Causal Mechanisms

Teppei Yamamoto

Keio University

Introduction to Causal Inference
Spring 2016

Motivation

- Randomized experiments and well-designed observational studies allow us to make inferences about whether X causes Y
- However, they normally don't tell us how and why X causes Y
- But this is often what political scientists are really interested in!

Motivation

- Randomized experiments and well-designed observational studies allow us to make inferences about whether X causes Y
- However, they normally don't tell us how and why X causes Y
- But this is often what political scientists are really interested in!
- Criticisms against the "experimental paradigm" are often about its black-box nature

Motivation

- Randomized experiments and well-designed observational studies allow us to make inferences about whether X causes Y
- However, they normally don't tell us how and why X causes Y
- But this is often what political scientists are really interested in!
- Criticisms against the "experimental paradigm" are often about its black-box nature
- Can we learn about causal mechanisms from quantitative data?
- Researchers typically do ad-hoc data analysis to justify their interpretation, or just give up and go qualitative
- Causal mediation analysis is a more consistent framework to think about causal mechanisms quantitatively

Goals

In this lecture, we will cover (as time permits):

- A quantitative definition of causal mechanisms
- Assumptions needed to identify a causal mechanism from data
- A general procedure to estimate a causal mechanism (given the assumptions)
- Methods for analyzing sensitivity to the violation of the assumptions
- Experimental designs to identify mechanisms with weaker assumptions

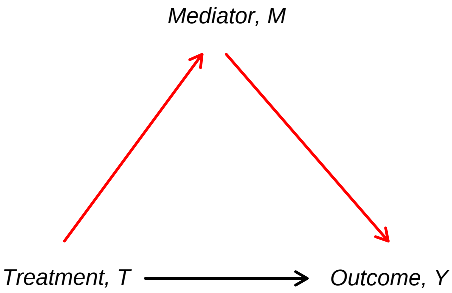
All the methods can be implemented in the R package `mediation`

What Is a Causal Mechanism?

- Mechanisms as causal pathways

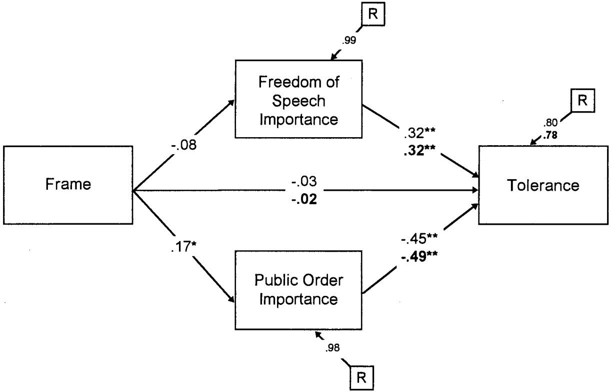
What Is a Causal Mechanism?

- Mechanisms as causal pathways
- Causal mediation analysis



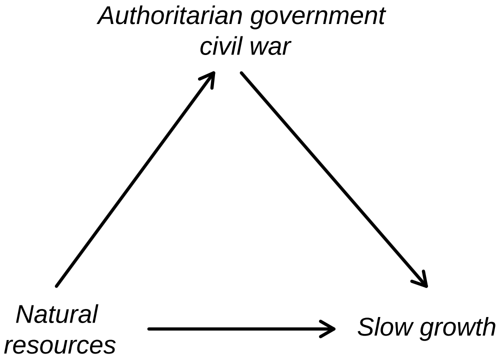
- Quantities of interest: Direct and indirect effects
- Fast growing methodological literature in the past 10–20 years

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



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 (*IO*, 1982)

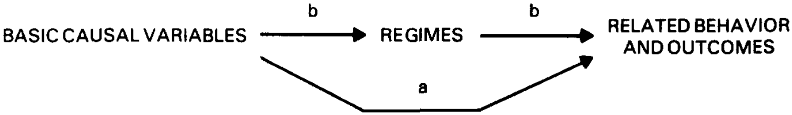


Figure 2

- Power and interests are mediated by regimes

Potential Outcomes for Mediation Analysis

- 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 Outcomes for Mediation Analysis

- 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

Potential Outcomes for Mediation Analysis

- 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

- In a standard experiment, only one potential outcome can be observed for each i

Potential Outcomes for Mediation Analysis

- 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

- In a standard experiment, only one potential outcome can be observed for each i
- Moreover, some potential outcomes can never be observed: $Y_i(t, M_i(t'))$ where $t \neq t'$

Causal Mediation Effects

- Total causal effect:

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

Causal Mediation Effects

- Total causal effect:

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

- Causal mediation effects (a.k.a. natural indirect effects):

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

Causal Mediation Effects

- Total causal effect:

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

- Causal mediation effects (a.k.a. natural 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
- Intuition: 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

- Natural direct effects:

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

Total Effect = Indirect Effect + Direct Effect

- Natural 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 $T_i = t$
- Intuition: Change the treatment from 0 to 1 while holding the mediator constant at its *natural value given t* for each i, $M_i(t)$
- Represents all mechanisms other than through M_i

Total Effect $=$ Indirect Effect $+$ Direct Effect

- Natural 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 $T_i = t$
- Intuition: Change the treatment from 0 to 1 while holding the mediator constant at its *natural value given t* for each i, $M_i(t)$
- Represents all mechanisms other than through M_i

Cf. Controlled direct effect:

$$\xi_i(m) \;\equiv\; Y_i(1, m) - Y_i(0, m)$$

- Causal effect of T_i on Y_i when mediator is manipulated at a *fixed value m* (regardless of unit i's natural response to T_i)

Total Effect = Indirect Effect + Direct Effect

- Natural 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 $T_i = t$
- Intuition: Change the treatment from 0 to 1 while holding the mediator constant at its *natural value given t* for each i, $M_i(t)$
- Represents all mechanisms other than through M_i

Cf. Controlled direct effect:

$$\xi_i(m) \equiv Y_i(1, m) - Y_i(0, m)$$

- Causal effect of T_i on Y_i when mediator is manipulated at a *fixed value m* (regardless of unit *i*'s natural response to T_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 Do the Observed Data Tell Us?

- Quantity of Interest: Average causal mediation effects

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

What Do the Observed Data Tell Us?

- Quantity of Interest: Average causal mediation effects

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

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

What Do the Observed Data Tell Us?

- Quantity of Interest: Average causal mediation effects

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

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

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

What Do the Observed Data Tell Us?

- Quantity of Interest: Average causal mediation effects

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

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

- Problem: $Y_i(t, M_i(t))$ 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 Standard Research Design

- Standard experiment: Randomize T_i and measure M_i and Y_i
- An identification assumption: Sequential Ignorability (SI)

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

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

Identification under Standard Research Design

- Standard experiment: Randomize T_i and measure M_i and Y_i
- An 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 standard experiments
- (2) does **not** hold if there exist:
  - unobserved pre-treatment M–Y confounders, **or**
  - any post-treatment M–Y confounding, *even if observed*
- In observational studies, neither (1) nor (2) is guaranteed

Identification under Standard Research Design

- Standard experiment: Randomize T_i and measure M_i and Y_i
- An 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 standard experiments
- (2) does **not** hold if there exist:
  - unobserved pre-treatment M–Y confounders, **or**
  - any post-treatment M–Y confounding, *even if observed*
- In observational studies, neither (1) nor (2) is guaranteed

Theorem (Imai et al. 2010): Under sequential ignorability, ACME and average direct effects are nonparametrically identified

A General Estimation Algorithm

Given the assumptions, the ACME and ADE can be estimated via the following imputation algorithm

A General Estimation Algorithm

Given the assumptions, the ACME and ADE can be estimated via the following imputation 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)

A General Estimation Algorithm

Given the assumptions, the ACME and ADE can be estimated via the following imputation 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

A General Estimation Algorithm

Given the assumptions, the ACME and ADE can be estimated via the following imputation 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 (Clarify) or bootstrapping to estimate uncertainty

Example: Anxiety, Group Cues and Immigration

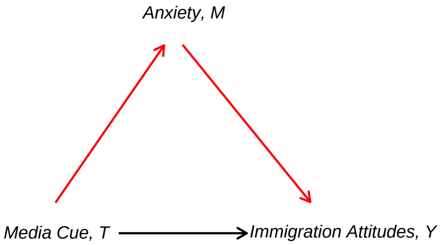
Brader, Valentino & Suhat (2008, AJPS)

- How and why do ethnic cues affect immigration attitudes?

Example: Anxiety, Group Cues and Immigration

Brader, Valentino & Suhat (2008, AJPS)

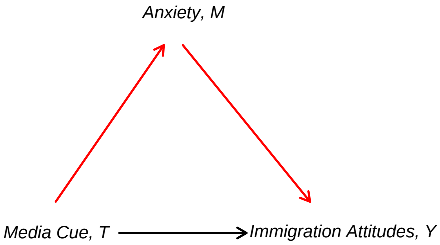
- How and why do ethnic cues affect immigration attitudes?
- Theory: Anxiety transmits the effect of cues on attitudes



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

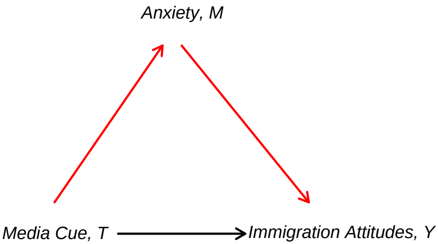


- ACME $=$ Average difference in immigration attitudes due to the change in anxiety induced by the media cue treatment

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

Reanalysis: Estimates under Sequential Ignorability

- Original method: Product of coefficients with the Sobel test

Reanalysis: Estimates 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)

Reanalysis: Estimates 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)
- This method: Calculate ACME using the general algorithm

Reanalysis: Estimates 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)
- This method: Calculate ACME using the 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]

Beyond Sequential Ignorability

- Assumption (2) is too strong in most scenarios
- Can we go beyond just making this assumption?

Beyond Sequential Ignorability

- Assumption (2) is too strong in most scenarios
- Can we go beyond just making this assumption?
- Sensitivity analysis: Assess the robustness of the estimates to the violation of sequential ignorability
- 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

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

Beyond Sequential Ignorability

- Assumption (2) is too strong in most scenarios
- Can we go beyond just making this assumption?
- Sensitivity analysis: Assess the robustness of the estimates to the violation of sequential ignorability
- 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

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

- Addresses the possible existence of unobserved *pre-treatment* confounders
- But not post-treatment confounders

Parametric Sensitivity Analysis

- Assume a linear structural equations model:

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

Parametric Sensitivity Analysis

- Assume a linear structural equations model:

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

- Sensitivity parameter: $\rho \equiv \mathrm{Corr}(\varepsilon_{i2}, \varepsilon_{i3})$

Parametric Sensitivity Analysis

- Assume a linear structural equations model:

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

- Sensitivity parameter: $\rho \equiv \mathrm{Corr}(\varepsilon_{i2}, \varepsilon_{i3})$

Result: ACME is identified given ρ as

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

where $\sigma_j^2 \equiv \mathbb{V}(\varepsilon_{ij})$ for $j = 1, 2$ and $\tilde{\rho} \equiv \mathrm{Corr}(\varepsilon_{i1}, \varepsilon_{i2})$ such that

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

Parametric Sensitivity Analysis

- Assume a linear structural equations model:

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

- Sensitivity parameter: $\rho \equiv \mathrm{Corr}(\varepsilon_{i2}, \varepsilon_{i3})$

Result: ACME is identified given ρ as

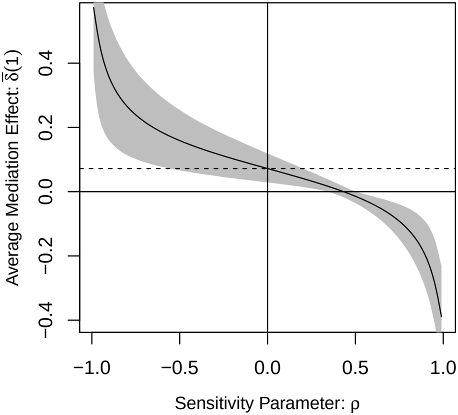
$$\bar{\delta}(0) = \bar{\delta}(1) = \frac{\beta_2 \sigma_1}{\sigma_2} \left\{ \tilde{\rho} - \rho \sqrt{(1 - \tilde{\rho}^2)/(1 - \rho^2)} \right\},$$

where $\sigma_j^2 \equiv \mathbb{V}(\varepsilon_{ij})$ for $j = 1, 2$ and $\tilde{\rho} \equiv \mathrm{Corr}(\varepsilon_{i1}, \varepsilon_{i2})$ such that

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

- Set ρ to different values and see how ACME changes
- Sequential ignorability implies $\rho = 0$

Anxiety Example: Sensitivity Analysis w.r.t. ρ



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

Alternative Formulation for Easier Interpretation

- Interpreting ρ: how small is too small?

Alternative Formulation for Easier Interpretation

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

$$\varepsilon_{i2} \; = \; \lambda_2 U_i + \varepsilon'_{i2} \quad \text{and} \quad \varepsilon_{i3} \; = \; \lambda_3 U_i + \varepsilon'_{i3}$$

- How much does U_i have to explain the variances of M_i and Y_i for our results to go away?

Alternative Formulation for Easier Interpretation

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

$$\varepsilon_{i2} \;=\; \lambda_2 U_i + \varepsilon'_{i2} \quad \text{and} \quad \varepsilon_{i3} \;=\; \lambda_3 U_i + \varepsilon'_{i3}$$

- How much does U_i have to explain the variances of M_i and Y_i for our results to go away?
- R squares as sensitivity parameters:
  1. Proportion of previously unexplained variance explained by U_i

$$R_M^{2*} \;\equiv\; 1 - \frac{\mathbb{V}(\varepsilon'_{i2})}{\mathbb{V}(\varepsilon_{i2})} \quad \text{and} \quad R_Y^{2*} \;\equiv\; 1 - \frac{\mathbb{V}(\varepsilon'_{i3})}{\mathbb{V}(\varepsilon_{i3})}$$

  2. Proportion of original variance explained by U_i

$$\widetilde{R}_M^2 \;\equiv\; \frac{\mathbb{V}(\varepsilon_{i2}) - \mathbb{V}(\varepsilon'_{i2})}{\mathbb{V}(M_i)} \quad \text{and} \quad \widetilde{R}_Y^2 \;\equiv\; \frac{\mathbb{V}(\varepsilon_{i3}) - \mathbb{V}(\varepsilon'_{i3})}{\mathbb{V}(Y_i)}$$

Alternative Formulation for Easier Interpretation

- Then ACME can be written in terms of these R squares as:

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

where R_M^2 and R_Y^2 are from the original mediator and outcome models

Alternative Formulation for Easier Interpretation

- Then ACME can be written in terms of these R squares as:

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

  where R_M^2 and R_Y^2 are from the original mediator and outcome models

- $\mathrm{sgn}(\lambda_2\lambda_3)$ indicates the direction of the effects of U_i on Y_i and M_i

Alternative Formulation for Easier Interpretation

- Then ACME can be written in terms of these R squares as:

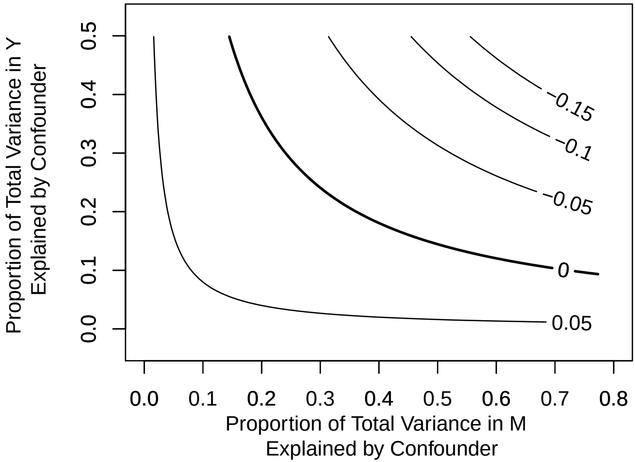
$$\rho \;=\; \mathrm{sgn}(\lambda_2\lambda_3) R_M^* R_Y^* \;=\; \frac{\mathrm{sgn}(\lambda_2\lambda_3)\widetilde{R}_M \widetilde{R}_Y}{\sqrt{(1-R_M^2)(1-R_Y^2)}},$$

  where R_M^2 and R_Y^2 are from the original mediator and outcome models

- $\mathrm{sgn}(\lambda_2\lambda_3)$ 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

Anxiety Example: Sensitivity Analysis w.r.t. $\tilde{R}^2_M$ and $\tilde{R}^2_Y$



- An unobserved confounder can account for up to 26.5% of the variation in both Y_i and M_i before ACME becomes zero

Experimental Designs for Identifying Mechanisms

- Sensitivity analysis may be unsatisfactory

Experimental Designs for Identifying Mechanisms

- Sensitivity analysis may be unsatisfactory
- What if we get rid of the assumption altogether?
- Under a standard design, even the sign of ACME is unidentified

Experimental Designs for Identifying Mechanisms

- Sensitivity analysis may be unsatisfactory
- What if we get rid of the assumption altogether?
- Under a standard design, even the sign of ACME is unidentified
- Can we do any better?

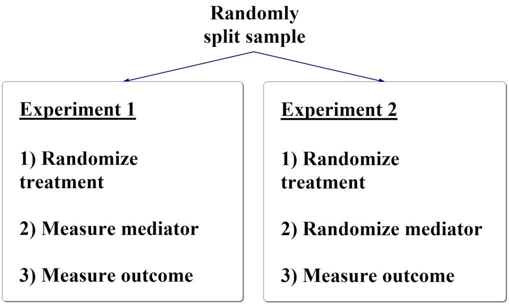
Experimental Designs for Identifying Mechanisms

- Sensitivity analysis may be unsatisfactory

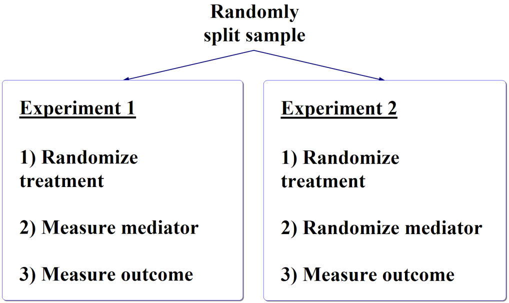
- What if we get rid of the assumption altogether?
- Under a standard design, even the sign of ACME is unidentified
- Can we do any better?

- Use alternative experimental designs for more credible yet powerful inference
- Designs feasible when the mediator can be directly or indirectly manipulated
- Experiments also serve as templates for observational studies

Parallel Design

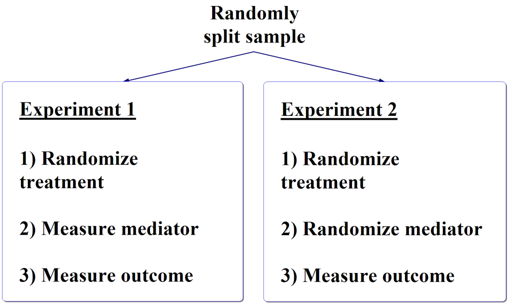


Parallel Design



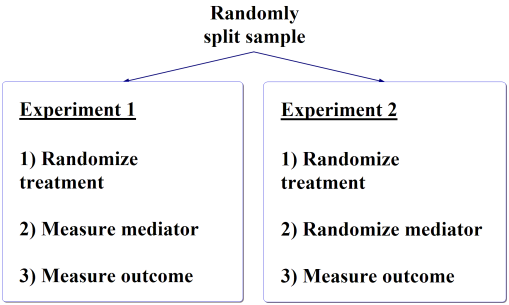
- Must assume consistency (i.e. no direct effect of manipulation on outcome)

Parallel Design



- Must assume consistency (i.e. no direct effect of manipulation on outcome)
- More informative than standard single experiment

Parallel Design



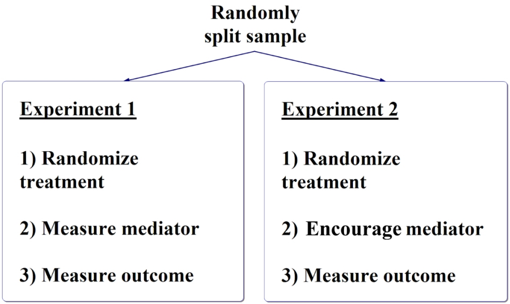
- Must assume consistency (i.e. no direct effect of manipulation on outcome)
- More informative than standard single experiment
- If we assume no T–M interaction, ACME is point identified
- Otherwise, we get bounds

Parallel Encouragement Design

- Direct manipulation of the mediator is often infeasible
- Even if feasible, more subtle form of intervention may be preferred to assure consistency

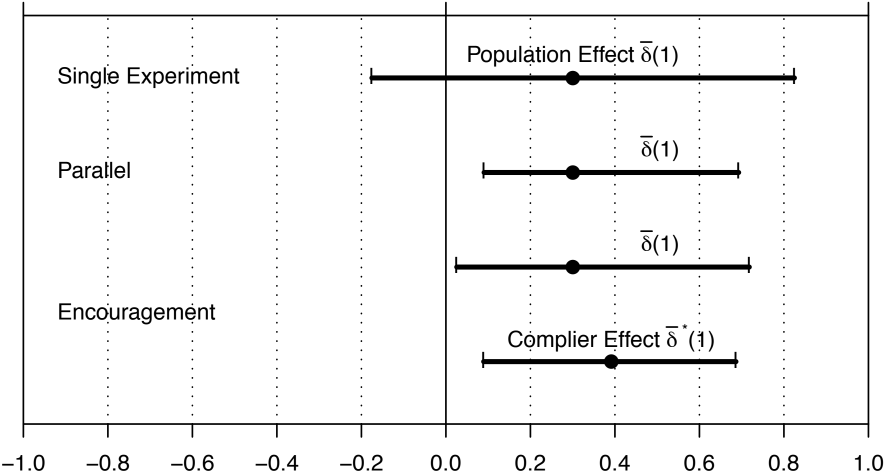
Parallel Encouragement Design

- Direct manipulation of the mediator is often infeasible
- Even if feasible, more subtle form of intervention may be preferred to assure consistency



- Parallel encouragement design: Randomly encourage subjects to take particular values of the mediator
- Standard instrumental variable assumptions (Angrist et al.)

Numerical Examples of the Bounds



Crossover 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 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

Crossover 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
- Crossover encouragement design:
  1. Round 1: Conduct a standard experiment
  2. Round 2: Same as crossover, except encourage subjects to take the mediator values

Crossover 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
- Crossover encouragement design:
  1. Round 1: Conduct a standard experiment
  2. Round 2: Same as crossover, except encourage subjects to take the mediator values
- Both must assume no carryover effect

Summary

- Even in a randomized experiment, a strong assumption is needed to identify causal mechanisms
- Analyzing mechanisms is, therefore, not so easy!
- Under the identification assumption, a general estimation procedure is available for various types of statistical models
- The violation of the assumption can be addressed by:
  - Analyzing sensitivity with respect to key assumptions
  - Creative research designs to avoid strong assumptions
- Therefore, progress can still be made!
- Extension to multiple mediators: Imai and Yamamoto (2013)
- Extension to instrumental variables: Yamamoto (2014)