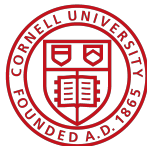


# Defence, PhD in Economics

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



---

ILR Doherty Lounge, Cornell University  
29 June 2026

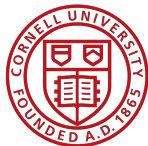
# About Me

I am an economics PhD student at Cornell University, with research interests in labour economics and applied econometrics.

1. Ok.

# Causal Mediation in Natural Experiments

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



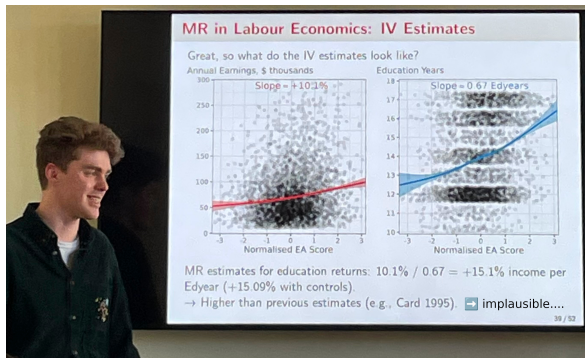
---

Cornell Economics, PhD Defence  
29 June 2026



## About Me

I am an economics PhD student at Cornell University, with research interests in labour economics and applied econometrics.



- Other applied research studying education polygenic index, and mediation of its causal effect through education.

# Introduction

Natural experiments are settings with credible estimates of causal effects

- Little information on the **mechanisms** through which they operate
- Limits understanding of the decisions and underlying economic system
- Causal Mediation (CM) is a framework for sufficiently analysing a causal effect along an observed mechanism, which is not widely used in applied economics.

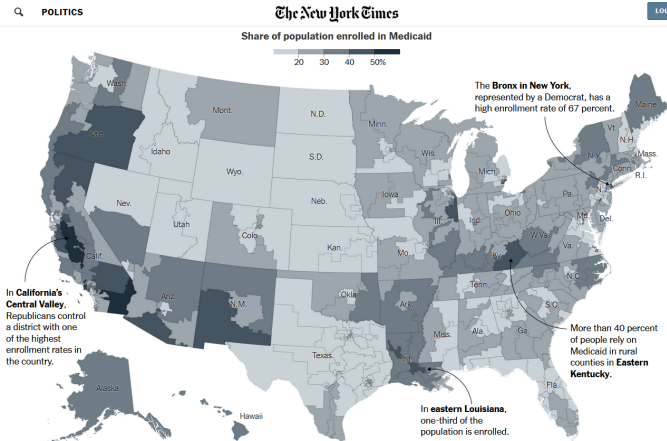
---

## This paper:

- ① Develop selection bias concept for CM in observational settings, with reasoning bias can crowd out inference
- ② Build an MTE-based model to tackle the identification problem
- ③ Illustrate my methods with decomposing causal effects in the Oregon Health Insurance Experiment.

# Oregon Health Insurance Experiment

In the USA, healthcare is only provided by the government in special cases  
→ Medicaid is the government programme which provides health insurance for those close to the poverty line (> 70 million people in 2025).



# Oregon Health Insurance Experiment

In the USA, healthcare is only provided by the government in special cases  
→ Medicaid is the government programme which provides health insurance for those close to the poverty line (> 70 million people in 2025).

2004: Oregon Medicaid enrolment frozen

2008: Enrolment reopens...  
90,000 sign up (far exceeding funding allocated)

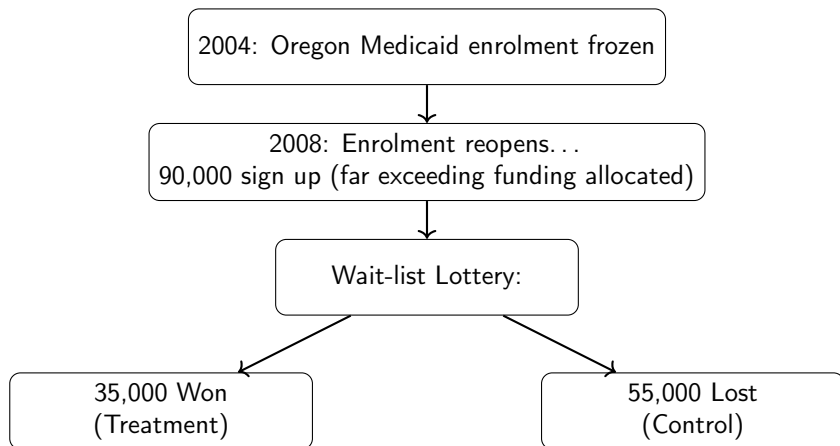
Wait-list Lottery:

35,000 Won  
(Treatment)

55,000 Lost  
(Control)

# Oregon Health Insurance Experiment

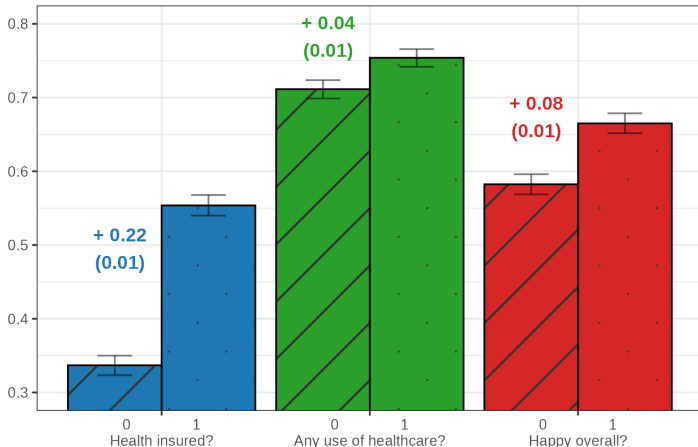
In the USA, healthcare is only provided by the government in special cases  
→ Medicaid is the government programme which provides health insurance for those close to the poverty line (> 70 million people in 2025).



# Oregon Health Insurance Experiment

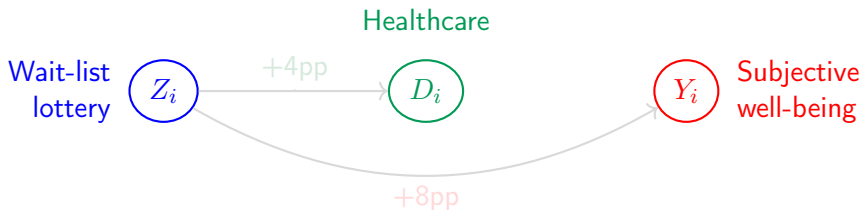
Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).

Mean Outcome, winning or losing the wait-list lottery.



## Oregon — Suggestive Evidence

Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



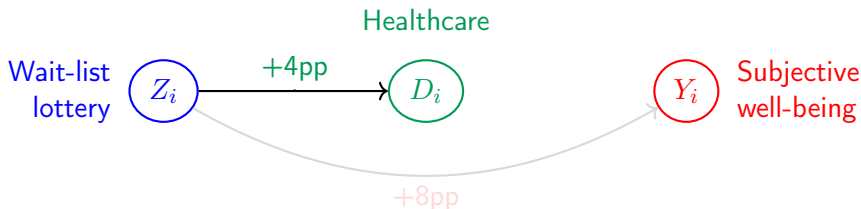
---

Suggestive evidence:

- If first-stage  $\neq 0$ , then healthcare may be a mediating mechanism
- This gives suggestive evidence for healthcare as mechanism.

## Oregon — Suggestive Evidence

Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



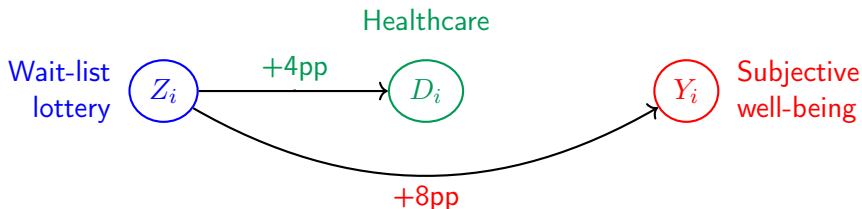
---

Suggestive evidence:

- If first-stage  $\neq 0$ , then healthcare may be a mediating mechanism
- This gives suggestive evidence for healthcare as mechanism.

## Oregon — Suggestive Evidence

Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



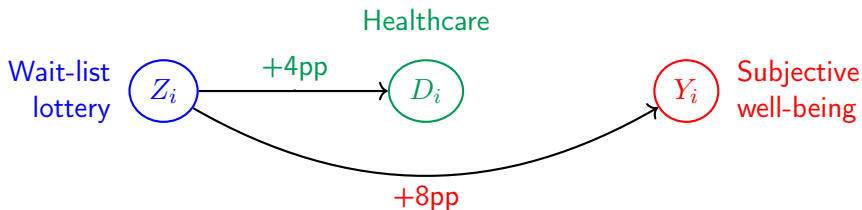
---

Suggestive evidence:

- If first-stage  $\neq 0$ , then healthcare may be a mediating mechanism
- This gives suggestive evidence for healthcare as mechanism.

## Oregon — Suggestive Evidence

Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



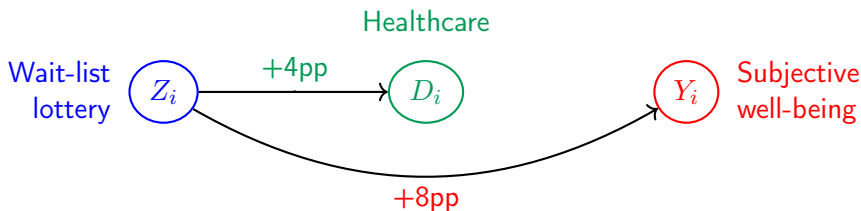

---

### Suggestive evidence:

- If first-stage  $\neq 0$ , then **healthcare** may be a mediating mechanism
- This gives **suggestive evidence for healthcare as mechanism.**

## Oregon — Suggestive Evidence

Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



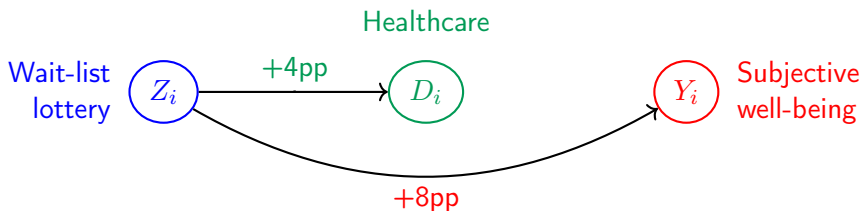
---

What about direct effects?

- Winning access to Medicaid means you can file for free health insurance (income effect)
- Less stress from no longer having to be uninsured (psychological gains).

## Oregon — Suggestive Evidence

Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).

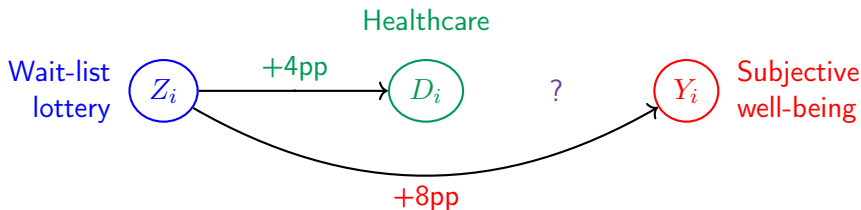


### What about direct effects?

- Winning access to Medicaid means you can file for free health insurance (income effect)
- Less stress from no longer having to be uninsured (psychological gains).

## Oregon — Suggestive Evidence

Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



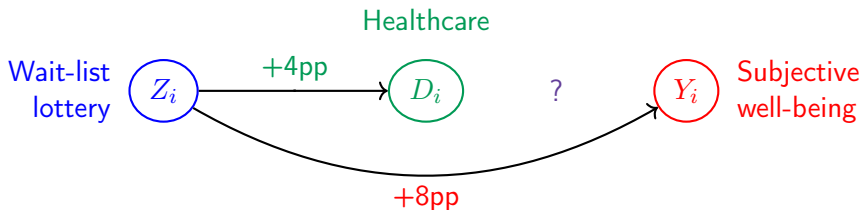
There is one missing piece to make a **definitive conclusion**:

Size of causal effect  $D_i \rightarrow Y_i \dots$

- If large, then healthcare explains all the lottery effect
- If small/zero then, then all direct (e.g., psychological) gains.

## Oregon — Suggestive Evidence

Winning this wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



There is one missing piece to make a **definitive conclusion**:

Size of causal effect  $D_i \rightarrow Y_i \dots$

- If large, then **healthcare** explains all the lottery effect
- If small/zero then, then all **direct** (e.g., psychological) gains.

## Causal Mediation (CM)

CM is an alternative framework to studying mechanisms, giving sufficient evidence on the mediating mechanism.



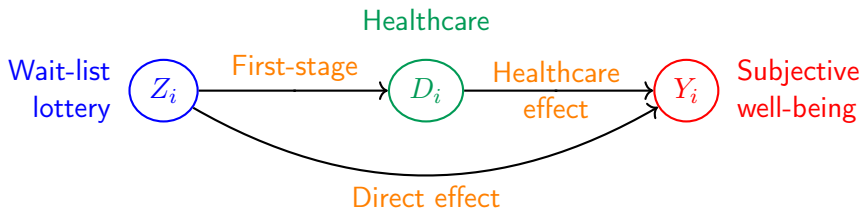
---

CM aims to decompose the ATE in two channels, direct and indirect effects

$$\text{ATE} = \text{ADE} + \text{AIE}.$$

## Causal Mediation (CM)

CM is an alternative framework to studying mechanisms, giving sufficient evidence on the mediating mechanism.

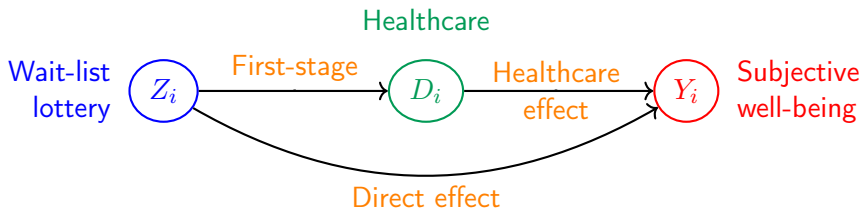


CM decomposes the ATE into two channels, direct and indirect effects

$$ATE = ADE + AIE.$$

## Causal Mediation (CM)

CM is an alternative framework to studying mechanisms, giving sufficient evidence on the mediating mechanism.



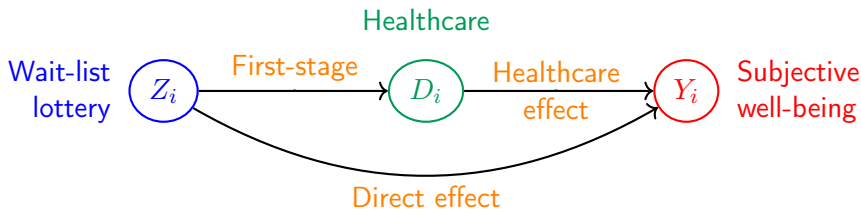
CM decomposes the ATE into

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

AIE represents the average effect going through healthcare.

## Causal Mediation (CM)

CM is an alternative framework to studying mechanisms, giving sufficient evidence on the mediating mechanism.



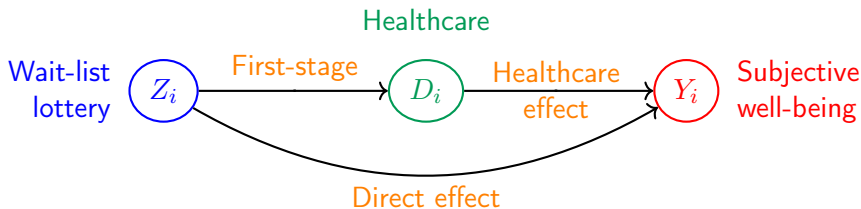
CM decomposes the ATE into

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

ADE represents the average effect going **absent healthcare**.

## Causal Mediation (CM)

Identification of the ADE + AIE needs further assumptions.



Conventional CM relies on two identifying assumptions,

- 1 Treatment  $Z_i$  is (quasi-)randomly assigned
- 2 Mediator  $D_i$  is (quasi-)randomly assigned, conditional on  $Z_i$  realisation and controls  $\mathbf{X}_i$ .

⇒ ADE + AIE are separately identified by two-stage regression (Imai Keele Yamamoto 2010).

## Causal Mediation (CM)

This approach (conventional CM) is used heavily in epidemiology and medicine to give evidence for the channels of a treatment effect, but there is a reason why this is not prominent in economics.

---

### Identifying assumptions:

- 1 Treatment  $Z_i$  is (quasi-)randomly assigned
- 2 Mediator  $D_i$  is (quasi-)randomly assigned, conditional on  $Z_i$  realisation (and covariates  $\mathbf{X}_i$ ).

**Translation:** Healthcare is a quasi-random choice, conditional on wait-list lottery realisation and demographic controls.

---

Would this be plausible in settings economists study?

## Causal Mediation (CM) — Roy Model

Consider the case that people, after the lottery, choose to **visit the doctor in the next 12 months** based on subjective costs and benefits,

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

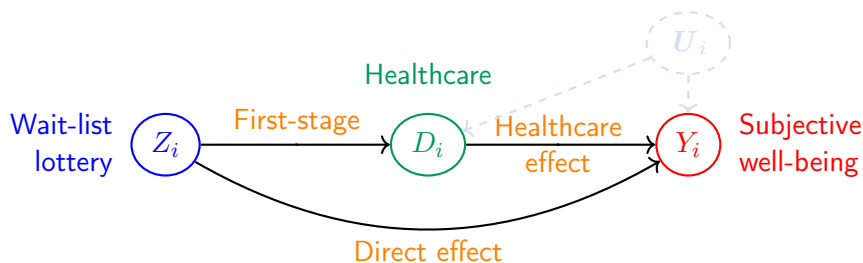
The **wait-list lottery** has no strategic selection, but then **visiting healthcare** is a choice.

---

**Theorem:** If choice to attend healthcare is based on costs and benefits (Roy model) and control variables do not explain all benefits  $\implies$  **mediator mechanism** is not random, there is unobserved confounding.

## Causal Mediation (CM) — Selection Bias

Individual unobserved benefits are an unobserved confounder  $U_i$  here,



In economic settings, Conventional CM analyses have bias similar to classical selection bias (Heckman Ichimura Smith Todd 1998).

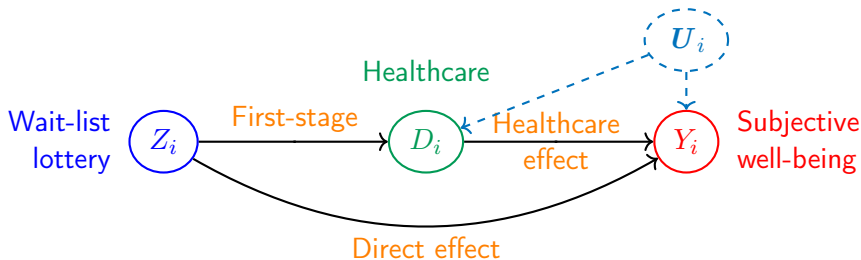
- Direct: CM Estimand = ADE + (Selection Bias + Group difference bias)
- Indirect: CM Estimand = AIE + (Selection Bias + Group difference bias)

▶ ADE biases

▶ AIE biases

## Causal Mediation (CM) — Selection Bias

Individual unobserved benefits are an unobserved confounder  $U_i$  here,



In economic settings, Conventional CM analyses have bias similar to classical selection bias (Heckman Ichimura Smith Todd 1998).

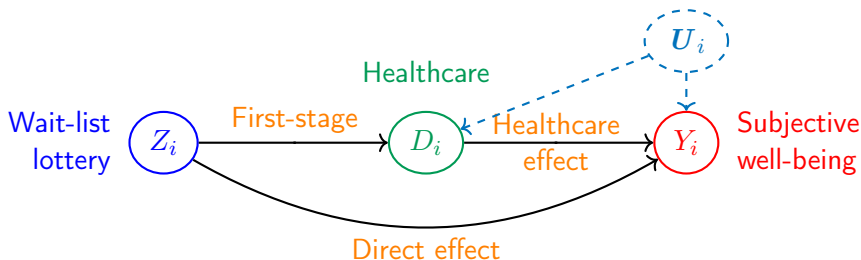
- Direct: CM Estimand = ADE + (Selection Bias + Group difference bias)
- Indirect: CM Estimand = AIE + (Selection Bias + Group difference bias)

▶ ADE biases

▶ AIE biases

## Causal Mediation (CM) — Selection Bias

Individual unobserved benefits are an unobserved confounder  $U_i$  here,



In economic settings, Conventional CM analyses have bias similar to classical selection bias (Heckman Ichimura Smith Todd 1998).

- Direct:  $\text{CM Estimand} = \text{ADE} + (\text{Selection Bias} + \text{Group difference bias})$
- Indirect:  $\text{CM Estimand} = \text{AIE} + (\text{Selection Bias} + \text{Group difference bias})$

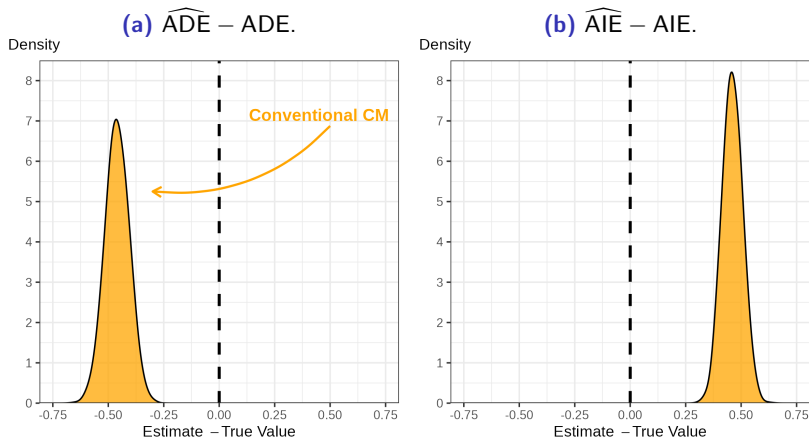
▶ ADE biases

▶ AIE biases

# Causal Mediation (CM) — Selection Bias

With strategic selection, the bias terms can be large and mislead inference on how much goes through the mediating channel.

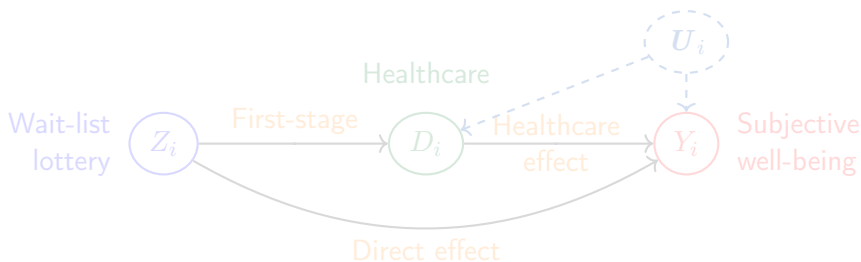
**Figure:** Simulated Distribution from 10,000 simulated datasets.



## CM with Selection

Conventional CM does not identify ADE + AIE in economic settings, so I build a structural model for natural experiment settings.

Take as given that  $Z_i$  is quasi-randomly assigned, but  $D_i$  is not:

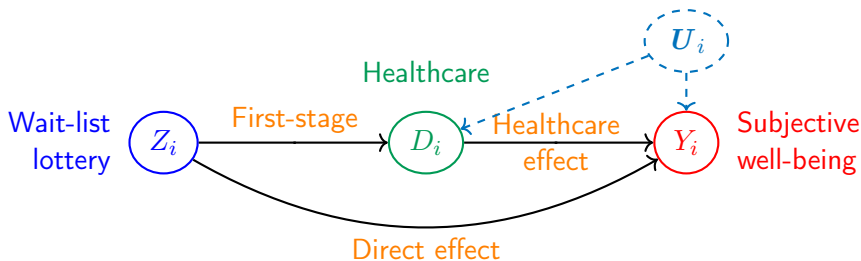


Intuition: model  $U_i$  via healthcare MTE to identify ADE + AIE.

## CM with Selection

Conventional CM does not identify ADE + AIE in economic settings, so I build a structural model for natural experiment settings.

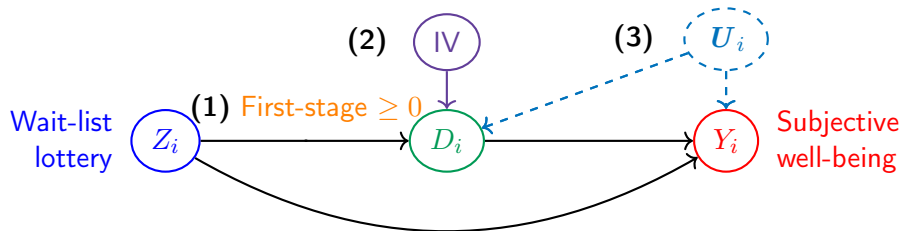
Take as given that  $Z_i$  is quasi-randomly assigned, but  $D_i$  is not:



**Intuition:** model  $U_i$  via healthcare MTE to identify ADE + AIE.

# MTE Model

The structural model is based on 3 assumptions.

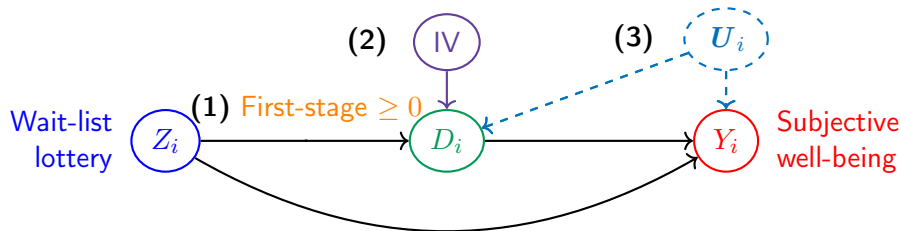


(1) First-stage monotonicity,

**Intuition:** No defiers — no one visits healthcare less if winning wait-list lottery, relative to losing.

# MTE Model

The structural model is based on 3 assumptions.



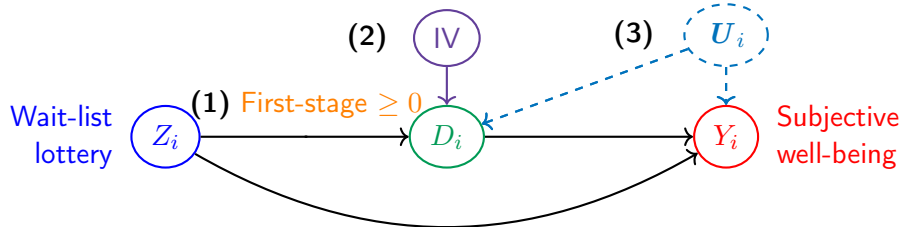
(1) First-stage monotonicity,

**Intuition:** No defiers — no one visits **healthcare** less if winning **wait-list lottery**, relative to losing.

# MTE Model

The structural model is based on 3 assumptions.

---



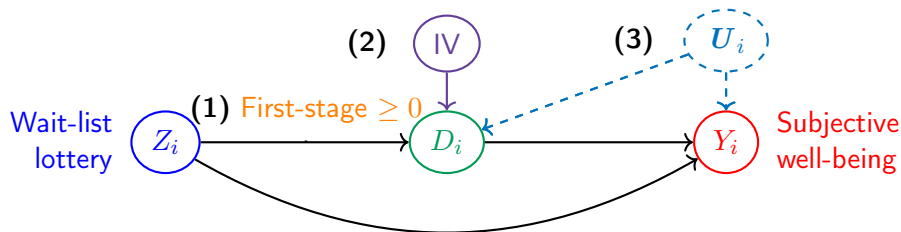
(2) Mediator take-up cost  $IV$

Requires an  $IV$ , which affects  $Y_i$  only via  $D_i$ .

**Key example:** Cost-shifting  $IV$  — random variation in **healthcare take-up** (not gains), e.g. **different healthcare costs**.

# MTE Model

The structural model is based on 3 assumptions.

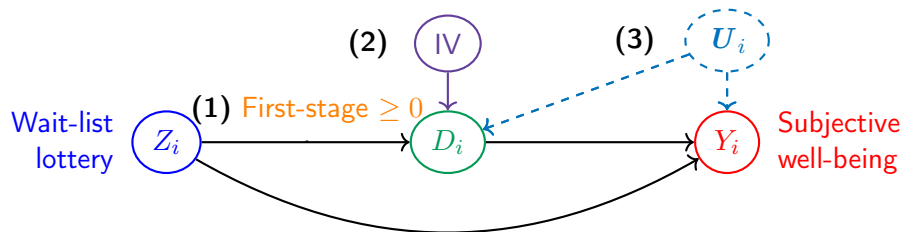


(3) Selection on benefits — unobserved selection in  $D_i$  is relevant

**Key example:** Roy model, people choose to take **healthcare** if internal **subjective gains** exceed costs.

## MTE Model — Identification

The structural model is based on 3 assumptions.



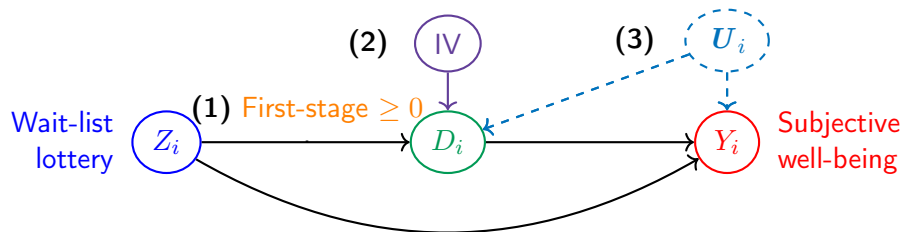
Theorem: Under assumptions (1), (2), (3) ADE + AIE are identified.

ADE Intuition: Control for unobserved confounding via healthcare MTE.

AIE Intuition: Extrapolate indirect effects across healthcare MTE.

## MTE Model — Identification

The structural model is based on 3 assumptions.



**Theorem:** Under assumptions (1), (2), (3) **ADE** + **AIE** are identified.

**ADE Intuition:** Control for unobserved confounding via **healthcare MTE**.

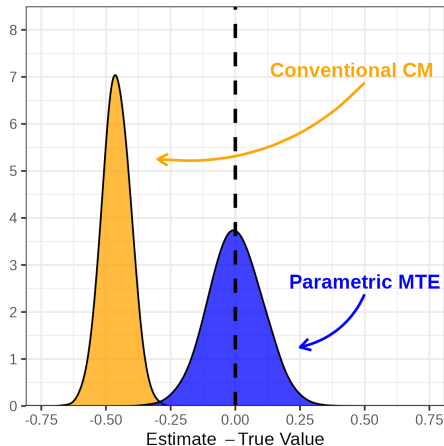
**AIE Intuition:** Extrapolate indirect effects across **healthcare MTE**.

# MTE Model — Estimation

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

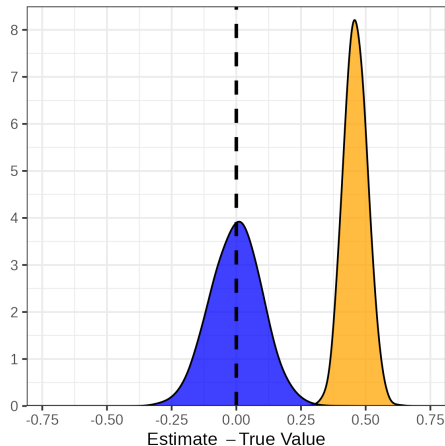
(a)  $\widehat{ADE} - ADE$ .

Density



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

Density



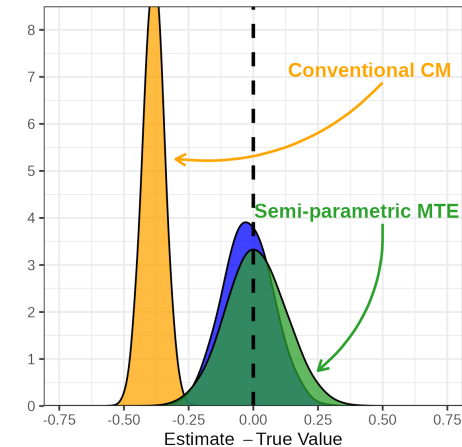
# MTE Model — Estimation

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

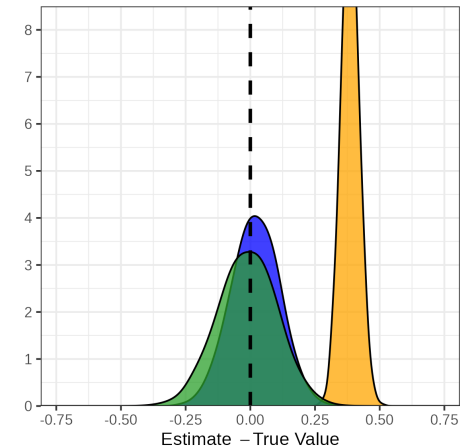
(a)  $\widehat{ADE} - ADE$ .

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

Density



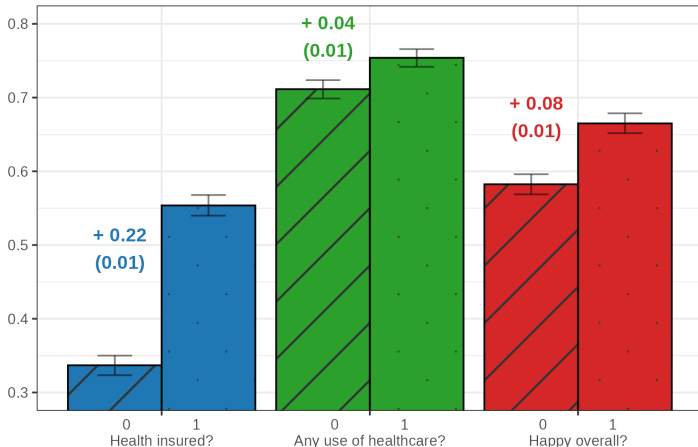
Density



# Return to Oregon

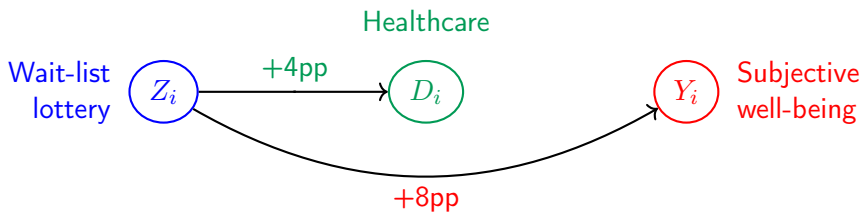
Winning the wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).

Mean Outcome, winning or losing the wait-list lottery.



## Return to Oregon

Winning the wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



Suggestive evidence:

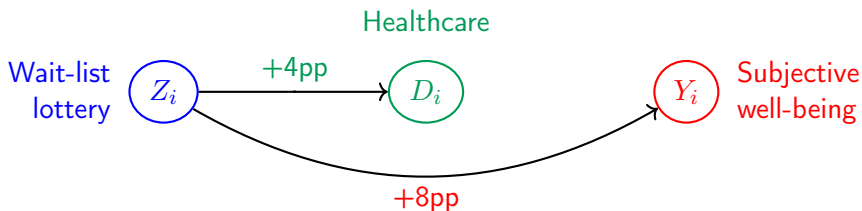
Healthcare is a mechanism.

Plausible direct effects:

Stress reduction and psychological gains.

## Return to Oregon

Winning the wait-list lottery significantly increased healthcare usage, plus subjective well-being (Finkelstein et al, 2012).



Suggestive evidence:

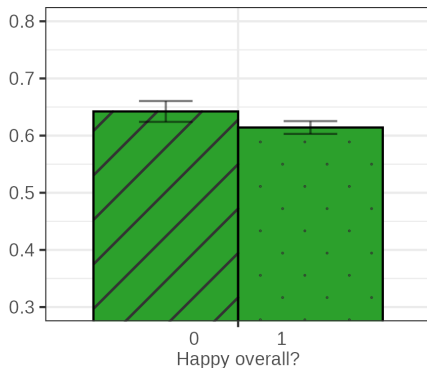
Healthcare is a mechanism.

Plausible direct effects:

Stress reduction and psychological gains.

## Oregon — Conventional CM

Does using healthcare improve subjective well-being?

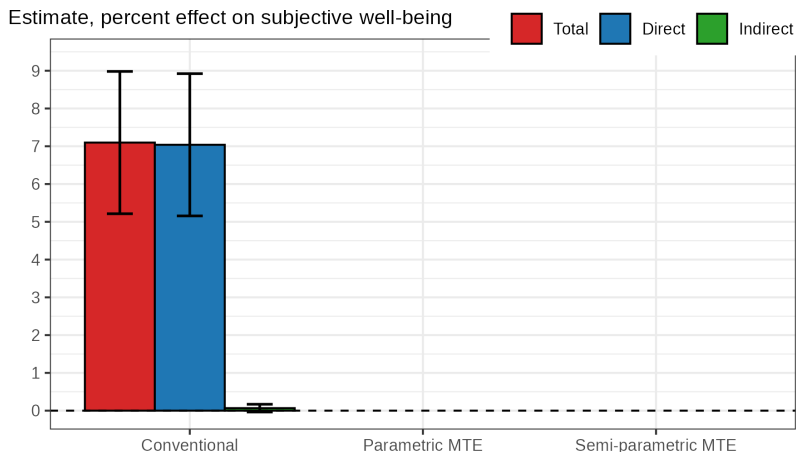


Outcome, attending healthcare or not.

- OLS estimate of healthcare  $D_i \rightarrow Y_i$  subjective well-being is  $-2.5\text{pp}$  (1.1).
- Controls for initial health conditions gives  $+2.8\text{pp}$  (1.1).

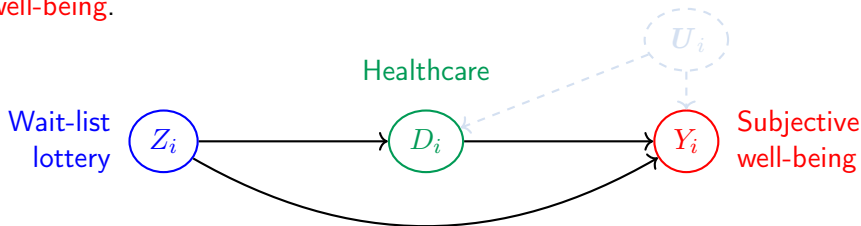
# Oregon — Conventional CM

Conventional CM estimates lottery **subjective well-being** effects as mostly direct,  $\approx 0$  **healthcare**.



## Oregon — Selection Bias

OLS Estimates had negative or little effects of **healthcare** → **subjective well-being**.

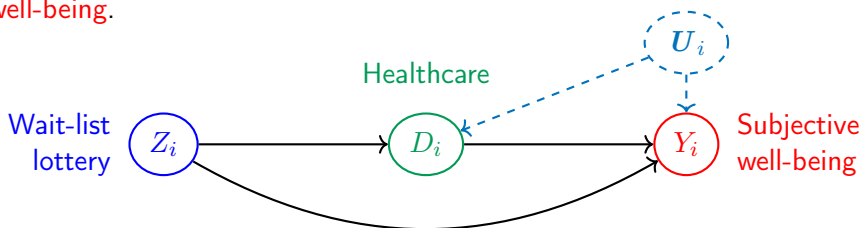


**Example confounder:** undiagnosed conditions among those without health insurance, near or below the poverty line.

**Implication:** negative selection bias in OLS estimates, and Conventional CM underestimated **indirect healthcare channel**.

## Oregon — Selection Bias

OLS Estimates had negative or little effects of healthcare → subjective well-being.



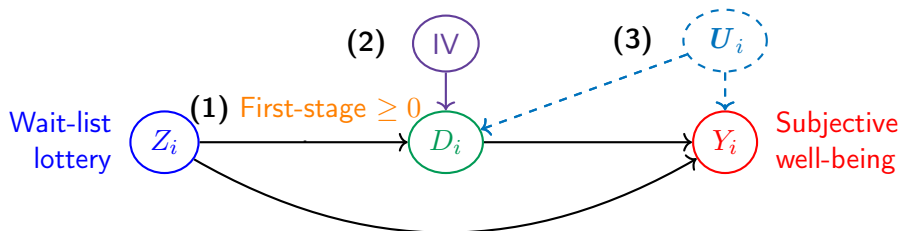
**Example confounder:** undiagnosed conditions among those without health insurance, near or below the poverty line.

**Implication:** negative selection bias in OLS estimates, and Conventional CM underestimated indirect healthcare channel.

## Oregon — MTE Model

I bring the MTE model to these data instead.

---



---

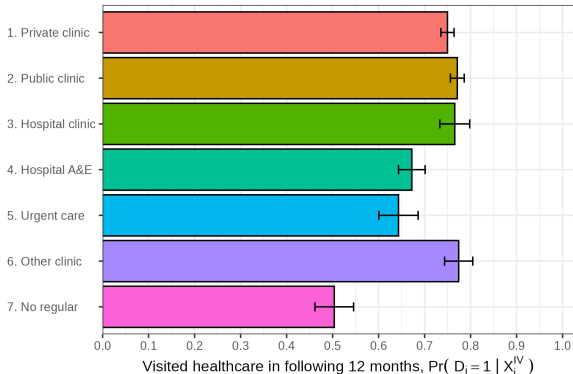
Healthcare IV: pre-lottery healthcare location.

**Intuition:** Different locations charge different prices for similar healthcare. Heading to A&E costs more than a local doctor's office.

# Oregon — MTE Model

Oregon Health Insurance applicants asked pre-lottery **healthcare location**.

Usual Healthcare Location

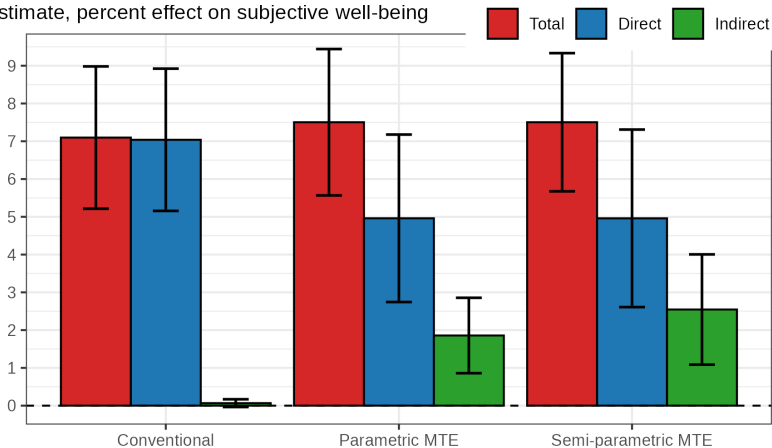


**IV assumption:** where the uninsured, near poverty line, participants visits is indicative of their local healthcare access and cost.

# Oregon — MTE Model

Using MTE approach, with regular healthcare location IV, restores indirect effect through increasing healthcare visitation.

Estimate, percent effect on subjective well-being



# Conclusion

## Overview:

- ① Selection bias in conventional CM analyses when mediator mechanism strategically selected into.
- ② Connect CM with labour theory + selection-into-treatment + MTEs.

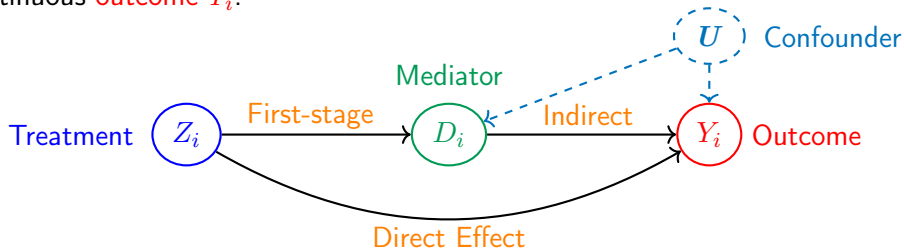
---

## Caveats and points to remember:

- Structural assumptions and IV for identification + estimation (not ideal)
- Application to Oregon Health Insurance Experiment, showing **subjective well-being** effects mediated by **healthcare take-up**
- **Credible** analyses of mechanisms are hard in practice, and wide confidence intervals show true uncertainty.

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

- ADE is causal effect  $Z \rightarrow Y$ , blocking the indirect  $D_i$  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  $D_i(Z_i) \rightarrow Y_i$ , blocking the direct  $Z_i$  path.

## Group Difference — ADE

CM effects contaminated by (less interpretable) bias terms.

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

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

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

⇒ consider this group bias, noting difference from true ADE. [▶ Back](#)

## Selection Bias — Direct Effect

CM Effects + contaminating bias.

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

► Model

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

## Group Difference — AIE

CM effects contaminated by (less interpretable) bias terms.

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

$$\begin{aligned} & \underbrace{\mathbb{E}_{Z_i} \left[ \left( \mathbb{E} [D_i | Z_i = 1] - \mathbb{E} [D_i | Z_i = 0] \right) \times \left( \mathbb{E} [Y_i | Z_i, D_i = 1] - \mathbb{E} [Y_i | Z_i, D_i = 0] \right) \right]}_{\text{Estimand, Indirect Effect}} \\ &= \underbrace{\mathbb{E} [Y_i(Z_i, D_i(1)) - Y_i(Z_i, D_i(0)) | D_i = 1]}_{\text{Average Indirect Effect on Mediated (AIEM) — i.e., } D_i = 1 \text{ weighted}} \\ &+ \underbrace{\pi \left( \mathbb{E} [Y_i(Z_i, 0) | D_i = 1] - \mathbb{E} [Y_i(Z_i, 0) | D_i = 0] \right)}_{\text{Selection Bias}} \\ &+ \underbrace{\pi \left[ \left( \frac{1 - \Pr(D_i(1) = 1, D_i(0) = 0)}{\Pr(D_i(1) = 1, D_i(0) = 0)} \right) \left( \mathbb{E} [Y_i(Z_i, 1) - Y_i(Z_i, 0) | D_i(1) = 0 \text{ or } D_i(0) = 1] - \mathbb{E} [Y_i(Z_i, 1) - Y_i(Z_i, 0)] \right) \right]}_{\text{Groups difference Bias}} \end{aligned}$$

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

→ consider this group bias, noting difference from true AIE.

## Selection Bias — Indirect Effect

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

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

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

Estimand, Indirect Effect

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

Average Indirect Effect

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

Selection Bias

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

Groups difference Bias

Group-diff

## Semi-parametric Control Functions

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

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

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

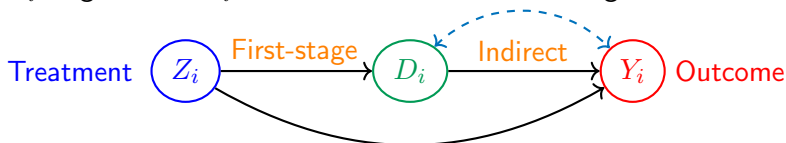
Intercepts,  $\alpha, (\alpha + \beta)$ , and relevance parameters  $\rho_0, \rho_1$  are not separately identified from the CFs  $\lambda_0(\cdot), \lambda_1(\cdot)$  so CF extrapolation term  $(\rho_1 - \rho_0)\Gamma(\pi(0; \mathbf{X}_i), \pi(1; \mathbf{X}_i))$  is not directly identified or estimable.

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

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

## Appendix: CM with Selection

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



Then this system has the following random coefficient equations:

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

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

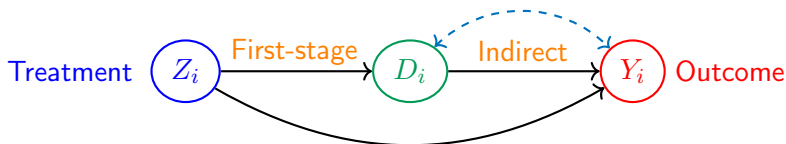
where  $\beta, \gamma, \delta$  are functions of  $\mu_{d'}(z'; \mathbf{X}_i)$ .

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

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

## Appendix: CM with Selection

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



Main problem, second-stage is not identified:

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

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

## Appendix: CM with Selection — Identification

Assume:

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

---

**Proposition:**

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

## Appendix: CM with Selection — Identification

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

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

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

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

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

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

## Appendix: 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.

$$D_i(z') = \mathbb{1} \{C_i \leq Y_i(z', 1) - Y_i(z', 0)\},$$

$$Y_i(z', d') = (z' + d' + z'd') + U_{d'} \quad \text{for } z', d' = 0, 1.$$

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

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

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

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

## Appendix: CM with Selection — Estimation

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

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

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

### Parametric Estimation Recipe:

- 1 Estimate first-stage  $\pi(Z_i; \mathbf{X}_i)$  with probit, including  $\mathbf{X}_i^{IV}$ .
- 2 Include  $\lambda_0, \lambda_1$  CFs in second-stage OLS estimation.
- 3 Compose CM estimates from two-stage plug-in estimates.

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

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

## Appendix: CM with Selection — Estimation

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

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

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

### Semi-parametric Estimation Recipe:

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

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

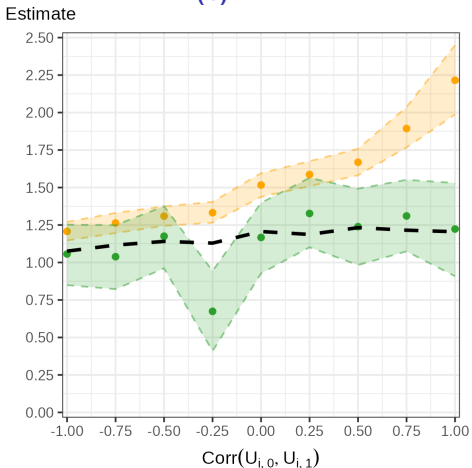
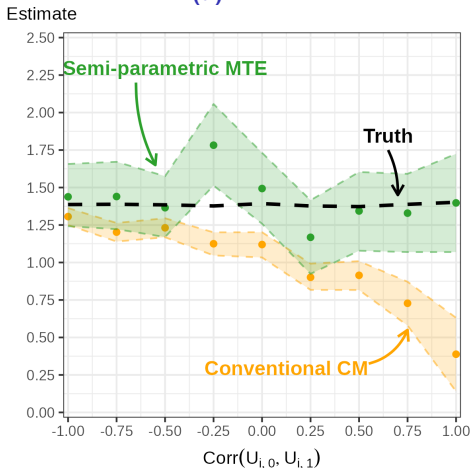
$$\widehat{\text{ADE}} = \mathbb{E}[\widehat{\gamma} + \widehat{\delta} D_i], \quad \widehat{\text{AIE}} = \mathbb{E}\left[\widehat{\pi}\left(\widehat{\beta} + \widehat{\delta} Z_i + (\widehat{\rho}_1 - \widehat{\rho}_0) \Gamma(\widehat{\pi}(0; \mathbf{X}_i), \widehat{\pi}(1; \mathbf{X}_i))\right)\right]$$

# Appendix: CM with Selection — Estimation

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

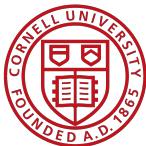
(a) ADE.

(b) AIE.



# The Direct and Indirect Effects of Genetics and Education

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



---

Cornell, Labor Work in Progress Seminar  
10 April 2026

# Introduction

The  
Economist

Menu



Try for free

Log in

Leaders | Modification revolution

## Science has made a new genetic revolution possible

Now let it flourish



## Nature vs Nurture

1. Genetics measure nature (initial endowment)
2. Find genes associated with outcomes
3. Deeper understanding of social outcomes, from genetics.

# Introduction

## Nature(?) vs Nurture

1. Genetics measure one component of nature (at most. . .)
2. Genes associated with outcomes are not the entire story
3. Genetics gives new findings, with caveats.

### INTERDISCIPLINARY RESEARCH

## Economics and genetics meet in uneasy union

*Use of population-genetic data to predict economic success sparks war of words.*

BY EWEN CALLAWAY

**T**he invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning." Evolutionary biologist Stephen Jay Gould was referring to purported links between genetics and an individual's intelligence when he made this familiar complaint in his 1981 book *The Mismeasure of Man*.

Fast-forward three decades, and leading geneticists and anthropologists are levelling a similar charge at economics researchers who claim that a country's genetic diversity can predict the success of its economy. To critics, the economists' paper seems to suggest that a country's poverty could be the result of its citizens' genetic make-up, and the paper is attracting charges of genetic determinism, and even racism. But the economists say that they have been misunderstood, and are merely using

genetics as a proxy for other factors that can drive an economy, such as history and culture. The debate holds cautionary lessons for a nascent field that blends genetics with economics, sometimes called *genoconomics*. The work could have real-world pay-offs, such as helping policy-makers "reduce barriers to the flows of ideas and innovations across populations", says Enrico Spolaore, an economist at Tufts University near Boston, Massachusetts, who has also used global genetic-diversity data in his research.

But the economists at the forefront of this field clearly need to be prepared for harsh scrutiny of their techniques and conclusions. At the centre of the storm is a 107-page paper by Oded Galor of Brown University in Providence, Rhode Island, and Quamrul Ashraf of Williams College in Williamstown, Massachusetts<sup>1</sup>. It has been peer-reviewed by economists and biologists, and will soon appear in *American Economic Review*, one of the most prestigious economics journals.

The paper argues that there are strong links between estimates of genetic diversity for 145 countries and per capita income, and



# Introduction

Genes associated with education causally improve labour market outcomes, but the **mechanism** is unclear

---

## This paper:

1. Estimate the causal effect of the Education PGI on education and income using quasi-random genetic variation from parents in UK Biobank (replicating Carvalho 2025)
  2. Decompose the labour market genetic effect into an **education channel** + **direct genetic effect** using *design-based* causal mediation framework
  3. Sensitivity analysis for different values for returns to an extra education year, transparently showing identification uncertainty.
- 

**Preview:** Education PGI causal effects transmit via **education channel**  
→ **Direct genetic effect** closes to zero at plausible education returns values.

# Introduction

How do genes associated with education affect labour outcomes?

## Introduction

### 1. Genetics & Education

Data: UK Biobank

### 2. Genetic Effects

Mendelian Independent Assortment

### 3. Causal Mediation Analysis

Sensitivity analysis

## Conclusion

## Introduction — Research Ethics

Genetic differences are real, but they are not fate.

- ▶ A higher Ed PGI reflects a *statistical tendency*, not a destiny — social policies and institutions can modify or eliminate the consequences of genetic variation
- ▶ These results provide no support for “natural” social hierarchies or genetic justifications for inequality
- ▶ Genetic associations estimated within one ancestry group **cannot** be meaningfully compared across groups; any such inference would be both invalid and ethically deficient

---

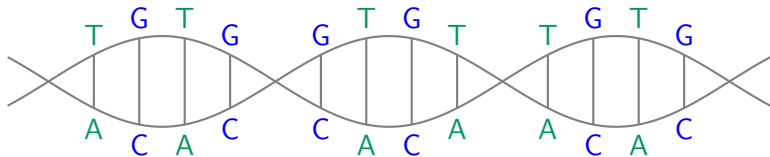
### The appropriate interpretation:

- ▶ Both social *and* genetic factors shape socioeconomic outcomes
- ▶ This should inspire support for policies promoting **equality of opportunity** — not acceptance of inequality as inevitable or “natural.”

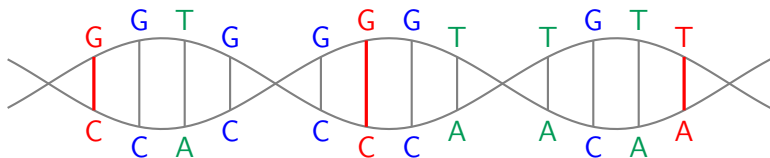
# Genetics: a Primer

- ▶ The human genome is a 3 billion-long sequence of base pairs
- ▶  $\approx 0.1\%$  these pairs differ between two average people

Person 1's DNA:



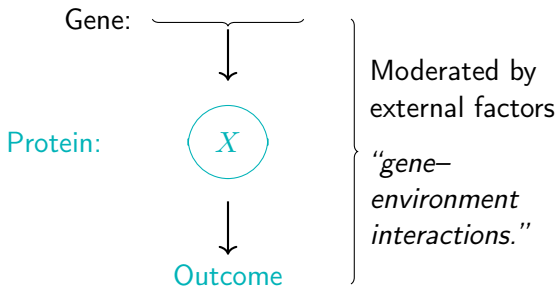
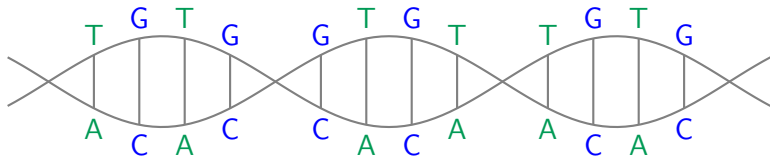
Person 2's DNA:



# Genetics: a Primer

- ▶ The human genome is a 3 billion-long sequence of base pairs
- ▶  $\approx 0.1\%$  these pairs differ between two average people

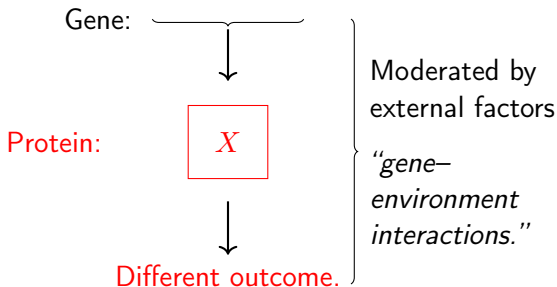
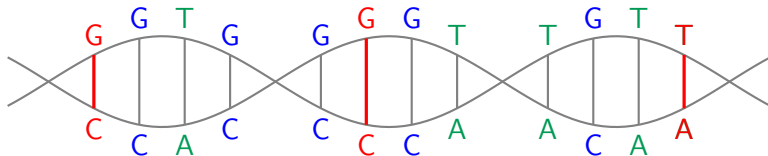
Person 1's DNA:



# Genetics: a Primer

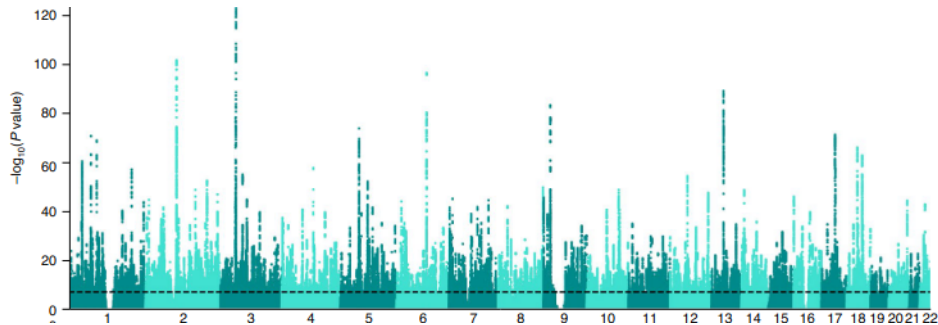
- ▶ The human genome is a 3 billion-long sequence of base pairs
- ▶  $\approx 0.1\%$  these **pairs differ** between two average people

Person 2's DNA:



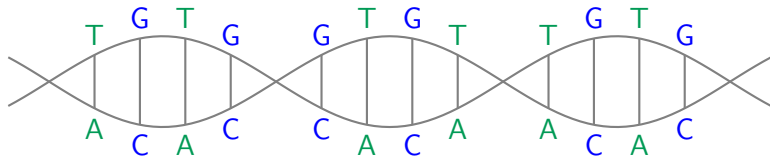
# Education PolyGenic Index (Ed PGI)

“We conduct a genome-wide association study (GWAS) of educational attainment (EA) in a sample of 3 million individuals and **identify 3,952 approximately uncorrelated genome-wide-significant single-nucleotide polymorphisms (SNPs)**. A genome-wide [...] polygenic index (PGI), explains 12–16% of EA variance and contributes to risk prediction for ten diseases.”  
*Abstract — Okbay et al. (2022, Nature Genetics)*

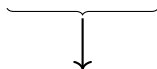


# Education PolyGenic Index (Ed PGI)

Person 1's DNA:



3,952 Different Genes:



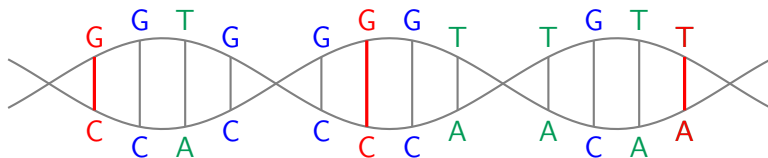
3,952 Different Proteins:



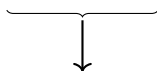
Education Years

# Education PolyGenic Index (Ed PGI)

Person 2's DNA:



3,952 Different Genes:



3,952 Different Proteins:

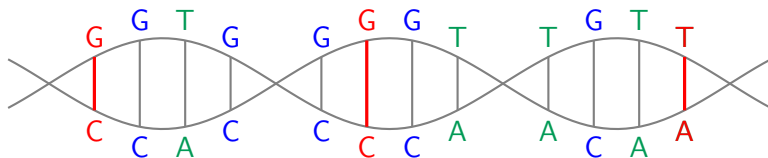


Education Years

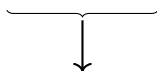
- ▶ In practice, we do not know exact location of relevant gene regions
- ▶ Nor the associated protein + how it works.

# Education PolyGenic Index (Ed PGI)

Person 2's DNA:



3,952 Different Genes:



3,952 Different Proteins:

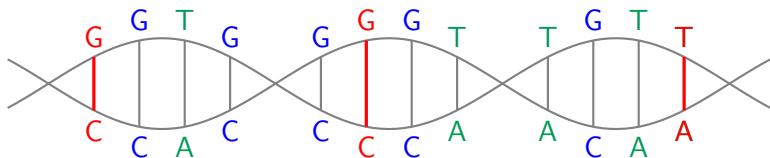


Education Years

- ▶ In practice, we do not know exact location of relevant gene regions
- ▶ Nor the associated protein + how it works.

# Education PolyGenic Index (Ed PGI)

Person 2's DNA:



Education Years

Calculate Polygenic Index (PGI):

1. Identify mutation regions (3,952 for Ed PGI)
2. Calculate weights for each
3. Assign a 1-dim genetic measure, summarising + dimension-reducing this topic.

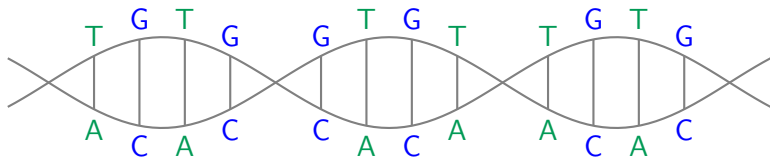
- ▶ In practice, we do not know exact location of relevant gene regions
- ▶ Nor the associated protein + how it works.

# Education PolyGenic Index (Ed PGI)

## Education Polygenic Index (Ed PGI):

A weighted sum of single letter gene changes predicting years of education.

Person 1's DNA:



Weights:  $w_1$  0 0 0      0  $w_2$  0 0      0 0 0  $w_3$

Person 1's Ed PGI:

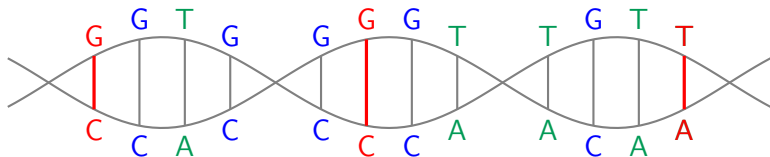
$$Z_i = \sum_{j=1}^J w_j \mathbb{1} \{ \text{person 1 has one-letter change at } j \} = 0$$

# Education PolyGenic Index (Ed PGI)

## Education Polygenic Index (Ed PGI):

A weighted sum of single letter gene changes predicting years of education.

Person 2's DNA:



Weights:  $w_1$  0 0 0      0  $w_2$  0 0      0 0 0  $w_3$

Person 2's Ed PGI:

$$Z_i = \sum_{j=1}^J w_j \mathbb{1} \{ \text{person 2 has one-letter change at } j \} = w_1 + w_2 + w_3$$

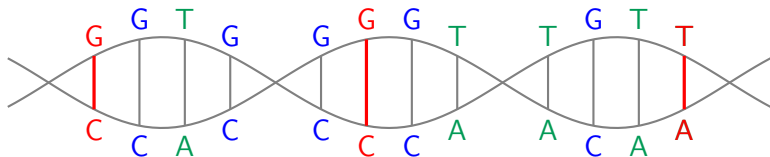
Note: Humans have two copies of each chromosome, so indicator is 0, 1, 2 corresponding to how many copies one has.

# Education PolyGenic Index (Ed PGI)

## Education Polygenic Index (Ed PGI):

A weighted sum of single letter gene changes predicting years of education.

Person 2's DNA:



Weights:  $w_1$  0 0 0      0  $w_2$  0 0      0 0 0  $w_3$

Person 2's Ed PGI:

$$Z_i = \sum_{j=1}^J w_j \mathbb{1} \{ \text{person 2 has one-letter change at } j \} = w_1 + w_2 + w_3$$

**Note:** Humans have two copies of each chromosome, so indicator is 0, 1, 2 corresponding to how many copies one has.

# Education PolyGenic Index (Ed PGI)

Previous economics + science research shows Ed PGI is strongly associated with multiple economics-relevant outcomes:

---

Retirement wealth	(Barth et al., 2020)
Life-cycle income	(Barth et al., 2022)
Longevity	(Marioni et al., 2016)
Primary school performance	(Aagaard-Houmark et al., 2022)
Parental-child care, nurture	(Aagaard-Houmark et al., 2025)
	(Muslimova et al. 2025)
...	
Labour market outcomes ( <b>causal</b> )	(Carvalho, 2025)

---

**Open question:** What are the mechanisms behind these relationships?

- ▶ Is **education the only way** it is related to outcomes?
- ▶ Is there a **direct genetic relationship**, absent returns to education?

## Data: UK Biobank

This is **the** dataset that genetics researchers have used for genetics research.

### The first “Biobank”

- ▶ Tissue samples (e.g., DNA) from 500,000 Britons
- ▶ Large-scale population study started in 2004, to investigate biological + genetic basis of low-frequency diseases
- ▶ Includes surveys for social outcomes, including education, occupation, etc.



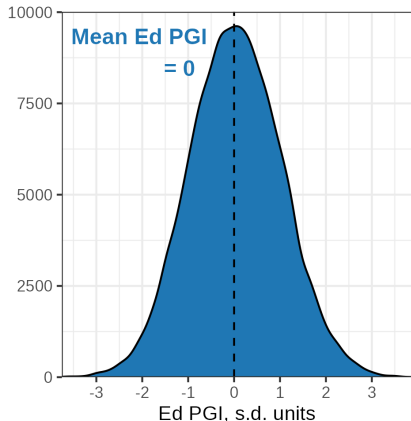
UK Biobank has  $N \approx 25,000$  subsample of people with genetic data from siblings included.<sup>1</sup>

<sup>1</sup>Limited to race = white, for whom Ed PGI is defined.

# Ed PGI and Education

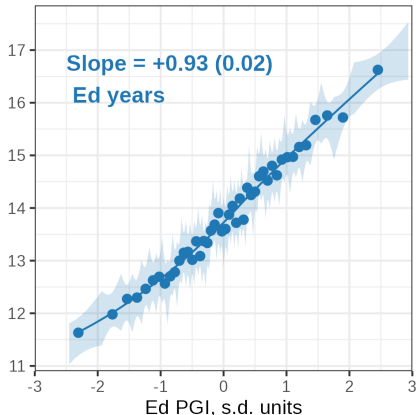
(a) Ed PGI.

Observations



(b) Ed PGI + Education Years, Binscatter.

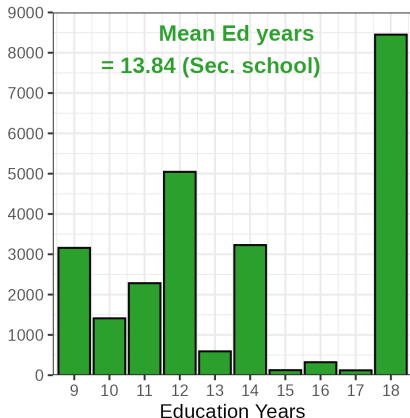
Education Years



# Education and Labour Market Outcomes

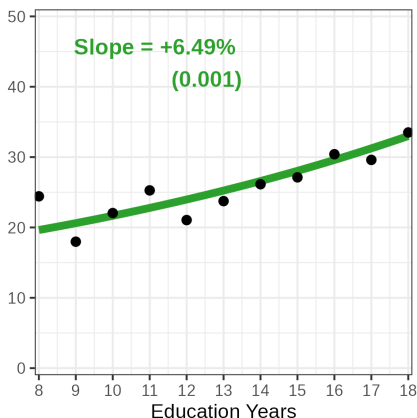
(a) Education Years.

Observations,  $N$



(b) Education Years + Income.

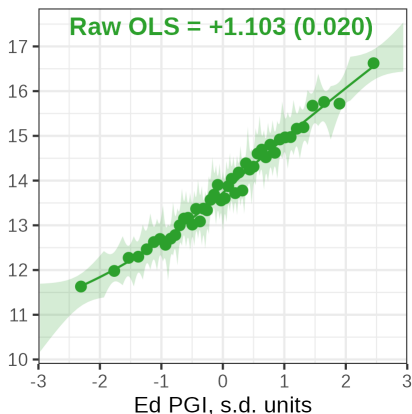
Annual Income, £ thousands



# Associations with the Ed PGI

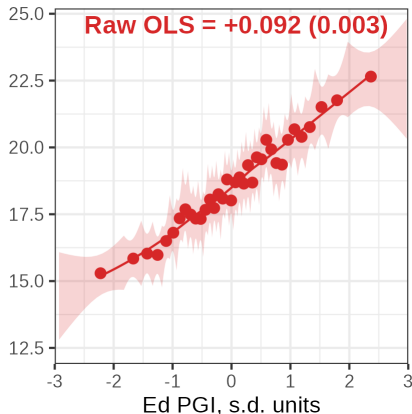
(a) Ed PGI  $\rightarrow$  Education Years.

Education Years



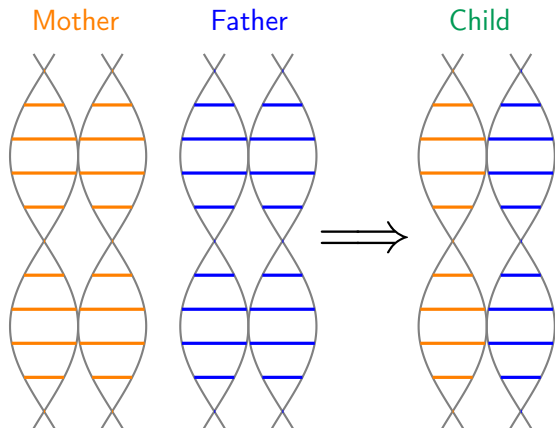
(b) Ed PGI  $\rightarrow$  Hourly Wages.

Occupation Hourly Wages, £



# Mendelian Independent Assortment

Question: Is  $Ed$  PGI randomly assigned?



Answer:

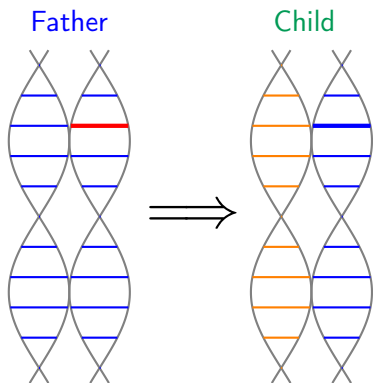
No, inherited from parents.

$$\mathbb{E} \left[ Z_i^{\text{Child}} \mid Z_{m(i)}^{\text{Mother}}, Z_{f(i)}^{\text{Father}} \right]$$

$$= \frac{1}{2} \left( Z_{m(i)}^{\text{Mother}} + Z_{f(i)}^{\text{Father}} \right)$$

# Mendelian Independent Assortment

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.



$$\Pr(\text{Child inherits } | \text{Father one copy}) = \frac{1}{2}$$

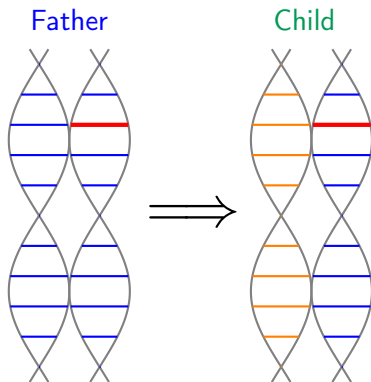
I.e., inheritance probability depends only on parents' genetics.

Propagate this reasoning across all **3,952 single-letter changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} = \mathbb{E} \left[ Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

# Mendelian Independent Assortment

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.



$$\Pr(\text{Child inherits} \mid \text{Father one copy}) = \frac{1}{2}$$

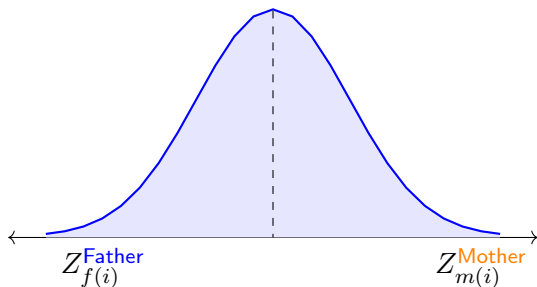
I.e., inheritance probability depends only on parents' genetics.

Propagate this reasoning across all **3,952 single-letter gene changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[ Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \rightarrow \text{Independently assigned.}$$

## Ed PGI Causal Effects: Identification.

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.

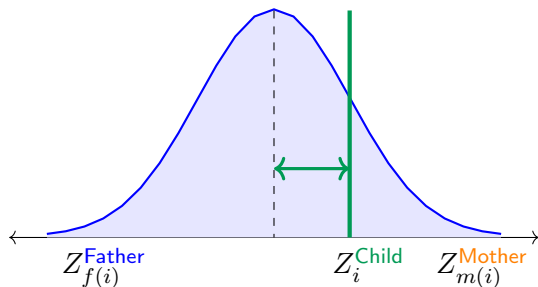


Propagate this reasoning across all **3,952 single-letter changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[ Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

## Ed PGI Causal Effects: Identification

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.

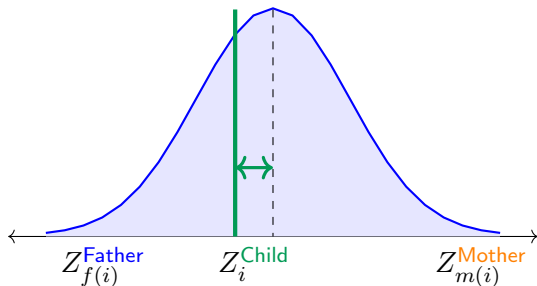


Propagate this reasoning across all **3,952 single-letter changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[ Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

## Ed PGI Causal Effects: Identification

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.

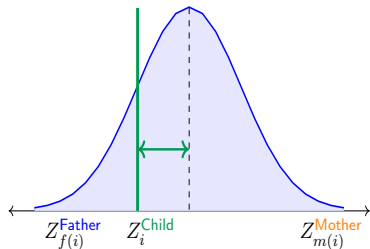


Propagate this reasoning across all **3,952 single-letter changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[ Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

## Ed PGI Causal Effects: Identification

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.  $\implies$  Use the difference from parents as **plausibly random genetic variation**.



### Remark on Estimation:

Control for imputed mean parental Ed PGI from siblings, because few parents observed in UK Biobank, losing statistical efficiency.

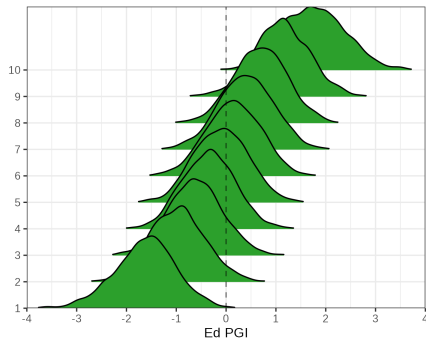
$$Z_i^{\text{Child}} - \mathbb{E} \left[ Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

# Ed PGI Causal Effects: Estimates

**Figure:** Distribution of the Ed PGI, Relative to Imputed Parental Decile.

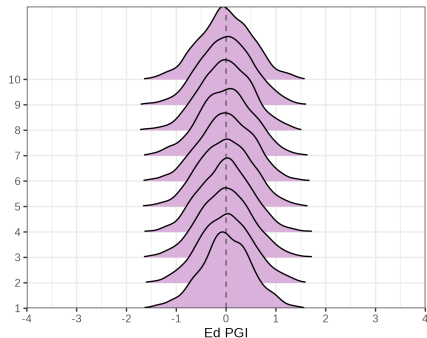
(a) Ed PGI.

Parental Quantile, 1st through 10th



(b) Ed PGI minus imputed parental mean.

Parental Quantile, 1st through 10th

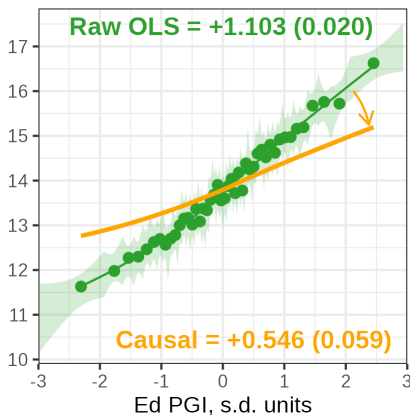


$$Z_i^{\text{Child}} - \mathbb{E} \left[ Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \rightarrow \text{Independently assigned.}$$

# Causal Effects of the Ed PGI

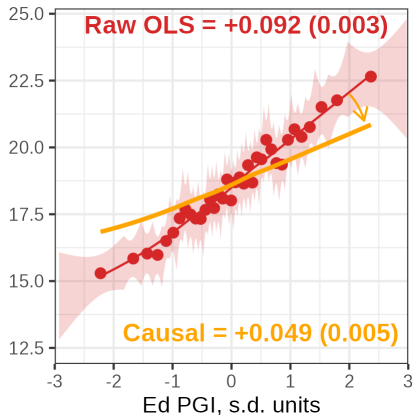
(a) Ed PGI  $\rightarrow$  Education Years.

Education Years



(b) Ed PGI  $\rightarrow$  Hourly Wages.

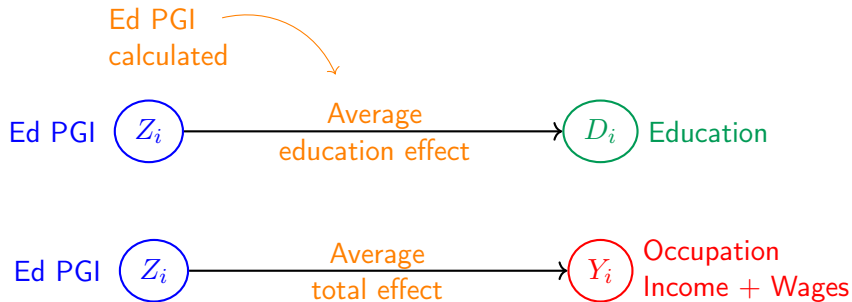
Occupation Hourly Wages, £



Replicates main estimates in Carvalho (2025).

## Direct and Indirect Effects

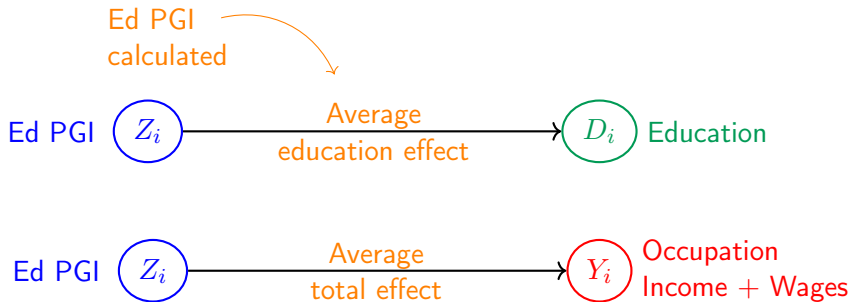
Ed PGI causally affects education years and labour market outcomes.  
 → Not surprising, Ed PGI computed from data **only** genes + education.



Usual approach: Observe education years → labour market outcomes, and conclude with suggestive evidence that education “is a mechanism.”

## Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.  
 → Not surprising, Ed PGI computed from data only genes + education.



**Usual approach:** Observe education years → labour market outcomes, and conclude with **suggestive evidence** that education “is a mechanism.”

# Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

⇒ Consider the mechanisms:

## Direct Genetic Channel

- ▶ Creativity
- ▶ Soft skills

## Returns to Education

- ▶ Classroom learning
- ▶ Hard skills
- ▶ Occupational qualifications

A direct effect implies Ed PGI genes have direct labour market benefits.

Motivated by literature noting that Ed PGI improves/correlated with

- ▶ fluid intelligence test score (Carvalho 2025)
- ▶ Income, after controlling for education (correlationally, Papageorge Thom 2020, Bryson et al. 2025)
- ▶ Financial literacy at retirement (Barth et al. 2020).

## Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

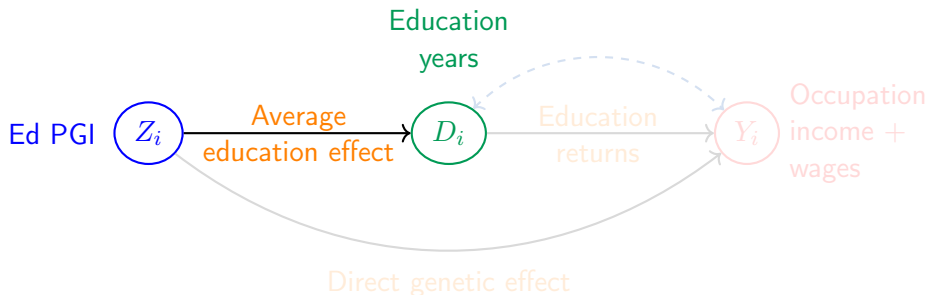
**Figure:** Model for Causal Effect of the Ed PGI.



# Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

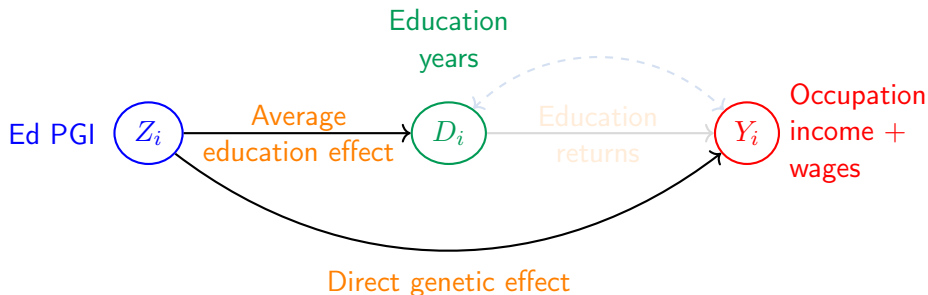
**Figure:** Model for Causal Mediation of the Ed PGI.



# Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

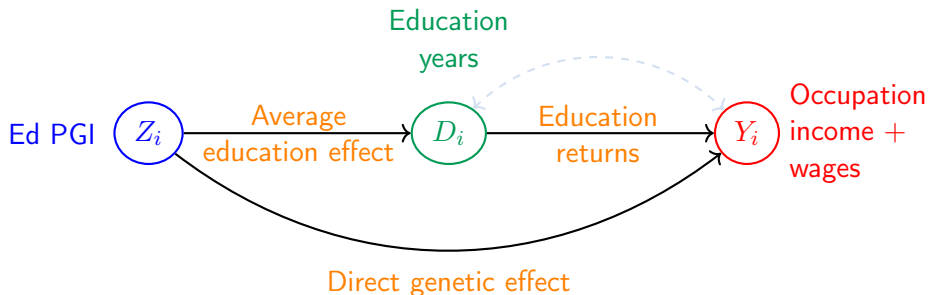
**Figure:** Model for Causal Mediation of the Ed PGI.



## Direct and Indirect Effects

Ed PGI causally affects **education years** and **labour market outcomes**.

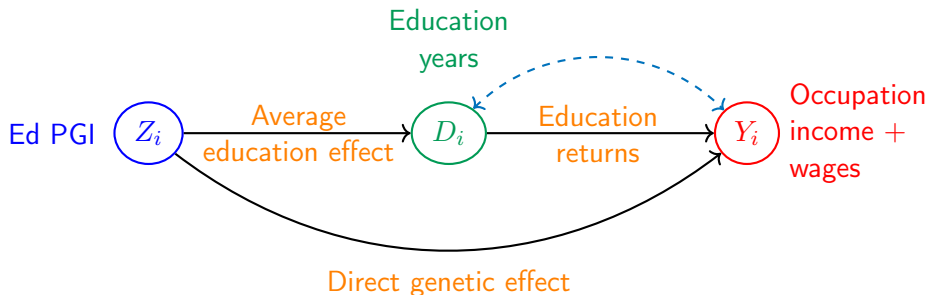
**Figure:** Model for Causal Mediation of the Ed PGI.



# Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

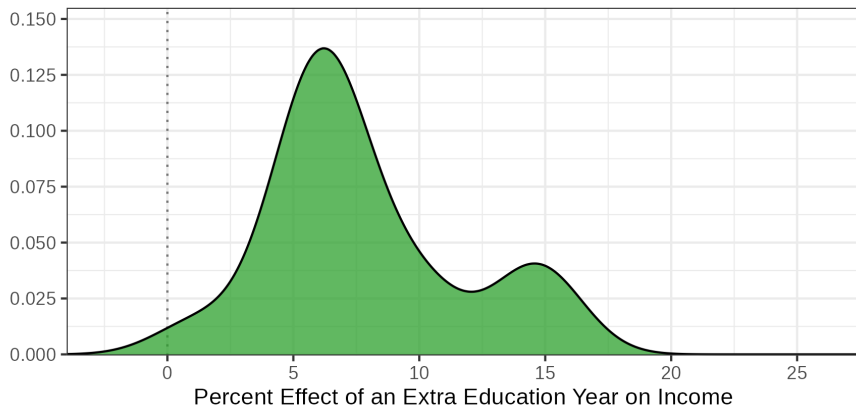
**Figure:** Model for Causal Mediation of the Ed PGI.



## Returns to Education

The only missing piece for a valid Causal Mediation analysis are causal estimates for returns to education (and statistical uncertainty).

Density of British IV Estimates

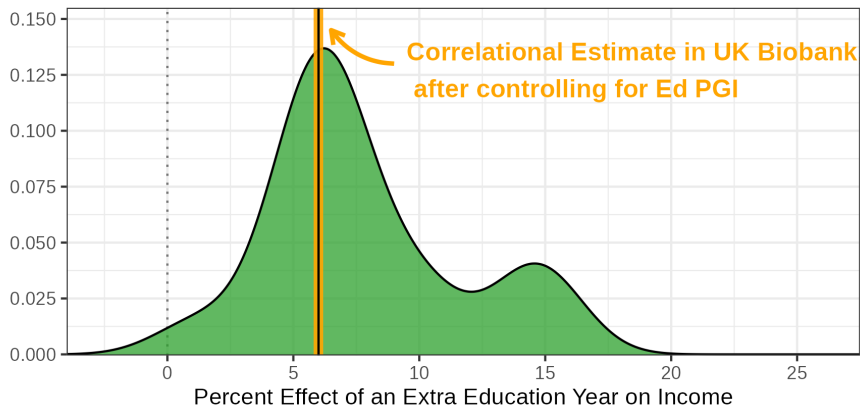


**Note:** Distribution from 17 IV estimates for Britain (Patrinos Psacharopoulos 2026).

## Returns to Education

The only missing piece for a valid Causal Mediation analysis are causal estimates for returns to education (and statistical uncertainty).

Density of British IV Estimates

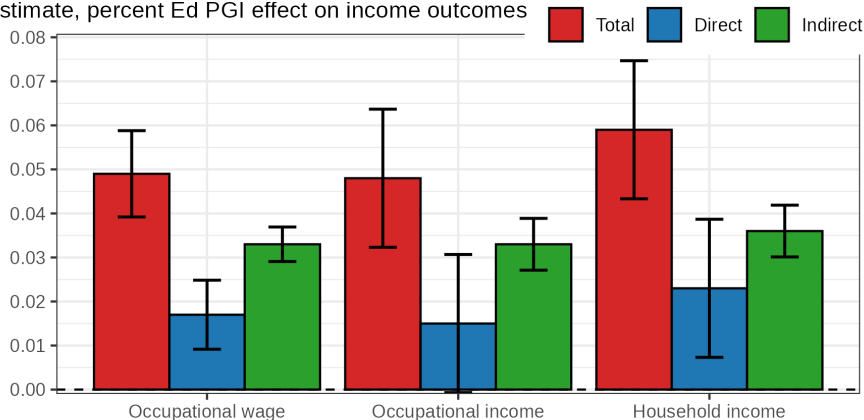


**Note:** Distribution from 17 IV estimates for Britain (Patrinos Psacharopoulos 2026).

## Direct and Indirect Effects — Estimates

Using Correlation return estimate of  $\approx 6\%$ (0.1 SE):

Estimate, percent Ed PGI effect on income outcomes



## Direct and Indirect Effects — Estimates

Using Correlation return estimate of  $\approx 6\%$ (0.1 SE):

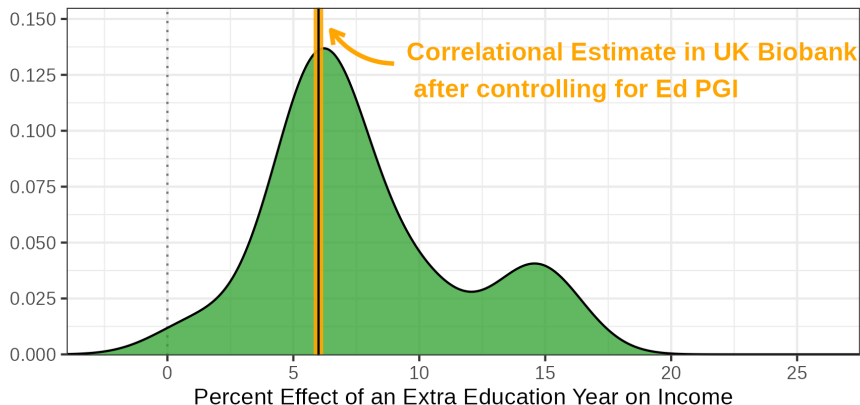
Estimate, Proportion mediated through education



## Direct and Indirect Effects — Sensitivity Analysis

The only missing piece for a valid Causal Mediation analysis are causal estimates for returns to education (and statistical uncertainty).

Density of British IV Estimates

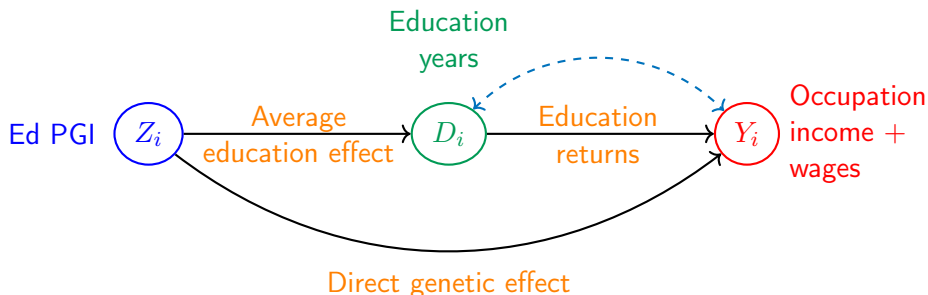


**Note:** Distribution from 17 IV estimates for Britain (Patrinos Psacharopoulos 2026).

## Direct and Indirect Effects — Sensitivity Analysis

The only missing piece for a valid Causal Mediation analysis are causal estimates for returns to education (and statistical uncertainty).

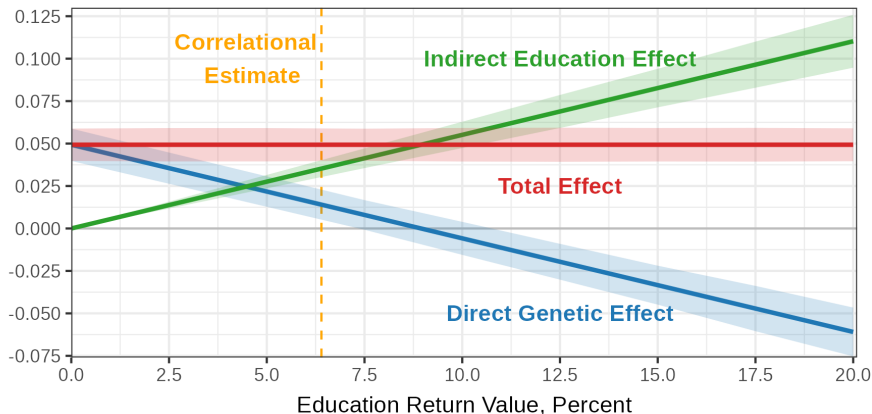
**Figure:** Model for Causal Mediation of the Ed PGI.



# Direct and Indirect Effects — Sensitivity Analysis

**Figure:** Sensitivity Analysis, Effect of Ed PGI on Log Occupational Wages via Education Years.

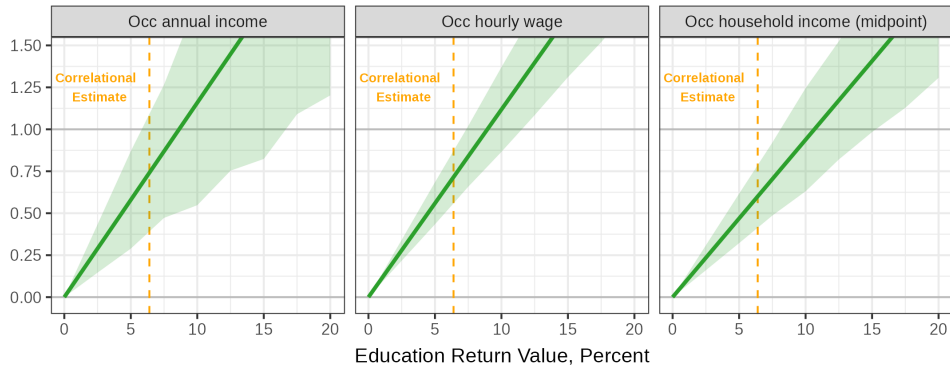
Direct and Indirect Effect Estimates



# Direct and Indirect Effects — Sensitivity Analysis

**Figure:** Sensitivity Analysis, Proportion of Ed PGI Effect on Income Outcomes via Education Years.

Indirect Effect / Total Effect

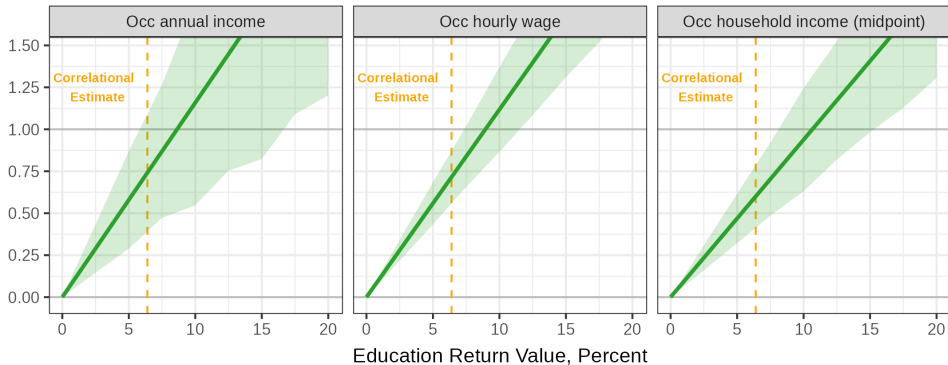


Direct genetic effect closes  $\rightarrow 0$  at plausible education returns values.

# Direct and Indirect Effects — Sensitivity Analysis

**Figure:** Sensitivity Analysis, Proportion of Ed PGI Effect on Income Outcomes via Education Years.

Indirect Effect / Total Effect



Direct genetic effect closes  $\rightarrow 0$  at plausible education returns values.

# Conclusion

Genes correlated with education impact income inequality primarily through education — implying this inequality gradient could be tackled via policies aimed at education access.

1. Ed PGI raises **education years** and **labour market outcomes** (replicating Carvalho 2025)
2. Economics literature has speculated (with suggestive evidence) on **direct genetic effects**, but lacked quantitative evidence
3. At plausible causal returns to education values, most of the total genetic income effect operates through the **education channel**, with a **direct genetic effect** that is small, and  $\rightarrow 0$  under higher return values.

“The difference of natural talents in different men, is, in reality, much less than we are aware of [...]. The difference between the most dissimilar characters, between a philosopher and a common street porter, seems to arise **not so much from nature, as from habit, custom, and education.**”

— A Smith (1776), *The Wealth of Nations*, Volume 1 p. 16.

# Conclusion

Genes correlated with education impact income inequality primarily through education — implying this inequality gradient could be tackled via policies aimed at education access.

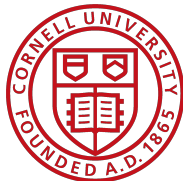
1. Ed PGI raises **education years** and **labour market outcomes** (replicating Carvalho 2025)
2. Economics literature has speculated (with suggestive evidence) on **direct genetic effects**, but lacked quantitative evidence
3. At plausible causal returns to education values, most of the total genetic income effect operates through the **education channel**, with a **direct genetic effect** that is small, and  $\rightarrow 0$  under higher return values.

**“The difference of natural talents in different men, is, in reality, much less than we are aware of [ . . . ]. The difference between the most dissimilar characters, between a philosopher and a common street porter, seems to arise **not so much from nature, as from habit, custom, and education.**”**

— A Smith (1776), *The Wealth of Nations*, Volume 1 p. 16.

# Less Funding, More Lecturers, Fewer Professors: Stagnating State Funding for Higher Education and its Effect on Faculty

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



---

Cornell Economics, PhD Defence  
29 June 2026

**INSIDE**  
Subscribe to our newsletter

Become an insider | Login

**Live Updates!** Annie Ernaux Wins Nobel in Literature

Publisher reinstates blocked ebooks, but librarians 'unsatisfied'

Lawsuits pile up against Biden's debt-relief plan

Patty Limerick speaks out on her dismissal from her center

Faculty members oppose trustees' plan to review DEI language

Study finds STEM could push out students

OPINION | **FACULTY**

## The End of Faculty Tenure

Marc Stein argues declines in tenure density constitute the greatest threat U.S. higher education has ever faced.

By Marc Stein  
Published April 25, 2022

# Is Tenure Really Dying? And Would That Be Good or Bad?

Is academic tenure in for major changes over the next decade?

Wednesday, September 21, 2022

## FACULTY & STAFF

# Report Reveals Biggest Decrease in Real Wages for Full-Time Faculty



Rebecca Kelliher  
Jun 22, 2022



The American Association of University Professors (AAUP) released its [2021-22 Annual Report on the Economic Status of the Profession](#), which details findings from the AAUP's annual Faculty Compensation Survey. One of the report's biggest takeaways is that real wages for full-time faculty in 2021 fell below Great Recession levels of the late 2000s, after adjusting for inflation.



## Funding Down, Tuition Up

State Cuts to Higher Education Threaten Quality and Affordability at Public Colleges

UPDATED AUGUST 15, 2016 | BY MICHAEL MITCHELL, MICHAEL LEACHMAN, AND KATHLEEN MASTERSON<sup>(1)</sup>

Years of cuts in state funding for public colleges and universities have driven up tuition and harmed students' educational experiences by forcing faculty reductions, fewer course offerings, and campus closings. These choices have made college less affordable and less accessible for students who need degrees to succeed in today's economy.

REPORT JUN 11, 2020

## Mounting Peril for Public Higher Education During the Coronavirus Pandemic

Absent federal action, state budget cuts are about to add to colleges' pain from the pandemic.

PBS  
NEWS  
HOUR



By –  
Jon Marcus,  
The  
Hechinger  
Report

Leave a  
comment

# Most Americans don't realize state funding for higher ed fell by billions

Education Feb 26, 2019 12:20 PM EDT

# Introduction — Preview & Road-Map

Three connected topics:

1. Document the substitution away from professors towards lecturers, at US public universities
  2. Link the substitution to stagnating state funding, with a causal approach
  3. Infer the mechanism for these changes, limited hiring, and implications for university decision-making.
- 

## Trends in US Higher Education

### Shifts in State Funding

Effects on Faculty

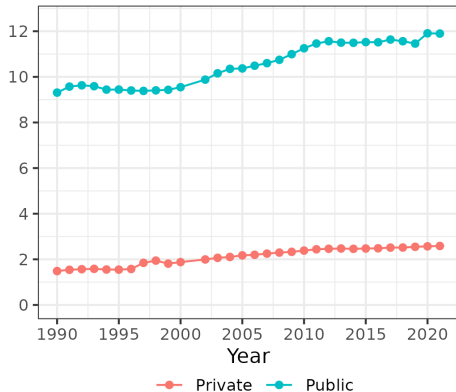
### Mechanism for Faculty Substitution

# Trends in US Higher Education

Despite rising enrolment at US public universities, **state funding has stagnated**.

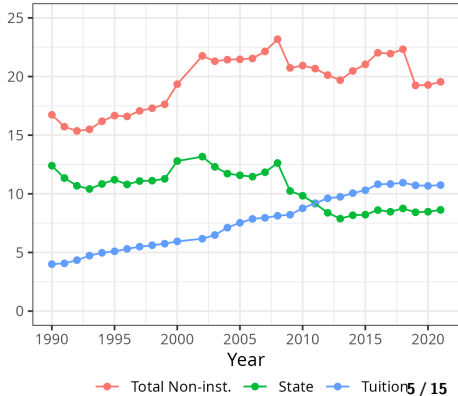
(a) Mean Enrolment, per University.

Student Enrolment, thousands



(b) Mean Funding per student among Public Universities.

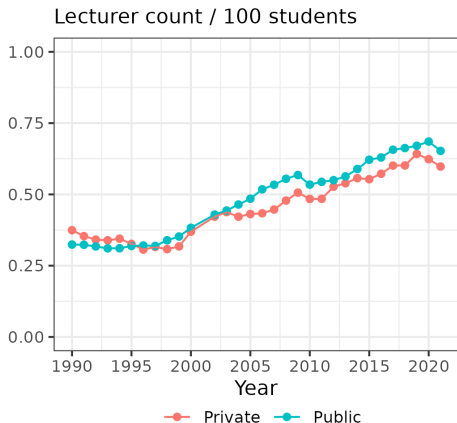
Mean Funding per Student, \$ thousands



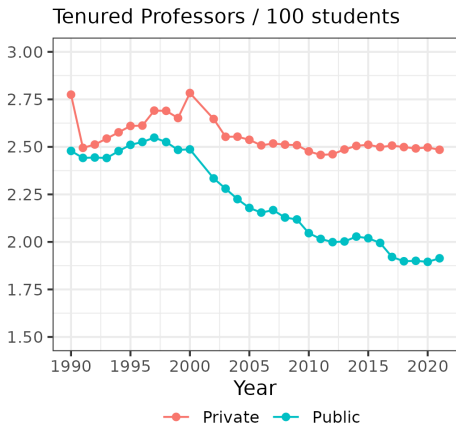
# Trends in US Higher Education

At the same time, lecturers employment has increased drastically, and tenure-track/tenured professors fallen at public universities.

(a) Non-tenure Track (lecturers).



(b) Tenured (Full Professors).

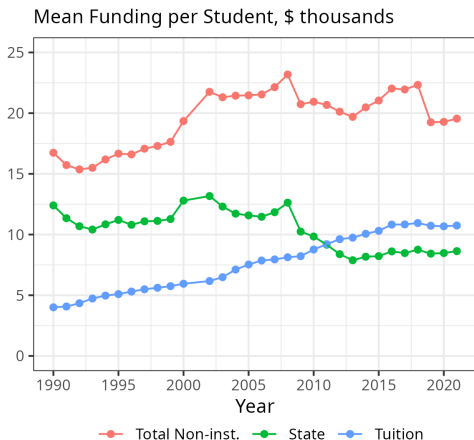


# Trends in US Higher Education

States decide an annual budget:

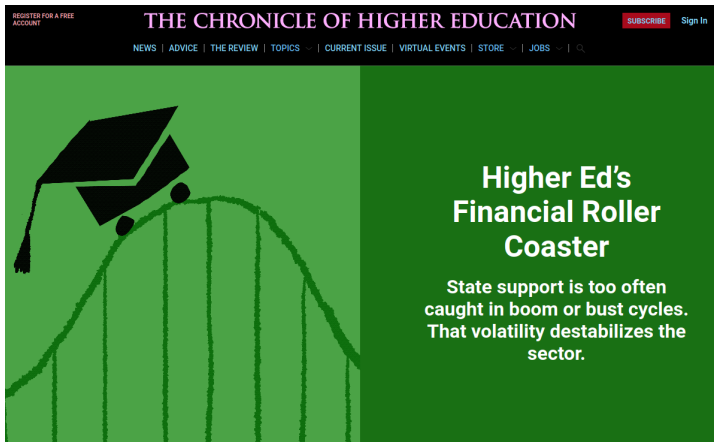
1. State legislature plans annual budget
2. Back-and-forth bargaining with governor, who signs off when they agree
3. Higher education is attractive to cut funding in this bargaining, thanks to low lobbying power.

**Figure:** Average Public University Funding.



⇒ institutional set-up leads to public universities being stiffed year-on-year.

# Trends in US Higher Education

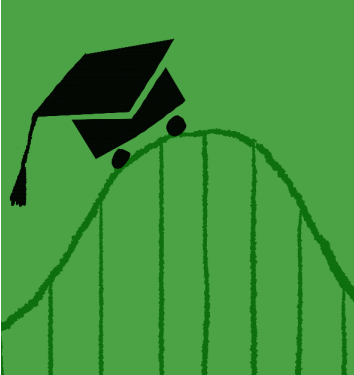


REGISTER FOR A FREE ACCOUNT

THE CHRONICLE OF HIGHER EDUCATION

SUBSCRIBE Sign In

NEWS | ADVICE | THE REVIEW | TOPICS | CURRENT ISSUE | VIRTUAL EVENTS | STORE | JOBS |

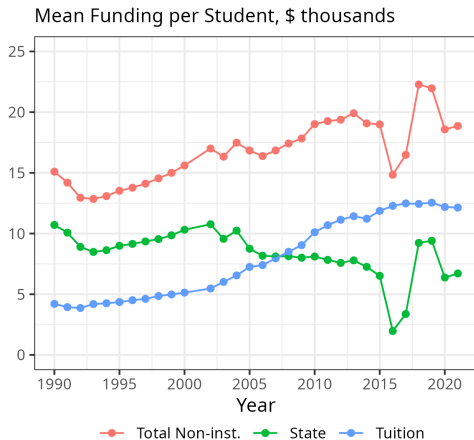


**Higher Ed's  
Financial Roller  
Coaster**

State support is too often caught in boom or bust cycles. That volatility destabilizes the sector.

# Trends in US Higher Education

**Figure:** Average Illinois University Funding.



This process in Illinois, 2015:

1. Democrat state legislature plans annual budget
2. Sends to Republican governor, then bargaining went nowhere
3. 2016 fiscal year starts with no budget, state staff furloughed etc.

⇒ Partisan disagreements led to Illinois public universities being stiffed in 2016.

# Shifts in State Funding

Estimate causal effects using **Shift-Share IV for state funding** → **faculty** (Deming Walters 2017, Chakrabarti Gorton Lovenheim 2020).

1. Exploit different rates of historical reliance on state funding between universities, in 1990–1993
2. Interacted with state-wide changes in higher education funding.

$$\mathbf{IV} = \underbrace{\sum_{\tau=0}^3 \frac{1}{4} \left( \frac{\text{State Funding}_{i,1990+\tau}}{\text{Total Revenues}_{i,1990+\tau}} \right)}_{1. \text{ Base reliance on state funding, 1990–1993}} \times \underbrace{\left( \frac{\text{Total State Funding}_{s(i),t}}{\text{Student Population}_{s(i),t}} \right)}_{2. \text{ State-wide higher ed funding shifts}}$$

---

## Intuition:

- ▶ State-wide shifts in higher ed funding affect universities that rely on state funding more
- ▶ State-wide shifts in higher ed funding come from population shifts + changing budget priorities — not one campus' actions.

# Shifts in State Funding

Estimate causal effects using **Shift-Share IV for state funding** → **faculty** (Deming Walters 2017, Chakrabarti Gorton Lovenheim 2020).

1. Exploit different rates of historical reliance on state funding between universities, in 1990–1993
2. Interacted with state-wide changes in higher education funding.

$$\mathbf{IV} = \underbrace{\sum_{\tau=0}^3 \frac{1}{4} \left( \frac{\text{State Funding}_{i,1990+\tau}}{\text{Total Revenues}_{i,1990+\tau}} \right)}_{1. \text{ Base reliance on state funding, 1990–1993}} \times \underbrace{\left( \frac{\text{Total State Funding}_{s(i),t}}{\text{Student Population}_{s(i),t}} \right)}_{2. \text{ State-wide higher ed funding shifts}}$$

---

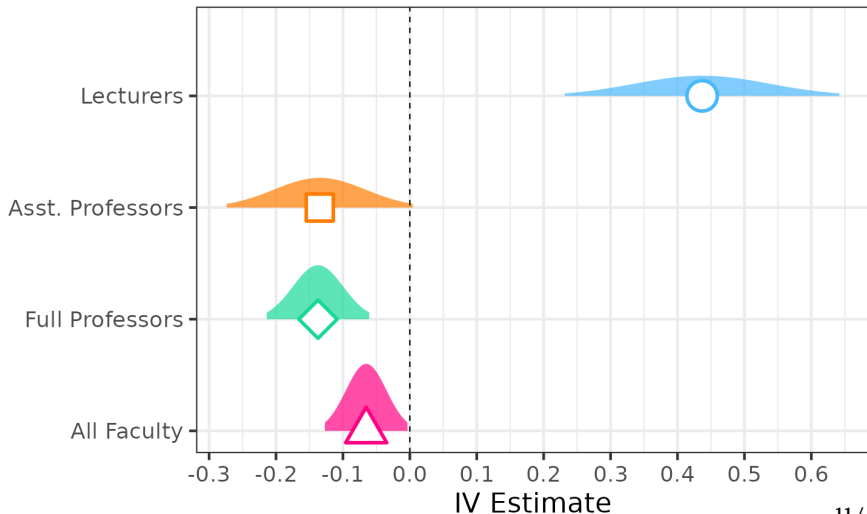
**Strong instrument** for yearly public university state funding:

- ▶  $-\$1,000$  per student funding shift-share →  $-\$1,176$  state funding
- ▶  $-10\%$  per student funding shift-share →  $-9.77\%$  state funding.

# Shifts in State Funding — Effects on Faculty

**Shift–Share IV Results** for the average public university:

- 10% state funding per student  $\rightarrow \approx 4.4\%$  more lecturers / students.
- 10% state funding per student  $\rightarrow \approx 1.4\%$  fewer professors / students.

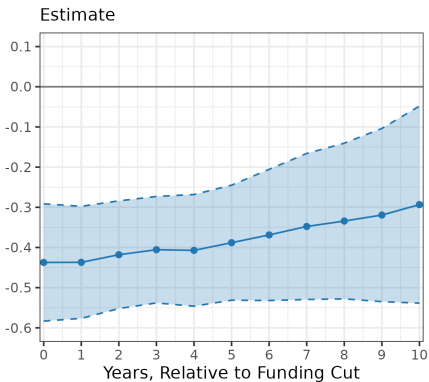


# Shifts in State Funding — Effects on Faculty

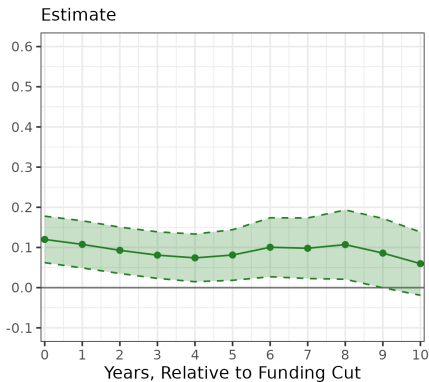
⇒ Stagnating state funding, and increasing reliance on lecturers.

**Substitution effects are long-lasting** → medium-term elasticities:

**(a)** Non-tenure Track (lecturers).



**(b)** Tenured (Full Professors).



# Mechanism for Faculty Substitution

Stagnating state funding led to substitution towards lecturers, at US public universities, but unclear how these effects come about.

- ▶ Are **faculty let-go** by their universities?
- ▶ Are **faculty leaving voluntarily**, following pay cuts?
- ▶ Are **faculty poached by other universities** e.g., after promotion limits at their current university?

---

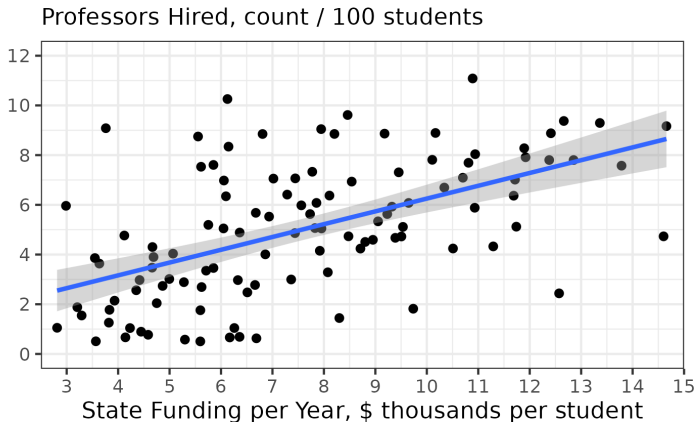
To briefly summarise effects of state funding cuts on faculty, using data on all Illinois public uni staff 2011–2021:

- ▶ **No detected effect of state funding cuts** on follow-on salaries, exit rates, or promotions among Illinois faculty
- ▶ These data represent those who are already faculty  
⇒ i.e., **incumbent faculty** unaffected by state funding cuts.

# Mechanism for Faculty Substitution

**Condundrum:** how can substitution effects happen?

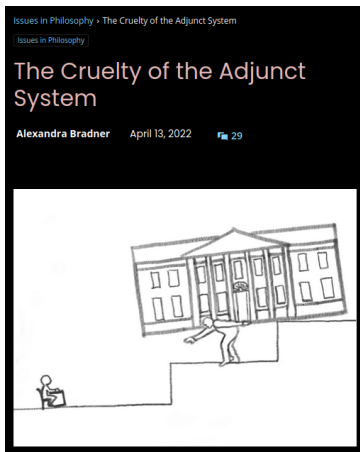
⇒ limits **faculty hiring**, supported by hiring data.



Auxiliary data on professor hiring from Wapman Zhang Clauset Larremore (2022) for USA public universities 2011–2021.

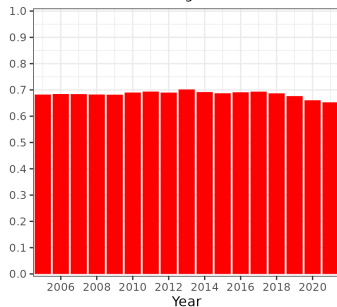
# Conclusion

Are lecturers a bad thing? No, but many are over-worked, under-paid. . .



(a) Part-time Rate, Lecturers at Public Unis

Percent of Lecturers Working Part-time



Implies that public universities (who educate the majority of US higher ed students) make long-lasting personnel decisions based on yearly funding cuts. While instruction salaries are  $< 30\%$  of total spending. . . .

# Acknowledgments

List acknowledgements with some photos.