Causal Mediation in Natural Experiments

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

Labour Work in Progress Seminar
6 March 2025

Introduction

Ever read an epidemiology/psychology/medicine paper's abstract, and seen claims of causal effects **mediated** through some mechanism?

Family communication patterns, family environment, and the impact of parental alcoholism on offspring self-esteem [PDF] sagep

S Rangarajan, L Kelly - Journal of Social and Personal ..., 2006 - journals.sagepub.com

This study examined the role of perceptions of family environment and family communication as mediators of the effects of parental alcoholism on the self-esteem of adult children of alcoholics. Participants (N= 227) completed self-reports of parental alcoholism, family environment, family communication patterns (FCP), and self-esteem. Results indicated a negative relationship between the seriousness of both maternal and paternal alcoholism and self-esteem. Paternal and maternal alcoholism were related to the two dimensions of ...

☆ Save 99 Cite Cited by 122 Related articles All 3 versions

Introduction

Ever read an epidemiology/psychology/medicine paper's abstract, and seen claims of causal effects **mediated** through some mechanism?

[HTML] Persistent depressive symptomatology and inflammation: to what extent do health behaviours and weight control mediate this relationship?

[HTML] scie

M Hamer, GJ Molloy, C de Oliveira... - Brain, Behavior, and ..., 2009 - Elsevier

We examined if persistent depressive symptoms are associated with markers of inflammation (C-Reactive Protein-CRP) and coagulation (fibrinogen), and if this association can be partly explained by weight control and behavioural risk factors (smoking, alcohol, physical activity). The study sample included 3609 men and women (aged 60.5±9.2 years) from The English Longitudinal Study of Ageing, a prospective study of community dwelling older adults. Depressive symptoms (using the 8-item CES-D scale), health behaviours ...

☆ Save 99 Cite Cited by 111 Related articles All 6 versions

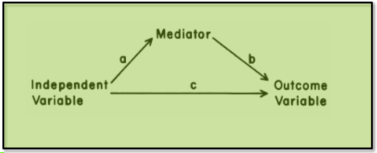
Introduction

Figure: Baron Kelly (1986), p. 1176.

1176 REUBEN M. BARON AND DAVID A. KENNY

gression equasion, as described by Cohen and Cohen (1983) and Cleary and Kessler (1982). So if the independent variable is denoted as X, the moderator as Z, and the dependent variable as Y, Y is regressed on X, Z, and XZ. Moderator effects are indicated by the significant effect of XZ while X and Z are controlled. The simple effects of the independent variable for different levels of the moderator can be measured and tested by procedures described by Aiken and West (1986). (Measurement error in the moderator requires the same remedies as measurement error in the independent variable under Case 2.)

The quadratic moderation effect can be tested by dichotomizing the moderator at the point at which the function is presumed to accelerate. If the function is quadratic, as in Figure 2, the effect of the independent variable should be greatest for those who are high on the moderator. Alternatively, quadratic moderation can be tested by hierarchical regression procedures described by Cohen and Cohen (1983). Using the same notation as in the previous paragraph, Y is regressed on X, Z, XZ, Z^2, and XZ^2. The test of quadratic moderation is given by the test



model, which recognizes that an active organism intervenes between stimulus and response, is perhaps the most generic formulation of a mediation hypothesis. The central idea in this model is that the effects of stimuli on behavior are mediated by various transformation processes internal to the organism. Theorists as diverse as Hull, Tolman, and Lewin shared a belief in the importance of postulating entities or processes that intervene between input and output. (Skinner's blackbox approach represents the notable exception.)

1980s: Psychometrics defined mediation (distinct from moderation).
Application of early econometric path analysis (Wright 1928).

2020s: Popular in epidemiology, medicine, psychology.

Introduction

Figure: Imai Keele Tingley Yamamoto (2010–).

A general approach to causal **mediation** analysis.
K **Imai**, L Keele, D Tingley - Psychological methods, 2010 - psycnet.apa.org
… Following prior work (eg, **Imai**, Keele, & Yamamoto, 2010; Pearl, 2001; Robins & Greenland…
mediation effects using the potential outcomes notation. We then review the key result of **Imai**…
☆ Save 99 Cite Cited by 3907 Related articles All 24 versions

Identification, inference and sensitivity analysis for causal mediation effects
K Imai, L Keele, T Yamamoto - 2010 - projecteuclid.org
Causal mediation analysis is routinely conducted by applied researchers in a variety of disciplines. The goal of such an analysis is to investigate alternative causal mechanisms by …
☆ Save 99 Cite Cited by 2078 Related articles

Mediation: R package for causal mediation analysis
D Tingley, T Yamamoto, K Hirose, L Keele… - Journal of statistical …, 2014 - jstatsoft.org
In this paper, we describe the R package mediation for conducting causal mediation analysis in applied empirical research. In many scientific disciplines, the goal of researchers …
☆ Save 99 Cite Cited by 4048 Related articles

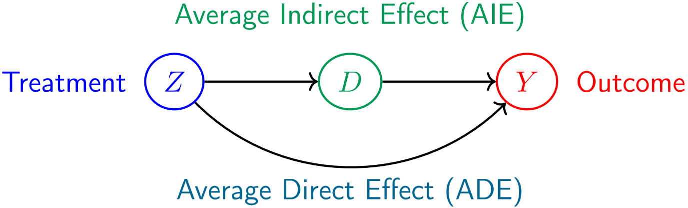
1980s: Psychometrics defined mediation (distinct from moderation).
Application of early econometric path analysis (Wright 1928).

2020s: Popular in epidemiology, medicine, psychology.

Introduction:

1. **[familiar]** Causal design to estimate a treatment effect.

2. **[unfamiliar]** CM decomposes ATE along a mechanism pathway.



	ATE	$\Longrightarrow$	Average causal effect $Z \rightarrow Y$
3.	**AIE**	$\Longrightarrow$	How much $Z \rightarrow Y$ effect through mediator D?
	ADE	$\Longrightarrow$	How much $Z \rightarrow Y$ effect is left over?

Introduction— CM Examples:

1. Lottery military draft 1969 (Angrist 1990).

Drafted Military Service Earnings

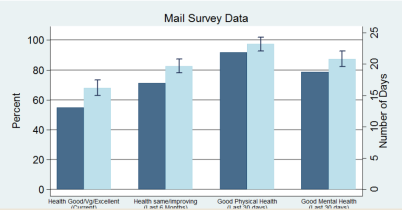
$$\begin{tikzcd} Z \arrow[r] \arrow[rr, bend right] & D \arrow[r] & Y \end{tikzcd}$$

Draft avoidance (education deferment)

Note: IV further assumes direct $= 0$ (exclusion restriction).

2. Oregon health insurance experiment (Finkelstein+ 2009).

Medicaid Improves Self-Reported Health



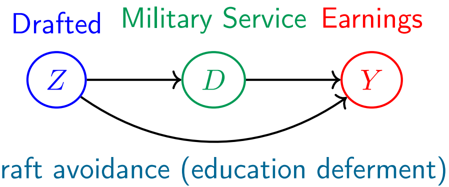
Medicaid Hospital use Health

$$\begin{tikzcd} Z \arrow[r] \arrow[rr, bend right] & D \arrow[r] & Y \end{tikzcd}$$

All else (e.g., less worry)

Introduction— CM Examples:

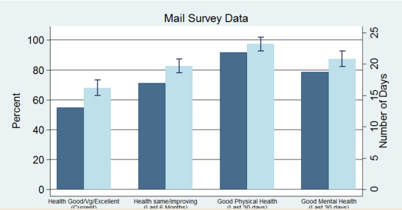
1. Lottery military draft 1969 (Angrist 1990).

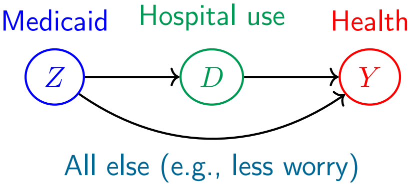


Note: IV further assumes direct $= 0$ (exclusion restriction).

2. Oregon health insurance experiment (Finkelstein+ 2009).

Medicaid Improves Self-Reported Health





Introduction

This project examines CM methods from an economic perspective:

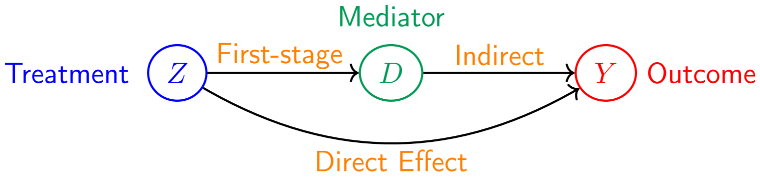
1. Problems with conventional, selection-on-observables, approach to CM in social science settings — including natural experiments.
   [Negative result]
2. Recovering valid CM effects under selection-into-mediator, using a selection model.
   [Positive result]

Brings together ideas from two different literatures:

- **Causal mediation.**
  Baron Kelly (1986), Imai Keele Yamamoto (2010), Flores Flores-Lagunes (2009), Frölich Huber (2017), Huber (2020), Kwon Roth (2024).
- **Labour theory, Selection-into-treatment.**
  Roy (1951), Heckman (1979), Heckman Honoré (1990), Florens Heckman Meghir Vytlacil (2008).

Direct & Indirect Effects — Model

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



Mediator D_i is a function of Z_i

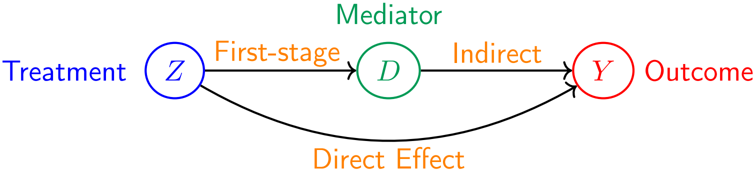
$$D_i = \begin{cases} D_i(0), & \text{if } Z_i = 0 \\ D_i(1), & \text{if } Z_i = 1. \end{cases}$$

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 controls $\boldsymbol{X}_i$.

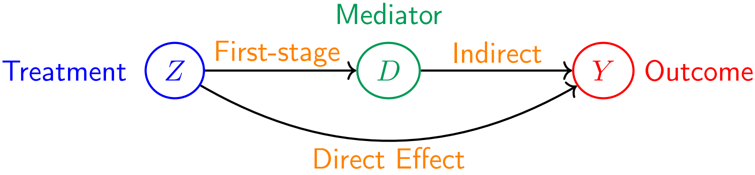
$$Z_i \perp\!\!\!\perp D_i(z), Y_i(z', d') \mid \boldsymbol{X}_i \text{ for } z, z', d' = 0, 1.$$

E.g., a natural experiment for Z_i disrupting open-world selection-into-Z_i

- Vietnam draft lottery for military conscription (Angrist, 1990).
- Oregon wait-list lottery for health insurance (Finkelstein+, 2009).

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 controls $\boldsymbol{X}_i$.

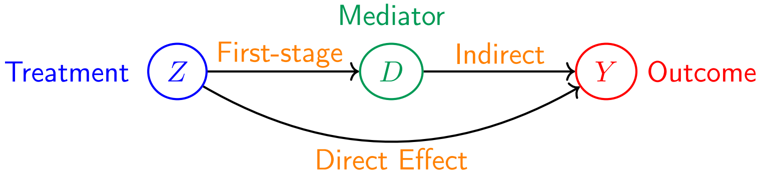
$$Z_i \perp\!\!\!\perp D_i(z), Y_i(z', d') \mid \boldsymbol{X}_i \text{ for } z, z', d' = 0, 1.$$

E.g., a natural experiment for Z_i disrupting open-world selection-into-Z_i

- Vietnam draft lottery for military conscription (Angrist, 1990).
- Oregon wait-list lottery for health insurance (Finkelstein+, 2009).

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 $\boldsymbol{X}_i$.

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

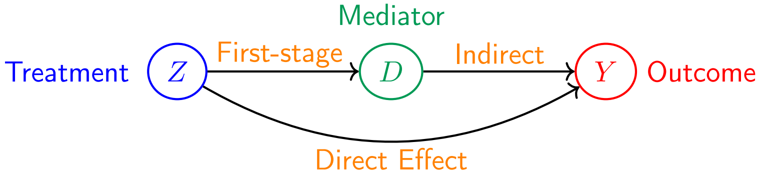
Only two causal effects are identified so far.

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

Direct & Indirect Effects — Model

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



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

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

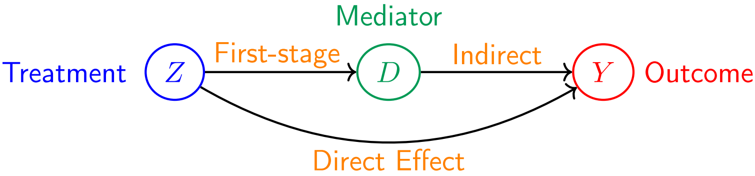
Only two causal effects are identified so far.

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

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:

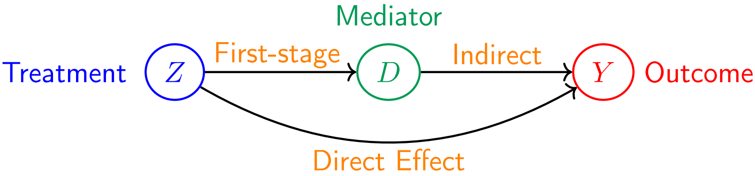
- Were drafted men poorer? Did they conform or dodge? (Angrist, 1990).
- Did socialised health insurance increase hospital use, and improve health? (Finkelstein+, 2009).

Unanswered questions about the mechanism(s):

- Did drafted men take education to dodge the draft? Was it entirely military service? (Angrist, 1990).
- Did health benefits come from using health care more? Health gains from reduced uncertainty, i.e., insurance? (Finkelstein+, 2009).

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:

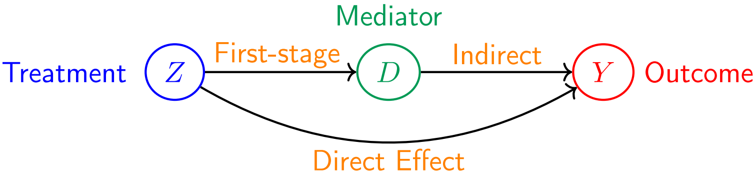
- Were drafted men poorer? Did they conform or dodge? (Angrist, 1990).
- Did socialised health insurance increase hospital use, and improve health? (Finkelstein+, 2009).

Unanswered questions about the mechanism(s):

- Did drafted men take education to dodge the draft? Was it entirely military service? (Angrist, 1990).
- Did health benefits come from using health care more? Health gains from reduced uncertainty, i.e., insurance? (Finkelstein+, 2009).

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:

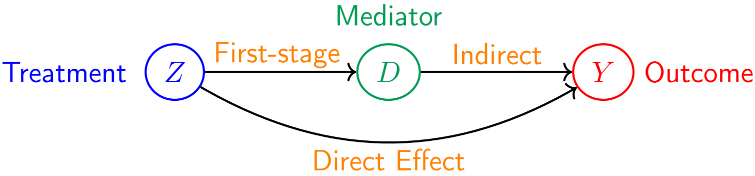
- Were drafted men poorer? Did they conform or dodge? (Angrist, 1990).
- Did socialised health insurance increase hospital use, and improve health? (Finkelstein+, 2009).

Unanswered questions about the mechanism(s):

- Did drafted men take education to dodge the draft? Was it entirely military service? (Angrist, 1990).
- Did health benefits come from using health care more? Health gains from reduced uncertainty — i.e., insurance? (Finkelstein+, 2009).

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

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

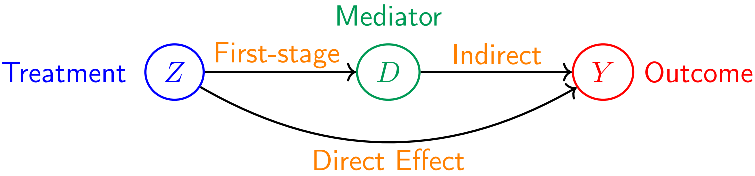
Average Indirect Effect (AIE): $\mathbb{E}\left[Y_i \left(Z_i, \boxed{D_i(1)} \right) - Y_i \left(Z_i, \boxed{D_i(0)} \right) \right]$

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

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

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

- 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.[1]

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

Direct & Indirect Effects — Identification

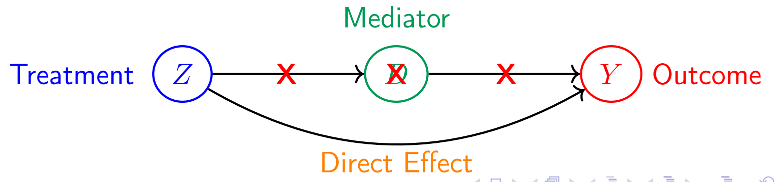
Sequential ignorability (**SI**, Imai Keele Yamamoto 2010):

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:

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



Direct & Indirect Effects — Identification

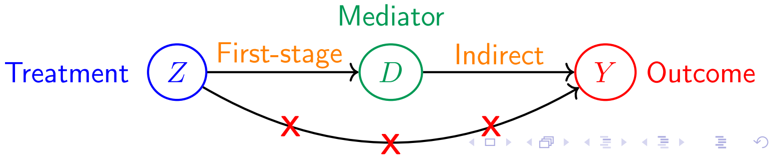
Sequential ignorability (SI, Imai Keele Yamamoto 2010):

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:

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



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

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

$Y_i \leftarrow$ Outcome

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

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

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

$\implies$ 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 $\to$ big gain!

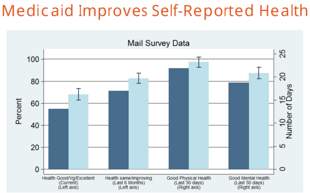
Before we import these methods to applied/labour economics and observational research, interrogate the **SI** assumption.

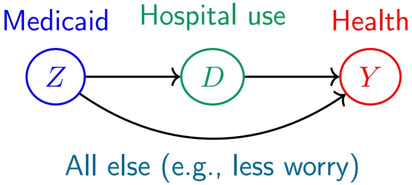
$$D_i \perp\!\!\!\perp Y_i(z', d') \mid \boldsymbol{X}_i, Z_i = z', \text{ for } z', d' = 0, 1.$$

Would this assumption hold true in settings economists study?

Direct & Indirect Effects — Selection

Oregon health insurance experiment (Finkelstein+ 2009).





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. Medicaid assigned randomly (ensured by studying the 2008 Oregon wait-list lottery).
2. Hospital usage is quasi-random, conditional on Medicaid assignment Z_i and demographics $\boldsymbol{X}_i$.

Direct & Indirect Effects — Selection

SI: Hospital usage is quasi-random, conditional on Medicaid assignment Z_i and demographics $\boldsymbol{X}_i$.

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

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

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: Hospital usage is quasi-random, conditional on Medicaid assignment Z_i and demographics $\boldsymbol{X}_i$.

Consider the case **individuals go to the hospital** 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

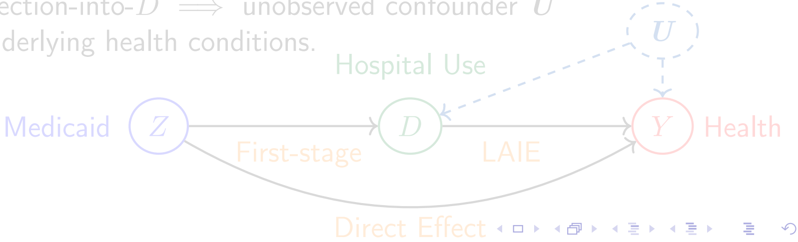
SI: Hospital usage is quasi-random, conditional on Medicaid assignment Z_i and demographics $\boldsymbol{X}_i$.

Consider the case **individuals go to the hospital** 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

SI: Hospital usage is quasi-random, conditional on Medicaid assignment Z_i and demographics $\boldsymbol{X}_i$.

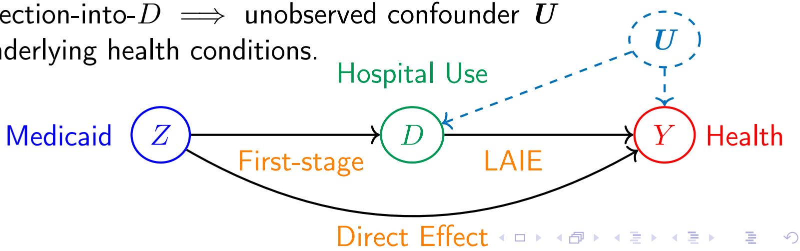
Consider the case **individuals go to the hospital** to maximise health.

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

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

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

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

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

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

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

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

▸ Group-diff

Direct & Indirect Effects — Selection Bias

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

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

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

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

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

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

▸ Group-diff

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).

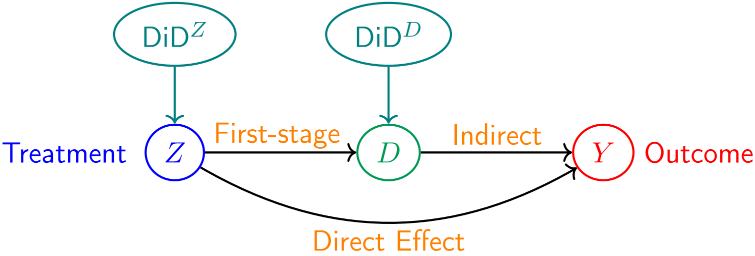
Note: assumes common trends across complier groups, identifies ADE + AIE local to complier groups.

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 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).

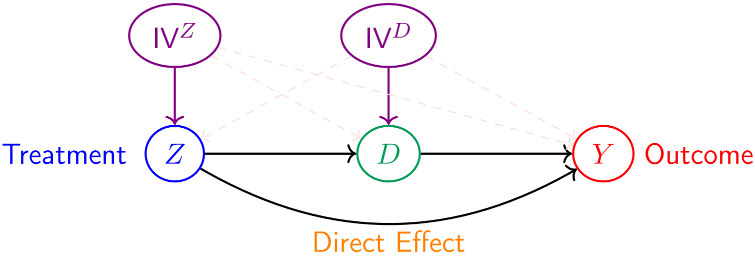
Note: two-way exclusion restriction, identifies ADE + AIE local to overlapping complier groups. Also avoid 2SLS (see Kim 2025)!

Identification Under Selection

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

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

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



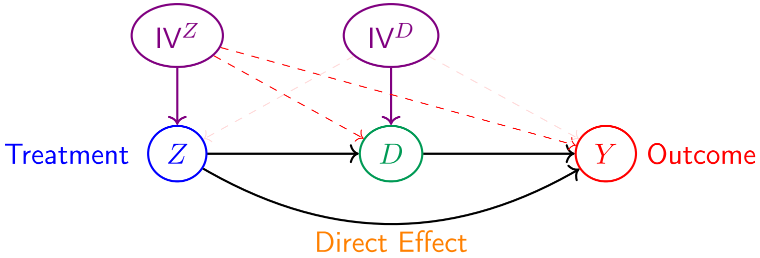
Note: two-way exclusion restriction, identifies ADE + AIE local to overlapping complier groups. Also avoid 2SLS (see Kim 2025)!

Identification Under Selection

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

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

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



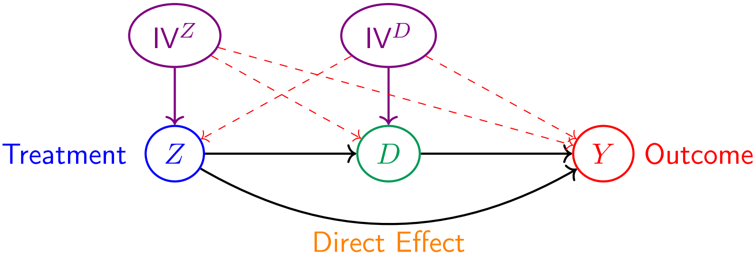
Note: two-way exclusion restriction, identifies ADE + AIE local to over-lapping complier groups. Also avoid 2SLS (see Kim 2025)!

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?

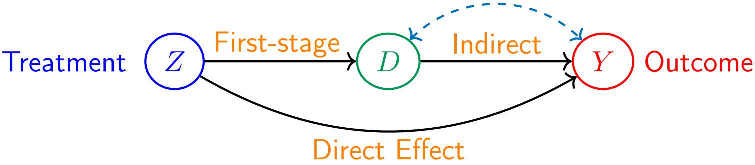
1. Estimate the ATE, and call it a day.
2. (optional) Present suggestive evidence of mechanisms. . . . ▸ Suggestive

---

New: selection model solution to identification.

Identification with a Selection Model

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:

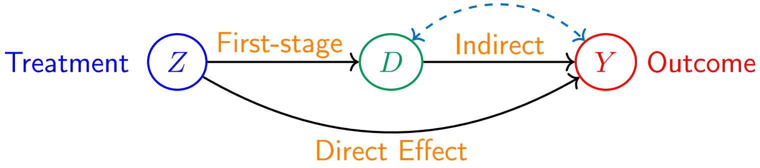
$$D_i = \phi + \pi Z_i + \varphi(\boldsymbol{X}_i) + U_i$$

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

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

Identification with a Selection Model

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



Then this system has the following regression equations:

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

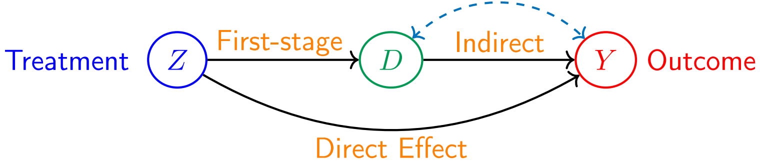
$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \zeta(\boldsymbol{X}_i) + \underbrace{(1 - D_i)\, U_{0,i} + D_i U_{1,i}}_{\text{Correlated error term.}}$$

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

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

Identification with a Selection Model

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



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

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

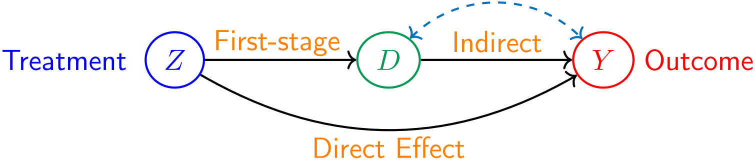
$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \zeta(\boldsymbol{X}_i) + \underbrace{(1-D_i)\,U_{0,i} + D_i U_{1,i}}_{\text{Correlated error term}}$$

where C_i are costs of taking D_i.

Selection model intuition: use first-stage errors to purge second-stage correlated errors.

Identification with a Selection Model

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



Heckman (1979) Selection model, 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,

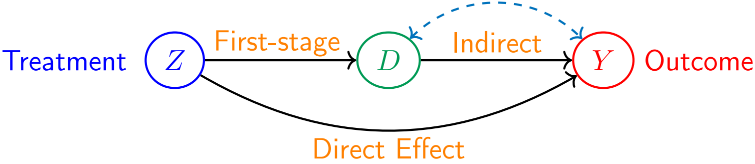
$$\text{Cov}\left[U_i, (1 - D_i)\, U_{0,i} + D_i U_{1,i}\right] \neq 0.$$

- 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 Selection Model

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

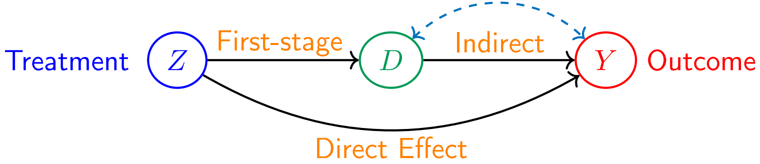


Heckman (1979) Selection model, in operation:

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 Selection Model

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



Semi-parametric control function (Newey Imbens 2009), 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,

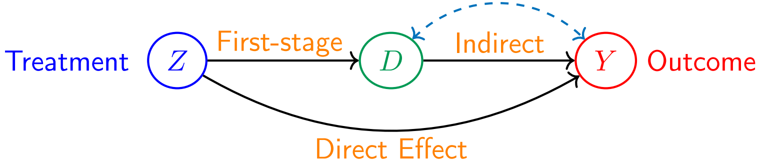
$$\text{Cov}\left[U_i, (1 - D_i)\, U_{0,i} + D_i U_{1,i}\right] \neq 0.$$

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 Selection Model

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



Semi-parametric control function (Newey Imbens 2009), 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,
$$\widehat{\text{ADE}} = \widehat{\gamma} + \widehat{\delta}\,\mathbb{E}\left[D_i\right], \quad \widehat{\text{AIE}} = \widehat{\pi}\left(\widehat{\beta} + \widehat{\delta}\,\mathbb{E}\left[Z_i\right] + \mathbb{E}\left[\widehat{\lambda}_0\left(\widehat{K}_i\right) - \widehat{\lambda}_1\left(-\widehat{K}_i\right) \right] \right).$$

Simulation Evidence

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

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 X_i^{IV} is an additive cost IV:

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

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

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

Simulation Evidence

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

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 X_i^{IV} is an additive cost IV:

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

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

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

Simulation Evidence

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

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:

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

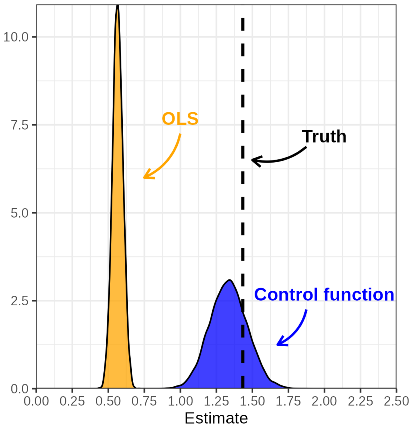
$\implies$ unobserved confounding by BivariateNormal $\left(U_{0,i}, U_{1,i}\right)$.

Simulation Evidence

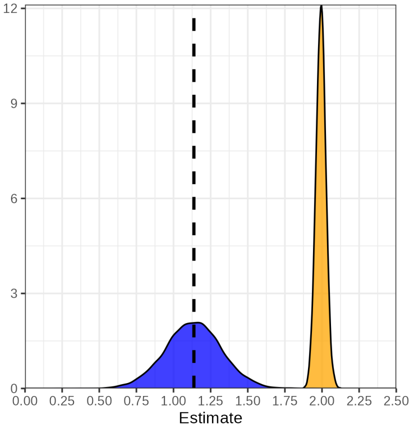
Simulation with Roy selection, BivariateNormal errors + unobserved costs.

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

(a) ADE.



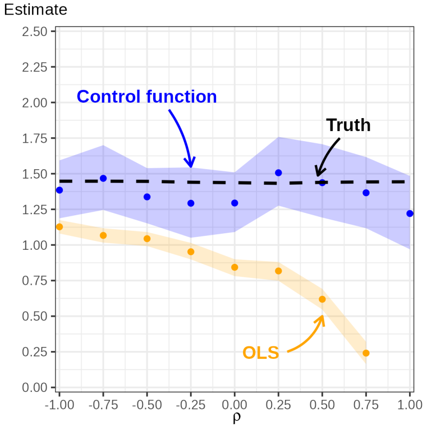
(b) AIE.



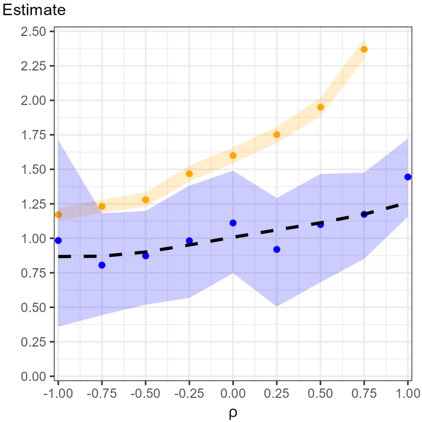
Simulation Evidence

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.

Conclusion

Overarching goals:

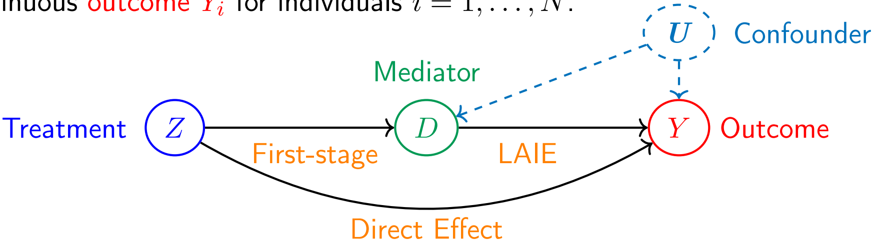
1. Ward economists away from using CM methods unabashedly.
   → Noted problems in the most popular methods for CM effects, pertinent for economic applications.
2. CM methods away from ignorability assumptions, inappropriate for economics (+ social science) settings.
   → Methods valid when selection-into-treatment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

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

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

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

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

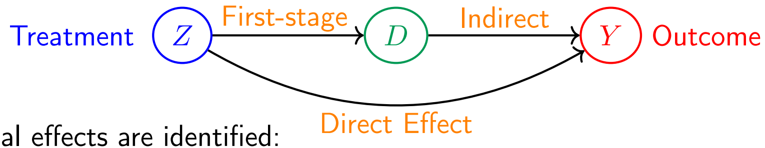
$$+ \underbrace{\pi\left[\left(\frac{1 - \Pr\left(D_i(1) = 1, D_i(0) = 0\right)}{\Pr\left(D_i(1) = 1, D_i(0) = 0\right)}\right)\left(\begin{aligned}&\mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0) \mid D_i(1) = 0 \text{ or } D_i(0)\right. \\ &- \mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0)\right]\end{aligned}\right.}_{\text{Groups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

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 $\to$ 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 selection model/CF, but AIE refers to mediator benefits only among mediator compliers.

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

Outline of MTE approach to identifying AIE:

1. Mediator monotonicity has a selection model 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).

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

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