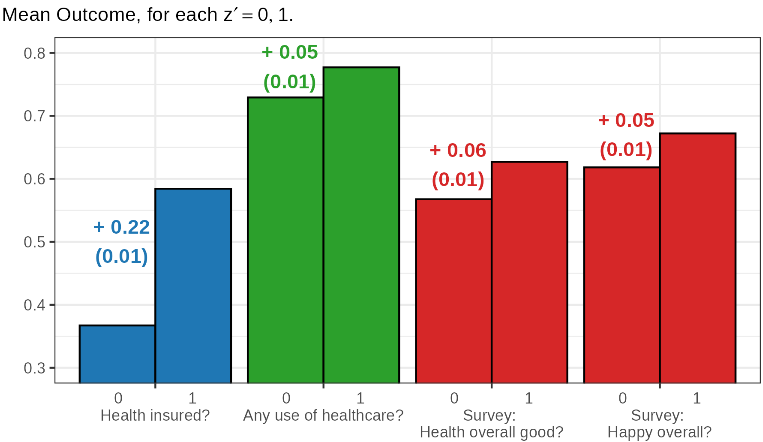
Causal Mediation in Natural Experiments

Senan Hogan-Hennessy
Economics Department, Cornell University
seh325@cornell.edu

Econometric Society World Congress, Seoul
22 August 2025

Intro: Oregon Health Insurance Experiment

In 2008, Oregon gave health insurance by lottery (Finkelstein et al, 2012).

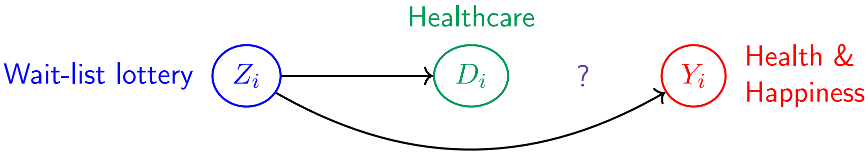


$\Rightarrow$ Suggestive evidence for healthcare as mechanism for wait-list lottery, access to health insurance. . . .

Intro: Oregon Health Insurance Experiment

Oregon gave health insurance by wait-list lottery (Finkelstein et al, 2012).

Figure: SCM for Suggestive Evidence of a Mechanism.



- Missing the $D_i \to Y_i$ edge of the triangular system…
- Is $D_i \to Y_i$ small, large, or even existent?
- Where else do we accept assumed causal effects without evidence?

Introduction

Causal Mediation (CM) is an alternative framework, which actually defines what is estimated, and assumptions under which they are identified.

This paper examines CM from an economic perspective:

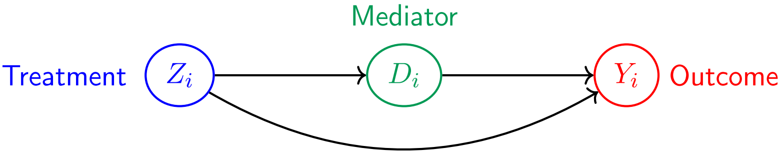
1. Problems with conventional approach to CM in observational settings. **[Negative result]**
2. Recovering valid CM effects, via MTE + control function modelling. **[Positive result]**

Brings together ideas from two different literatures:

- **CM.**
  Imai Keele Yamamoto (2010), Frölich Huber (2017), Deuchert Huber Schelker (2019), Huber (2020), Kwon Roth (2024).
- **Labour theory, Selection-into-treatment, MTEs.**
  Roy (1951), Heckman (1979), Heckman Honoré (1990), Vycatil (2002), Heckman Vycatil (2005), Kline Walters (2019).

1. CM — Model

Consider binary treatment $Z_i = 0, 1$, binary mediator $D_i = 0, 1$, and continuous outcome Y_i for individuals $i = 1, \ldots, n$.



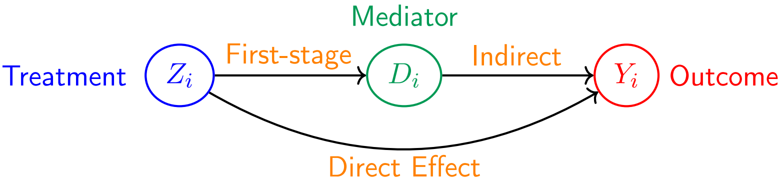
Assume Z_i is ignorable, $Z_i \perp\!\!\!\perp D_i(z'), Y_i(z, d') \mid \boldsymbol{X}_i, \quad$ for $z', z, d' = 0, 1$.

Mediator D_i is a function of Z_i. Outcome Y_i is a function of both Z_i, D_i.

$$D_i = \begin{cases} D_i(0), & \text{if } Z_i = 0 \\ D_i(1), & \text{if } Z_i = 1. \end{cases} \qquad Y_i = \begin{cases} Y_i(0, D_i(0)), & \text{if } Z_i = 0 \\ Y_i(1, D_i(1)), & \text{if } Z_i = 1. \end{cases}$$

1. CM — Model

Consider binary treatment $Z_i = 0, 1$, binary mediator $D_i = 0, 1$, and continuous outcome Y_i for individuals $i = 1, \ldots, n$.



Assume Z_i is ignorable (conditional on $\boldsymbol{X}_i$).

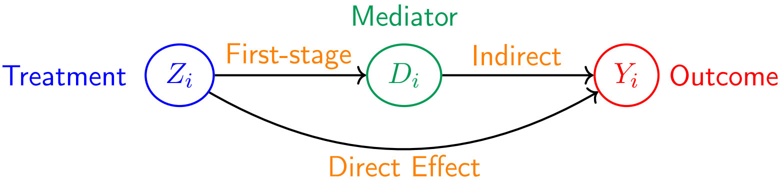
Only two causal effects are identified so far.

ATE: $\mathbb{E}\left[Y_i(1, D_i(1)) - Y_i(0, D_i(0))\right] = \mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right]$

Average first-stage: $\mathbb{E}\left[D_i(1) - D_i(0)\right] = \mathbb{E}\left[D_i \mid Z_i = 1\right] - \mathbb{E}\left[D_i \mid Z_i = 0\right]$

1. CM — Model

Consider binary treatment $Z_i = 0, 1$, binary mediator $D_i = 0, 1$, and continuous outcome Y_i for individuals $i = 1, \ldots, n$.



Average Direct Effect (ADE) : $\mathbb{E}\left[Y_i\left(\boxed{1}, D_i(Z_i)\right) - Y_i\left(\boxed{0}, D_i(Z_i)\right)\right]$

- ADE is causal effect $Z \to Y$, blocking the indirect D_i path.

Average Indirect Effect (AIE): $\mathbb{E}\left[Y_i\left(Z_i, \boxed{D_i(1)}\right) - Y_i\left(Z_i, \boxed{D_i(0)}\right)\right]$

- AIE is causal effect of $D(Z) \to Y$, blocking the direct Z_i path.

1. CM — Identification

Mediator Ignorability (**MI**, Imai Keele Yamamoto 2010):

Assume mediator D_i is *also* ignorable, conditional on $\boldsymbol{X}_i$ and Z_i realisation

$$D_i \perp\!\!\!\perp Y_i(z', d') \mid \boldsymbol{X}_i, Z_i = z', \text{ for } z', d' = 0, 1.$$

If **MI** holds then ADE and AIE are identified by two-stage regression:

$$\mathbb{E}_{D_i, \boldsymbol{X}_i}\left[\underbrace{\mathbb{E}\left[Y_i \mid Z_i = 1, D_i, \boldsymbol{X}_i\right] - \mathbb{E}\left[Y_i \mid Z_i = 0, D_i, \boldsymbol{X}_i\right]}_{\text{Second-stage regression, } Y_i \text{ on } Z_i \text{ holding } D_i, \boldsymbol{X}_i \text{ constant}}\right] = \text{ADE}$$

$$\mathbb{E}_{Z_i, \boldsymbol{X}_i}\left[\begin{array}{l}\underbrace{\left(\mathbb{E}\left[D_i \mid Z_i = 1, \boldsymbol{X}_i\right] - \mathbb{E}\left[D_i \mid Z_i = 0, \boldsymbol{X}_i\right]\right)}_{\text{First-stage regression, } D_i \text{ on } Z_i} \\ \times \underbrace{\left(\mathbb{E}\left[Y_i \mid Z_i, D_i = 1, \boldsymbol{X}_i\right] - \mathbb{E}\left[Y_i \mid Z_i, D_i = 0, \boldsymbol{X}_i\right]\right)}_{\text{Second-stage regression, } Y_i \text{ on } D_i \text{ holding } Z_i, \boldsymbol{X}_i \text{ constant}}\end{array}\right] = \text{AIE}$$

2. Selection Bias

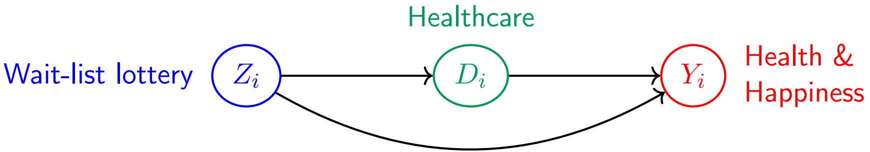
Mediator ignorability (**MI**, Imai Keele Yamamoto 2010):

Assume mediator D_i is ***also* ignorable**, conditional on $\boldsymbol{X}_i$, Z_i realisation

$$D_i \perp\!\!\!\perp Y_i(z', d') \mid \boldsymbol{X}_i, Z_i = z', \text{ for } z', d' = 0, 1.$$

Would this assumption hold true in settings economists study?

E.g., Oregon Health Insurance Experiment.



1. Treatment is as-good-as random (2008 Oregon wait-list lottery).
2. Healthcare is quasi-random, conditional on lottery realisation Z_i and demographic controls $\boldsymbol{X}_i$ (no natural experiment...).

2. Selection Bias

Assume: Healthcare is quasi-random, conditional on lottery realisation Z_i and demographic controls $\boldsymbol{X}_i$ (no natural experiment...).

Consider the case **individuals visit the doctor** to maximise health.

$$D_i\left(z'\right) = \mathbb{1}\left\{\underbrace{C_i}_{\text{Costs}} \leq \underbrace{Y_i\left(z', 1\right) - Y_i\left(z', 0\right)}_{\text{Benefits}}\right\}, \quad \text{for } z' = 0, 1$$

i.e., Roy (1951) selection-into-D_i.

Theorem: If selection is Roy-style, and benefits are not 100% explained by $Z_i, \boldsymbol{X}_i$, then **MI** does not hold.

Proof sketch: suppose D_i is ignorable $\implies$ selection-into-D_i is explained 100% by $\{C_i, Z_i, \boldsymbol{X}_i\}$, while unobserved benefits explain 0%.

2. Selection Bias

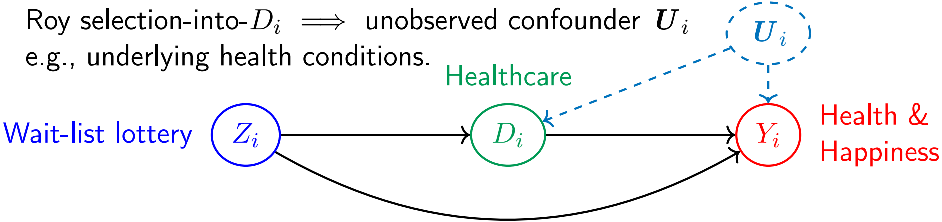
Assume: Healthcare is quasi-random, conditional on lottery realisation Z_i and demographic controls $\boldsymbol{X}_i$ (no natural experiment...).

Consider the case **individuals visit the doctor** to maximise health.

$$D_i\left(z'\right) = \mathbb{1}\left\{ \underbrace{C_i}_{\text{Costs}} \leq \underbrace{Y_i\left(z', 1\right) - Y_i\left(z', 0\right)}_{\text{Benefits}} \right\}, \quad \text{for } z' = 0, 1.$$

i.e., Roy (1951) selection-into-D_i.

Roy selection-into-$D_i \implies$ unobserved confounder $\boldsymbol{U}_i$
e.g., underlying health conditions.



2. Selection Bias

In observational setting, must have an additional credible research design for **Mediator Ignorability** to believe this assumption.

(a) Cells in a lab → **MI** believable. (b) People choosing healthcare → **MI** not.

2. Selection Bias

- What happens if you go ahead and estimate CM anyway?
- Would this be problematic?
- Estimating causal effects with an unobserved confounder is usually bad. . . .

Definition: Selection bias (Heckman Ichimura Smith Todd, 1998).

Estimating $D_i \to Y_i$, if D_i not ignorable:

$$\begin{aligned}\mathbb{E}\left[Y_i \mid D_i = 1\right] - \mathbb{E}\left[Y_i \mid D_i = 0\right] = \text{ATE} \\ + \underbrace{\Big(\mathbb{E}\left[Y_i(.,0) \mid D_i = 1\right] - \mathbb{E}\left[Y_i(.,0) \mid D_i = 0\right] \Big)}_{\text{Selection Bias}} \\ + \underbrace{\Pr\left(D_i = 0\right)\left(\text{ATT} - \text{ATU}\right)}_{\text{Group difference bias}}.\end{aligned}$$

2. Selection Bias — Direct Effect

CM Effects have this same flavour, causal effects + contaminating bias.

$$\text{CM Estimand} = \text{ADE} + \Big(\text{Selection Bias} + \text{Group difference bias}\Big)$$ ▸ Model

$$\underbrace{\mathbb{E}_{D_i=d'}\Big[\mathbb{E}\left[Y_i \,\middle|\, Z_i=1, D_i=d'\right] - \mathbb{E}\left[Y_i \,\middle|\, Z_i=0, D_i=d'\right]\Big]}_{\text{Estimand, Direct Effect}}$$

$$= \underbrace{\mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i))\right]}_{\text{Average Direct Effect}}$$

$$+ \underbrace{\mathbb{E}_{D_i=d'}\Big[\mathbb{E}\left[Y_i(0, D_i(Z_i)) \,\middle|\, D_i(1)=d'\right] - \mathbb{E}\left[Y_i(0, D_i(Z_i)) \,\middle|\, D_i(0)=d'\right]\Big]}_{\text{Selection Bias}}$$

$$+ \underbrace{\mathbb{E}_{D_i=d'}\left[\begin{array}{l}\Big(1 - \Pr\left(D_i(1)=d'\right)\Big) \\ \times \left(\begin{array}{l}\mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) \,\middle|\, D_i(1)=1-d'\right] \\ - \mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) \,\middle|\, D_i(0)=d'\right]\end{array}\right)\end{array}\right]}_{\text{Group difference bias}}$$ ▸ Group-diff

2. Selection Bias — Direct Effect

CM Effects have this same flavour, causal effects + contaminating bias.

CM Estimand = ADE + (Selection Bias + Group difference bias) ▸ Model

$$\underbrace{\mathbb{E}_{D_i=d'}\left[\mathbb{E}\left[Y_i \,\middle|\, Z_i=1, D_i=d'\right] - \mathbb{E}\left[Y_i \,\middle|\, Z_i=0, D_i=d'\right]\right]}_{\text{Estimand, Direct Effect}}$$

$$= \underbrace{\mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i))\right]}_{\text{Average Direct Effect}}$$

$$+ \underbrace{\mathbb{E}_{D_i=d'}\left[\mathbb{E}\left[Y_i(0, D_i(Z_i)) \,\middle|\, D_i(1)=d'\right] - \mathbb{E}\left[Y_i(0, D_i(Z_i)) \,\middle|\, D_i(0)=d'\right]\right]}_{\text{Selection Bias}}$$

$$+ \underbrace{\mathbb{E}_{D_i=d'}\left[\begin{array}{l}\left(1 - \Pr\left(D_i(1)=d'\right)\right) \\ \times \left(\begin{array}{l}\mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) \,\middle|\, D_i(1)=1-d'\right] \\ - \mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) \,\middle|\, D_i(0)=d'\right]\end{array}\right)\end{array}\right]}_{\text{Group difference bias}}$$

▸ Group-diff

2. Selection Bias — Indirect Effect

CM Effects have this same flavour, causal effects + contaminating bias.[1]

$$\text{CM Estimand} = \text{AIE} + \Big(\text{Selection Bias} + \text{Group difference bias} \Big)$$ ▸ Model

---

$$\underbrace{\mathbb{E}_{Z_i}\left[\Big(\mathbb{E}\left[D_i \mid Z_i = 1 \right] - \mathbb{E}\left[D_i \mid Z_i = 0 \right] \Big) \times \Big(\mathbb{E}\left[Y_i \mid Z_i, D_i = 1 \right] - \mathbb{E}\left[Y_i \mid Z_i, D_i = 0 \right] \Big) \right]}_{\text{Estimand, Indirect Effect}}$$

$$= \underbrace{\mathbb{E}\left[Y_i(Z_i, D_i(1)) - Y_i(Z_i, D_i(0)) \right]}_{\text{Average Indirect Effect}}$$

$$+ \underbrace{\overline{\pi} \Big(\mathbb{E}\left[Y_i(Z_i, 0) \mid D_i = 1 \right] - \mathbb{E}\left[Y_i(Z_i, 0) \mid D_i = 0 \right] \Big)}_{\text{Selection Bias}}$$

$$+ \underbrace{\overline{\pi} \left[\Big(1 - \Pr\left(D_i = 1 \right) \Big) \begin{pmatrix} \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i = 1 \right] \\ - \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i = 0 \right] \end{pmatrix} + \left(\frac{1 - \Pr\left(D_i(1) = 1, D_i(0) = 0 \right)}{\Pr\left(D_i(1) = 1, D_i(0) = 0 \right)} \right) \begin{pmatrix} \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i(Z_i) \neq Z_i \right] \\ - \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \right] \end{pmatrix} \right]}_{\text{Groups difference Bias}}$$

▸ Group-diff

2. Selection Bias

$\Longrightarrow$ Unless mediator D_i is also randomly assigned, then controlling for it does not lead to interpretable causal effects.

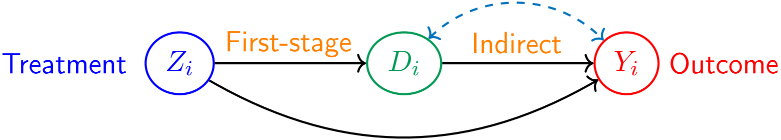
---

1. Counter to accepted intuitive reasoning in applied economics, which oftens controls for plausible mediators
2. Invalidates conclusions in fields that uncritically apply CM methods with no case for mediator ignorability (common in some fields of epidemiology, psychology, medicine, and sociology)
3. Warning sign for economists to avoid picking up this practice, and stop using it as a "robustness check."

Applied economics would do better by thinking deeper on mediation.

3. Saving CM Effects

Suppose Z_i is ignorable, D_i is not, so we have the following causal model.



Write POs as sum of PO means and mean-zero errors, $U_{d',i}$.

$$Y_i(Z_i, 0) = \mu_0\big(Z_i; \boldsymbol{X}_i\big) + U_{0,i}, \quad Y_i(Z_i, 1) = \mu_1\big(Z_i; \boldsymbol{X}_i\big) + U_{1,i}.$$

Then this system has the following random coefficient equations:

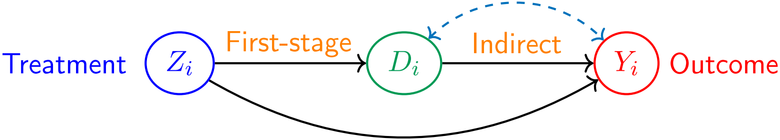
$$D_i = \phi + \overline{\pi} Z_i + \varphi(\boldsymbol{X}_i) + U_i$$

$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \zeta(\boldsymbol{X}_i) + \underbrace{(1 - D_i)\, U_{0,i} + D_i U_{1,i}}_{\text{Correlated error term}}$$

where β, γ, δ are functions of $\mu_{d'}(z'; \boldsymbol{X}_i)$.

3. Saving CM Effects

Suppose Z_i is ignorable, D_i is not, so we have the following causal model.



Then this system has the following random coefficient equations:

$$D_i = \phi + \overline{\pi} Z_i + \varphi(\boldsymbol{X}_i) + U_i$$

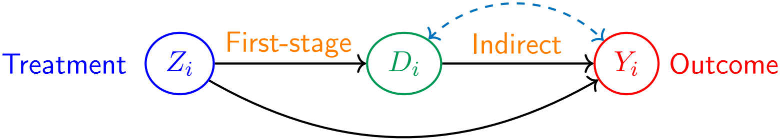
$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \zeta(\boldsymbol{X}_i) + \underbrace{(1 - D_i)\, U_{0,i} + D_i U_{1,i}}_{\text{Correlated error term}}$$

$$\text{ADE} = \mathbb{E}\left[\gamma + \delta D_i\right], \quad \text{AIE} = \mathbb{E}\left[\overline{\pi}\left(\beta + \delta Z_i + \widetilde{U}_i\right)\right]$$

with $\widetilde{U}_i = \mathbb{E}\left[U_{1,i} - U_{0,i} \,|\, \boldsymbol{X}_i, D_i(0) \neq D_i(1)\right]$ unobserved complier gains.

3. Saving CM Effects

Suppose Z_i is ignorable, D_i is not, so we have the following causal model.



Main problem, second-stage is not identified:

$$\begin{aligned}\mathbb{E}\left[Y_i \,\middle|\, Z_i, D_i, \boldsymbol{X}_i\right] &= \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \varphi(\boldsymbol{X}_i) \\ &\quad + \underbrace{(1 - D_i)\,\mathbb{E}\left[U_{0,i} \,\middle|\, D_i = 0, \boldsymbol{X}_i\right] \\ &\quad + D_i \mathbb{E}\left[U_{1,i} \,\middle|\, D_i = 1, \boldsymbol{X}_i\right]}_{\text{Unobserved } D_i \text{ confounding.}}\end{aligned}$$

Identification intuition: Identify second-stage via MTE control function (CF).

3. Saving CM Effects — Identification

Assume:

1. Mediator monotonicity, $\Pr\left(D_i(0) \leq D_i(1) \,|\, \boldsymbol{X}_i\right) = 1$

$$\implies D_i(z') = \mathbb{1}\left\{U_i \leq \pi(z'; \boldsymbol{X}_i)\right\}, \quad \text{for } z' = 0, 1 \text{ (Vycatil 2002).}$$

2. Selection on mediator benefits, $\text{Cov}\left(U_i,\, U_{0,i}\right),\ \text{Cov}\left(U_i,\, U_{1,i}\right) \neq 0$

$\implies$ First-stage take-up informs second-stage confounding.

3. There is an IV for the mediator, $\boldsymbol{X}_i^{\text{IV}}$ among control matrix $\boldsymbol{X}_i$.

$$\implies \pi(z'; \boldsymbol{X}_i) = \Pr\left(D_i = 1 \,\middle|\, Z_i = z', \boldsymbol{X}_i\right) \text{ is separately identified.}$$

Proposition: Under assumptions (1), (2), (3) the marginal indirect effect is identified,

$$\mathbb{E}\left[Y_i(z', 1) - Y_i(z', 0) \,\middle|\, Z_i = z', \boldsymbol{X}_i, U_i = u'\right]$$
$$= \beta + \delta z' + \mathbb{E}\left[U_{1,i} - U_{0,i} \,\middle|\, \boldsymbol{X}_i, U_i = u'\right].$$

3. Saving CM Effects — Identification

The marginal effect has corresponding Control Functions (CFs), describing unobserved selection-into-D_i,

$$\lambda_0(p') = \mathbb{E}\left[U_{0,i} \,\middle|\, p' \leq U_i\right], \quad \lambda_1(p') = \mathbb{E}\left[U_{1,i} \,\middle|\, U_i \leq p'\right].$$

These CFs restore second-stage identification, by extrapolating from $\boldsymbol{X}_i^{\text{IV}}$ compliers to $D_i(Z_i)$ mediator compliers,

$$\begin{aligned}\mathbb{E}\left[Y_i \,|\, Z_i, D_i, \boldsymbol{X}_i\right] &= \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \varphi(\boldsymbol{X}_i) \\ &\quad + \underbrace{\rho_0 \left(1 - D_i\right) \lambda_0\big(\pi(Z_i; \boldsymbol{X}_i)\big) + \rho_1 D_i \lambda_1\big(\pi(Z_i; \boldsymbol{X}_i)\big)}_{\text{CF adjustment.}}\end{aligned}$$

This adjusted second-stage re-identifies the ADE and AIE,

$$\text{ADE} = \mathbb{E}\left[\gamma + \delta D_i\right], \ \text{AIE} = \mathbb{E}\left[\overline{\pi}\left(\beta + \delta Z_i + \underbrace{\left(\rho_1 - \rho_0\right)\Gamma\big(\pi(0; \boldsymbol{X}_i), \, \pi(1; \boldsymbol{X}_i)\big)}_{\text{Mediator compliers extrapolation.}}\right)\right]$$

3. Saving CM Effects — Estimation

Will explain how estimation works, with simulation evidence $n = 5,000$.

1. Random treatment $Z_i \sim \text{Binom}(0.5)$
2. $(U_{0,i}, U_{1,i}) \sim \text{BivariateNormal}(0, 0, \sigma_0, \sigma_1, \rho)$, Costs $C_i \sim N(0, 0.5)$.

Roy selection-into-D_i, with constant partial effects + interaction term.

$$D_i(z') = \mathbb{1}\left\{ Y_i(z', 1) - Y_i(z', 0) \geq C_i \right\},$$
$$Y_i(z', d') = \left(z' + d' + z'd' \right) + U_{d'} \qquad \text{for } z', d' = 0, 1.$$

Following the previous, these data have the following first and second-stage equations, where $\boldsymbol{X}_i^{\text{IV}}$ is an additive cost IV:

$$D_i = \mathbb{1}\left\{ Z_i - \boldsymbol{X}_i^{\text{IV}} \geq C_i - \left(U_{1,i} - U_{0,i} \right) \right\}$$
$$Y_i = Z_i + D_i + Z_i D_i + (1 - D_i) U_{0,i} + D_i U_{1,i}.$$

$\implies$ unobserved confounding by BivariateNormal $\left(U_{0,i}, U_{1,i} \right)$.

3. Saving CM Effects — Estimation

Errors are normal, so system is Heckman (1979) selection model.

CFs are the inverse Mills ratio, with $\phi(.)$ normal pdf and $\Phi(.)$ normal cdf,

$$\lambda_0(p') = \frac{\phi(-\Phi^{-1}(p'))}{\Phi(-\Phi^{-1}(p'))}, \quad \lambda_1(p') = \frac{\phi(\Phi^{-1}(p'))}{\Phi(\Phi^{-1}(p'))}, \quad \text{for } p' \in (0,1).$$

Parametric Estimation Recipe:

1. Estimate first-stage $\pi(Z_i; \boldsymbol{X}_i)$ with probit, including $\boldsymbol{X}_i^{\text{IV}}$.
2. Include λ_0, λ_1 CFs in second-stage OLS estimation.
3. Compose CM estimates from two-stage plug-in estimates.

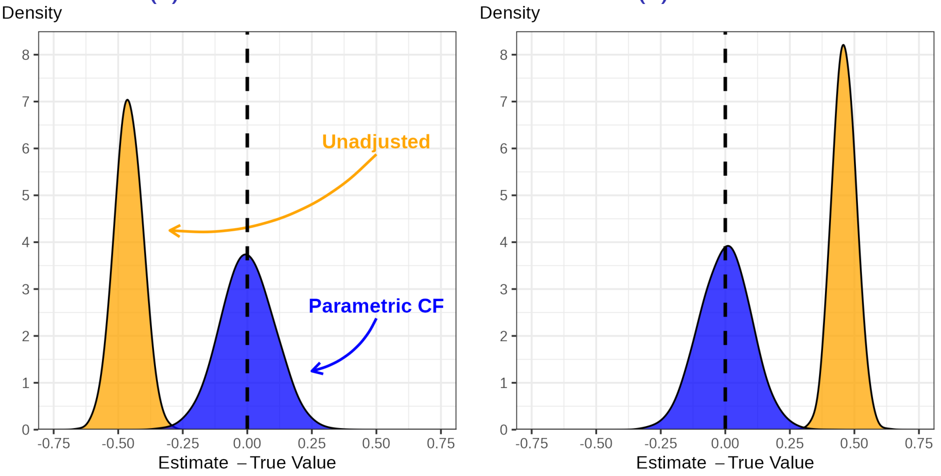
→ Same as conventional CM estimates (two-stages), with CFs added.

$$\widehat{\text{ADE}} = \mathbb{E}\left[\widehat{\gamma} + \widehat{\delta} D_i\right], \quad \widehat{\text{AIE}} = \mathbb{E}\left[\widehat{\overline{\pi}}\left(\widehat{\beta} + \widehat{\delta} Z_i + \underbrace{\left(\widehat{\rho}_1 - \widehat{\rho}_0\right)\Gamma\left(\widehat{\pi}(0; \boldsymbol{X}_i), \widehat{\pi}(1; \boldsymbol{X}_i)\right)}_{\text{Mediator compliers extrapolation.}}\right)\right]$$

3. Saving CM Effects — Estimation

Figure: Simulated Distribution of CM Effect Estimates from 10,000 DGPs.

(a) $\widehat{\text{ADE}} - \text{ADE}$. **(b)** $\widehat{\text{AIE}} - \text{AIE}$.



3. Saving CM Effects — Estimation

If errors are not normal, then CFs do not have a known form, so semi-parametrically estimate them (e.g., splines).

$$\mathbb{E}\left[Y_i \,|\, Z_i, D_i = 0, \boldsymbol{X}_i\right] = \alpha + \gamma Z_i + \varphi\big(\boldsymbol{X}_i\big) + \rho_0 \lambda_0\big(\pi(Z_i; \boldsymbol{X}_i)\big),$$

$$\mathbb{E}\left[Y_i \,|\, Z_i, D_i = 1, \boldsymbol{X}_i\right] = (\alpha + \beta) + (\gamma + \delta) Z_i + \varphi\big(\boldsymbol{X}_i\big) + \rho_1 \lambda_1\big(\pi(Z_i; \boldsymbol{X}_i)\big)$$

Semi-parametric Estimation Recipe:

1. Estimate first-stage $\pi(Z_i; \boldsymbol{X}_i)$, including $\boldsymbol{X}_i^{\mathrm{IV}}$.
2. Estimate second-stage separately for $D_i = 0$ and $D_i = 1$, with regressors $\lambda_0(p'), \lambda_1(p')$, semi-parametric in $p' = \widehat{\pi}(Z_i; \boldsymbol{X}_i)$.
3. Compose CM estimates from two-stage plug-in estimates.

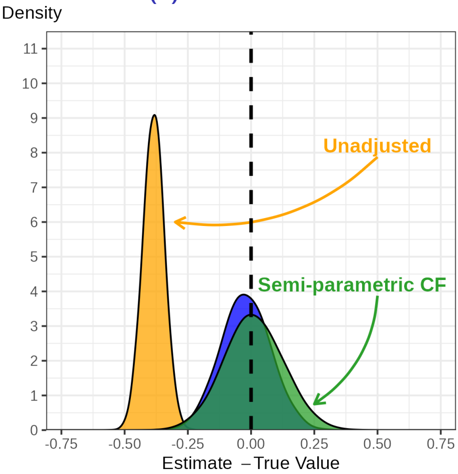
$\rightarrow$ Same as conventional CM estimates, with semi-parametric CFs. CFs.

$$\widehat{\mathrm{ADE}} = \mathbb{E}\left[\widehat{\gamma} + \widehat{\delta} D_i\right], \quad \widehat{\mathrm{AIE}} = \mathbb{E}\left[\widehat{\overline{\pi}}\left(\widehat{\beta} + \widehat{\delta} Z_i + \left(\widehat{\rho}_1 - \widehat{\rho}_0\right) \Gamma\big(\widehat{\pi}(0; \boldsymbol{X}_i), \widehat{\pi}(1; \boldsymbol{X}_i)\big)\right)\right]$$

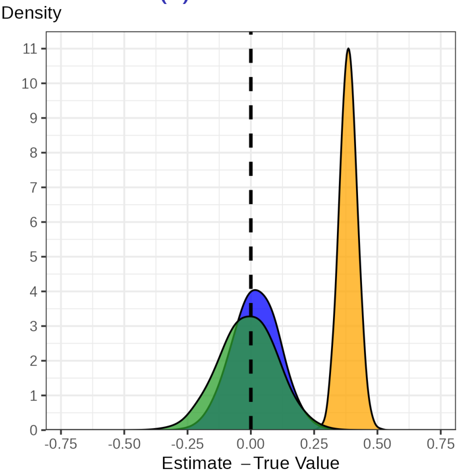
3. Saving CM Effects — Estimation

Figure: Simulated Distribution of CM Effect Estimates with Uniform Errors.

(a) $\widehat{\text{ADE}} - \text{ADE}$.

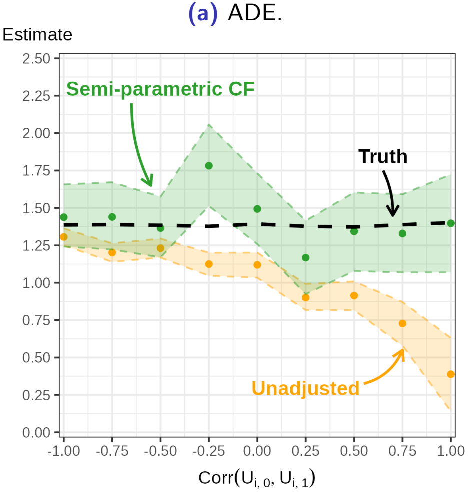


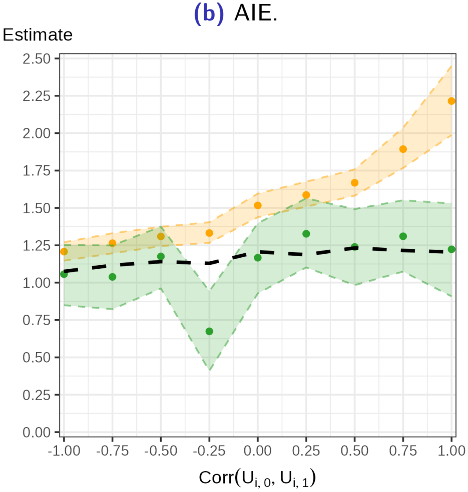
(b) $\widehat{\text{AIE}} - \text{AIE}$.



3. Saving CM Effects — Estimation

Figure: CF Adjusted Estimates Work with Different Error Term Parameters.





Conclusion

Overarching goals:

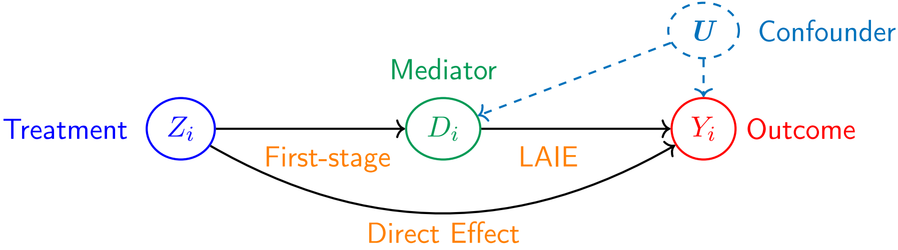
1. Alternative to practice of "suggestive evidence of mechanisms."
2. Selection bias in conventional CM analyses with no case for mediator ignorability
   → Noted problems in the most popular methods for CM effects, pertinent for economic applications.
3. Connect CM with labour theory + selection-into-treatment + MTEs
   → Valid CM identification in these settings.

Caveats and points to remember:

- Structural assumptions and IV for identification + estimation (not ideal).
- Application to Oregon Health Insurance Experiment in the paper, showing health + well-being effects mediated by healthcare (wide confidence intervals).
- **Credible** CM analyses are hard in practice.

Appendix: CM Guiding Model

Consider binary treatment $Z_i = 0, 1$, binary mediator $D_i = 0, 1$, and continuous outcome Y_i for individuals $i = 1, \ldots, n$.



Average Direct Effect (ADE) : $\mathbb{E}\left[Y_i\left(\boxed{1}, D_i(Z_i)\right) - Y_i\left(\boxed{0}, D_i(Z_i)\right)\right]$

- ADE is causal effect $Z \to Y$, blocking the indirect D_i path.

Average Indirect Effect (AIE): $\mathbb{E}\left[Y_i\left(Z_i, \boxed{D_i(1)}\right) - Y_i\left(Z_i, \boxed{D_i(0)}\right)\right]$

- AIE is causal effect of $D(Z) \to Y$, blocking the direct Z_i path.

Group Difference — ADE

CM effects contaminated by (less interpretable) bias terms.

CM Estimand = ADEM + Selection Bias

$$\underbrace{\mathbb{E}_{D_i}\left[\mathbb{E}\left[Y_i \,|\, Z_i = 1, D_i\right] - \mathbb{E}\left[Y_i \,|\, Z_i = 0, D_i\right]\right]}_{\text{Estimand, Direct Effect}}$$

$$= \underbrace{\mathbb{E}_{D_i=d'}\left[\mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) \,\middle|\, D_i(1) = d'\right]\right]}_{\text{Average Direct Effect on Mediator (ADEM) take-up — i.e., } D_i(1) \text{ weighted}}$$

$$+ \underbrace{\mathbb{E}_{D_i}\left[\mathbb{E}\left[Y_i(0, D_i(Z_i)) \,\middle|\, D_i(1) = d'\right] - \mathbb{E}\left[Y_i(0, D_i(Z_i)) \,\middle|\, D_i(0) = d'\right]\right]}_{\text{Selection Bias}}$$

The weighted ADE you get here is a positive weighted sum of local ADEs, but with policy irrelevant weights $D_i(1) = d'$.

$\Longrightarrow$ consider this group bias, noting difference from true ADE. Back

Group Difference — AIE

CM effects contaminated by (less interpretable) bias terms.

$$\text{CM Estimand} = \text{AIEM} + \Big(\text{Selection Bias} + \text{Group difference bias}\Big)$$

$$\underbrace{\mathbb{E}_{Z_i}\left[\Big(\mathbb{E}\left[D_i \mid Z_i = 1\right] - \mathbb{E}\left[D_i \mid Z_i = 0\right]\Big) \times \Big(\mathbb{E}\left[Y_i \mid Z_i, D_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i, D_i = 0\right]\Big)\right]}_{\text{Estimand, Indirect Effect}}$$

$$= \underbrace{\mathbb{E}\left[Y_i(Z_i, D_i(1)) - Y_i(Z_i, D_i(0)) \mid D_i = 1\right]}_{\text{Average Indirect Effect on Mediated (AIEM) — i.e., } D_i = 1 \text{ weighted}}$$

$$+ \underbrace{\overline{\pi}\Big(\mathbb{E}\left[Y_i(Z_i, 0) \mid D_i = 1\right] - \mathbb{E}\left[Y_i(Z_i, 0) \mid D_i = 0\right]\Big)}_{\text{Selection Bias}}$$

$$+ \underbrace{\overline{\pi}\left[\left(\frac{1 - \Pr\left(D_i(1) = 1, D_i(0) = 0\right)}{\Pr\left(D_i(1) = 1, D_i(0) = 0\right)}\right)\Big(\mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i(1) = 0 \text{ or } D_i(0) \right. \\ \left. - \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0)\right]\right.}_{\text{Groups difference Bias}}$$

The weighted AIE you get here is not a positive weighted sum of local AIEs, because the AIE is only about $D(Z)$ compliers. Model.

$\implies$ consider this group bias, noting difference from true AIE. Back

Semi-parametric Control Functions

Semi-parametric specifications for the CFs λ_0, λ_1 bring some complications to estimating the AIE.

$$\mathbb{E}\left[Y_i \,|\, Z_i, D_i = 0, \boldsymbol{X}_i\right] = \alpha + \gamma Z_i + \varphi\big(\boldsymbol{X}_i\big) + \boxed{\rho_0 \lambda_0\big(\pi(Z_i; \boldsymbol{X}_i)\big)},$$

$$\mathbb{E}\left[Y_i \,|\, Z_i, D_i = 1, \boldsymbol{X}_i\right] = (\alpha + \beta) + (\gamma + \delta) Z_i + \varphi\big(\boldsymbol{X}_i\big) + \boxed{\rho_1 \lambda_1\big(\pi(Z_i; \boldsymbol{X}_i)\big)}$$

Intercepts, $\alpha, (\alpha + \beta)$, and relevance parameters ρ_0, ρ_1 are not separately identified from the CFs $\lambda_0(.), \lambda_1(.)$.

These problems can be avoided by estimating the AIE using its relation to the ATE, $\widehat{\text{AIE}}^{\text{CF}} =$

$$\widehat{\text{ATE}} - (1 - \overline{Z}) \underbrace{\left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\gamma} + \widehat{\delta}\, \widehat{\pi}(1; \boldsymbol{X}_i) \right)}_{\widehat{\text{ADE}} \text{ given } Z_i = 1} - \overline{Z} \underbrace{\left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\gamma} + \widehat{\delta}\, \widehat{\pi}(0; \boldsymbol{X}_i) \right)}_{\widehat{\text{ADE}} \text{ given } Z_i = 0}.$$