Identification and Inference in Causal Mediation Analysis

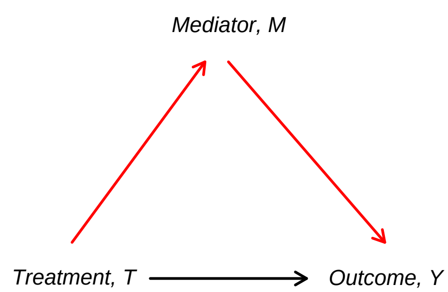
Kosuke Imai **Luke Keele** **Teppei Yamamoto**

Princeton University Ohio State University

November 12, 2008

Causal Mediation Analysis

- Investigation of causal mechanisms via intermediate variables
- *How* does the treatment alter the outcome?
- Direct and indirect effects



- Popular among epidemiologists, psychologists, political scientists
- Fast growing methodological literature

Overview

1. Identification under sequential ignorability
   - Nonparametric identification without an additional assumption
   - Parametric identification under the linear structural equation model
2. Estimation and inference under sequential ignorability
   - Parametric estimation
   - Nonparametric estimator and its asymptotic variance
3. Sensitivity analysis for the sequential ignorability assumption
   - Nonparametric sensitivity analysis
   - Parametric sensitivity analysis
4. Empirical illustration
   - A randomized experiment from political psychology
   - The treatment is randomized but the mediator is not

Definition of Causal Mediation Effects

- 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 $Y_i = Y_i(T_i, M_i(T_i))$
- Total causal effect: $\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0))$
- Causal mediation effects: $\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0))$
- Natural (pure) direct effects: $\zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t))$
- The relationship: $\tau_i = \delta_i(t) + \zeta_i(1 - t)$

Interpretation of Causal Mediation Effects

- $\delta_i(t)$ is the indirect causal effect of the treatment on the outcome through the mediator under treatment status t
- *Controlled* indirect effects, $Y_i(1, m) - Y_i(0, m)$, for the mediator that can be manipulated and/or randomized
- Observational studies and experiments with non-random M
- descriptive vs. prescriptive effects
- $Y_i(t, M_i(t))$ is observable but $Y_i(t, M_i(1-t))$ is not
- $\delta_i(t) = 0$ if $M_i(1) = M_i(0)$
- Quantity of interest:

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

Sequential Ignorability

Assumption 1 (Sequential Ignorability)

$$\{Y_i(t, m), M_i(t)\} \perp\!\!\!\perp T_i \mid X_i,$$

$$Y_i(t, m) \perp\!\!\!\perp M_i \mid T_i, X_i$$

for $t = 0, 1$ and all $m \in \mathcal{M}$

- The second equation can be rewritten as,

$$Y_i(t, m) \perp\!\!\!\perp M_i(t^*) \mid T_i = t^*, X_i$$

Nonparametric Identification

Theorem 1 (Nonparametric Identification)

Under Assumption 1, for $t = 0, 1$,

$$\bar{\delta}(t) = (-1)^t \int \left\{ \int \mathbb{E}(Y_i \mid M_i, T_i = t, X_i)\, dP(M_i \mid T_i = 1 - t, X_i) - \mathbb{E}(Y_i \mid T_i = t, X_i) \right\} dP(X_i)$$

Proof (Discrete Mediator with No Observed Covariates):

$$\begin{aligned}
& \bar{\zeta}(t^*) \\
= & \sum_{t=0}^{1} \sum_{m=0}^{J-1} \mathbb{E}(Y_i(1,m) - Y_i(0,m) \mid M_i(t^*) = m, T_i = t) \Pr(M_i(t^*) = m, T_i = t) \\
= & \sum_{m=0}^{J-1} \Big\{ \mathbb{E}(Y_i(1,m) - Y_i(0,m) \mid T_i = t^*) \Pr(M_i(t^*) = m \mid T_i = t^*) \Pr(T_i = t^*) \\
& + \mathbb{E}(Y_i(1,m) - Y_i(0,m) \mid M_i(t^*) = m, T_i = 1 - t^*) \Pr(M_i(t^*) = m, T_i = 1 - t^*) \Big\} \\
= & \sum_{m=0}^{J-1} \mathbb{E}(Y_i(1,m) - Y_i(0,m)) \Pr(M_i = m \mid T_i = t^*) \Pr(T_i = t^*) \\
& + \mathbb{E}(Y_i(1, M_i(t^*)) - Y_i(0, M_i(t^*)) \mid T_i = 1 - t^*) \Pr(T_i = 1 - t^*) \\
= & \sum_{m=0}^{J-1} \Big\{ \mathbb{E}(Y_i \mid M_i = m, T_i = 1) - \mathbb{E}(Y_i \mid M_i = m, T_i = 0) \Big\} \Pr(M_i = m \mid T_i = t^*) \\
& \times \Pr(T_i = t^*) + \bar{\zeta}(t^*) \Pr(T_i = 1 - t^*).
\end{aligned}$$

Thus, we have $\bar{\zeta}(t^*) = \sum_{m=0}^{J-1} \{\mathbb{E}(Y_i \mid M_i = m, T_i = 1) - \mathbb{E}(Y_i \mid M_i = m, T_i = 0)\} \Pr(M_i = m \mid T_i = t^*)$. □

Comparison with the Existing Identification Results

- The literature insists that an additional assumption is required
- Pearl's assumption for the identification of $\bar{\delta}(t^*)$:

$$Y_i(t,m) \perp\!\!\!\perp M_i(t^*) \mid X_i,$$

  in place of $Y_i(t,m) \perp\!\!\!\perp M_i \mid T_i, X_i$
- Robins' no-interaction assumption about *controlled* direct effects:

$$Y_i(1,m) - Y_i(0,m) = B_i$$

  where B_i is a random variable that does not depend on m

- Sequential ignorability *alone* is sufficient

Linear Structural Equation Model (LSEM)

- The Model:

$$\begin{aligned} Y_i &= \alpha_1 + \beta_1 T_i + \epsilon_{1i}, \\ M_i &= \alpha_2 + \beta_2 T_i + \epsilon_{2i}, \\ Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i}, \end{aligned}$$

  where $\mathbb{E}(\epsilon_{1i} \mid T_i) = \mathbb{E}(\epsilon_{2i} \mid T_i) = \mathbb{E}(\epsilon_{3i} \mid M_i, T_i) = 0$.

- Baron and Kenny (1986):
  1. the association between Y_i and T_i exists
  2. the association between M_i and T_i exists
  3. the conditional association between Y_i and M_i given T_i exists
  4. $\beta_2\gamma$ as the causal mediation effect

- One equation is redundant:

$$Y_i = (\alpha_3 + \alpha_2\gamma) + (\beta_3 + \beta_2\gamma)T_i + (\gamma\epsilon_{2i} + \epsilon_{3i})$$

  where $\gamma\mathbb{E}(\epsilon_{2i} \mid T_i) + \mathbb{E}\{\mathbb{E}(\epsilon_{3i} \mid M_i, T_i) \mid T_i\} = 0$.

Parametric Identification under Sequential Ignorability

Theorem 2 (Identification under LSEM)

Consider the following linear structural equation model

$$\begin{aligned} M_i &= \alpha_2 + \beta_2 T_i + \epsilon_{2i}, \\ Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i}. \end{aligned}$$

Under Assumption 1, the average causal mediation effects are identified as $\bar{\delta}(0) = \bar{\delta}(1) = \beta_2\gamma$.

- Assumption 1 implies $\epsilon_{2i} \perp\!\!\!\perp \epsilon_{3i}$ as well as $\epsilon_{2i} \perp\!\!\!\perp T_i$, $\epsilon_{3i} \perp\!\!\!\perp T_i$, and $\epsilon_{3i} \perp\!\!\!\perp M_i \mid T_i$.
- Contrary to the literature, sequential ignorability *alone* is sufficient
- β_3 is the average natural direct effect

Identification without the No-interaction Assumption

Assumption 2 (No-interaction)

$$\bar{\delta}(0) = \bar{\delta}(1)$$

- Assumption 2 is unnecessary
- The LSEM with an interaction term:

$$\begin{aligned} M_i &= \alpha_2 + \beta_2 T_i + \epsilon_{2i}, \\ Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \kappa T_i M_i + \epsilon_{3i}. \end{aligned}$$

- Under Assumption 1, $\bar{\delta}(t) = \beta_2(\gamma + t\kappa)$ for $t = 0, 1$.

Parametric Estimation and Inference

- Under sequential ignorability, equation-by-equation least squares
- Asymptotic variance via the Delta method:
  1. No-interaction:

$$\text{Var}(\hat{\delta}(t)) \approx \beta_2^2 \text{Var}(\hat{\gamma}) + \gamma^2 \text{Var}(\hat{\beta}_2)$$

  2. With-interaction:

$$\text{Var}(\hat{\delta}(t)) \approx (\gamma + t\kappa)^2 \text{Var}(\hat{\beta}) + \beta_2^2 \{\text{Var}(\hat{\gamma}) + t\text{Var}(\hat{\kappa}) + 2t\text{Cov}(\hat{\gamma}, \hat{\kappa})\}$$

Nonparametric Estimation and Inference (Discrete M)

- A simple nonparametric estimator $\hat{\delta}(t)$:

$$(-1)^t \left(\sum_{m=0}^{J-1} \frac{\sum_{i=1}^n \mathbf{1}\{T_i = 1-t, M_i = m\} \sum_{i=1}^n Y_i \mathbf{1}\{T_i = t, M_i = m\}}{n_{1-t} \sum_{i=1}^n \mathbf{1}\{T_i = t, M_i = m\}} - \frac{1}{n_t} \sum_{i=1}^n \mathbf{1}\{T_i = t\} Y_i \right)$$

  where $n_t = \sum_{i=1}^n \mathbf{1}\{T_i = t\}$.
- Estimate within each strata defined by X, and then aggregate

Theorem 3 (Asymptotic Variance)

Under Assumption 1, the asymptotic variance of the nonparametric estimator is

$$\begin{aligned}\mathrm{Var}(\hat{\delta}(t)) \approx\ & \frac{1}{n_t}\sum_{m=0}^{J-1}\lambda_{1-t,m}\left\{\left(\frac{\lambda_{1-t,m}}{\lambda_{tm}}-2\right)\mathrm{Var}(Y_i \mid M_i=m, T_i=t)\right.\\ & \left.+\frac{n_t(1-\lambda_{1-t,m})\mu_{tm}^2}{n_{1-t}}\right\}+\frac{1}{n_t}\mathrm{Var}(Y_i \mid T_i=t)\\ & -\frac{2}{n_{1-t}}\sum_{m'=m+1}^{J-1}\sum_{m=0}^{J-2}\lambda_{1-t,m}\lambda_{1-t,m'}\mu_{tm}\mu_{tm'},\end{aligned}$$

where $\lambda_{tm} \equiv \Pr(M_i = m \mid T_i = t)$ *and* $\mu_{tm} \equiv \mathbb{E}(Y_i \mid M_i = m, T_i = t)$.

Using Nonparametric Regressions

- Fit two nonparametric regressions:
  1. $\mu_{tm}(x) \equiv \mathbb{E}(Y_i \mid T_i = t, M_i = m, X_i = x)$
  2. $\lambda_{tm}(x) \equiv \Pr(M_i = m \mid T_i = t, X_i = x)$
- An estimator:

$$(-1)^t\left\{\sum_{m=0}^{J-1}\frac{\sum_{i=1}^n \mathbf{1}\{T_i=1-t\}\hat{\lambda}_{1-t,m}(X_i)\sum_{i=1}^n \mathbf{1}\{T_i=t\}\hat{\mu}_{tm}(X_i)\hat{\lambda}_{tm}(X_i)}{n_{1-t}\sum_{i=1}^n \mathbf{1}\{T_i=t\}\hat{\lambda}_{tm}(X_i)} - \frac{1}{n_t}\sum_{i=1}^n \mathbf{1}\{T_i=t\}\left(\sum_{m=0}^{J-1}\hat{\mu}_{tm}(X_i)\hat{\lambda}_{tm}(X_i)\right)\right\}.$$

- Nonparametric or parametric bootstrap for uncertainty estimates

A Simulation Study

- Binary mediator, lognormal outcome
- $Y_i(t, m) \perp\!\!\!\perp M_i(t') \mid T_i = t'$ but $Y_i(t, m) \not\perp\!\!\!\perp M_i(t') \mid T_i = 1 - t'$
- True values: $\bar{\delta}(0) \approx 0.67$ and $\bar{\delta}(1) \approx 3.95$

Estimator	n	Bias	RMSE	90% CI	95% CI
$\hat{\delta}(0)$	50	0.013	1.05	0.77	0.83
	100	0.014	0.69	0.83	0.87
	250	0.014	0.42	0.86	0.91
	500	0.013	0.29	0.88	0.93
	1000	0.013	0.20	0.89	0.94
	2000	0.016	0.14	0.90	0.95
$\hat{\delta}(1)$	50	0.088	2.07	0.85	0.89
	100	0.080	1.46	0.87	0.92
	250	0.071	0.92	0.89	0.94
	500	0.080	0.65	0.90	0.95
	1000	0.079	0.46	0.90	0.95
	2000	0.094	0.34	0.90	0.95

Need for Sensitivity Analysis

- The sequential ignorability assumption is often too strong!
- Need to assess the robustness of findings via sensitivity analysis

Assumption 3 (Ignorability of Treatment Assignment)

$$\{Y_i(t, m), M_i(t)\} \perp\!\!\!\perp T_i \mid X_i$$

- Parametric and nonparametric sensitivity analysis under Assumption 3 alone
- Maximal degree of departure from Assumption 1 while maintaining the original conclusion

Parametric Sensitivity Analysis

- Assumption 3 implies $\epsilon_{2i} \perp\!\!\!\perp T_i$ and $\epsilon_{3i} \perp\!\!\!\perp T_i$ but *not* $\epsilon_{2i} \perp\!\!\!\perp \epsilon_{3i}$
- Sensitivity parameter: $\rho \equiv \mathrm{Corr}(\epsilon_{2i}, \epsilon_{3i})$

Theorem 4 (Identification with a Known Error Correlation)

Under Assumption 3,

$$\bar{\delta}(0) = \bar{\delta}(1) = \beta_2 \left(\frac{\sigma_{23}^*}{\sigma_2^2} - \frac{\rho}{\sigma_2} \sqrt{\frac{1}{1-\rho^2} \left({\sigma_3^*}^2 - \frac{{\sigma_{23}^*}^2}{\sigma_2^2} \right)} \right),$$

where $\sigma_j^2 \equiv \mathrm{Var}(\epsilon_{ji})$ *for* $j = 2, 3$, ${\sigma_3^*}^2 \equiv \mathrm{Var}(\epsilon_{3i}^*)$, $\sigma_{23}^* \equiv \mathrm{Cov}(\epsilon_{2i}, \epsilon_{3i}^*)$, *and* $\epsilon_{3i}^* = \gamma \epsilon_{2i} + \epsilon_{3i}$.

- Fit the following LSEM via eq.-by-eq. least squares or SUR

$$\begin{aligned} M_i &= \alpha_2 + \beta_2 T_i + \epsilon_{2i} \\ Y_i &= \alpha_3^* + \beta_3^* T_i + \epsilon_{3i}^* \end{aligned}$$

- Monotone function of ρ

$$\frac{\partial}{\partial \rho} \bar{\delta}(t) = -\frac{\beta_2}{\sigma_2(1-\rho^2)} \sqrt{\frac{1}{1-\rho^2} \left({\sigma_3^*}^2 - \frac{{\sigma_{23}^*}^2}{\sigma_2^2} \right)}$$

- $\bar{\delta}(t) = 0$ if and only if $\rho = \mathrm{Corr}(\epsilon_{2i}, \epsilon_{3i}^*)$ (easy to compute!)
- For confidence intervals, apply the iterative FGLS algorithm to

$$\begin{aligned} M_i &= \alpha_2 + \beta_2 T_i + \epsilon_{2i} \\ Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i} \end{aligned}$$

Large Sample Nonparametric Bounds

- Balke and Pearl (1997)'s strategy: discrete outcome and mediator
- Binary case: population probabilities of 64 types

$$\pi^{m_1 m_0}_{y_{11} y_{10} y_{01} y_{00}} \equiv \Pr(Y_i(1,1) = y_{11}, Y_i(1,0) = y_{10}, Y_i(0,1) = y_{01}, Y_i(0,0) = y_{00}, M_i(1) = m_1, M_i(0) = m_0)$$

- Mediation effects as a linear function of π

$$\bar{\delta}(t) = \sum_{m=0}^{1} \sum_{y_{1-t,m}=0}^{1} \sum_{y_{1,1-m}=0}^{1} \sum_{y_{0,1-m}=0}^{1} \left(\sum_{m_0=0}^{1} \pi^{m m_0}_{y_{11} y_{10} y_{01} y_{00}} - \sum_{m_1=0}^{1} \pi^{m_1 m}_{y_{11} y_{10} y_{01} y_{00}} \right)$$

- Assumption 3 implies linear restrictions

$$\Pr(Y_i = y, M_i = m \mid T_i = t) = \sum_{y_{1-t,m}=0}^{1} \sum_{y_{t,1-m}=0}^{1} \sum_{y_{1-t,1-m}=0}^{1} \sum_{m_{1-t}=0}^{1} \pi^{m_1 m_0}_{y_{11} y_{10} y_{01} y_{00}},$$

where $m_t = m$ and $y_{tm} = y$.

- Symbolic linear programming

Theorem 5 (Sharp Large Sample Bounds)

Under Assumption 3 the sharp large sample bounds of the average causal mediation effects are given by,

$$\max \left\{ \begin{array}{c} -\Pr(Y_i = 1-t \mid T_i = t) \\ -\Pr(M_i = 1-t \mid T_i = 1-t) - \Pr(Y_i = M_i = 1-t \mid T_i = t) \\ -\Pr(M_i = t \mid T_i = 1-t) - \Pr(Y_i = 1-t, M_i = t \mid T_i = t) \end{array} \right\} \leq$$

$$\bar{\delta}(t) \leq \min \left\{ \begin{array}{c} \Pr(Y_i = t \mid T_i = t) \\ \Pr(M_i = 1-t \mid T_i = 1-t) + \Pr(Y_i = t, M_i = 1-t \mid T_i = t) \\ \Pr(M_i = t \mid T_i = 1-t) + \Pr(Y_i = M_i = t \mid T_i = t) \end{array} \right\},$$

for $t = 0, 1$.

- $[\alpha, \beta]$ always improves upon $[-1, 1]$; $\beta - \alpha \leq 1$
- Not very informative $-1 \leq \alpha \leq 0 \leq \beta \leq 1$
- Possible to impose the no-interaction assumption $\bar{\delta}(1) = \bar{\delta}(0)$

Nonparametric Sensitivity Analysis

- Bounds are not informative even under additional assumptions
- Ignorability of the mediator implies

$$\begin{aligned} &\Pr(Y_i(1,1)=y_{11}, Y_i(1,0)=y_{10}, Y_i(0,1)=y_{01}, Y_i(0,0)=y_{00} \mid M_i=1, T_i=t') \\ =&\Pr(Y_i(1,1)=y_{11}, Y_i(1,0)=y_{10}, Y_i(0,1)=y_{01}, Y_i(0,0)=y_{00} \mid M_i=0, T_i=t') \end{aligned}$$

- Sensitivity parameter:

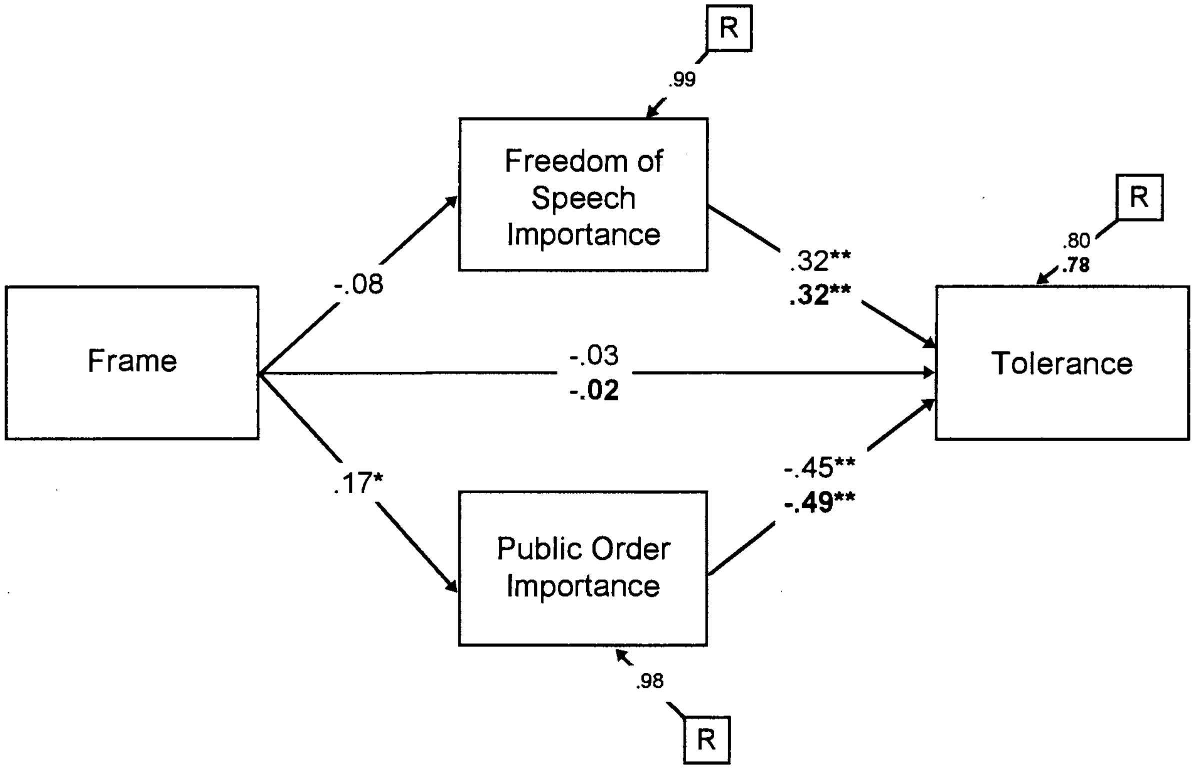
$$\begin{aligned} \left| \frac{\sum_{m_0=0}^{1} \pi^{1m_0}_{y_{11}y_{10}y_{01}y_{00}}}{\Pr(M_i=1 \mid T_i=1)} - \frac{\sum_{m_0=0}^{1} \pi^{0m_0}_{y_{11}y_{10}y_{01}y_{00}}}{\Pr(M_i=0 \mid T_i=1)} \right| &\leq \rho, \\ \left| \frac{\sum_{m_1=0}^{1} \pi^{m_1 1}_{y_{11}y_{10}y_{01}y_{00}}}{\Pr(M_i=1 \mid T_i=0)} - \frac{\sum_{m_1=0}^{1} \pi^{m_1 0}_{y_{11}y_{10}y_{01}y_{00}}}{\Pr(M_i=0 \mid T_i=0)} \right| &\leq \rho, \end{aligned}$$

  where $0 \leq \rho \leq 1$
- Compute the sharp bounds for various values of ρ

Political Psychology Experiment: Nelson *et al.* (*APSR*)

- How does media framing affect citizens' political opinions?
- News stories about the Ku Klux Klan rally in Ohio
- Free speech frame ($T_i = 0$) and public order frame ($T_i = 1$)
- Randomized experiment with the sample size $= 136$

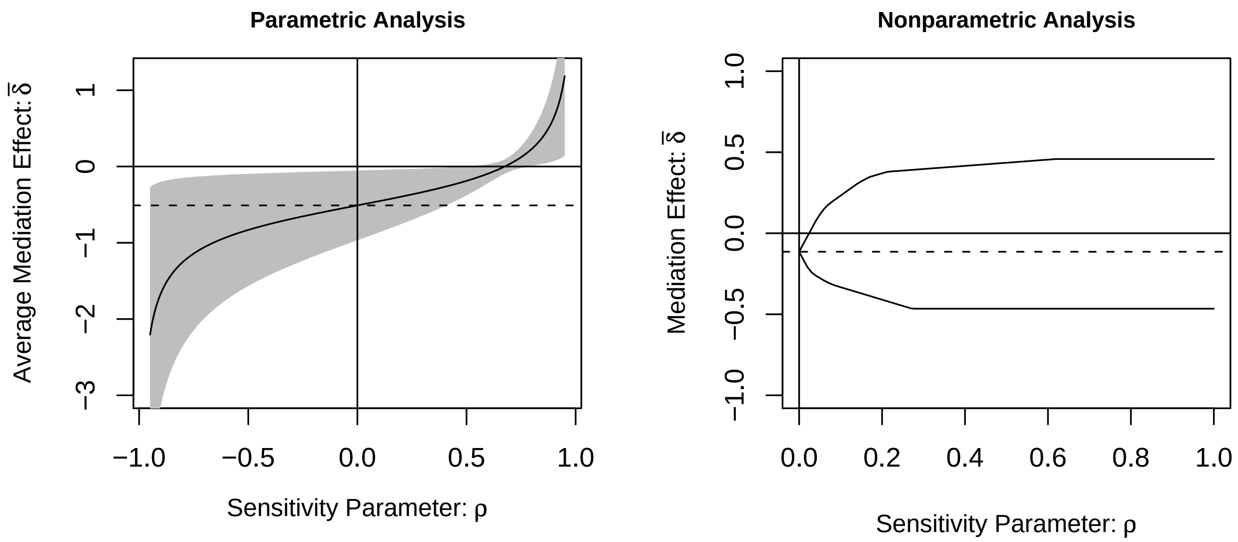
- Mediators: general attitudes (12 point scale) about the importance of free speech and public order
- Outcome: tolerance (7 point scale) for the Klan rally
- Expected findings: negative mediation effects



Analysis under Sequential Ignorability

	Mediator	
Estimator	Public Order	Free Speech
Parametric		
No-interaction	-0.510	-0.126
	$[-0.969, -0.051]$	$[-0.388, 0.135]$
$\hat{\delta}(0)$	-0.451	-0.131
	$[-0.871, -0.031]$	$[-0.404, 0.143]$
$\hat{\delta}(1)$	-0.566	-0.122
	$[-1.081, -0.050]$	$[-0.380, 0.136]$
Nonparametric		
$\hat{\delta}(0)$	-0.374	-0.094
	$[-0.823, 0.074]$	$[-0.434, 0.246]$
$\hat{\delta}(1)$	-0.596	-0.222
	$[-1.168, -0.024]$	$[-0.662, 0.219]$

Sensitivity Analysis



Concluding Remarks and Future Work

- Nonparametric identification under sequential ignorability
- Parametric identification under LSEM
- Nonparametric estimator and its asymptotic variance
- Nonparametric and parametric sensitivity analysis

- Nonparametric sensitivity analysis in a more general setting
- Nonparametric estimation under the no-interaction assumption
- Use of parametric/nonparametric regressions in practical causal mediation analysis
- Extension to multiple mediators