

Statistical Analysis of Causal Mechanisms

Kosuke Imai[†] **Luke Keele**[§]
Dustin Tingley[†] **Teppei Yamamoto**[†]

[†]Princeton University [§]Ohio State University

July 25, 2009
Summer Political Methodology Conference

Statistics and Causal Mechanisms

- Causal inference is a central goal of social science
- Challenge is how to identify causal mechanisms
- Even randomized experiments can only determine **whether** the treatment causes changes in 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?

Goals of the Talk

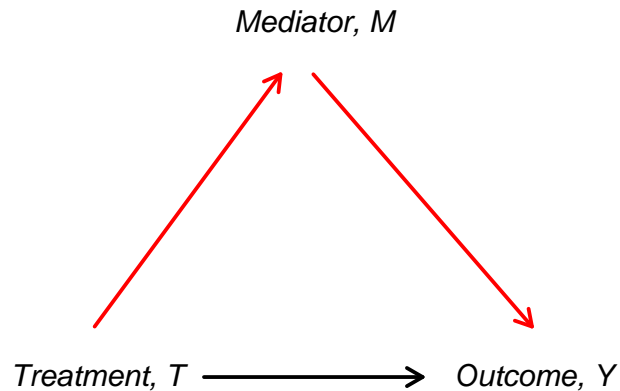
- 1 Convince you that statistics can play a role in identifying causal mechanisms via **causal mediation analysis**
- 2 Offer a general **identification and estimation strategy** that is widely applicable to variety of settings
- 3 Propose **sensitivity analyses** to assess the robustness to potential violations of the key identifying assumption
- 4 Provide an **easy-to-use R package** for implementing all of the proposed methods

Papers and Software

- “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/>

Causal Mediation Analysis

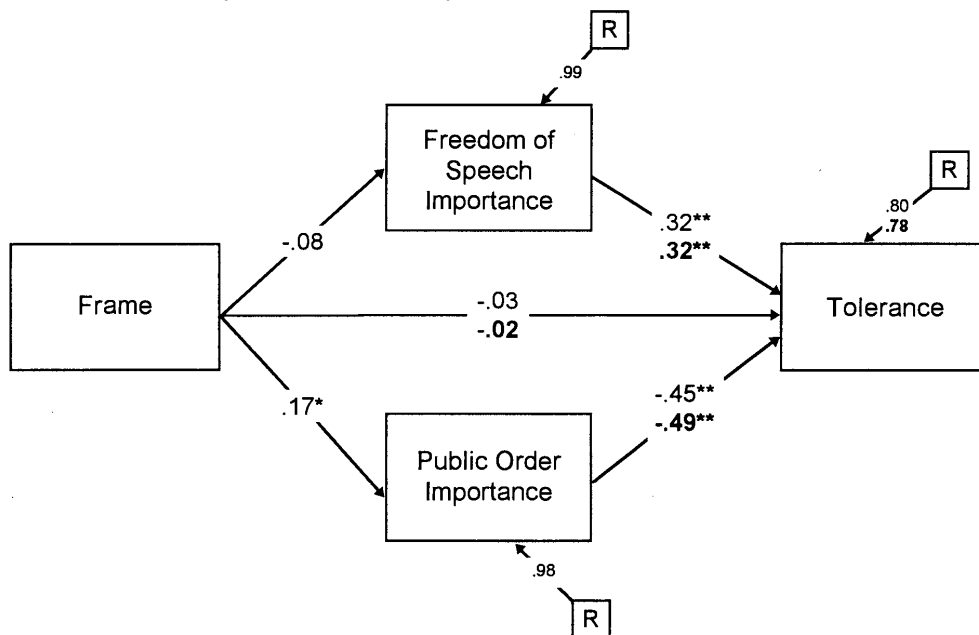
- Graphical representation



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

Causal Mediation Analysis in American Politics

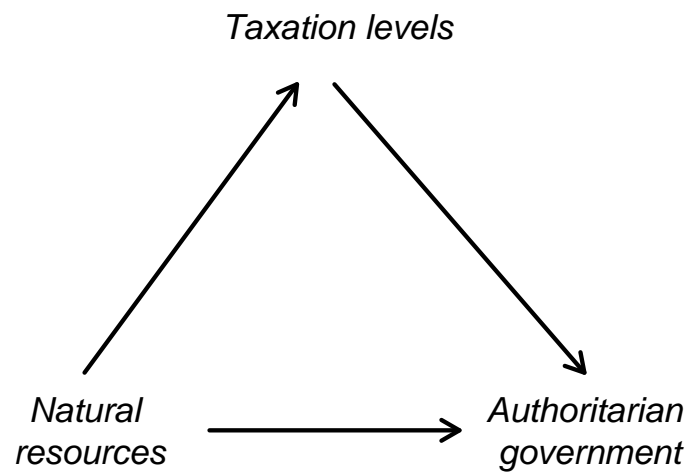
- The political psychology literature on media framing
- Nelson *et al.* (APSR, 1997)



- Popular in social psychology

Causal Mediation Analysis in Comparative Politics

- Resource curse thesis (Ross *WP*, 2001)



Causal Mediation Analysis in International Relations

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

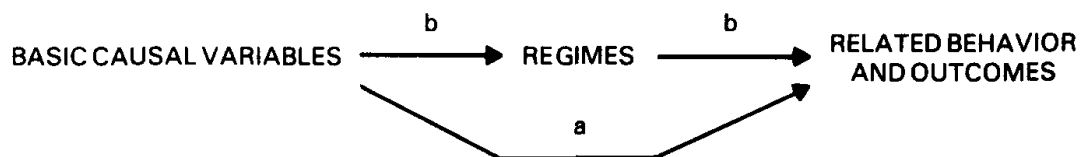


Figure 2

- Power and interests are mediated by regimes

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

Notation for Causal Mediation Analysis

- 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 $Y_i = Y_i(T_i, M_i(T_i))$

Defining and Interpreting Causal Mediation Effects

- Total causal effect:

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

- 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

- 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 realize when $T_i = t$
- Change the treatment from 0 to 1 while holding the mediator constant at $M_i(t)$

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

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

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**

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

Theorem 1 (Nonparametric Identification)

Under sequential ignorability,

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

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)**

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

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.

Algorithm for Estimating Causal Mediation Effects

Parametric models: **Quasi-Bayesian approximation** (King et al. 2000)

- Step 1: Estimate models for outcome and mediator
- Step 2: Take J draws from the asymptotic sampling distribution of model parameters
- Step 3: Repeat the following steps for each $j = 1, 2, \dots, J$
 - 1: Sample K copies of $M_i(t)$ from the mediator model
 - 2: Given this draw, sample one copy of $Y_i(t', M_i(t))$ from the outcome model
 - 3: Compute QoI based on these K sets of draws
- Step 4: Compute the point estimate and uncertainty estimates from the resulting J draws of QoI

Nonparametric/semiparametric models: **Nonparametric bootstrap**

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:** $\rho \equiv \text{Corr}(\epsilon_{i2}, \epsilon_{i3})$
- 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 $\sigma_j^2 \equiv \text{var}(\epsilon_{ij})$ 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 :

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

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

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

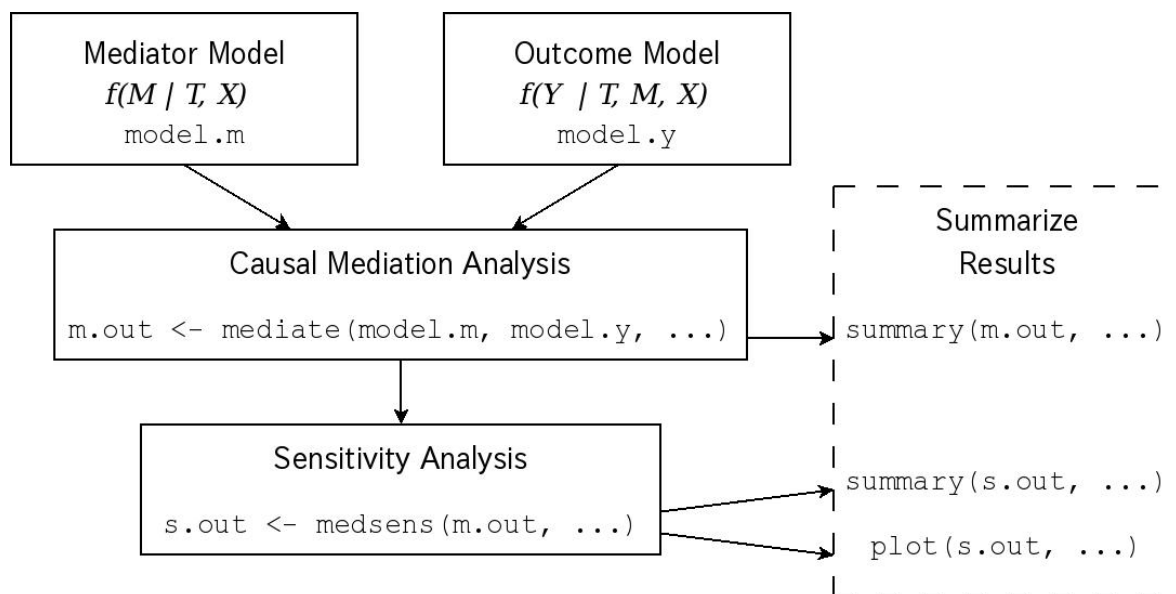
- Then reparameterize ρ using (R_M^{2*}, R_Y^{2*}) (or $(\tilde{R}_M^2, \tilde{R}_Y^2)$):

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

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 $(\tilde{R}_M^2, \tilde{R}_Y^2)$) to different values and see how mediation effects change

Overview of R Package **mediation**



- 1 Fit models for the mediator and outcome variable and store these models.

```
> m <- lm(Mediator ~ Treat + X)
> y <- lm(Y ~ Treat + Mediator + X)
```

- 2 **Mediation analysis:** Feed model objects into the `mediate` function. Call a summary of results.

```
> m.out <- mediate(m, y, treat="Treat", mediator="Mediator")
> summary(m.out)
```

- 3 **Sensitivity analysis:** Feed the output into the `medsens` function. Summarize and plot.

```
> s.out <- medsens(m.out)
> summary(s.out)
> plot(s.out, "rho")
> plot(s.out, "R2")
```

Data Types Available via **mediation** (For Now)

- For the `mediate` function:

<i>Mediator</i>	<i>Outcome</i>		
	Continuous	Ordered	Binary
Continuous	Yes	No	Yes
Ordered	Yes	No	No
Binary	Yes	No	Yes

- Can also deal with interaction between treatment and mediator, semiparametric regression, and quantile regression

- For the `medsens` function:

<i>Mediator</i>	<i>Outcome</i>		
	Continuous	Ordered	Binary
Continuous	Yes	No	Yes
Ordered	No	No	No
Binary	Yes	No	No

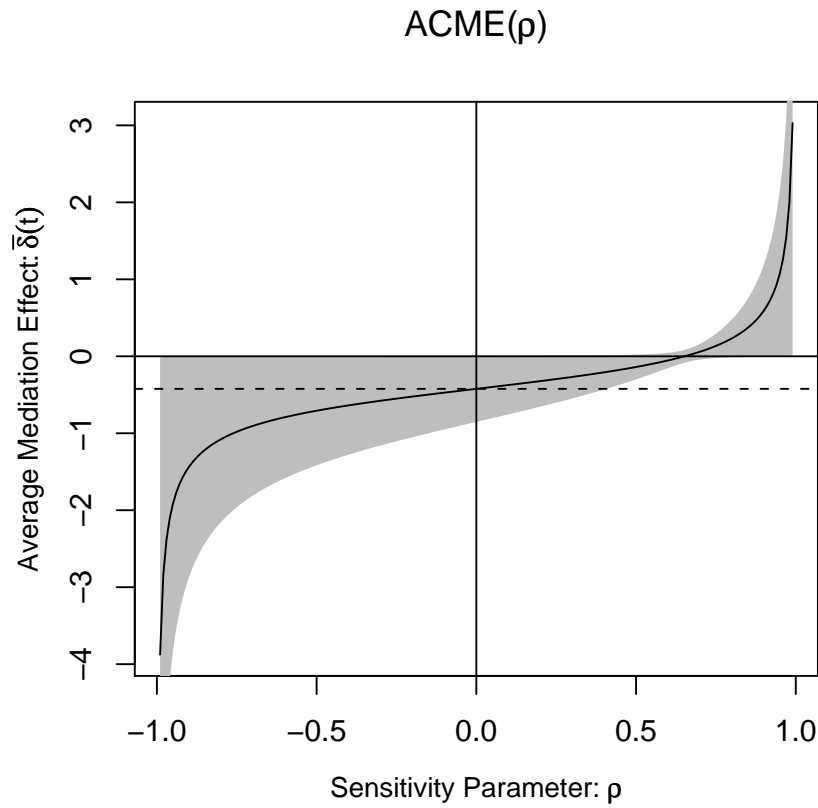
Empirical Illustration: Nelson *et al.* (APSR)

- 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

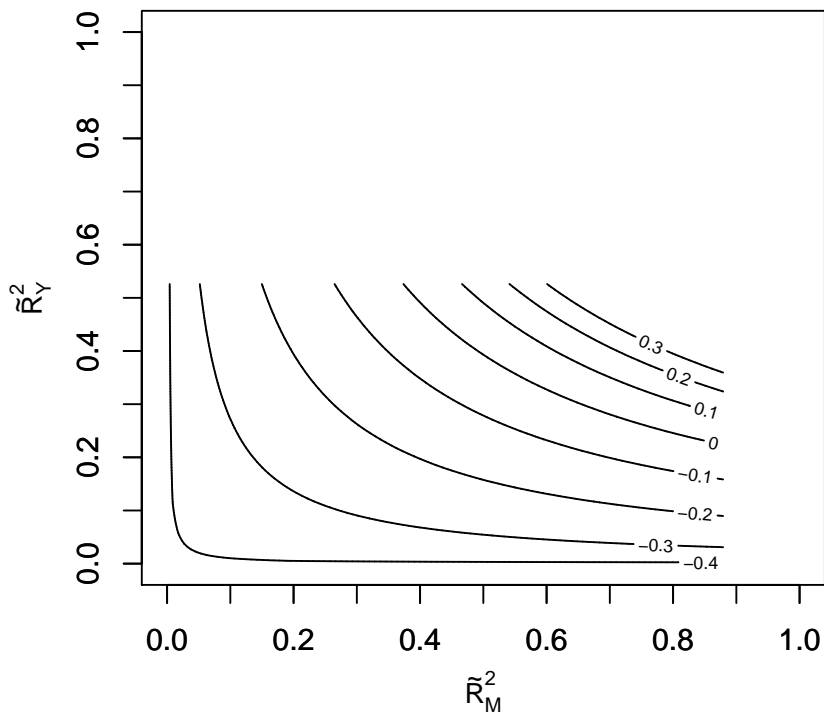
Average Mediation Effects $\hat{\delta}(0) = \hat{\delta}(1)$	-0.44 [-0.87, -0.01]
Average Direct Effects $\hat{\zeta}(0) = \hat{\zeta}(1)$	-0.02 [-0.49, 0.47]
Average Total Effect $\hat{\tau}$	-0.46 [-1.11, 0.23]

Sensitivity Analysis with Respect to ρ



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

ACME($\tilde{R}_M^2, \tilde{R}_Y^2$), $\text{sgn}(\lambda_2\lambda_3) = 1$



Summary

- Quantitative analysis can be used to identify causal mechanisms!
- Estimate causal mediation effects rather than marginal effects
- Generalize existing LSEM technique and put everything within causal inference framework
- Provide statistical software to implement our techniques
- Current and future work: multiple mediators, sensitivity analysis for other data types