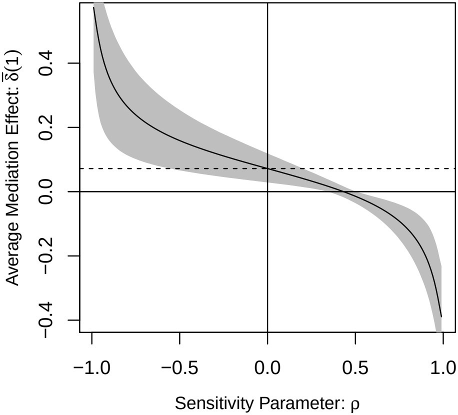
- When do my results go away completely?
- $\bar{\delta}(t) = 0$ if and only if $\rho = \mathrm{Corr}(\epsilon_{i1}, \epsilon_{i2})$ where

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

- Easy to estimate from the regression of Y_i on T_i:

- Alternative interpretation based on R^2:
  How big does the effects of unobserved confounders have to be in order for my results to go away?

Example: Sensitivity Analysis



- ACME $>$ 0 as long as the error correlation is less than 0.39 (0.30 with 95% CI)

Beyond Sequential Ignorability

- Without sequential ignorability, standard experimental design lacks identification power
- Even the sign of ACME is not identified

- Need to develop alternative research design strategies for more credible inference
- New experimental designs: Possible when the mediator can be directly or indirectly manipulated
- Observational studies: use experimental designs as templates

Crossover Design

- 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 doesn't affect Round 2
- Can be made plausible by design

Example: Labor Market Discrimination 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

Designing Observational Studies

- Key difference between experimental and observational studies: treatment assignment
- Sequential ignorability:
  1. Ignorability of treatment given covariates
  2. Ignorability of mediator given treatment and covariates
- Both (1) and (2) are suspect in observational studies

- Statistical control: matching, propensity scores, etc.
- Search for quasi-randomized treatments: “natural” experiments

- How can we design observational studies?
- Experiments can serve as templates for observational studies

Example: Incumbency Advantage

- Estimation of incumbency advantages goes back to 1960s
- Why incumbency advantage? Scaring off quality challenger

- Use of cross-over design (Levitt and Wolfram, LSQ)
  1. 1st Round: two non-incumbents in an open seat
  2. 2nd Round: same candidates with one being an incumbent

- Assumption: challenger quality (mediator) stays the same
- Estimation of direct effect is possible

Concluding Remarks

- Quantitative analysis can be used to identify causal mechanisms!
- Estimate causal mediation effects rather than marginal effects
- Wide applications across social and natural science disciplines

- Under standard research designs, sequential ignorability must hold for identification of causal mechanisms
- Under SI, a general, flexible estimation method is available
- SI can be probed via sensitivity analysis
- Easy-to-use software **mediation** is available in R and STATA
- Credible inference is possible under alternative research designs

- Ongoing research: multiple mediators, instrumental variables

The project website for papers and software:

http://imai.princeton.edu/projects/mechanisms.html

Email for comments and suggestions:

kimai@Princeton.Edu