Causal Mediation in Natural Experiments

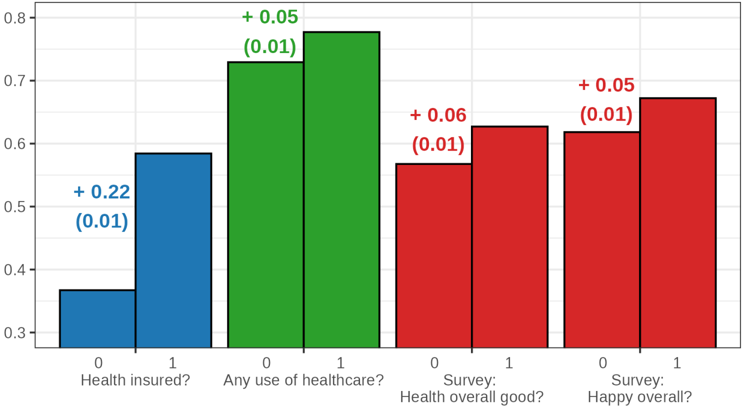
Senan Hogan-Hennessy
Economics Department, Cornell University
seh325@cornell.edu

Econometric Society World Congress, Seoul
22 August 2025

Intro: Oregon Health Insurance Experiment

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

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



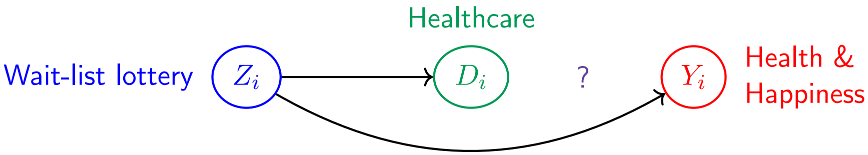
Applied practice:

⇒ 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 suggestive evidence of mechanisms:

- 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 to studying mechanisms, and assumptions for identification.

---

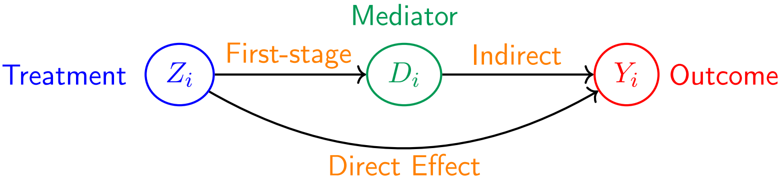
1. Problems with conventional approach to CM in observational settings. **[Negative result]**
2. Recovering valid CM effects, via MTE + control function modelling. **[Positive result]**

New insights from intersection of two fields:

- **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).

Introduction – CM

Consider ignorable treatment $Z_i = 0, 1$, binary mediator $D_i = 0, 1$, and continuous outcome Y_i.



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

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

Average Direct Effect (ADE) and Average Indirect Effect (AIE) are identified by two-stage regression

- ADE is causal effect $Z_i \to Y_i$, blocking the indirect D_i path
- AIE is causal effect of $D_i(Z_i) \to Y_i$, blocking the direct Z_i path.

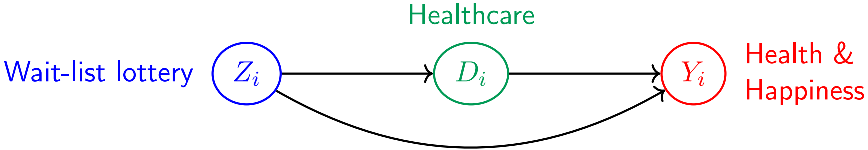
1. Selection Bias

Assumption: **Mediator ignorability** (**MI**, Imai Keele Yamamoto 2010)
mediator D_i is ***also*** **ignorable**, conditional on $\boldsymbol{X}_i$, Z_i realisation

---

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 $\boldsymbol{X}_i$.

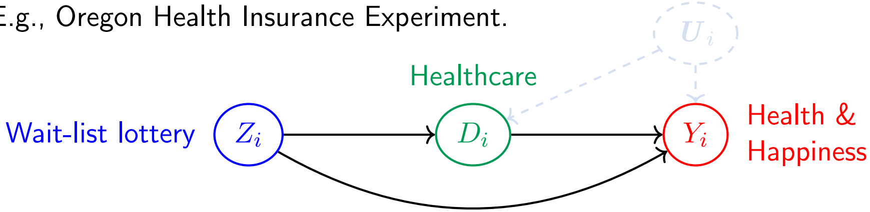
1. Selection Bias

Assumption: **Mediator ignorability** (**MI**, Imai Keele Yamamoto 2010)
mediator D_i is ***also*** **ignorable**, conditional on $\boldsymbol{X}_i$, Z_i realisation

---

Would this assumption hold true in settings economists study?

E.g., Oregon Health Insurance Experiment.



Theorem: If choice to attend healthcare is unconstrained, based on costs and benefits (Roy model) and demographics do not explain all benefits $\implies$ **MI** does not hold, unobserved confounding.

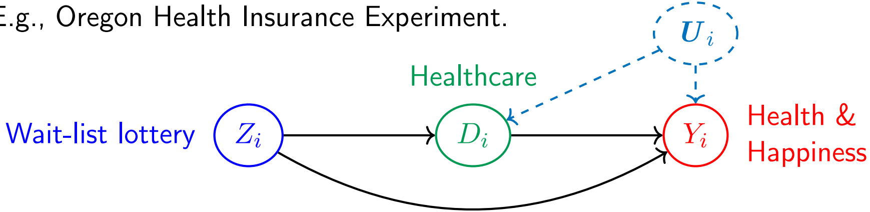
1. Selection Bias

Assumption: **Mediator ignorability** (**MI**, Imai Keele Yamamoto 2010)
mediator D_i is ***also* ignorable**, conditional on $\boldsymbol{X}_i$, Z_i realisation

---

Would this assumption hold true in settings economists study?

E.g., Oregon Health Insurance Experiment.



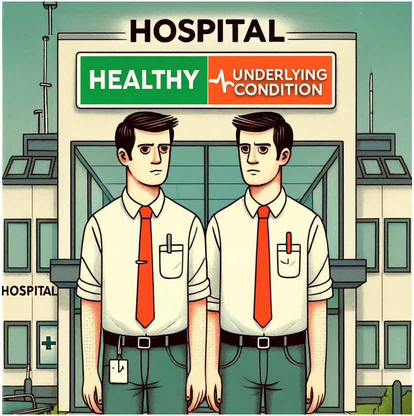
Theorem: If choice to attend healthcare is unconstrained, based on costs and benefits (Roy model) and demographics do not explain all benefits $\implies$ **MI** does not hold, unobserved confounding.

1. Selection Bias

In an observational setting, need an additional credible research design for **Mediator Ignorability (MI)** to be credible.

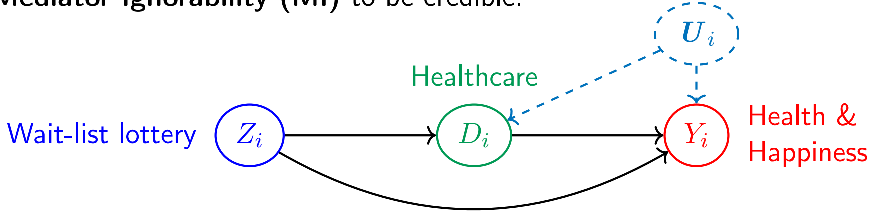
(a) Cells in a lab → **MI** believable.

(b) People choosing healthcare → **MI** not.



1. Selection Bias

In an observational setting, need an additional credible research design for **Mediator Ignorability (MI)** to be credible.



If not, then CM effects are contaminated by bias terms, similar to classical selection bias (e.g., Heckman Ichimura Smith Todd 1998).

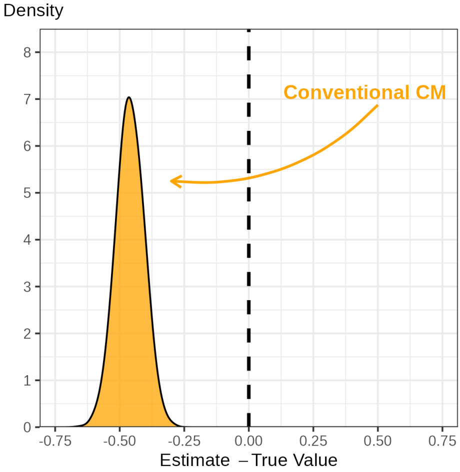
- ADE: CM Estimand $=$ ADE $+$ $\Big($Selection Bias $+$ Group difference bias$\Big)$
- AIE: CM Estimand $=$ AIE $+$ $\Big($Selection Bias $+$ Group difference bias$\Big)$

ADE biases AIE biases

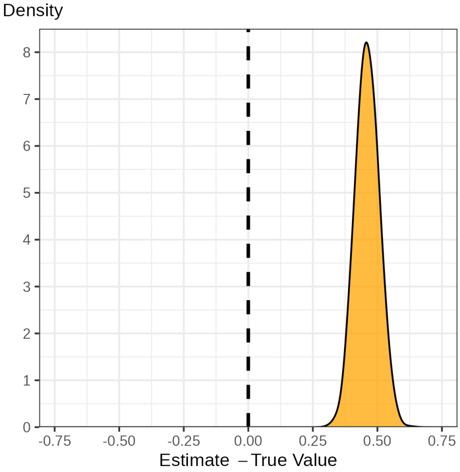
1. Selection Bias

In a simulation with Roy selection-into-D_i, CM estimates are biased.

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



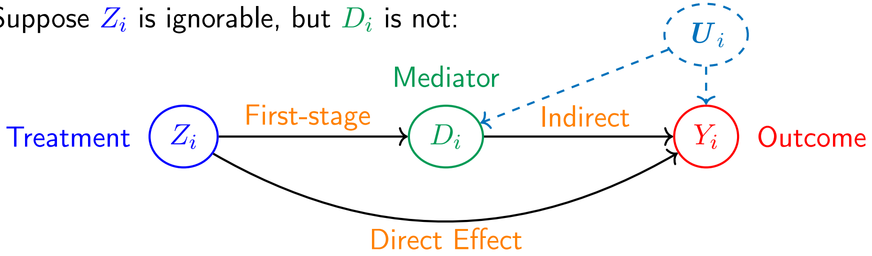
(b) $\widehat{\text{AIE}} - \text{AIE}$.



2. CM with Selection

Conventional CM methods do not identify ADE + AIE in a natural experiment setting, but can we make adjustments?

Suppose Z_i is ignorable, but D_i is not:



---

1. Average first-stage, $Z_i \to D_i$, is identified
2. Average second-stage, $Z_i, D_i \to Y_i$, is not — represented by $\boldsymbol{U}_i$.

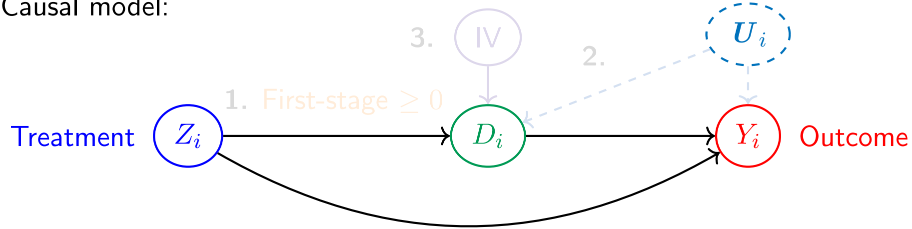
Intuition: model $\boldsymbol{U}_i$ via mediator MTE to identify ADE + AIE.

2. CM with Selection — Identification

MTE assumptions:

1. Mediator monotonicity
2. Selection on mediator benefits
3. IV for mediator take-up cost.

Causal model:



Proposition: Under MTE assumptions, the mediator MTE is identified.

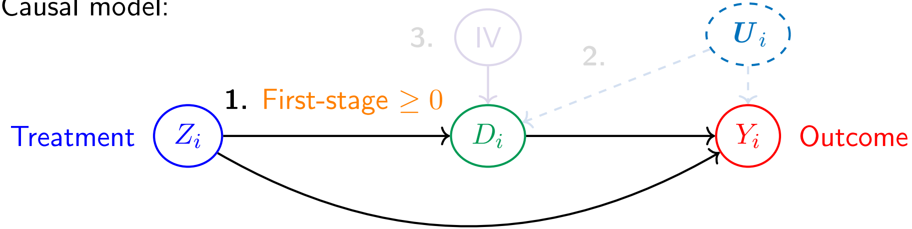
Theorem: Mediation second-stage effects, $Z_i, D_i \to Y_i$, are identified by the MTE associated Control Functions (CFs).

2. CM with Selection — Identification

MTE assumptions:

1. Mediator monotonicity
2. Selection on mediator benefits
3. IV for mediator take-up cost.

Causal model:



Proposition: Under MTE assumptions, the mediator MTE is identified.

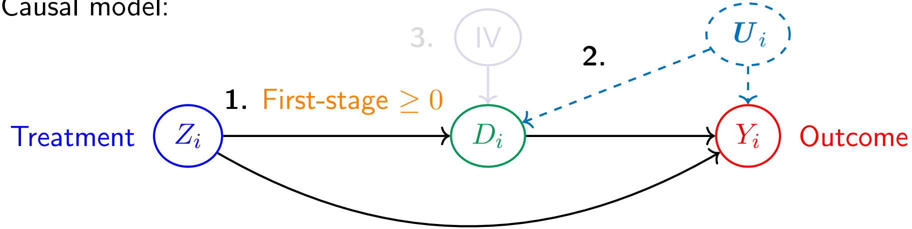
Theorem: Mediation second-stage effects, $Z_i, D_i \to Y_i$, are identified by the MTE associated Control Functions (CFs).

2. CM with Selection — Identification

MTE assumptions:

1. Mediator monotonicity
2. Selection on mediator benefits
3. IV for mediator take-up cost.

Causal model:



Proposition: Under MTE assumptions, the mediator MTE is identified.

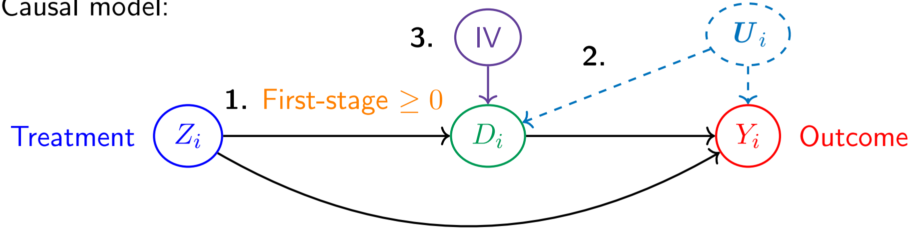
Theorem: Mediation second-stage effects, $Z_i, D_i \to Y_i$, are identified by the MTE associated Control Functions (CFs).

2. CM with Selection — Identification

MTE assumptions:

1. Mediator monotonicity
2. Selection on mediator benefits
3. IV for mediator take-up cost.

Causal model:



Proposition: Under MTE assumptions, the mediator MTE is identified.

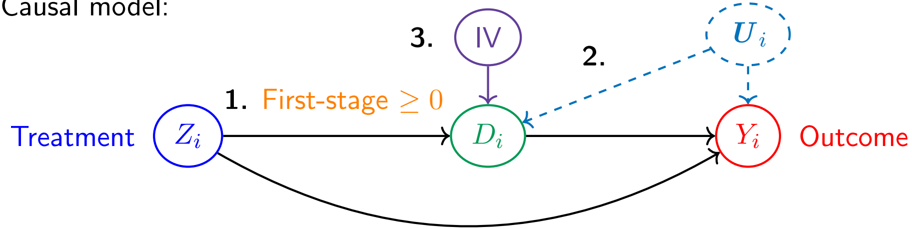
Theorem: Mediation second-stage effects, $Z_i, D_i \to Y_i$, are identified by the MTE associated Control Functions (CFs).

2. CM with Selection — Identification

MTE assumptions:

1. Mediator monotonicity
2. Selection on mediator benefits
3. IV for mediator take-up cost.

Causal model:



Proposition: Under MTE assumptions, the mediator MTE is identified.

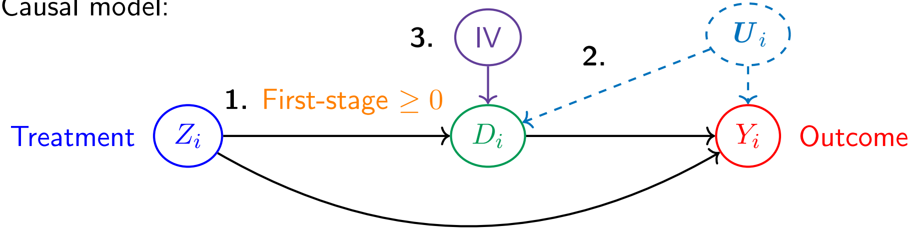
Theorem: Mediation second-stage effects, $Z_i, D_i \to Y_i$, are identified by the MTE associated Control Functions (CFs).

2. CM with Selection — Identification

MTE assumptions:

1. Mediator monotonicity
2. Selection on mediator benefits
3. IV for mediator take-up cost.

Causal model:



Proposition: Under MTE assumptions, the mediator MTE is identified.

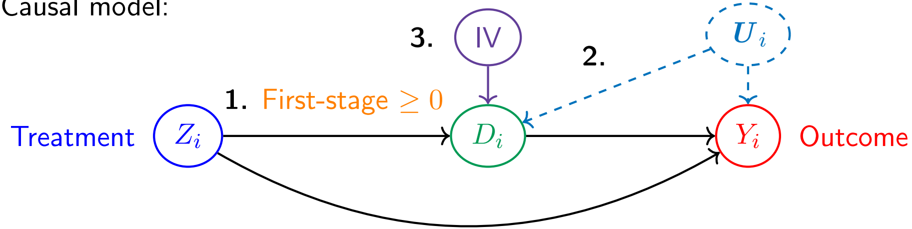
Intuition: Identifies ADE + AIE by extrapolating from IV compliers to mediator compliers (MTE extrapolation e.g., Mogstad Torgovitsky 2018).

2. CM with Selection — Identification

MTE assumptions:

1. Mediator monotonicity
2. Selection on mediator benefits
3. IV for mediator take-up cost.

Causal model:



Proposition: Under MTE assumptions, the mediator MTE is identified.

Intuition: Identifies ADE + AIE by extrapolating from IV compliers to mediator compliers (MTE extrapolation e.g., Mogstad Torgovitsky 2018).

2. CM with Selection — Estimation

In practice, this means two-stage CM estimation, with CF in second-stage.

Parametric CF Estimation Recipe:

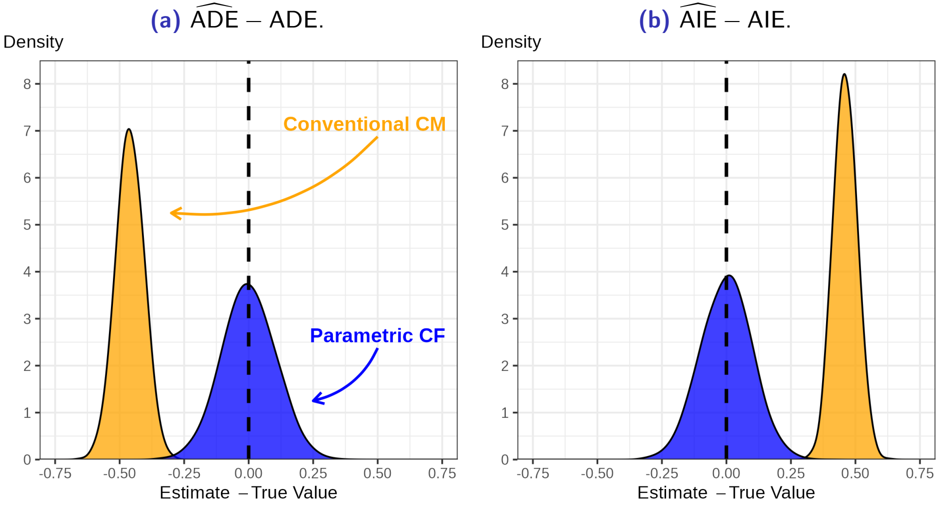
1. Estimate mediation first-stage with probit, including the IV.
2. Estimate mediation second-stage by OLS, with Mills ratio CF terms (Heckman 1979).
3. Compose CM estimates from two-stage plug-in estimates.

Semi-parametric CF Estimation Recipe:
Same as before, but replace **2.** with semi-parametric CFs.

$\rightarrow$ Same as conventional CM estimates (two-stages), with CFs added.

2. CM with Selection — Estimation

Figure: CM Estimates from 10,000 DGPs with **Normal** Errors.



2. CM with Selection — Estimation

In practice, this means two-stage CM estimation, with CF in second-stage.

Parametric CF Estimation Recipe:

1. Estimate mediation first-stage with probit, including the IV.
2. Estimate mediation second-stage by OLS, with Mills ratio CF terms (Heckman 1979).
3. Compose CM estimates from two-stage plug-in estimates.

Semi-parametric CF Estimation Recipe:
Same as before, but replace **2.** with semi-parametric CFs.

$\rightarrow$ Same as conventional CM estimates (two-stages), with CFs added.

2. CM with Selection — Estimation

In practice, this means two-stage CM estimation, with CF in second-stage.

Parametric CF Estimation Recipe:

1. Estimate mediation first-stage with probit, including the IV.
2. Estimate mediation second-stage by OLS, with Mills ratio CF terms (Heckman 1979).
3. Compose CM estimates from two-stage plug-in estimates.

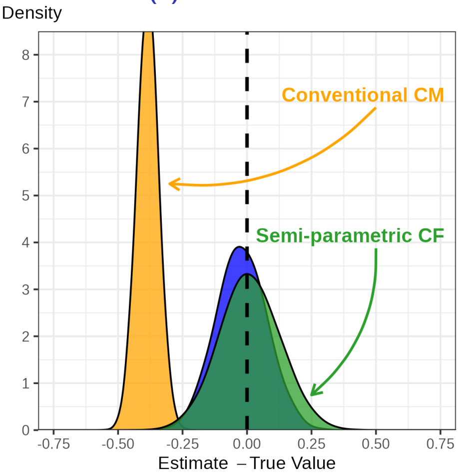
Semi-parametric CF Estimation Recipe:
Same as before, but replace **2.** with semi-parametric CFs.

$\rightarrow$ Same as conventional CM estimates (two-stages), with CFs added.

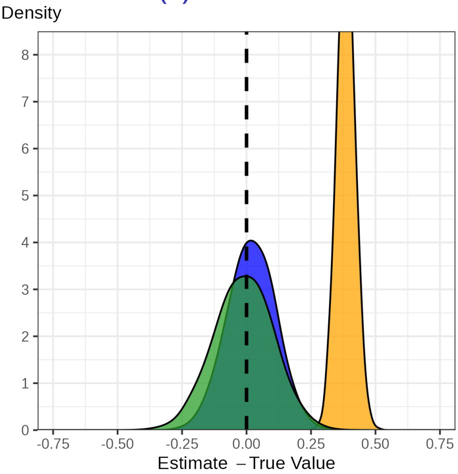
2. CM with Selection — Estimation

Figure: CM Estimates from 10,000 DGPs with **Uniform** Errors.

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

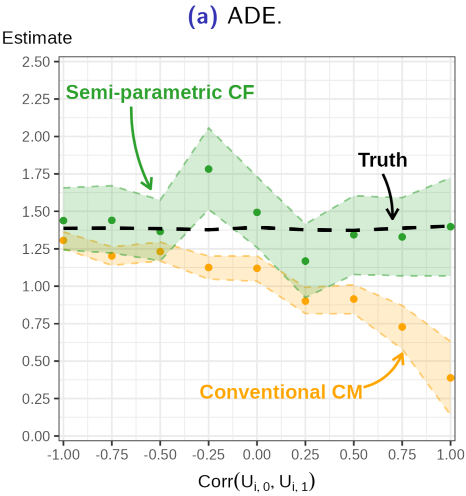


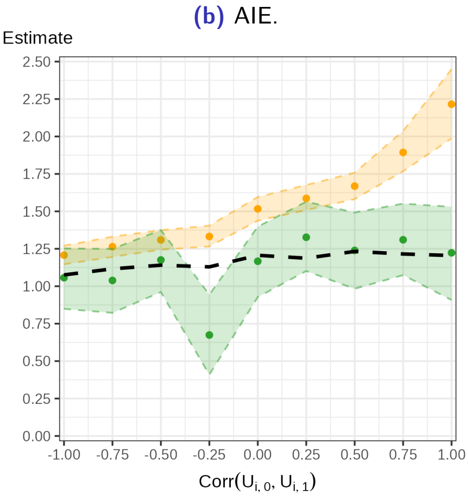
(b) $\widehat{\text{AIE}} - \text{AIE}$.



2. CM with Selection — Estimation

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





Conclusion

Overview:

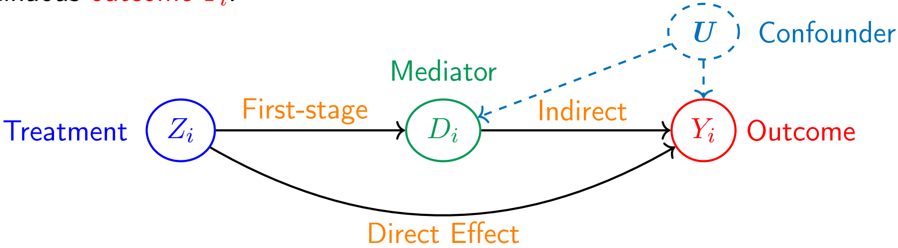
1. CM as alternative to "suggestive evidence for mechanisms."
2. Selection bias in conventional CM analyses with no case for mediator ignorability
   $\rightarrow$ Noted problems in the most popular methods for CM, pertinent for economic applications.
3. Connect CM with labour theory + selection-into-treatment + MTEs
   $\rightarrow$ 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.



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.

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

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

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

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

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

Selection Bias — Direct Effect

CM Effects + contaminating bias.

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

Model

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

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

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

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

Group-diff

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

Selection Bias — Indirect Effect

CM Effects + contaminating bias, where $\overline{\pi} = \Pr\left(D_i(0) \neq D_i(1)\right)$.

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

Semi-parametric Control Functions

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

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

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

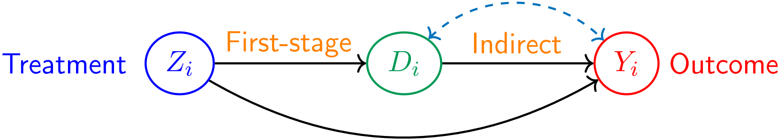
Intercepts, $\alpha, (\alpha + \beta)$, and relevance parameters ρ_0, ρ_1 are not separately identified from the CFs $\lambda_0(.), \lambda_1(.)$ so CF extrapolation term $(\rho_1 - \rho_0)\Gamma\big(\pi(0; \boldsymbol{X}_i), \, \pi(1; \boldsymbol{X}_i)\big)$ is not directly identified or estimable.

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

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

Appenidx: 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:

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

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

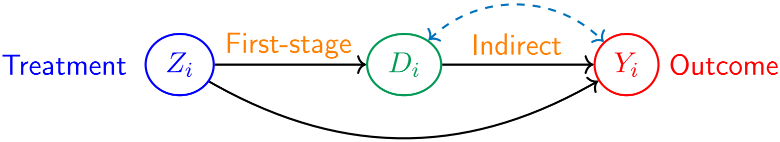
where β, γ, δ are functions of $\mu_{d'}(z'; \boldsymbol{X}_i)$.

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

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

Appenidx: 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:

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

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

Appenidx: CM with Selection — 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, Cov $\left(U_i,\, U_{0,i}\right)$, 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 variables $\boldsymbol{X}_i$.

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

Proposition:

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

Appenidx: CM with Selection — Identification

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

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

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,

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

Appenidx: 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.

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

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

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

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

Appenidx: CM with Selection — 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.

$\rightarrow$ Same as conventional CM estimates (two-stages), with CFs added.

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

Appenidx: 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).

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

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

Semi-parametric Estimation Recipe:

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

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

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