Quantitative Methodology and Causal Mechanisms

- Investigation of causal mechanisms via intermediate variables
- Randomized experiments can only determine whether the treatment causes changes in the outcome
- Not how and why the treatment affects the outcome
- Social scientists use qualitative methods (e.g. process tracing) to answer these questions
- How can quantitative research be used to identify causal mechanisms?
Causal Mediation Analysis

- Quantities of interest: Direct and indirect effects
- Traditional tools: Path analysis, structural equation modeling
- Fast growing methodological literature

Common Practice

- 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!
Defining Causal Mediation Effects

- Binary treatment (can be generalized): \( T_i \in \{0, 1\} \)
- 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 \( Y_i = Y_i(T_i, M_i(T_i)) \)

Total causal effect: \( \tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \)

Causal mediation effects: \( \delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \)

Direct effects: \( \zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t)) \)

Total effect = Mediation (indirect) effect + Direct effect:

\[
\tau_i = \delta_i(t) + \zeta_i(1 - t) = \frac{1}{2} \sum_{t=0}^{1} \{\delta_i(t) + \zeta_i(t)\}
\]

Interpreting Causal Mediation Effects

- \( \delta_i(t) \): Causal effect of the change in \( M_i \) on \( Y_i \) that would be induced by \( T_i \), holding actual treatment constant at \( t \)
- \( \zeta_i(t) \): Causal effect of \( T_i \) on \( Y_i \), holding mediator constant at its potential value that would realize when \( T_i = t \)

Different from controlled direct effects: \( Y_i(t, m) - Y_i(t, m') \)

Mediation effects — identify causal paths from \( T_i \) to \( Y_i \)

Controlled effects — study how \( T_i \) moderates the effect of \( M_i \) on \( Y_i \)

Average Causal Mediation Effects:

\[
\bar{\delta}(t) \equiv \mathbb{E}(\delta_i(t)) = \mathbb{E}\{ Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \}
\]
Nonparametric Identification

- 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\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!"
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

\[ Y_i(t', m) \not\perp \perp M_i \mid T_i = t, X_i = x \]

Possible existence of unobserved *pre-treatment* confounder

---

**Parametric Sensitivity Analysis**

Consider LSEM (aka Baron-Kenny procedure):

\[
M_i = \alpha_2 + \beta_2 T_i + \epsilon_{2i}, \quad Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i}.
\]

**Sensitivity parameter**: \( \rho \equiv \text{Corr}(\epsilon_{2i}, \epsilon_{3i}) \)

Sequential ignorability implies \( \rho = 0 \)

Set \( \rho \) to different values and see how mediation effects change

An alternative explanation of \( \rho \) based on \( R^2 \)

Work for probit models – binary outcome, binary mediator, etc.

Difficult to construct a more general sensitivity analysis
Facilitating Interpretation

- How big is $\rho$?
- An unobserved (pre-treatment) confounder formulation:
  \begin{align*}
  \epsilon_{2i} &= \lambda_2 U_i + \epsilon'_{2i} \quad \text{and} \quad \epsilon_{3i} = \lambda_3 U_i + \epsilon'_{3i},
  \end{align*}
- Assume $Y_i(t', m) \perp M_i \mid T_i = t, U_i = u$
- Assume also $\epsilon'_{2i} \perp U_i$ and $\epsilon'_{3i} \perp U_i$
- Proportion of previously unexplained variance explained by the unobserved confounder
  \begin{align*}
  R^2_M &= 1 - \frac{\text{var}(\epsilon'_{2i})}{\text{var}(\epsilon_{2i})}, \quad \text{and} \quad R^2_Y \equiv 1 - \frac{\text{var}(\epsilon'_{3i})}{\text{var}(\epsilon_{3i})}
  \end{align*}
Proportion of original variance explained by the unobserved confounder

\[ \tilde{R}_M^2 = \frac{\text{var}(\epsilon_2) - \text{var}(\epsilon'_2)}{\text{var}(M_i)} \quad \text{and} \quad \tilde{R}_Y^2 = \frac{\text{var}(\epsilon_3) - \text{var}(\epsilon'_3)}{\text{var}(Y_i)} \]

Specify \(\text{sgn}(\lambda_2\lambda_3)\) and \(R^*_M, R^*_Y\) (or \(\tilde{R}_M^2, \tilde{R}_Y^2\))

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

where \(R^2_M\) and \(R^2_Y\) are based on

\[ M_i = \alpha_2 + \beta_2 T_i + \epsilon_2 \]
\[ Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_3 \]
Overview of **R Package mediation**

- Object-oriented nature of **R** made this design possible

### An Illustrative Example

- **Job Search Intervention Study (JOBS II)**
- A randomized evaluation of a job training program
- **Treatment**: Job-skills workshop
- **Mediator**: a continuous measure of job-search self-efficacy
- **Outcome**: a binary measure of employment
- **Question**: Does the workshop improve the prospect of future employment by increasing the level of job-search self-efficacy?
Step 1: Fitting the Outcome and Mediator Models

`# load the library`
`library(mediation)`
`# load the data set`
`data(jobs)`
`# fit the mediator model`
`model.m <- lm(job_seek ~ treat + depress1 + econ_hard + sex + age + occp + marital + nonwhite + educ + income, data = jobs)`
`# fit the outcome model`
`model.y <- glm(work1 ~ treat + job_seek + depress1 + econ_hard + sex + age + occp + marital + nonwhite + educ + income, family = binomial(link="probit"), data = jobs)`

Step 2: Conducting Causal Mediation Analysis

`# mediation analysis`
`m.out <- mediate(model.m, model.y, sims = 1000, T = "treat", M = "job_seek")`
`# summary of the analysis`
`summary(m.out)`

Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

Mediation Effect: 0.003558 95% CI -0.001074 0.010679
Direct Effect: 0.05455 95% CI -0.006838 0.116466
Total Effect: 0.0581 95% CI -0.003083 0.119178
Proportion of Total Effect via Mediation: 0.05687 95% CI -0.2028 0.4490
Step 3: Conducting Sensitivity Analysis

```r
> s.out <- medsens(model.m, model.y, sims = 1000,
>                   T = "treat", M = "job_seek", INT = FALSE,
>                   DETAIL = FALSE)
> summary(s.out)

Mediation Sensitivity Analysis

Sensitivity Region

<table>
<thead>
<tr>
<th>Rho</th>
<th>Med. Eff.</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,] -0.9</td>
<td>0.1183</td>
<td>-0.0274</td>
<td>0.2640</td>
</tr>
<tr>
<td>[2,] -0.8</td>
<td>0.0715</td>
<td>-0.0166</td>
<td>0.1597</td>
</tr>
<tr>
<td>[3,] -0.7</td>
<td>0.0489</td>
<td>-0.0115</td>
<td>0.1093</td>
</tr>
</tbody>
</table>

... output truncated

> plot(s.cont)
```

Kosuke Imai (Princeton)  
Causal Mediation Analysis  
June 18, 2009  
19 / 22
Concluding Remarks

- Quantitative analysis can be used to identify causal mechanisms!
- Wide applications in many social scientific disciplines
- Sensitivity analysis is critical
- Development of easy-to-use software mediation
- Object-oriented nature of R facilitated this development
- Future extensions: multiple mediators, sensitivity analysis for other models

Papers and Software

- All are available at http://imai.princeton.edu/