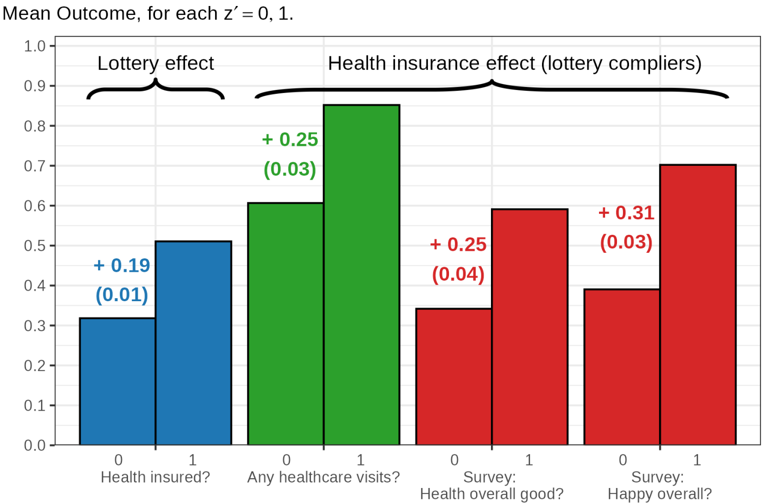
Causal Mediation in Natural Experiments

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

Econometric Society World Congress, Seoul
22 August 2025

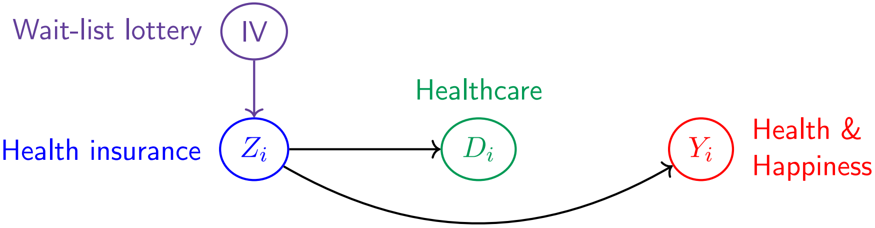
Oregon Health Insurance Experiment

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



Oregon Health Insurance Experiment

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

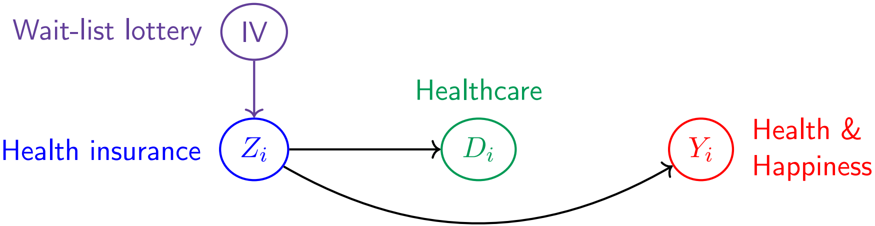


$\Rightarrow$ suggestive evidence of healthcare visits as a mechanism for health insurance gains.

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

Oregon Health Insurance Experiment

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



⇒ suggestive evidence of healthcare visits as a mechanism for health insurance gains.

1. This paper considers an alternative approach, Causal Mediation (CM)
2. CM explicitly states its estimands + identifying assumptions
3. Hugely popular in other fields, but not so in quas-experimental economics (for good reason...)

Introduction

This project examines Causal Mediation (CM) with economic perspective:

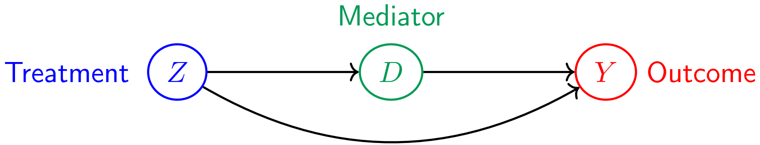
1. Problems with conventional approach to CM (and informal mechanism analyses) in social science settings — focusing on natural experiments.
   [Negative result]
2. Recovering valid CM effects under selection-into-mediator, with modelling asumptions.
   [Positive result]

Brings together ideas from two different literatures:

- **Causal Mediation (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).

Direct & Indirect Effects — Model

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



Mediator D_i is a function of Z_i.

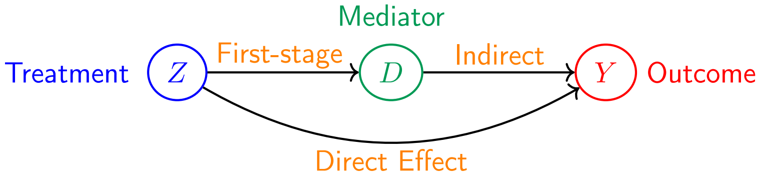
$$D_i = \begin{cases} D_i(0), & \text{if } Z_i = 0 \\ D_i(1), & \text{if } Z_i = 1. \end{cases}$$

Outcome Y_i is a function of both Z_i, D_i.

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

Direct & Indirect Effects — Model

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



Suppose Z_i is ignorable, conditional on $\boldsymbol{X}_i$.

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

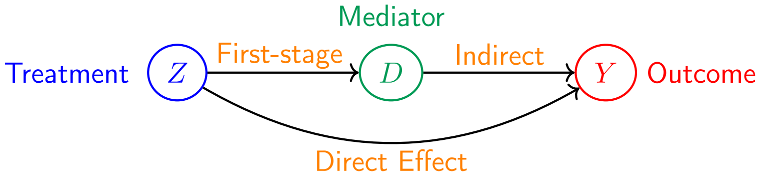
Only two causal effects are identified so far.

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

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

Direct & Indirect Effects — Model

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



First-stage and ATE answer important questions:

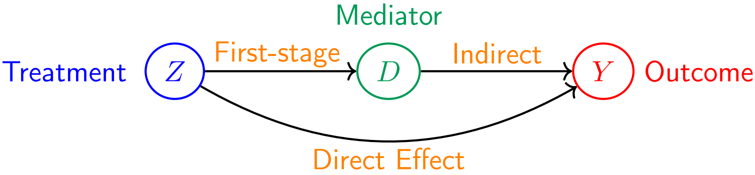
- Did socialised health insurance increase healthcare use, and improve health? (Finkelstein et al, 2012).

Unanswered questions about the mechanism(s):

- Did health benefits come from using health care more? Health gains from reduced uncertainty — i.e., insurance?
- Is health insurance more about the health or more about the insurance?

Direct & Indirect Effects — 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) : $\mathbb{E}\left[Y_i\left(1, D_i(Z_i) \right) - Y_i\left(0, D_i(Z_i) \right) \right]$

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

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

- AIE is causal effect of $D(Z) \to Y$, blocking the direct Z path.[1]

[1] Note: AIE = fraction of $D(Z)$ compliers $\times$ average effect $D \to Y$ among compliers.

Direct & Indirect Effects — Identification

Sequential ignorability (**SI**, 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 **SI** holds then ADE and AIE are identified by two-stage regression:

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

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

Direct & Indirect Effects — Identification

Sequential ignorability (**SI**, Imai Keele Yamamoto 2010):

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

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

E.g., OLS simultaneous regression (Imai Keele Yamamoto, 2010):

$Z_i \leftarrow$ Treatment

$D_i \leftarrow$ Mediator

$Y_i \leftarrow$ Outcome

First-stage: $D_i = \phi + \pi Z_i + \boldsymbol{\psi}_1' \boldsymbol{X}_i + U_i$

Second-stage: $Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \boldsymbol{\psi}_2' \boldsymbol{X}_i + \varepsilon$

$$\implies \text{ADE} = \gamma + \delta \mathbb{E}\left[D_i\right]$$

$$\text{AIE} = \pi \left(\beta + \delta \mathbb{E}\left[Z_i\right]\right)$$

i.e., a regression decomposition.
Other estimation methods do the same decomposition, avoiding linearity assumptions (see Huber 2020 for an overview).

Direct & Indirect Effects — Selection

$\implies$ Great, we can use the Imai Keele Yamamoto (2010) approach to CM in all our respective applied projects.
$\implies$ Learn the mechanism pathways in causal research $\rightarrow$ big gain!

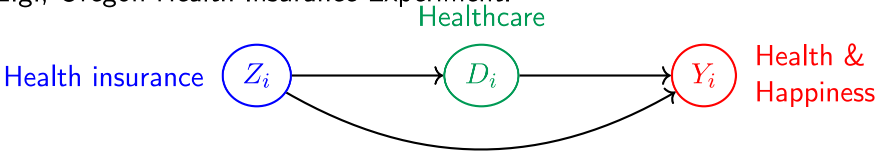
Before we import these methods to applied/labour economics and observational research, interrogate the **SI** assumption.

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

Would this assumption hold true in settings economists study?

Direct & Indirect Effects — Selection

E.g., Oregon Health Insurance Experiment.



SI in practice:

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

1. Health insurance assigned randomly (e.g., the 2008 Oregon wait-list lottery).
2. Healthcare is quasi-random, conditional on health insurance Z_i and demographics $\boldsymbol{X}_i$.

Direct & Indirect Effects — Selection

SI: healthcare usage is quasi-random, conditional on Health insurance assign Z_i and demographics $\boldsymbol{X}_i$.

Consider the case **individuals go to the healthcare** to maximise health.

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

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 **SI** 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%.

Direct & Indirect Effects — Selection

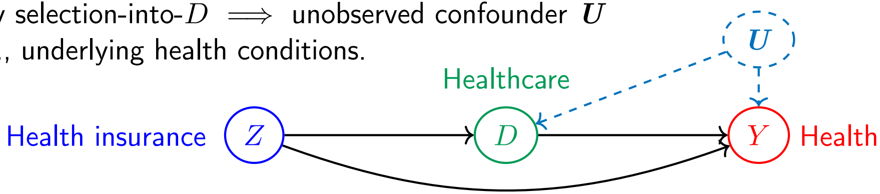
SI: healthcare usage is quasi-random, conditional on Health insurance assign
Z_i and demographics $\boldsymbol{X}_i$.

Consider the case **individuals go to the healthcare** to maximise health.

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

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

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



Direct & Indirect Effects — Selection

In practice, the only way to believe the **SI** assumption (selection-on-observables) is to study a case with another natural experiment for D_i — in addition to the one that guaranteed Z_i is ignorable.

(a) Cells in a lab $\rightarrow$ **SI** believable. **(b)** People choosing healthcare $\rightarrow$ **SI** not.

Direct & Indirect Effects — 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 $D \to Y$, if D not ignorable:

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

Direct & Indirect Effects — 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 $D \to Y$, if D not ignorable:

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

Direct & Indirect Effects — Selection Bias

$\Longrightarrow$ CM Effects have this same flavour, causal effects contaminated by (less interpretable) bias terms. ▸ Model

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

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

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

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

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

Direct & Indirect Effects — Selection Bias

$\implies$ CM Effects have this same flavour, causal effects contaminated by (less interpretable) bias terms. Model Put $\pi = \Pr\left(D_i(1) = 1, D_i(0) = 0\right)$.

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

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

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

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

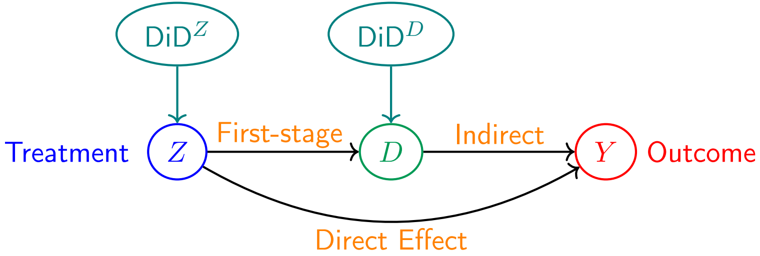
$$+ \underbrace{\pi\left[\Big(1 - \Pr\left(D_i = 1\right)\Big)\begin{pmatrix}\mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i = 1\right] \\ - \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i = 0\right]\end{pmatrix} + \left(\frac{1 - \Pr\left(D_i(1) = 1, D_i(0) = 0\right)}{\Pr\left(D_i(1) = 1, D_i(0) = 0\right)}\right)\begin{pmatrix}\mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i(1) = 0 \text{ or } D_i(\\ - \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0)\right]\end{pmatrix}\right.}$$

Identification Under Selection

That was a long way of giving negative results. Is there any hope?

If you can use a two-way research design, then please do!

Figure: Two-way Diff-in-Diff (see Deuchert Huber Schelker, 2019).



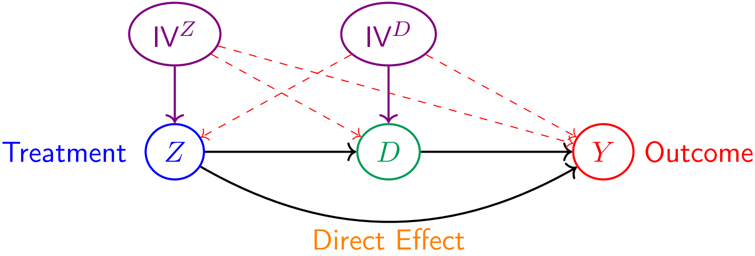
Note: assumes common trends across complier groups, identifies ADE + AIE local to complier groups.

Identification Under Selection

That was a long way of giving negative results. Is there any hope?

If you can use a two-way research design, then please do!

Figure: Two-way IV (see Frlölich Huber, 2017).



Note: two-way exclusion restriction, identifies ADE + AIE local to overlapping complier groups. Also avoid 2SLS (see Kim 2025)!

Identification Under Selection

That was a long way of giving negative results. Is there any hope?

What about the mainstream case, with research design for only Z?
How do economists do causal effects in these systems?

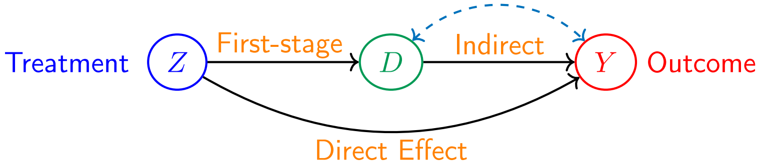
1. Estimate the ATE, and call it a day.
2. (optional) Present suggestive evidence of mechanisms. . . . ▸ Suggestive

---

New: Control Function solution to identification.

Identification with a Control Function

Suppose Z is ignorable, D is not, so we have the following causal model.



Write outcomes as sum of means and mean-zero errors, $U_{D_i,i}$.

$$Y_i(Z_i,0) = \mathbb{E}\left[Y_i(Z_i,0) \mid \boldsymbol{X}_i\right] + U_{0,i}, \quad Y_i(Z_i,1) = \mathbb{E}\left[Y_i(Z_i,1) \mid \boldsymbol{X}_i\right] + U_{1,i}.$$

Then this system has the following regression equations:

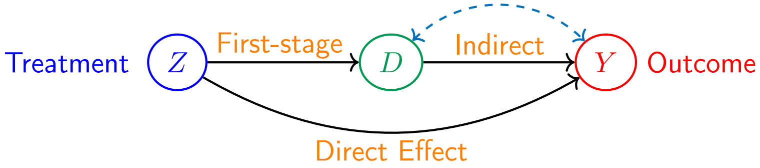
$$D_i = \phi + \pi Z_i + \varphi(\boldsymbol{X}_i) + U_i$$

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

Where $\beta, \gamma, \delta, \pi$ comprise the ADE and AIE.

Identification with a Control Function

Suppose Z is ignorable, D is not, so we have the following causal model.



Then this system has the following regression equations:

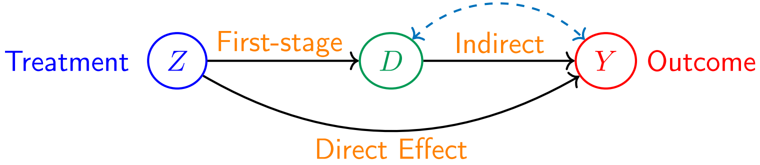
$$D_i = \phi + \pi Z_i + \varphi(\boldsymbol{X}_i) + U_i$$
$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \zeta(\boldsymbol{X}_i) + \underbrace{(1 - D_i)\, U_{0,i} + D_i U_{1,i}}_{\text{Correlated error term.}}$$

Where $\beta, \gamma, \delta, \pi$ comprise the ADE and AIE.

Control Function intuition: Identify second-stage (despite correlated error term), to get ADE + AIE.

Identification with a Control Function

Suppose Z is ignorable, D is not, so we have the following causal model.



Note: Roy selection has first- and second-stage errors correlated.

$$D_i = \mathbb{1}\left\{ Z_i(\delta+\beta) + (1-Z_i)\beta \geq C_i - \left(U_{1,i} - U_{0,i} \right) \right\}$$

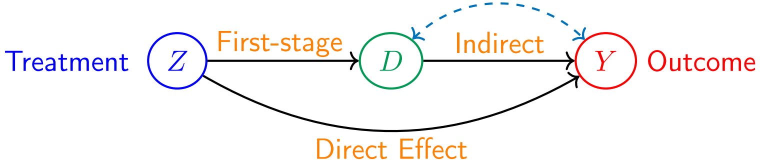
$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \zeta(\boldsymbol{X}_i) + \underbrace{(1-D_i)\,U_{0,i} + D_i U_{1,i}}_{\text{Correlated error term}}$$

where C_i are costs of taking D_i.

Control Function intuition: use first-stage errors to purge second-stage correlated errors.

Identification with a Control Function

Suppose Z is ignorable, D is not, so we have the following causal model.



Heckman (1979) Control Function, assumptions:

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

$$\implies D_i(z') = \mathbb{1}\left\{\mu(z'; \boldsymbol{X}_i) \geq U_i\right\}.$$

- First-stage errors inform second-stage errors,

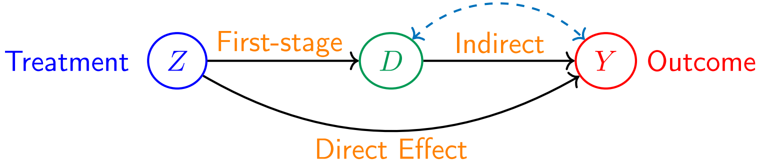
$$\text{Cov}\left[U_i, (1 - D_i)\, U_{0,i} + D_i U_{1,i}\right] \neq 0.$$

- Error-term distribution, $U_i, U_{0,i}, U_{1,i} \sim \text{TriNormal}\left(\boldsymbol{M}, \boldsymbol{\Sigma}\right)$.

$\implies$ identify second-stage, and thus ADE + AIE.

Identification with a Control Function

Suppose Z is ignorable, D is not, so we have the following causal model.



Heckman (1979) Control Function, in operation:

1. Back out Control Function (CF) in first-stage (probit, normal errors),

$$\widehat{K}_i = D_i - \widehat{\mathbb{E}}\left[D_i | Z_i, \boldsymbol{X}_i\right].$$

2. Include Mills ratio CF in OLS estimates of the second-stage,

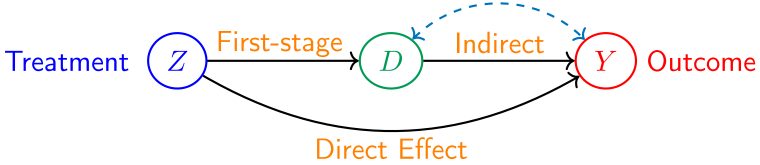
$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \boldsymbol{\zeta}' \boldsymbol{X}_i + \underbrace{(1 - D_i)\,\lambda\left(-\widehat{K}_i\right) + D_i \lambda\left(\widehat{K}_i\right)}_{\text{CF correction, } \lambda(.) \text{ inv Mills ratio.}} + \varepsilon_i$$

3. Compose estimates from second-stage,

$$\widehat{\text{ADE}} = \widehat{\gamma} + \widehat{\delta}\mathbb{E}\left[D_i\right], \quad \widehat{\text{AIE}} = \widehat{\pi}\left(\widehat{\beta} + \widehat{\delta}\mathbb{E}\left[Z_i\right] + \mathbb{E}\left[\lambda\left(\widehat{K}_i\right) - \lambda\left(-\widehat{K}_i\right)\right]\right)$$

Identification with a Control Function

Suppose Z is ignorable, D is not, so we have the following causal model.



Semi-parametric control function (Newey Imbens 2012), assumptions:

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

$$\implies D_i(z') = \mathbb{1}\left\{\mu(z'; \boldsymbol{X}_i) \geq U_i\right\}.$$

2. First-stage errors inform second-stage errors,

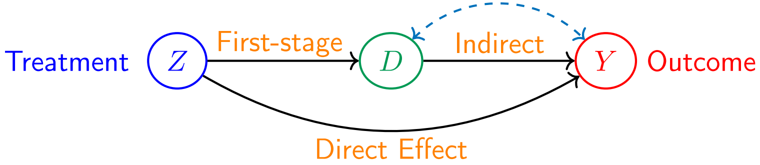
$$\text{Cov}\Big[U_i, (1 - D_i)\, U_{0,i} + D_i U_{1,i}\Big] \neq 0.$$

3. Valid instrument $\boldsymbol{X}_i^{\text{IV}}$ for D_i, to separate CF functional form. ▸ MTEs

$\implies$ identifies second-stage, ADE + AIE (w.out error dist assumption).

Identification with a Control Function

Suppose Z is ignorable, D is not, so we have the following causal model.



Semi-parametric control function (Newey Imbens 2012), in operation:

1. Back out Control Function (CF) in first-stage (semi/non-parametric), with IV $\boldsymbol{X}_i^{\text{IV}}$,
$$\widehat{K}_i = D_i - \widehat{\mathbb{E}}\left[D_i \middle| Z_i, \boldsymbol{X}_i^{\text{IV}}, \boldsymbol{X}_i \right].$$
2. Include semi-parametric CF in OLS estimates of the second-stage,
$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \boldsymbol{\zeta}' \boldsymbol{X}_i + \underbrace{(1 - D_i)\, \lambda_0\left(-\widehat{K}_i\right) + D_i \lambda_1\left(\widehat{K}_i\right)}_{\text{CF correction, } \lambda_0(.), \lambda_1(.) \text{ splines.}} + \varepsilon_i$$
3. Compose estimates from second-stage,
$$\widehat{\text{ADE}} = \widehat{\gamma} + \widehat{\delta}\mathbb{E}[D_i] \qquad \widehat{\text{AIE}} = \widehat{\pi}\Big(\widehat{\beta} + \widehat{\delta}\mathbb{E}[Z_i] + \mathbb{E}\Big[\widehat{\lambda}_0\Big(\widehat{K}_i\Big) - \widehat{\lambda}_1\Big(-\widehat{K}_i\Big)\Big]\Big)$$

Simulation Evidence

Simulation with trivariate normal errors + unobserved costs, $N = 10,000$.

1. Random treatment $Z_i \sim$ 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 $\left(U_{0,i}, U_{1,i}\right)$.

Simulation Evidence

Simulation with Roy selection, BivariateNormal errors + unobserved costs.

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

(a) ADE. **(b)** AIE.

Simulation Evidence

Simulation with Roy selection, trivariate normal errors, unobserved costs.

Figure: Point Estimates of CM Effects, OLS versus Control Function, varying ρ values with $\sigma_0 = 1, \sigma_1 = 2$ fixed.

(a) ADE.

(b) AIE.

Conclusion

Overarching goals:

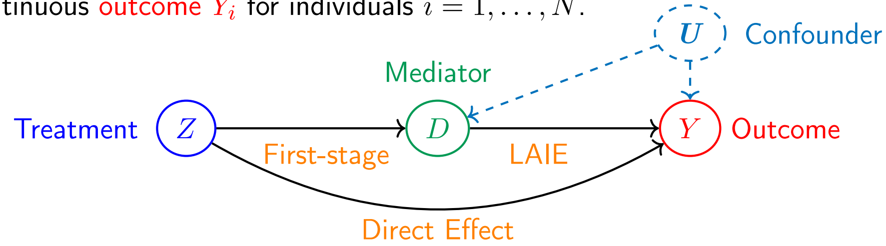
1. Ward economists away from using CM methods unabashedly.
   → Noted problems in the most popular methods for CM effects, pertinent for economic applications.
2. CM methods away from ignorability assumptions, inappropriate for economics (+ social science) settings.
   → Methods valid when selection-into-treatment theory relevant.

Work-in-progress part of LWIPS:

- Connect the control function approach to MTE methods ▸ MTEs
- Large sample properties + analytical SEs
- Use this approach to estimate direct and indirect effects of genetics and education (companion paper)
- (eventually) *R* package for selection-adjusted CM effects, by Heckman model and IV-assisted CF/MTE.

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 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 path.[2]

[2] Note: AIE = fraction of $D(Z)$ compliers $\times$ average effect $D \to Y$ among compliers.

Group Difference — ADE

CM effects contaminated by (less interpretable) bias terms.

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

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

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

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

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

Group Difference — AIE

CM effects contaminated by (less interpretable) bias terms.

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

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

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

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

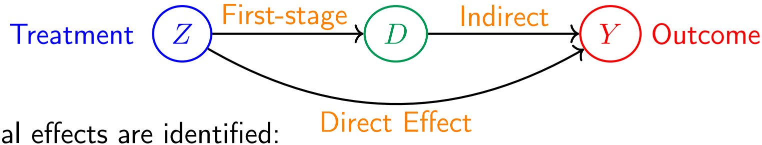
$$+ \underbrace{\pi\left[\left(\frac{1 - \Pr\left(D_i(1) = 1, D_i(0) = 0\right)}{\Pr\left(D_i(1) = 1, D_i(0) = 0\right)}\right)\Big(\mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i(1) = 0 \text{ or } D_i(0) \right. - \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0)\right]}_{\text{Groups difference Bias}}$$

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

Appendix: Suggestive Evidence of Mechanisms

How empirical economists currently give evidence for mechanisms/mediators in causal effects.



Two causal effects are identified:

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

$\implies$ Show results of these two effects and assume indirect effect is positive, constant $\rightarrow$ suggestive evidence of mechanisms!

See Blackwell Matthew Ruofan Opacic (2024) for this in full, and a partial identification approach to avoid its unrealistic assumptions.

Appendix: Connection to MTEs

The ADE is fine to estimate with a Control Function/CF, but AIE refers to mediator benefits only among mediator compliers.

$$\text{AIE} = \mathbb{E}\left[D_i(1) \neq D_i(0)\right] \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \,\middle|\, D_i(1) \neq D_i(0)\right].$$

Outline of MTE approach to identifying AIE:

1. Mediator monotonicity has a Control Function for D_i (Vycatil 2002).

$$D_i(z') = \mathbb{1}\left\{\mu(z'; \boldsymbol{X}_i) \geq U_i\right\}$$

2. Identify Marginal Indirect Effect (MIE), with instrument by LIV.

$$\mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \,\middle|\, U_i = u'\right]$$

3. AIE among compliers is an integral of the MIE (Mogstad Santos Torgovitsky, 2017).

$$\int \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \,\middle|\, U_i = u'\right] dF_W(u'),$$

$$\text{for } W = \left\{ i \,\middle|\, D_i(1) = 1, D_i(0) = 0 \right\}.$$