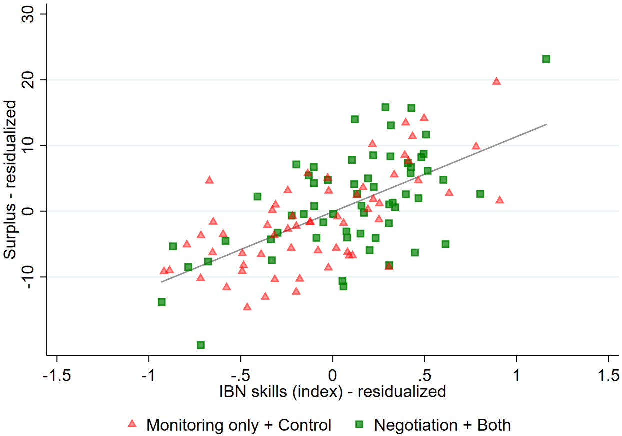
Lecture 15: Moderators, Mediators, and Causal Explanation

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

March 29, 2022


Surplus - residualized
30
20
10
0
-10
-1.5
-1
-.5
0
.5
1
1.5
IBN skills (index) - residualized
Monitoring only + Control
Negotiation + Both

Motivation

- A single effect may give rise to different interpretations.

Motivation

- A single effect may give rise to different interpretations.
  - GDP/capita $\rightarrow$ conflict: opportunity costs of labor? state policing capacity?

Motivation

- A single effect may give rise to different interpretations.
  - GDP/capita $\rightarrow$ conflict: opportunity costs of labor? state policing capacity?
  - Ethnic/racial diversity $\rightarrow$ lower public goods provision: discrimination? communication barriers?

Motivation

- A single effect may give rise to different interpretations.
  - GDP/capita $\rightarrow$ conflict: opportunity costs of labor? state policing capacity?
  - Ethnic/racial diversity $\rightarrow$ lower public goods provision: discrimination? communication barriers?
- Different interpretations have different implications for theory and policy.
- How can we sort between different interpretations?

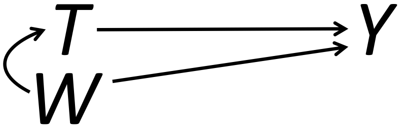
Motivation

- Different interpretations have different observable implications beyond the reduced form cause-effect relationship.
- "$\Rightarrow$ effects should be *stronger* for certain types."
- "$\Rightarrow$ effects should be *transmitted through* certain pathways."

Motivation

$$T \longrightarrow Y$$

Motivation



Motivation

$$T \longrightarrow Y$$

$W=1$

$$T \qquad\qquad Y$$

$W=0$

Motivation

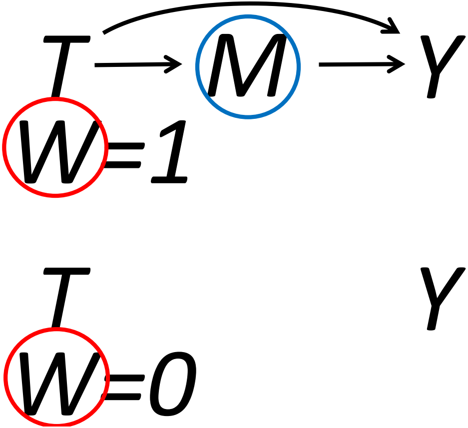
$$\begin{tikzcd} T \arrow[r] \arrow[rr, bend left] & M \arrow[r] & Y \end{tikzcd}$$

$W=1$

$$\begin{tikzcd} T & & Y \end{tikzcd}$$

$W=0$

Motivation



- M =“Mediator”—M *mediates* the effect of T on Y.
- W =“Moderator”—W *moderates* the effect of T on Y.

Motivation

- Today: mediation & mechanisms.
- Next time: moderators and effect heterogeneity.

Mediation

$$\begin{tikzcd} T \arrow[r] \arrow[rr, bend left] & M \arrow[r] & Y \end{tikzcd}$$

- Basic mediation graph.
- To what extent does M mediate the effect of T on Y?

Mediation effect vs. front door

$$\begin{tikzcd} T \arrow[r] \arrow[rr, bend left] & M \arrow[r] & Y \end{tikzcd}$$

$$\begin{tikzcd} T \arrow[r] & M \arrow[r] & Y \\ & U \arrow[ul, bend left] \arrow[ur, bend right] & \end{tikzcd}$$

Mediation

$$\begin{tikzcd} T \arrow[r] \arrow[rr, bend left] & M \arrow[r] & Y \end{tikzcd}$$

- Basic mediation graph.
- To what extent does M mediate the effect of T on Y?

Defining effects

- Suppose a sample indexed by i.
- Realized and potential outcomes for unit i:
  - $Y_i = Y_i(t, m)$ when $T_i = t$ and $M_i = m$.
  - $M_i = M_i(t)$ when $T_i = t$.
- For simplicity, suppose $T_i = 0, 1$.
- Definitions below follow Robins and Greenland (1992), Pearl (2001), Imai et al. (2010, 2011), and Vanderweele (2015).

Defining effects

- Total effect: $\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0))$
  "Overall effect of treatment on outcome."

Defining effects

- Total effect: $\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0))$
  "Overall effect of treatment on outcome."
- Controlled direct effect: $\kappa_i(m) = Y_i(1, m) - Y_i(0, m)$
  "Effect of treatment if we block path through M."

Defining effects

- Total effect: $\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0))$
  "Overall effect of treatment on outcome."
- Controlled direct effect: $\kappa_i(m) = Y_i(1, m) - Y_i(0, m)$
  "Effect of treatment if we block path through M."
- Natural direct effect: $\zeta_i(t) = Y_i(1, M_i(t)) - Y_i(0, M_i(t))$
  "Share of total effect that doesn't go through M."

Defining effects

- Total effect: $\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0))$
  "Overall effect of treatment on outcome."
- Controlled direct effect: $\kappa_i(m) = Y_i(1, m) - Y_i(0, m)$
  "Effect of treatment if we block path through M."
- Natural direct effect: $\zeta_i(t) = Y_i(1, M_i(t)) - Y_i(0, M_i(t))$
  "Share of total effect that doesn't go through M."
- Natural mediation effect: $\delta_i(t) = Y_i(t, M_i(1)) - Y_i(t, M_i(0))$
  "Share of total effect that goes through M."

Defining effects

- Total effect: $\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0))$
  "Overall effect of treatment on outcome."
- Controlled direct effect: $\kappa_i(m) = Y_i(1, m) - Y_i(0, m)$
  "Effect of treatment if we block path through M."
- Natural direct effect: $\zeta_i(t) = Y_i(1, M_i(t)) - Y_i(0, M_i(t))$
  "Share of total effect that doesn't go through M."
- Natural mediation effect: $\delta_i(t) = Y_i(t, M_i(1)) - Y_i(t, M_i(0))$
  "Share of total effect that goes through M."

Note the decomposition of the total effect:

$$\begin{aligned}
\tau_i &= Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \\
&= \underbrace{Y_i(1, M_i(1)) - Y_i(1, M_i(0))}_{\delta_i(1)} + \underbrace{Y_i(1, M_i(0)) - Y_i(0, M_i(0))}_{\zeta_i(0)} \\
&= \underbrace{Y_i(1, M_i(1)) - Y_i(0, M_i(1))}_{\zeta_i(1)} + \underbrace{Y_i(0, M_i(1)) - Y_i(0, M_i(0))}_{\delta_i(0)}
\end{aligned}$$

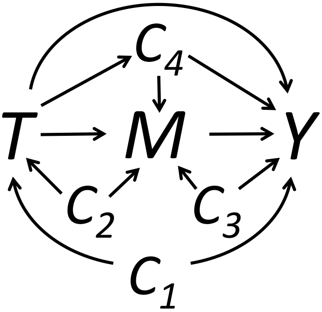
$\Rightarrow \tau_i = \delta_i(t) + \zeta_i(1-t)$ for $t = 0, 1$.

Identification

- We focus on identifying average causal effects:
  - Average total effect:
    $\tau = \mathrm{E}\left[Y_i(1, M_i(1)) - Y_i(0, M_i(0))\right]$
  - Average controlled direct effect: $\kappa(m) = \mathrm{E}\left[Y_i(1, m) - Y_i(0, m)\right]$
  - Average natural direct effect: $\zeta(t) = \mathrm{E}\left[Y_i(1, M_i(t)) - Y_i(0, M_i(t))\right]$
  - Average natural mediation effect*:
    $\delta(t) = \mathrm{E}\left[Y_i(t, M_i(1)) - Y_i(t, M_i(0))\right]$
    (*Imai et al.: "avg. causal mediation effect", ACME).
- From above: $\tau = \delta(t) + \zeta(1-t)$ for t=0,1
- Identification of 2 implies the third is also identified.

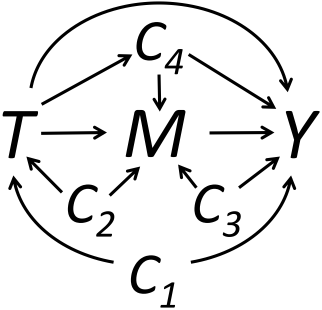
Identification

- Consider a richer DAG:



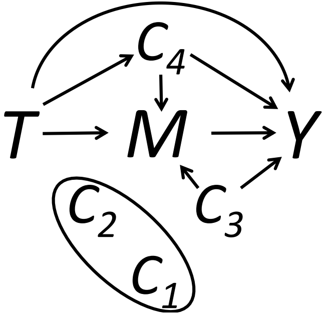
- C_1: usual confounding for effect of T on Y.
- C_2: confounding for effect of T on M.
- C_3, C_4: confounding for M on Y, with C_4 endogenous to T.

Identification



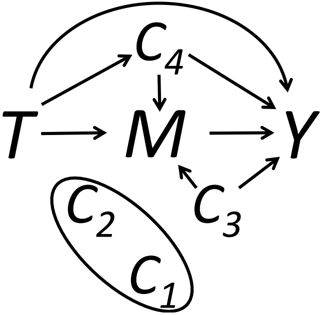
- $T \perp\!\!\!\perp Y(m,t) | C_1$ for all m, t.
- Also, $T \perp\!\!\!\perp M(t) | C_2$ for all m, t.
- Altogether: $T \perp\!\!\!\perp (Y(m,t), M(t)) | (C_1, C_2)$.

Identification



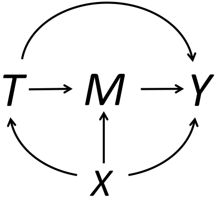
- $T \perp\!\!\!\perp (Y(m,t), M(t)) | (C_1, C_2)$.
- Randomization means (C_1, C_2) are not even on the graph.
- W/o randomization, CIA wrt C_1, C_2, so must be measured.
- $\Rightarrow$ with randomization (or CIA) you can
  - Estimate $T \rightarrow Y$.
  - Estimate $T \rightarrow M$ to see if mediation via M is *plausible*.

Identification



- $M \perp\!\!\!\perp Y(m,t) | (C_1, C_2, T, C_3, C_4)$.
- Even w/ randomization, have to deal with C_3 and C_4.
- If C_3 or C_4 contain unobservables, $M \to Y$ not identified.
- $\Rightarrow$ randomization does not identify mediation effects.

Identification



- Imai et al. (2010, 2011) consider this scenario.
- Covariate vector X contains C_1, C_2, C_3 confounders; no C_4.

Identification

$$\begin{tikzcd} T \arrow[r] \arrow[rr, bend left] & M \arrow[r] & Y \\ & X \arrow[ul, bend left] \arrow[u] \arrow[ur, bend right] & \end{tikzcd}$$

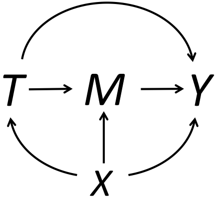
- Imai et al. (2010, 2011) consider this scenario.
- Covariate vector X contains C_1, C_2, C_3 confounders; no C_4.
- $\Rightarrow$ "sequential ignorability" (SI) condition:

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

$$M_i(t) \perp\!\!\!\perp Y_i(t',m) | T_i = t, X_i = x \tag{2}$$

for $t, t' = 0, 1$ & $x \in \mathscr{X}$, w/ $0 < \Pr[T_i = 1 | X_i = x]$ &
$0 < p(M_i(t) = m | T_i = t, X_i = x)$ for $t = 0, 1$ & all $x \in \mathscr{X}$, $m \in \mathscr{M}$.

Identification



- Imai et al. (2010, 2011) consider this scenario.
- Covariate vector X contains C_1, C_2, C_3 confounders; no C_4.
- $\Rightarrow$ "sequential ignorability" (SI) condition:

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

$$M_i(t) \perp\!\!\!\perp Y_i(t', m) | T_i = t, X_i = x \tag{2}$$

  for $t, t' = 0, 1$ & $x \in \mathscr{X}$, w/ $0 < \Pr[T_i = 1 | X_i = x]$ &
  $0 < p(M_i(t) = m | T_i = t, X_i = x)$ for $t = 0, 1$ & all $x \in \mathscr{X}$, $m \in \mathscr{M}$.
- This SI condition identifies natural direct & mediation effects.

Identification

Lemma:

- We have

$$(Y_i(t',m), M_i(t)) \perp\!\!\!\perp T_i | X_i$$
$$\Rightarrow Y_i(t',m) \perp\!\!\!\perp T_i | X_i \text{ and } M_i(t) \perp\!\!\!\perp T_i | X_i$$

  As such,

$$p[T_i|X_i] = p[T_i|Y_i(t',m), M_i(t), X_i]$$

  and

$$p[T_i|X_i] = p[T_i|M_i(t), X_i]$$

  in which case

$$p[T_i|Y_i(t',m), M_i(t), X_i] = p[T_i|M_i(t), X_i]$$

  and so

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

Identification

$$\text{ACME: } \delta(t) = \mathrm{E}\left[Y_i(t, M_i(1)) - Y_i(t, M_i(0))\right].$$

- Under SI, $Y_i(t, M_i(t'))$ counterfactual obeys:

$$\begin{aligned}
\mathrm{E}\left[Y_i(t, M_i(t')) | X_i = x\right] &= \int \mathrm{E}\left[Y_i(t,m) | M_i(t') = m, X_i = x\right] dF_{M_i(t')|X_i = x}(m) \\
(lemma) &= \int \mathrm{E}\left[Y_i(t,m) | M_i(t') = m, T_i = t', X_i = x\right] dF_{M_i(t')|X_i = x}(m) \\
(2) &= \int \mathrm{E}\left[Y_i(t,m) | T_i = t', X_i = x\right] dF_{M_i(t')|T_i = t', X_i = x}(m) \\
(1), (lemma) &= \int \mathrm{E}\left[Y_i(t,m) | T_i = t, X_i = x\right] dF_{M_i(t')|T_i = t', X_i = x}(m) \\
(2) &= \int \mathrm{E}\left[Y_i(t,m) | M_i(t) = m, T_i = t, X_i = x\right] dF_{M_i(t')|T_i = t', X_i = x}(m) \\
&= \int \mathrm{E}\left[Y_i | M_i = m, T_i = t, X_i = x\right] dF_{M_i|T_i = t', X_i = x}(m)
\end{aligned}$$

- That is, weighted average of outcomes in t but with weights from dist. of m in group t'.

Identification

$$\text{ACME: } \delta(t) = \mathrm{E}\left[Y_i(t, M_i(1)) - Y_i(t, M_i(0))\right].$$

- Plugging this in, ACME is identified as,

$$\delta(t) = \mathrm{E}_X\Big\{ \int \mathrm{E}\left[Y_i | M_i = m, T_i = t, X_i = x\right] \\ \times \left(dF_{M_i|T_i=1,X_i=x}(m) - dF_{M_i|T_i=0,X_i=x}(m)\right)\Big\}$$

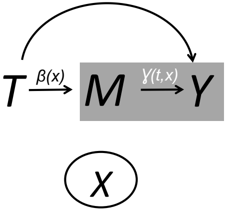
- Using outcomes from group t.
- Weighting by distributions of M in groups t and t'
- Taking the difference.

Identification

- When $M_i = 0, 1$:

$$
\begin{aligned}
\delta(t) &= \mathrm{E}_X\{\mathrm{E}[Y_i|M_i=0,T_i=t,X](\Pr[M_i=0|T_i=1,X]-\Pr[M_i=0|T_i=0,X]) \\
&\qquad + \mathrm{E}[Y_i|M_i=1,T_i=t,X](\Pr[M_i=1|T_i=1,X]-\Pr[M_i=1|T_i=0,X])\} \\
&= \mathrm{E}_X\{\underbrace{(\mathrm{E}[Y_i|M_i=1,T_i=t,X]-\mathrm{E}[Y_i|M_i=0,T_i=t,X])}_{\gamma(t,x)} \\
&\qquad \times \underbrace{(\Pr[M_i=1|T_i=1,X]-\Pr[M_i=1|T_i=0,X])}_{\beta(x)}\} \\
&= \beta\gamma(t).
\end{aligned}
$$

(where X is shorthand for $X_i = x$)



Identification

Average Natural Direct Effect: $\zeta(t) = \mathrm{E}[Y_i(1, M_i(t)) - Y_i(0, M_i(t))]$.

- By similar arguments

$$\begin{aligned}\zeta(t) = &\mathrm{E}_X[\int \{\mathrm{E}[Y_i | M_i = m, T_i = 1, X_i = x] \\ &- \mathrm{E}[Y_i | M_i = m, T_i = 0, X_i = x]\} dF_{M_i | T_i = t, X_i = x}(m)]\end{aligned}$$

- Differences across treatment and control, weighting by distribution of M in group t.

Estimation

- Classical approach uses linear structural equation models (Barron & Kenny, 1986):

$$
\begin{aligned}
M_i &= \alpha + \beta T_i + X_i'\delta + \varepsilon_i \\
Y_i &= \lambda + \omega T_i + \gamma M_i + X_i'\zeta + \nu_i,
\end{aligned}
$$

  with $\delta(0) = \delta(1) = \beta\gamma$ and $\zeta = \omega$.
- Fit via OLS. Standard errors easy to derive.
- Consistency requires homogeneity, functional form (nb: no interaction), and SI to be true.
- Modest generalization adds the interaction:

$$
Y_i = \gamma_b + \omega_b T_i + \gamma_b M_i + X_i'\zeta_b + \kappa T_i M_i + \nu_{bi},
$$

  with $\delta(t) = \beta(\gamma_b + t\kappa_b)$ and $\zeta = \omega_b + \kappa(\alpha + t\beta)$.

Estimation

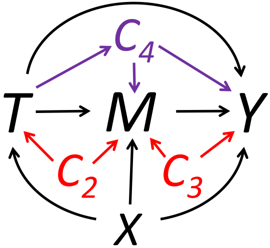
- Non-parametric approach generalizes wrt effect heterogeneity, non-linear X.
- E.g., for $M_i = 0, 1$, $T_i = 0, 1$, within strata defined by $X_i = x$, compute:

$$\hat{\gamma}(t) = \frac{\sum_{i=1}^{N} Y_i I(M_i = 1, T_i = t)}{\sum_{i=1}^{N} I(M_i = 1, T_i = t)} - \frac{\sum_{i=1}^{N} Y_i I(M_i = 0, T_i = t)}{\sum_{i=1}^{N} I(M_i = 0, T_i = t)}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{N} M_i I(T_i = 1)}{\sum_{i=1}^{N} I(T_i = 1)} - \frac{\sum_{i=1}^{N} M_i I(T_i = 0)}{\sum_{i=1}^{N} I(T_i = 0)}$$

  (could be done with series of simple regressions or a single interacted regression)
- Then $\hat{\delta}(t) = \hat{\beta}\hat{\gamma}(t)$. Similar for $\zeta(t)$.
- Standard errors from delta method or bootstrap.
- Imai et al. demonstrate approaches for general T_i and M_i.

Limitations

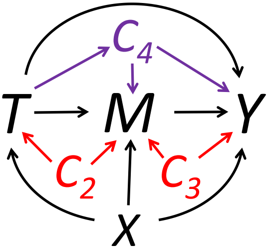


- Limits of Imai et al.'s results:
- Cannot have C_2 or C_3 type confounders outside X.
- Cannot include C_4 type confounders in X. More complex adjustment strategies needed.
  - "Effect of ethnic diversity on conflict mediated through economic growth?"
  - Diversity lowers growth directly through communication barriers but also indirectly through mistrust, but the latter also affects conflict in a more direct way...

Limitations

- Typically, neither the data nor design provide immediate ways to judge the plausibility of SI, absence of post-treatment confounding, or even causal ordering.
- Has to come from other substantive information.

Sensitivity Analysis



- Imai et al. develop methods for sensitivity analysis for C_2 and C_3 confounders that are not in X (e.g., unmeasured).
- Helpful to a certain extent, but leaves open C_4.

Experimental Designs for ACME?

- Suppose you randomize T_i on a population and estimate effect on M_i, and then randomize M_i on that population and estimate effects on Y_i.
- Does this identify ACME?

Experimental Designs for ACME?

- Suppose you randomize T_i on a population and estimate effect on M_i, and then randomize M_i on that population and estimate effects on Y_i.
- Does this identify ACME?
- No. E.g., ACME accounts for the fact that...
  - ...for some people, T has no effect on M. For such people, the effect of M on Y is not part of the ACME.
  - ...or, T may have a negative effect on M for some, and positive effect on others.
  - We would need to match up such heterogeneous effects on M to corresponding effects of M on Y to identify the ACME. *This is not straightforward.*
- Experimental designs and associated assumptions are quite subtle (Imai et al. 2011).

Other causal quantities

- Recall the average controlled direct effect (ACDE):

$$\kappa(m) = \mathrm{E}\left[Y_i(1,m) - Y_i(0,m)\right].$$

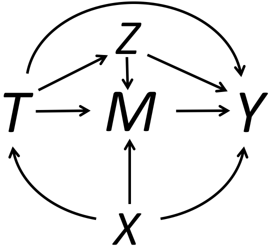
- Identified with a weaker sequential ignorability assumption (Robins & Greenland 1992; Pearl 2001; Acharya et al. 2015; Vanderweele 2015): Suppose X contains all C_1, C_2, C_3 confounders, and Z contains all C_4 confounders.

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

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

for $t, t' = 0, 1$, all $x \in \mathscr{X}$, with $0 < \Pr[T_i = 1 | X_i = x]$ and $0 < p(M_i(t) = m | T_i = t, X_i = x, Z_i = z)$ for $t = 0, 1$ and all $x \in \mathscr{X}$ and $m \in \mathscr{M}$.

Other causal quantities



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

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

for $t, t' = 0, 1$, all $x \in \mathscr{X}$, with $0 < \Pr[T_i = 1 | X_i = x]$ and $0 < p(M_i(t) = m | T_i = t, X_i = x, Z_i = z)$ for $t = 0, 1$ and all $x \in \mathscr{X}$ and $m \in \mathscr{M}$.

Other causal quantities

- To make estimation simpler, consider "no interactions" assumption:

$$\begin{aligned}&\mathrm{E}\left[Y_i(t,m) - Y_i(t,m')|X_i = x, T_i = t, Z_i = z\right]\\ &\qquad = \mathrm{E}\left[Y_i(t,m) - Y_i(t,m')|X_i = x, T_i = t\right]\end{aligned}$$

  for $t = 0,1$, all $x \in \mathscr{X}$, $m \in \mathscr{M}$, and $z \in \mathscr{Z}$.
- Then, ACDE estimation algorithm:
  1. estimate effect of M_i versus $M_i = 0$ conditional on (T_i, X_i, Z_i),
  2. "demediate" Y_i by subtracting off the relevant mediator effect,
  3. estimate effect of T_i on the demediated Y_i.
- Without the no interactions assumption, ACDE is still identified, but estimation is more complicated.

Discussion

Yes, But What's the Mechanism? (Don't Expect an Easy Answer)

John G. Bullock and Donald P. Green
Yale University

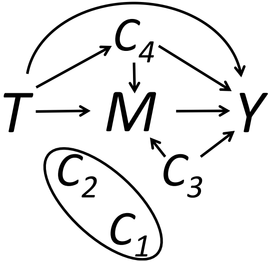
Shang E. Ha
Brooklyn College of the City University of New York

Enough Already about "Black Box" Experiments: Studying Mediation Is More Difficult than Most Scholars Suppose

Donald P. Green, Shang E. Ha and John G. Bullock

The ANNALS of the American Academy of Political and Social Science 2010 628: 200

Discussion



- With randomization (or CIA) you can
  - Estimate $T \rightarrow Y$.
  - Estimate $T \rightarrow M$ to see if mediation via M is *plausible*.
- Beyond that requires making commitments to untestable assumptions on the DGP.
- Can give you *suggestions* about how to interpret effects, perhaps with some care by using sensitivity analysis
- But should not be given the same standing as design-justified inferences.