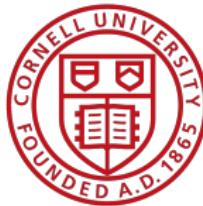


# Causal Mediation in Natural Experiments

Senan Hogan-Hennessy  
Economics Department, Cornell University  
[seh325@cornell.edu](mailto:seh325@cornell.edu)



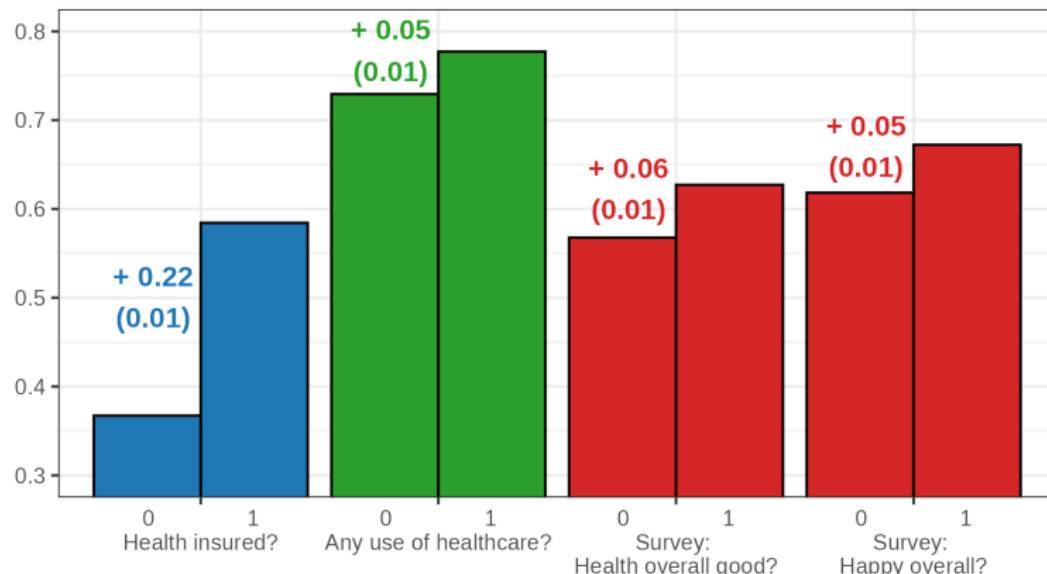
---

Econometric Society World Congress, Seoul  
22 August 2025

## Intro: Oregon Health Insurance Experiment

In 2008, Oregon gave access to socialised health insurance by wait-list lottery (Finkelstein et al, 2012).

Mean Outcome, for each  $z' = 0, 1$ .



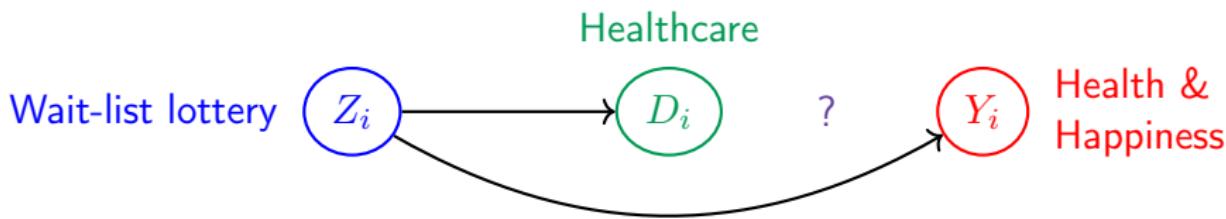
## Applied practice:

⇒ Suggestive evidence for healthcare as mechanism for wait-list lottery. . .

## Intro: Oregon Health Insurance Experiment

In 2008, Oregon gave access to socialised health insurance by wait-list lottery (Finkelstein et al, 2012).

**Figure:** Model for Suggestive Evidence of a Mechanism.



#### Inconsistencies in this conclusion:

- A triangular system missing the  $D_i \rightarrow Y_i$  edge...
  - Is  $D_i \rightarrow Y_i$  small, large, or even nonexistent?
  - 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:

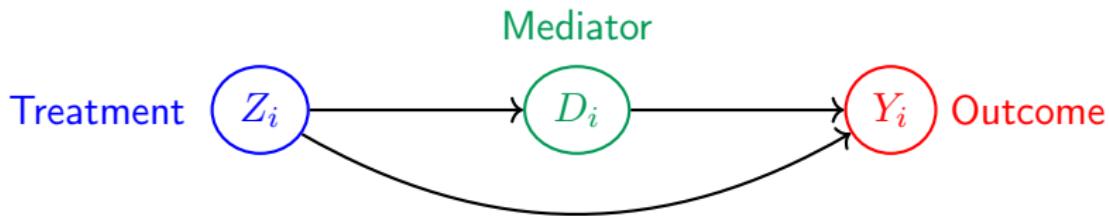
- ① Problems with conventional approach to CM in observational settings.  
**[Negative result]**
  - ② Recovering valid CM effects, via MTE + control function modelling.  
**[Positive result]**
- 

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), Brinch Mogstad Wiswall (2017), 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  $i = 1, \dots, n$ .



Assume  $Z_i$  is ignorable (conditional on  $X_i$ ).

---

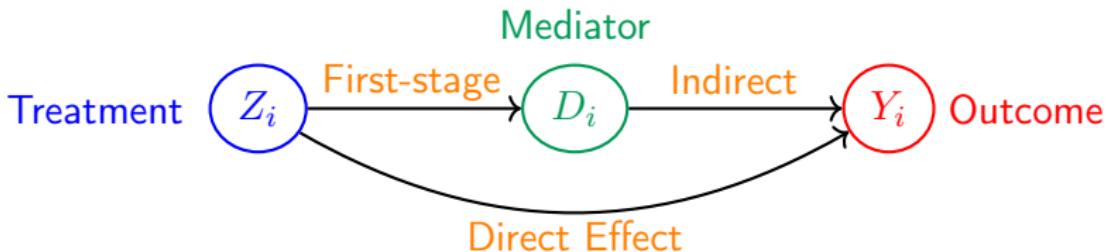
Mediator  $D_i$  is a function of  $Z_i$ . Outcome  $Y_i$  is a function of both  $Z_i, D_i$ .

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

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

# 1. CM — Model

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



Assume  $Z_i$  is ignorable (conditional on  $X_i$ ).

---

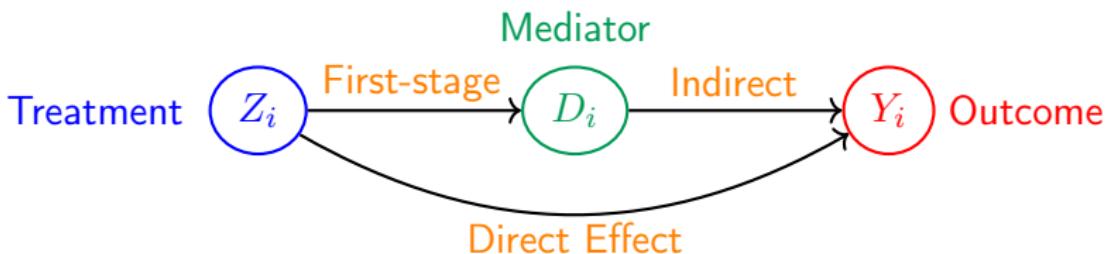
Only two causal effects are identified so far.

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

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

# 1. CM — Model

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



---

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

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

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

- AIE is causal effect of  $D(Z) \rightarrow Y$ , 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  $\mathbf{X}_i$  and  $Z_i$  realisation

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

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

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

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

## 2. Selection Bias

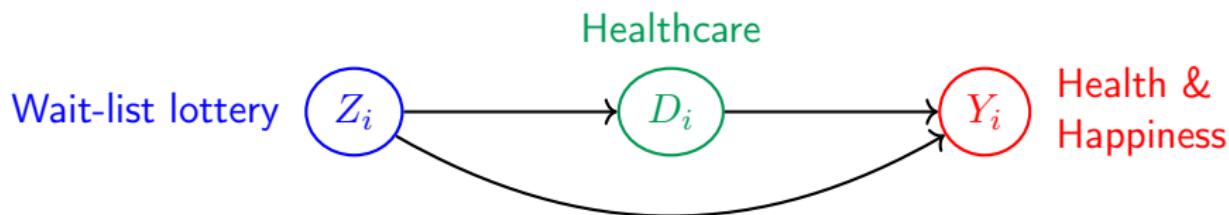
Mediator ignorability (MI, Imai Keele Yamamoto 2010):

Assume mediator  $D_i$  is **also ignorable**, conditional on  $X_i, Z_i$  realisation

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

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  $X_i$  (no natural experiment...).

## 2. Selection Bias

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

---

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

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

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

---

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

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

## 2. Selection Bias

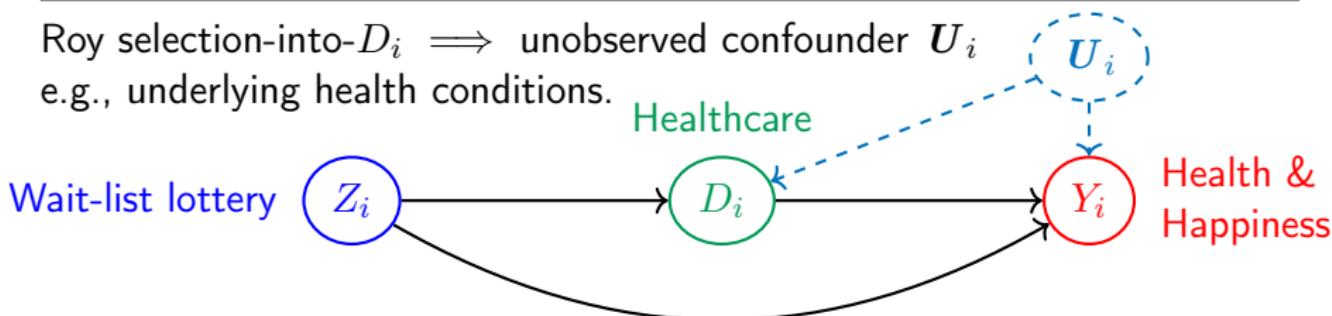
Assume: Healthcare is quasi-random, conditional on lottery realisation  $Z_i$  and demographic controls  $X_i$  (no natural experiment...).

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

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

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

Roy selection-into- $D_i \implies$  unobserved confounder  $U_i$   
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  $D_i \rightarrow Y_i$ , if  $D_i$  not ignorable:

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

## 2. Selection Bias — Direct Effect

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

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


---

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

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

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

$$+ \underbrace{\mathbb{E}_{D_i=d'} \left[ \begin{aligned} & \left( 1 - \Pr(D_i(1) = d') \right) \\ & \times \left( \mathbb{E} [Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) | D_i(1) = 1 - d'] \right. \\ & \left. - \mathbb{E} [Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) | D_i(0) = d'] \right) \end{aligned} \right]}_{\text{Group difference bias}} \quad \blacktriangleright \text{Group-diff}$$

## 2. Selection Bias — Indirect Effect

CM Effects have this same flavour, causal effects + contaminating bias.<sup>1</sup>

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

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

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

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

$$+ \bar{\pi} \left[ \left( 1 - \Pr(D_i = 1) \right) \begin{pmatrix} \mathbb{E}[Y_i(Z_i, 1) - Y_i(Z_i, 0) | D_i = 1] \\ - \mathbb{E}[Y_i(Z_i, 1) - Y_i(Z_i, 0) | D_i = 0] \end{pmatrix} \right. \\ \left. + \left( \frac{1 - \Pr(D_i(1) = 1, D_i(0) = 0)}{\Pr(D_i(1) = 1, D_i(0) = 0)} \right) \begin{pmatrix} \mathbb{E}[Y_i(Z_i, 1) - Y_i(Z_i, 0) | D_i(Z_i) \neq Z_i] \\ - \mathbb{E}[Y_i(Z_i, 1) - Y_i(Z_i, 0)] \end{pmatrix} \right]$$

Groups difference Bias ▶ Group-diff

## 2. Selection Bias

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

---

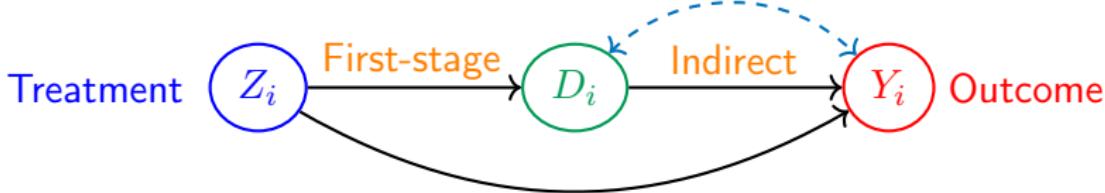
- ① Counter to accepted intuitive reasoning in applied economics, which often controls for plausible mediators
- ② Invalidates conclusions in fields that uncritically apply CM methods with no case for mediator ignorability (common in some fields of epidemiology, psychology, medicine, and sociology)
- 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. CM with Selection

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,  $U_{d',i}$ .

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

Then this system has the following random coefficient equations:

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

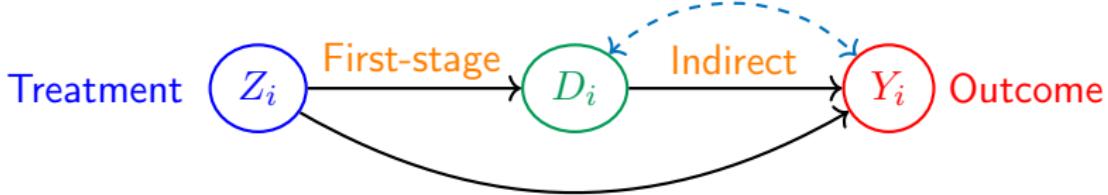
$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \zeta(\mathbf{X}_i) + \underbrace{(1 - D_i) U_{0,i} + D_i U_{1,i}}_{\text{Correlated error term}}$$

where  $\beta, \gamma, \delta$  are functions of  $\mu_{d'}(z'; \mathbf{X}_i)$ .

Correlated error term

### 3. CM with Selection

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:

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

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

where  $\beta, \gamma, \delta$  are functions of  $\mu_{d'}(z'; \mathbf{X}_i)$ .

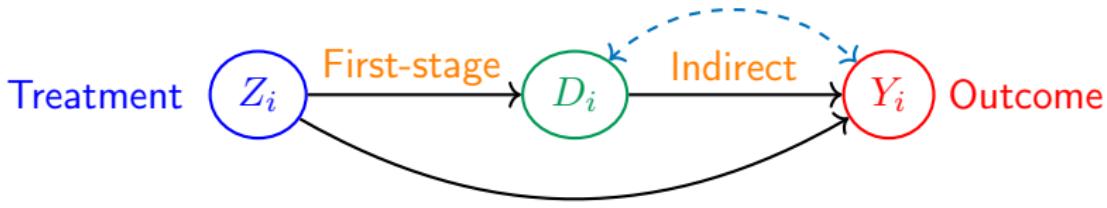
Correlated error term

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

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

### 3. CM with Selection

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



Main problem, second-stage is not identified:

$$\begin{aligned}
 \mathbb{E} [Y_i | Z_i, D_i, \mathbf{X}_i] &= \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \varphi(\mathbf{X}_i) \\
 &+ (1 - D_i) \mathbb{E} [U_{0,i} | D_i = 0, \mathbf{X}_i] \\
 &+ D_i \mathbb{E} [U_{1,i} | D_i = 1, \mathbf{X}_i]
 \end{aligned}$$

Unobserved  $D_i$  confounding.

**Identification intuition:** Identify second-stage via MTE control function.

### 3. CM with Selection — Identification

Assume:

- ① Mediator monotonicity,  $\Pr(D_i(0) \leq D_i(1) | \mathbf{X}_i) = 1$   
 $\implies D_i(z') = \mathbb{1}\{U_i \leq \pi(z'; \mathbf{X}_i)\}, \text{ for } z' = 0, 1$  (Vycatil 2002).
- ② Selection on mediator benefits,  $\text{Cov}(U_i, U_{0,i}), \text{Cov}(U_i, U_{1,i}) \neq 0$   
 $\implies$  First-stage take-up informs second-stage confounding.
- ③ There is an IV for the mediator,  $\mathbf{X}_i^{\text{IV}}$  among control variables  $\mathbf{X}_i$ .  
 $\implies \pi(Z_i | \mathbf{X}_i) = \Pr(D_i = 1 | Z_i, \mathbf{X}_i)$  is separately identified.

---

**Proposition:** Under assumptions (1), (2), (3) the marginal effect of the mediator is identified,

$$\begin{aligned} & \mathbb{E}[Y_i(z', 1) - Y_i(z', 0) | Z_i = z', \mathbf{X}_i, U_i = p'] \\ &= \beta + \delta z' + \mathbb{E}[U_{1,i} - U_{0,i} | \mathbf{X}_i, U_i = p'], \quad \text{for } p' \in (0, 1). \end{aligned}$$

### 3. CM with Selection — Identification

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

$$\rho_0 \lambda_0(p') = \mathbb{E} [U_{0,i} \mid p' \leq U_i], \quad \rho_1 \lambda_1(p') = \mathbb{E} [U_{1,i} \mid U_i \leq p'].$$

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

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

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

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

### 3. CM with Selection — Estimation

Will explain how estimation works, with simulation evidence.

- ① Random treatment  $Z_i \sim \text{Binom}(0.5)$ , for  $n = 5,000$ .
- 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.

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

---

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

$$D_i = \mathbb{1} \left\{ C_i - \left( U_{1,i} - U_{0,i} \right) \leq Z_i - X_i^{\text{IV}} \right\}$$

$$Y_i = Z_i + D_i + Z_i D_i + (1 - D_i) U_{0,i} + D_i U_{1,i}.$$

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

### 3. CM with Selection — Estimation

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

CFs are the inverse Mills ratio, with  $\phi(\cdot)$  normal pdf and  $\Phi(\cdot)$  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  $\pi(Z_i; \mathbf{X}_i)$  with probit, including  $\mathbf{X}_i^{\text{IV}}$ .
- ② Include  $\lambda_0, \lambda_1$  CFs in second-stage OLS estimation.
- ③ Compose CM estimates from two-stage plug-in estimates.

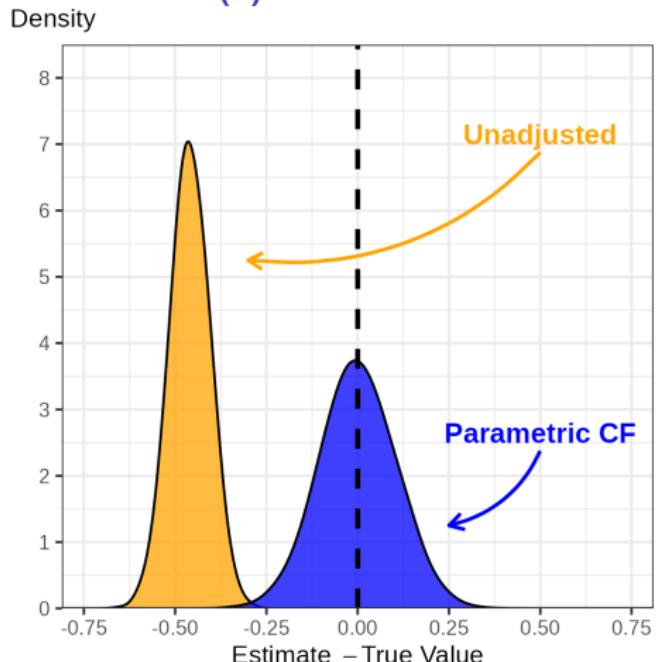
→ Same as conventional CM estimates (two-stages), with CFs added.

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

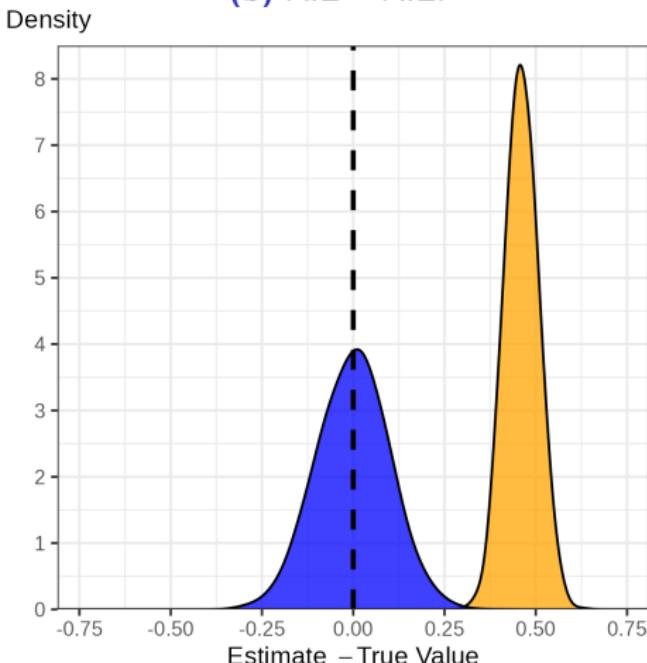
### 3. CM with Selection — Estimation

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

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



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



### 3. CM with Selection — Estimation

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

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

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

#### Semi-parametric Estimation Recipe:

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

→ Same as conventional CM estimates, with semi-parametric CFs.

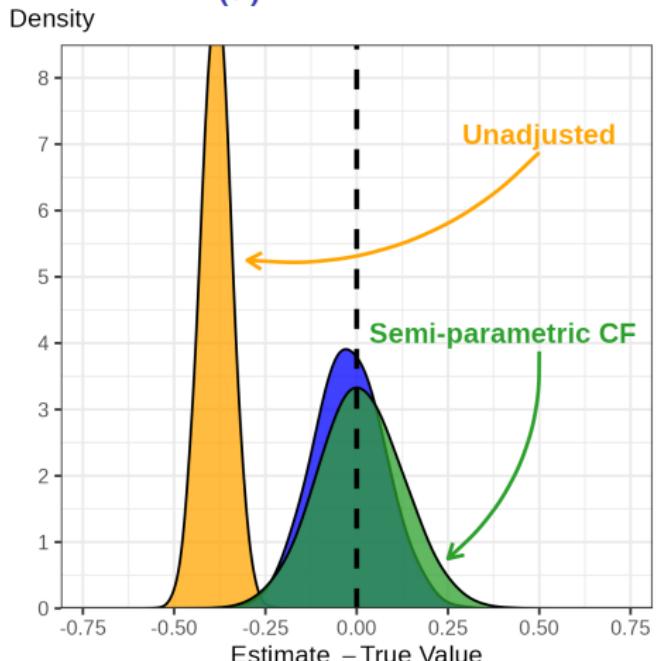
► CFs.

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

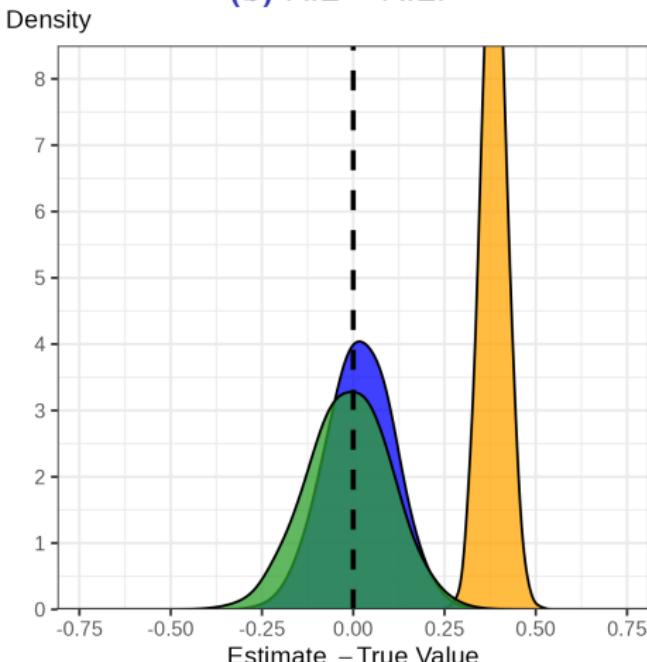
### 3. CM with Selection — Estimation

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

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



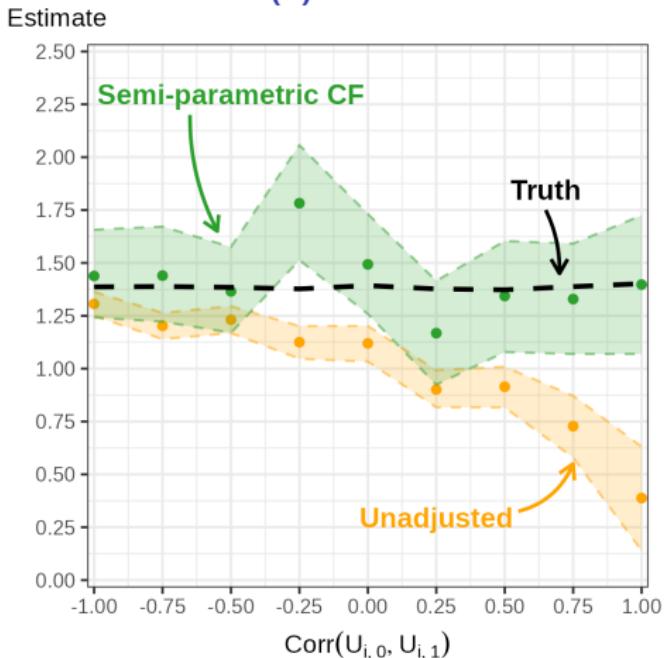
(b)  $\widehat{AIE} - AIE$ .



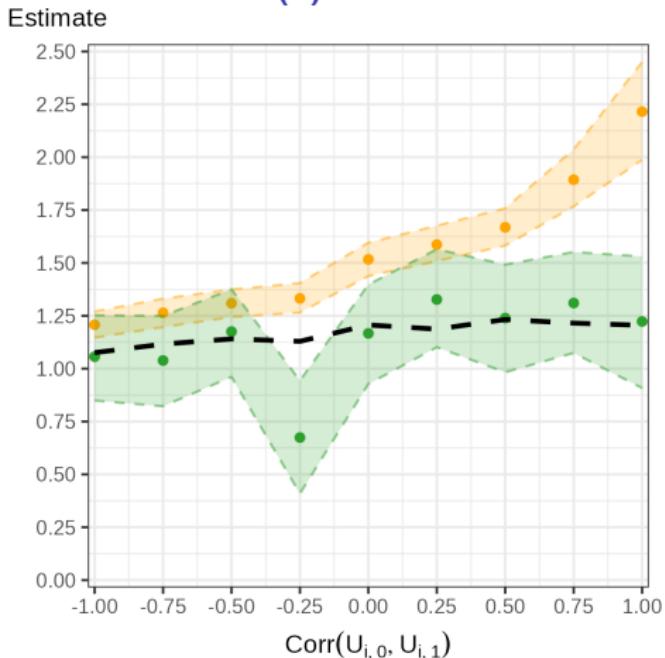
### 3. CM with Selection — Estimation

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

(a) ADE.



(b) AIE.



# Conclusion

## Overarching goals:

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

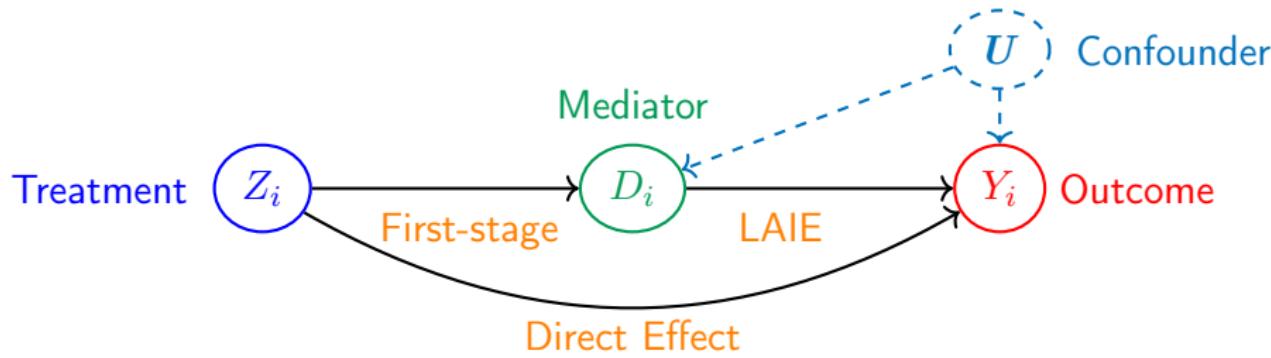
---

## 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  $i = 1, \dots, n$ .



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

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

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

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

# Group Difference — ADE

CM effects contaminated by (less interpretable) bias terms.

$$\text{CM Estimand} = \text{ADEM} + \text{Selection Bias}$$

$$\begin{aligned}
 & \underbrace{\mathbb{E}_{D_i} \left[ \mathbb{E}[Y_i | Z_i = 1, D_i] - \mathbb{E}[Y_i | Z_i = 0, D_i] \right]}_{\text{Estimand, Direct Effect}} \\
 &= \underbrace{\mathbb{E}_{D_i=d'} \left[ \mathbb{E}[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) | D_i(1) = d'] \right]}_{\text{Average Direct Effect on Mediator (ADEM) take-up — i.e., } D_i(1) \text{ weighted}} \\
 & \quad + \underbrace{\mathbb{E}_{D_i} \left[ \mathbb{E}[Y_i(0, D_i(Z_i)) | D_i(1) = d'] - \mathbb{E}[Y_i(0, D_i(Z_i)) | D_i(0) = d'] \right]}_{\text{Selection Bias}}
 \end{aligned}$$

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

⇒ consider this group bias, noting difference from true ADE.

[Back](#)

# Group Difference — AIE

CM effects contaminated by (less interpretable) bias terms.

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

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

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

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

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

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](#).

→ consider this group bias noting difference from true AIE.

[► Back](#)

## Semi-parametric Control Functions

Semi-parametric specifications for the CFs  $\lambda_0, \lambda_1$  bring some complications to estimating the AIE.

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

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

Intercepts,  $\alpha$ ,  $(\alpha + \beta)$ , and relevance parameters  $\rho_0, \rho_1$  are not separately identified from the CFs  $\lambda_0(\cdot), \lambda_1(\cdot)$  so CF extrapolation term  $(\rho_1 - \rho_0)\Gamma(\pi(0; \mathbf{X}_i), \pi(1; \mathbf{X}_i))$  is not directly identified or estimable.

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

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