# Uncovering Causal Mechanisms

## Mediation Analysis and Surrogate Indices

Raj Chetty and Kosuke Imai
Harvard University

NBER Methods Lecture

July 24, 2025

These slides have been posted on the NBER Methods Lecture website:
https://www.nber.org/conferences/si-2025-methods-lecture-uncovering-causal-mechanisms-mediation-analysis-and-surrogate-indices
Thanks to Juhui Jin and Yechan Park for excellent research assistance in preparing this presentation.

OPPORTUNITY INSIGHTS

# A Recent Motivating Inquiry

An inquiry regarding a new workforce training program in NYC that builds connections between people to obtain jobs, motivated by research on social capital and upward mobility:

> *We would love to gather any ideas to improve predictions we can make [about the program's impacts]. As you know, in government, we* ***don't have the luxury of waiting 10 years*** *to report solid outcomes. We need to be able to tell the program's story of impact on an ongoing basis. We are continuously* ***innovating on how to be more efficient****.*

- Diana Franco, NYC Economic Development Corporation
July 16, 2025

# Two Questions

1. Can we measure impacts of interventions more quickly rather than waiting to observe long-term outcomes (e.g., changes in career trajectories)?
    - More generally, when we can't measure primary outcomes of interest directly, can we use other variables as proxies?
2. What are the mechanisms through which an intervention matters (e.g., job referrals from connections vs. information about career paths)?
    - Understanding mechanisms critical for "innovating to improve efficiency"

- Both questions involve the analysis of **intermediate outcomes** – either as predictors of long-term outcomes or as mediators
- Goal of this lecture: give an applied introduction to methods to answer these questions

# Outline

1. Motivating Applications
2. Mediation Analysis
3. Surrogate Indices
4. Application and Recommendations for Empirical Practice

# Motivating Example 1: Peer Effects on College Attendance

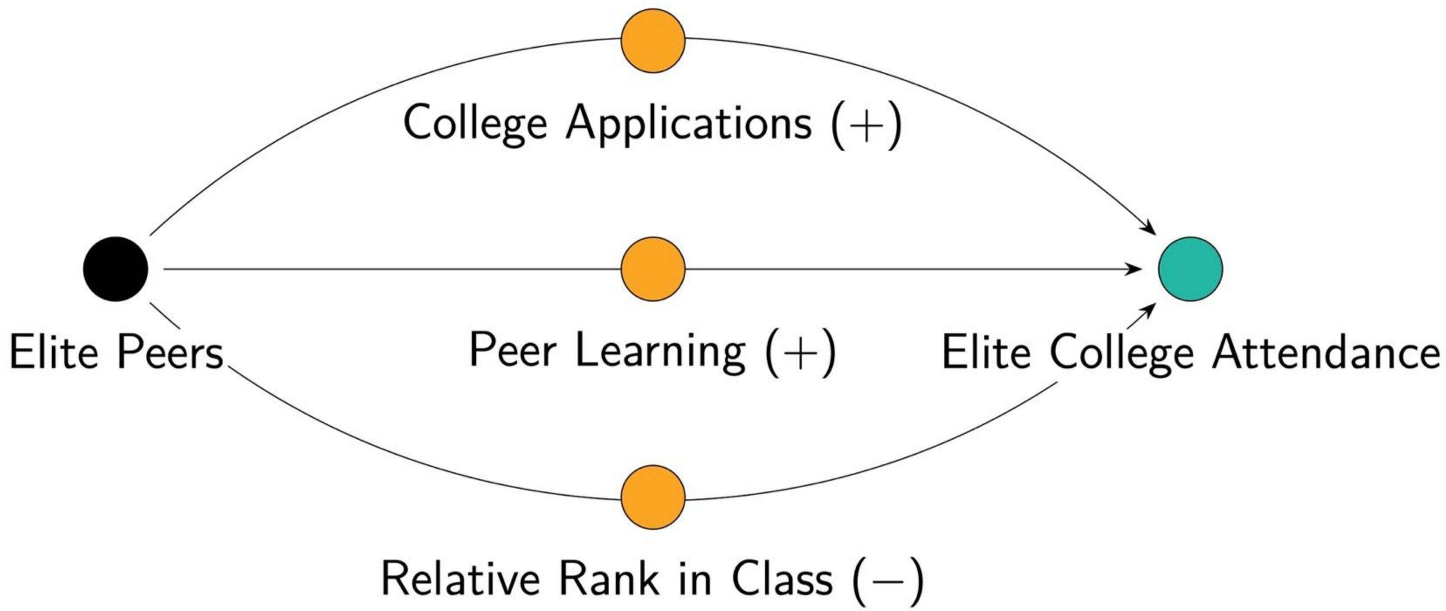
- Cattan, Salvanes, and Tominey (2025) analyze the effects of having high-school peers whose parents attended elite colleges on elite college attendance rates in Norway

- Research Design: random variation in peer composition arising from finite-sample fluctuations across cohorts within schools

# Cattan, Salvanes, and Tominey (2025): Reduced-Form Treatment Effect


1 SD increase in share of peers with elite-educated parents in school cohort →
**2.6 pp** increase in probability of enrolling in elite college (relative to mean of 10 pp)

# Cattan, Salvanes, and Tominey (2025): Causal Mechanisms

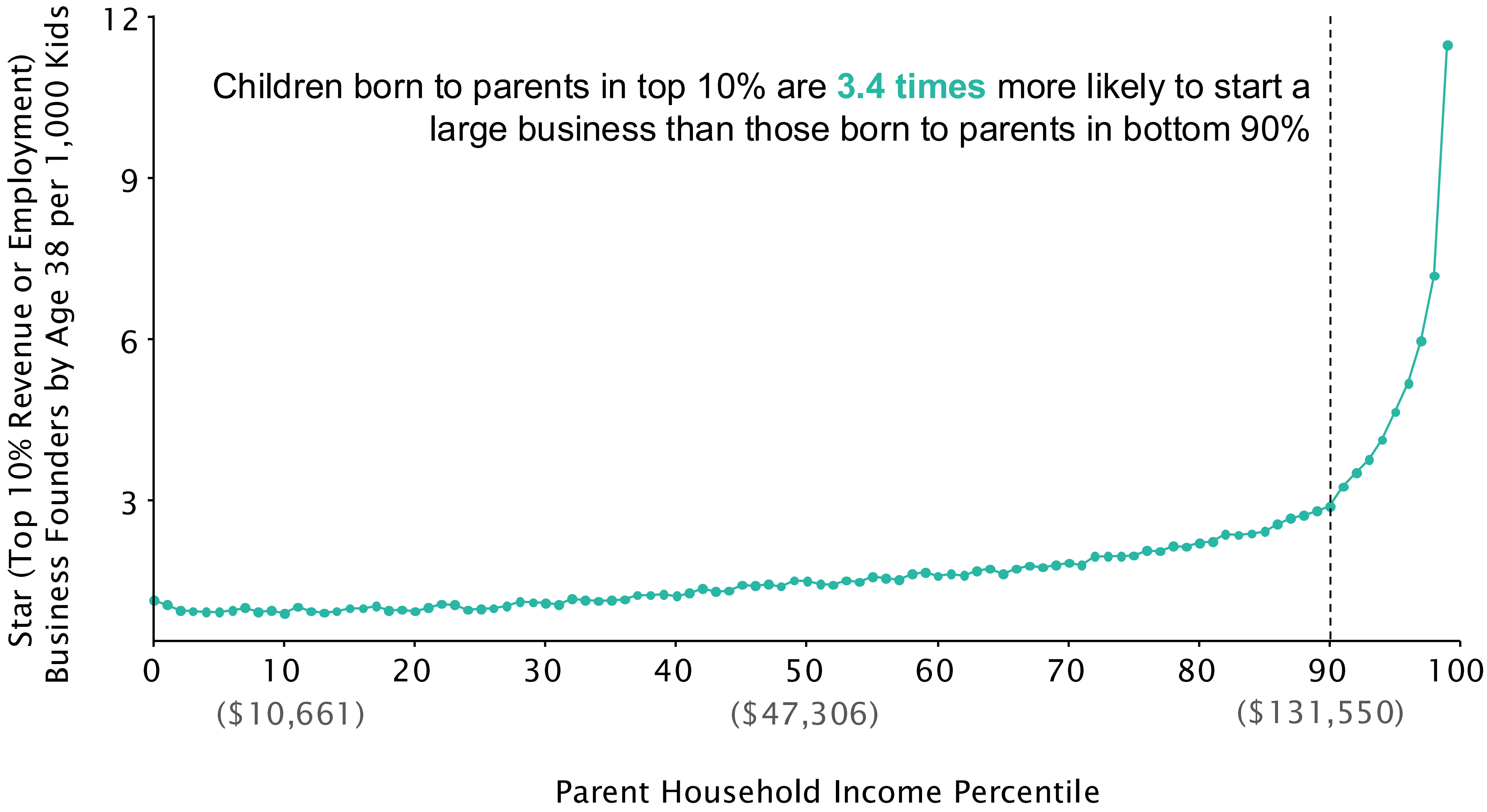



Direct Acyclic Graph (DAG) Representation of Causal Mechanisms [Pearl 2000]

# Motivating Example 2: Determinants of Entrepreneurship

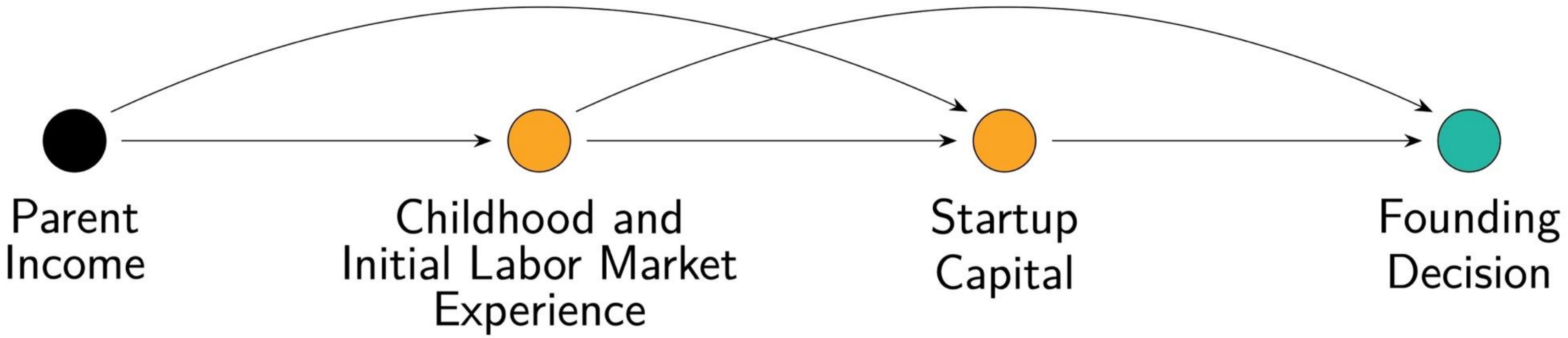
- Chetty, Dossi, Smith, van Reenen, Zidar, Zwick (2025) analyze determinants of and returns to entrepreneurship in the U.S. using data from anonymized tax records

# Entrepreneurship Rates by Parental Income



# The Entrepreneurial Pipeline: Causal Mechanisms

# The Entrepreneurial Pipeline: Causal Mechanisms



# A Common Approach to Mediation Analysis: Controlling for Mediators

- How much of the difference in entrepreneurship rates between children from low vs. high income families is explained by difference in access to capital at startup?

- Intuitive, commonly used approach: control for wealth at startup and measure how much of the raw gap in founding by parental income is “explained” by wealth

# A Common Approach to Mediation Analysis: Controlling for Mediators

- Regress indicator for business startup on having high-income (top decile) parents:

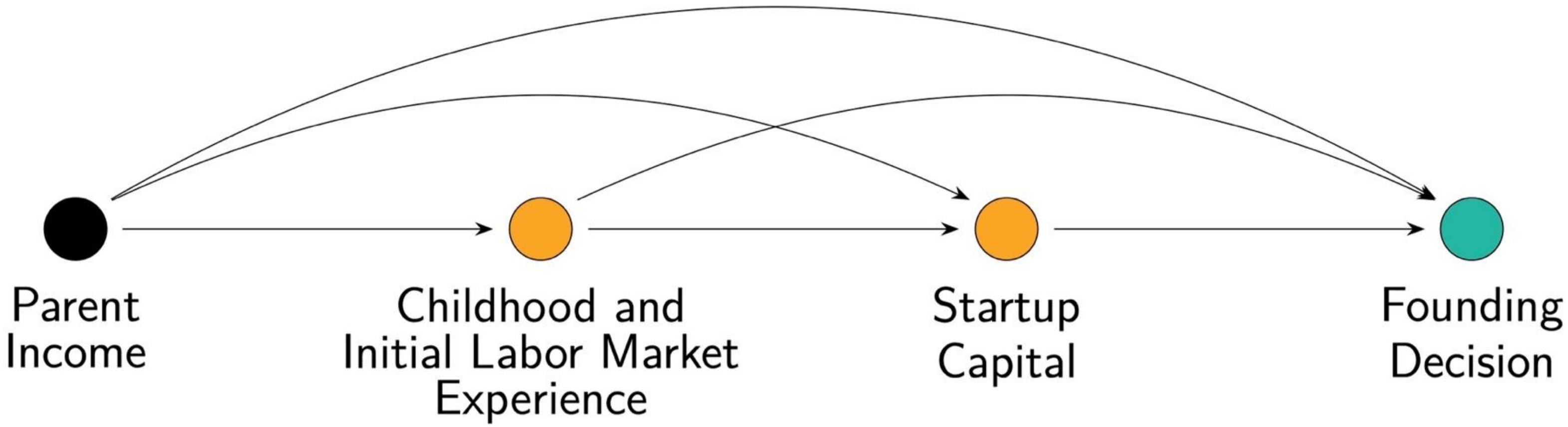
$$Y_i = \alpha_i + \beta \cdot (\text{Parent in Top 10\%})_i + \gamma \cdot (\text{Wealth at Startup})_i + \epsilon_i$$

| | Dep. Variable: Founded a Large Firm (per 1,000 kids) |
|---|---|
| Parents in Top Decile | 7.69 |
| | (0.04) |
| Wealth ($1 mil) at Startup | |

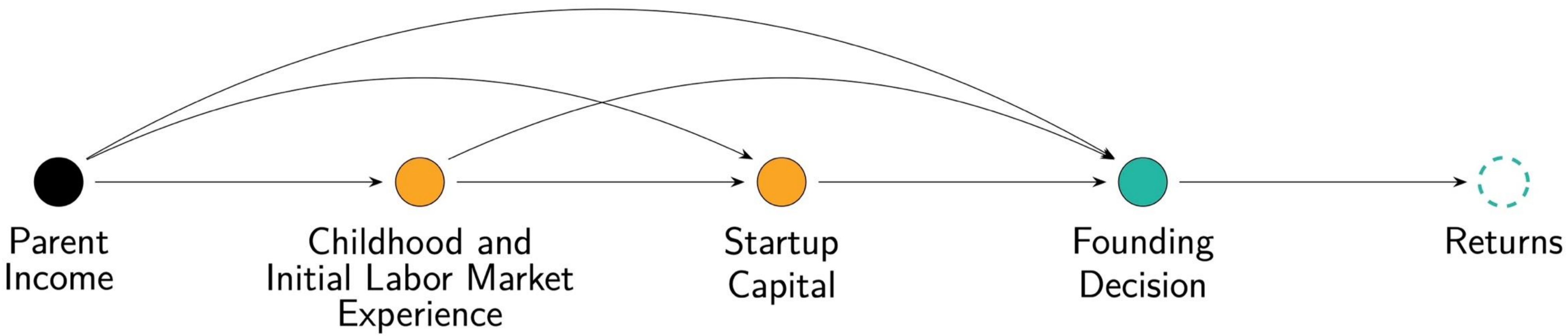
→ Wealth accounts for 1-0.18/0.769 = **77%** of effect of parent income on entrepreneurship

- **Part 1** of this lecture (mediation analysis): under what assumptions is this approach valid? How can we identify the role of mediators when those assumptions fail?

# The Entrepreneurial Pipeline

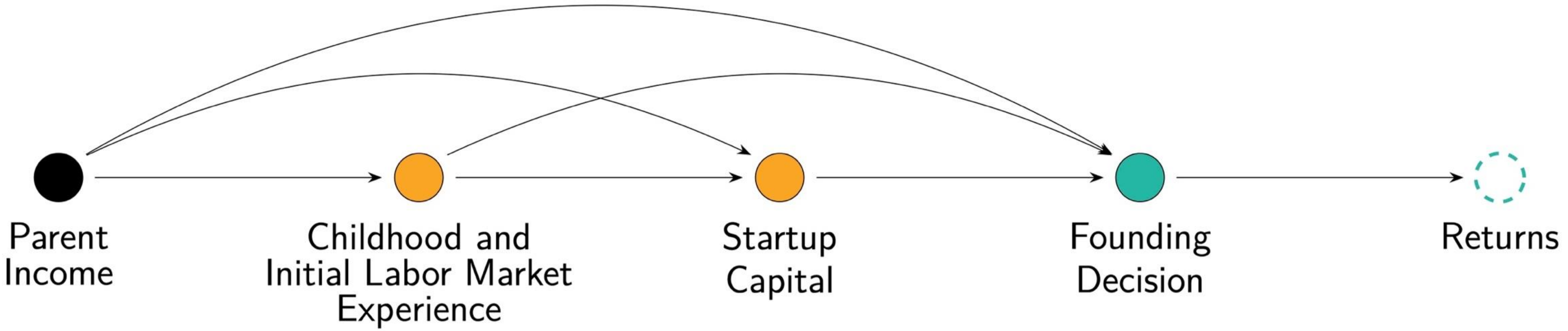


# The Entrepreneurial Pipeline: Predicting Returns




- What are the marginal returns to inducing new business startups?
- Challenge: takes many years to measure returns → difficult to answer directly given censored data (need data from childhood to a decade after business startup)

# The Entrepreneurial Pipeline: Predicting Returns




- Common heuristic practice: predict long-term returns using proxies measured at earlier stages (e.g., employment, revenues)
- **Part 2** of this lecture (surrogate indices): under what assumptions is this approach valid? How can we predict long-term outcomes when those assumptions fail?

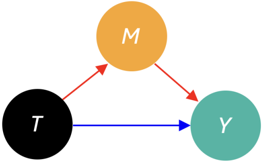
# Part 1: Mediation Analysis

## Identifying Mechanisms Underlying Treatment Effects on Primary Outcomes

# Part I. Introduction to Mediation

# Causal Mechanism as Direct and Indirect Effects

- Directed Acyclic Graph (DAG; Pearl, 2000)
  - $T \in \mathcal{T} = \{0, 1\}$: treatment
  - $M \in \mathcal{M}$: mediator (mechanism variable)
  - $Y \in \mathcal{Y}$: observed outcome




- Direct effect: Effect of $T$ on $Y$ while holding $M$ constant
- Indirect effect: Effect of $T$ on $Y$ through $M$

- DAG = Nonparametric Structural Equation Model (NPSEM)

$$Y = f_Y(M, T, \epsilon)$$
$$M = f_M(T, \eta)$$

where $\epsilon$ and $\eta$ are i.i.d. and are usually omitted from DAG

# Controlled Direct Effect (CDE)

- $Y(t, m) \in \mathcal{Y}$: potential outcome when $T = t$ and $M = m$
- Definition

$$\begin{aligned} \text{Individual:} \quad \mathrm{CDE}_i(m) &:= Y_i(1, m) - Y_i(0, m) \\ \text{Average:} \quad \overline{\mathrm{CDE}}(m) &:= \mathbb{E}[Y(1, m) - Y(0, m)] \end{aligned}$$

  for a given mediator value $m \in \mathcal{M}$

- Interpretation
  - direct effect of treatment while holding the mediator constant at $m$
  - effect of joint intervention on $T$ and $M$

- If $M$ fully captures treatment effect, CDEs will be zero for all $m$
- Potential interaction effects:

$$\mathrm{CDE}_i(m) \neq \mathrm{CDE}_i(m') \quad \text{for some } i \text{ and } m \neq m'$$

# Natural Indirect Effect (NIE)

- Definition (Robins and Greenland, 1992; Pearl, 2001)

$$\begin{aligned} \text{Individual:} \quad \text{NIE}_i(t) &:= Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \\ \text{Average:} \quad \overline{\text{NIE}}(t) &:= \mathbb{E}[Y(t, M(1)) - Y(t, M(0))] \end{aligned}$$

- Interpretation
  - effect of change in $M$ on $Y$ induced by $T$
  - change $M$ from $M(0)$ to $M(1)$ while holding $T$ at $t = 0$ or $t = 1$
  - zero treatment effect on $M$ implies zero NIE

- Represents the causal effect of $T$ on $Y$ through $M$
- Complete mediation $\rightsquigarrow$ $\text{NIE}_i = \text{TE}_i := Y_i(1, M_i(1)) - Y_i(0, M_i(0))$

# Treatment Effect Decomposition

- Natural direct effect (NDE):

$$\begin{aligned}\text{Indiviual:}\quad \text{NDE}_i(t) &:= Y_i(1, M_i(t)) - Y_i(0, M_i(t)) \\ \text{Average:}\quad \overline{\text{NDE}}(t) &:= \mathbb{E}[Y(1, M(t)) - Y(0, M(t))]\end{aligned}$$

  - change $T$ from 0 to 1 while holding $M$ constant at $M(t)$
  - causal effect of $T$ on $Y$, holding $M$ constant at its potential value that would be realized when $T = t$
- Represents all mechanisms other than through $M$
  - Complete mediation $\rightsquigarrow$ $\text{NDE}_i(t) = 0$
  - No mediation $\rightsquigarrow$ $\text{NDE}_i = \text{TE}_i$
- Effect decomposition:

$$\begin{aligned}\underbrace{Y_i(1, M_i(1)) - Y_i(0, M_i(0))}_{=\text{total effect } (\text{TE}_i)} &= \text{NIE}_i(t) + \text{NDE}_i(1-t) \\ &= \frac{1}{2}\sum_{t=0}^{1}\{\text{NIE}_i(t) + \text{NDE}_i(t)\}\end{aligned}$$

# Gender Bias and Educational Attainment (Chen et al. 2019)

- Data on Taiwanese families
  - $Y$: educational attainment of the oldest child who is female
  - $T$: gender of the second oldest child
  - $M$: number of siblings
- Gender bias
  - Direct effect: having a brother takes away resources from a female child
  - Indirect effect: having a brother leads to a smaller number of siblings and hence more resources
  - Direct and indirect effects may have opposite signs
- Causal effects of interest
  - CDE: effect of having a brother while keeping sibling size constant at a fixed value, e.g., 2
  - NDE: effect of having a brother while keeping sibling size constant at a value that would result, e.g., if the second child were male
  - NIE: effect of having a brother through sibling size

# Take-aways I

- Causal mechanism
  - how and why (not just whether) treatment affects outcome
  - understanding of causal structure (DAG $=$ NPSEM)

- Causal quantities of interest
  - Controlled direct effect (CDE)
  - Natural direct and indirect effects (NDE, NIE)
  - Effect decomposition: $\text{TE} = \text{NDE} + \text{NIE}$
  - No similar decomposition for CDE
  - Complete mediation: $\text{CDE} = \text{NDE} = 0$ and $\text{NIE} = \text{TE}$
  - No mediation: $\text{NIE} = 0$ and $\text{NDE} = \text{TE}$

# Part II. Mediation Analysis Under Pretreatment Confounding

# Linear Structural Equation Model (LSEM)

- Let's build some intuition with LSEM
- Homogeneous effects without interaction:

$$Y_i = \alpha_Y + \beta_Y T_i + \gamma_Y M_i + \epsilon_i$$
$$M_i = \alpha_M + \beta_M T_i + \eta_i$$

  - $\overline{\text{CDE}}(m) = \overline{\text{NDE}}(t) = \beta_Y$ for any $m$ and $t$
  - $\overline{\text{NIE}}(t) = \beta_M \times \gamma_Y$ for any $t$
  - CDE and NDE are identical

- Homogeneous effects with interaction:

$$Y_i = \alpha_Y + \beta_Y T_i + \gamma_Y M_i + \delta_Y T_i M_i + \epsilon_i$$

  - $\overline{\text{CDE}}(m) = \beta_Y + m\delta_Y$
  - $\overline{\text{NDE}}(t) = \beta_Y + \delta_Y(\alpha_M + t\beta_M)$
  - $\overline{\text{NIE}}(t) = \beta_M \times \gamma_Y + t\beta_M \times \delta_Y$
  - CDE is different from NDE

# LSEM with Heterogeneous Effects and Interaction

- Model

$$Y_i = \alpha_Y + \beta_Y^{(i)} T_i + \gamma_Y^{(i)} M_i + \delta_Y^{(i)} T_i M_i + \epsilon_i$$
$$M_i = \alpha_M + \beta_M^{(i)} T_i + \eta_i$$

  - $\overline{\text{CDE}}(m) = \bar{\beta}_Y + m\bar{\delta}_Y$ where $\bar{\beta}_Y = \mathbb{E}[\beta_Y^{(i)}]$ and $\bar{\delta}_Y = \mathbb{E}[\delta_Y^{(i)}]$
  - $\overline{\text{NDE}}(t) = \bar{\beta}_Y + \alpha_M \times \bar{\delta}_Y + \mathbb{E}[\delta_Y^{(i)}(t\beta_M^{(i)} + \eta_i)]$
  - $\overline{\text{NIE}}(t) = \mathbb{E}[\beta_M^{(i)} \times (\gamma_Y^{(i)} + t\delta_Y^{(i)})]$

- Heterogeneous effects may be correlated with one another
  - For example, $\mathbb{E}[\beta_M^{(i)} \times \gamma_Y^{(i)}] \neq \bar{\beta}_M \times \bar{\gamma}_Y$
  - Possible to have $\bar{\beta}_M, \bar{\gamma}_Y > 0$ but $\mathbb{E}[\beta_M^{(i)} \times \gamma_Y^{(i)}] < 0$ or vice versa

- $\bar{\beta}_M$, $\bar{\gamma}_Y$, $\bar{\delta}_Y$, etc. are identifiable under exogeneity
- But, $\mathbb{E}[\beta_M^{(i)} \times \gamma_Y^{(i)}]$, $\mathbb{E}[\beta_M^{(i)} \times \delta_Y^{(i)}]$, etc. are unidentifiable
- This is essentially a problem of unobserved pre-treatment confounding

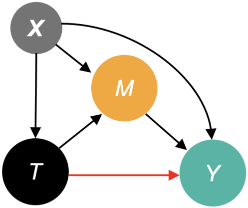
# Identification of CDE with Pre-treatment Confounding

- Assumptions:
  1. Unconfoundedness

$$\{Y_i(t,m), M_i(t)\}_{t,m} \perp\!\!\!\perp T_i \mid \boldsymbol{X}_i = \boldsymbol{x}$$
$$\{Y_i(t,m)\}_m \perp\!\!\!\perp M_i \mid T_i = t, \boldsymbol{X}_i = \boldsymbol{x}$$

  2. Overlap

$$P(T_i = t \mid \boldsymbol{X}_i = \boldsymbol{x}) > 0$$
$$P(M_i = m \mid T_i = t, \boldsymbol{X}_i = \boldsymbol{x}) > 0$$



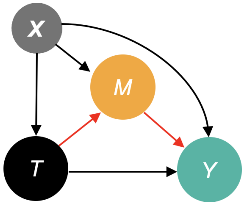

- Identification:

$$\overline{\text{CDE}}(m) = \sum_{\boldsymbol{X}} \left(\mathbb{E}[Y \mid T=1, M=m, \boldsymbol{X}] - \mathbb{E}[Y \mid T=0, M=m, \boldsymbol{X}]\right) P(\boldsymbol{X})$$

# Identification of NDE/NIE with Pretreatment Confounding

- Replace the following assumption

$$\{Y_i(t,m)\}_m \perp\!\!\!\perp \underbrace{M_i}_{=M_i(t)} \mid T_i = t, \boldsymbol{X}_i = \boldsymbol{x}$$




with the cross-world independence

$$\{Y_i(t',m)\}_{t',m} \perp\!\!\!\perp M_i(t) \mid T_i = t, \boldsymbol{X}_i = \boldsymbol{x}$$

- Additional conditional independence between $Y_i(t',m)$ and $M_i(t)$
- Identification (Imai et al. 2010)

$$\begin{aligned}
\overline{\text{NDE}}(t) &= \sum_{M,\boldsymbol{X}} (\mathbb{E}[Y \mid M, T=1, \boldsymbol{X}] - \mathbb{E}[Y \mid M, T=0, \boldsymbol{X}]) \\
&\qquad \times P(M \mid T=t, \boldsymbol{X}) P(\boldsymbol{X}) \\
\overline{\text{NIE}}(t) &= \sum_{M,\boldsymbol{X}} \mathbb{E}[Y \mid M, T=t, \boldsymbol{X}] \\
&\qquad \times \{P(M \mid T=1, \boldsymbol{X}) - P(M \mid T=0, \boldsymbol{X})\} P(\boldsymbol{X})
\end{aligned}$$

# Experimental Identification (Imai et al. 2013)

- Parallel design
  1. Randomize $T$ and observe $M$ and $Y$
  2. Randomize $T$ and $M$ and observe $Y$
- We can identify $P(M(t))$, $P(Y(t, M(t))$, and $P(Y(t, m))$
- CDE is identified
- NDE/NIE is still not identifiable:
  - randomization cannot break correlation between $Y(t', m)$ and $M(t)$
  - partial identification: sharp bounds contain zero
- Crossover design
  1. Randomize $T$ and observe $M$ and $Y$
  2. On the same sample, change $T$ to the opposite condition while holding $M$ at the same value and observe $Y$
- $Y(t, M(t))$, $M(t)$, and $Y(1 - t, M(t))$ are observable
- Additional assumption: no carryover effects
- NDE/NIE is identifiable

# No Interaction Assumption

- No individual-level interaction

$$Y_i(1,m) - Y_i(0,m) = Y_i(1,m') - Y_i(0,m')$$

  - $\text{NDE}_i(t) = \text{CDE}_i(m) = \text{CDE}_i$
  - $\overline{\text{NDE}}(t) = \overline{\text{CDE}}(m) = \overline{\text{CDE}}$
  - $\overline{\text{NIE}}(t) = \text{ATE} - \overline{\text{NDE}}$

- Testable implication:

$$\mathbb{E}[Y_i(1,m) - Y_i(0,m) \mid \boldsymbol{X}_i = \boldsymbol{x}] = \mathbb{E}[Y_i(1,m') - Y_i(0,m') \mid \boldsymbol{X}_i = \boldsymbol{x}]$$

  for all $\boldsymbol{x}$

- NDE/NIE is identifiable so long as CDE can be identified
- Experimental identification, and identification with pretreatment and posttreatment confounding are all possible

# Estimation of Natural Direct and Indirect Effects

- Recall the identification formula (NIE)

$$\overline{\mathsf{NIE}}(t) = \sum_{M,\boldsymbol{X}} \mathbb{E}[Y \mid M, T = t, \boldsymbol{X}] \times \{P(M \mid T = 1, \boldsymbol{X}) - P(M \mid T = 0, \boldsymbol{X})\} P(\boldsymbol{X})$$

1. predict $M$ given each treatment value: $\{M_i(1), M_i(0)\}$
2. predict $Y$ by first setting $T_i = t$ and $M_i = M_i(0)$, and then $T_i = t$ and $M_i = M_i(1)$: $\{Y_i(t, M_i(0)), Y_i(t, M_i(1))\}$
3. compute the average difference between two predicted outcomes

- Estimation of NDE is similar

$$\overline{\mathsf{NDE}}(t) = \sum_{M,\boldsymbol{X}} (\mathbb{E}[Y \mid M, T = 1, \boldsymbol{X}] - \mathbb{E}[Y \mid M, T = 0, \boldsymbol{X}]) \times P(M \mid T = t, \boldsymbol{X}) P(\boldsymbol{X})$$

- One can also do: $\overline{\mathsf{NDE}}(t) = \mathsf{ATE} - \overline{\mathsf{NIE}}(1 - t)$

# Weighting Methods for NDE and NIE

- Three weighting formulae:

$$\mathbb{E}[Y(t, M(t'))] = \mathbb{E}\left[\underbrace{\frac{1\{T = t'\}}{\Pr(T = t' \mid \boldsymbol{X})}}_{\text{weighting to get } P(M(t')|\boldsymbol{X})} \times \mathbb{E}[Y \mid M, T = t, \boldsymbol{X}]\right]$$

$$= \mathbb{E}\left[\underbrace{\frac{1\{T = t\}}{\Pr(T = t \mid \boldsymbol{X}_i)}}_{\text{treatment weighting}} \times \underbrace{\frac{P(M \mid T = t', \boldsymbol{X})}{P(M \mid T_i = t, \boldsymbol{X}_i)}}_{\text{mediator weighting}} \times Y\right]$$

$$= \mathbb{E}\left[\frac{1\{T = t\}}{\Pr(T = t \mid M, \boldsymbol{X})} \times \frac{\Pr(T = t' \mid M, \boldsymbol{X})}{\Pr(T = t' \mid \boldsymbol{X})} \times Y\right]$$

  - The third expression follows from Bayes rule
  - Useful when the mediator is high-dimensional

- Multiply-robust semiparametric estimator (Tchetgen Tchetgen and Shpitser, 2012); Double machine learning (Farbmacher et al. 2022)

# Sensitivity Analysis

- Examine the robustness of empirical findings to the violation of untestable assumptions
- How large a departure from the key identification assumption must occur for the conclusions to no longer hold?
- Potential existence of unobserved pretreatment confounding ($T$ is assumed to be unconfounded)

$$\{Y_i(t', m)\}_{t',m} \not\perp\!\!\!\perp M_i(t) \mid T_i = t, \boldsymbol{X}_i = \boldsymbol{x}$$

- Recall LSEM (or more generally, additive semiparametric model)

$$Y_i = \alpha_Y + \beta_Y T_i + \gamma_Y M_i + \underbrace{\lambda_\epsilon U_i + \tilde{\epsilon}_i}_{=\epsilon_i}$$

$$M_i = \alpha_M + \beta_M T_i + \underbrace{\lambda_\eta U_i + \tilde{\eta}_i}_{=\eta_i}$$

- How much does $U_i$ have to matter for the results to go away?

# Sensitivity Parameters

- $R^2$ parameterization
  1. Proportion of previously unexplained variance explained by $U_i$

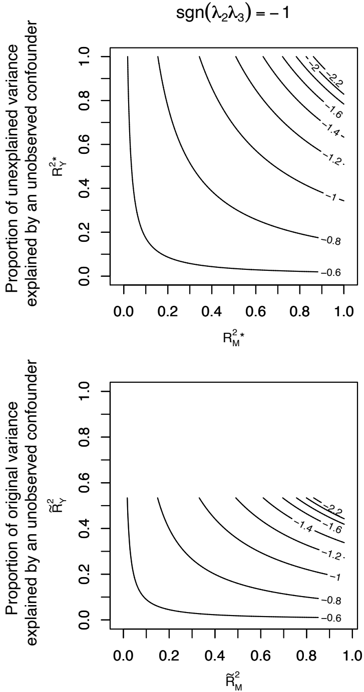
$$R_M^{2*} \equiv \frac{\mathbb{V}(\lambda_\eta U_i)}{\mathbb{V}(\eta_i)} \quad \text{and} \quad R_Y^{2*} \equiv \frac{\mathbb{V}(\lambda_\epsilon U_i)}{\mathbb{V}(\epsilon_i)}$$

  2. Proportion of original variance explained by $U_i$

$$\widetilde{R}_M^2 \equiv \frac{\mathbb{V}(\lambda_\eta U_i)}{\mathbb{V}(M_i)} \quad \text{and} \quad \widetilde{R}_Y^2 \equiv \frac{\mathbb{V}(\lambda_\epsilon U_i)}{\mathbb{V}(Y_i)}$$

- We also need to specify the direction of effects:

$$\operatorname{sgn}(\lambda_\eta \lambda_\epsilon) = \begin{cases} 1 & \text{if same direction} \\ -1 & \text{if opposite directions} \end{cases}$$


sgn(λ2λ3) = −1
sgn(λ2λ3) = 1
Proportion of unexplained variance explained by an unobserved confounder
Proportion of original variance explained by an unobserved confounder
R2Y*
R2M*
R̃2Y
R̃2M

# Gender Bias Application: Standard Mediation Analysis

- The original analysis fits LSEM with interaction

$$
\begin{aligned}
Y_i &= \alpha_Y + \beta_Y T_i + \gamma_Y M_i + \delta_Y T_i M_i + \boldsymbol{\xi}_Y^\top \mathbf{X}_i + \epsilon_i \\
M_i &= \alpha_M + \beta_M T_i + \boldsymbol{\xi}_M^\top \mathbf{X}_i + \eta_i
\end{aligned}
$$

  - $Y_i$: university admission
  - $T_i$: the second child is male
  - $M_i$: sibling size is greater than two

- Estimates:

| | |
|---|---|
| $\widehat{\text{ATE}}$ | 0.0020 (0.0013) |
| $\widehat{\overline{\text{CDE}}(\overline{M})}$ | −0.0010 (0.0014) |
| $\widehat{\overline{\text{NDE}}(1)}$ | −0.0001 (0.0014) |
| $\widehat{\overline{\text{NIE}}(0)}$ | 0.0022 (0.0005) |

- Also, fits a random coefficient model to address heterogeneity
- Sensitivity analysis based on a semiparametric random coefficient model (Imai and Yamamoto, 2013)

# Take-aways II

- Linear structural equation model
  - two key assumptions beyond exogeneity:
    1. homogeneous effects
    2. no interaction
  - CDE $=$ NDE under those assumptions
  - Relaxing these assumptions lead to different interpretations and identification issues

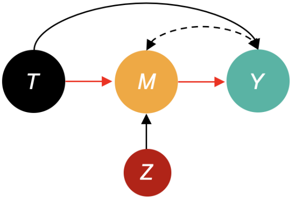
- Nonparametric identification analysis under pretreatment confounding
  - CDE is identifiable under standard exogeneity
  - NDE/NIE requires cross-world independence
  - alternatively, CDE $=$ NDE if we assume no individual-level interaction

- Difficulty of identification
  - even when $M$ is randomized, NIE/NDE are unidentifiable
  - sensitivity analysis plays an important role for assessing robustness

# Part III. Coping with Identification Difficulties

# Instrumenting the Mediator

- Instrument: $Z_i$
- Mediator: $M_i(t, z)$
- Exclusion restriction
$$Y_i(t, m, z) = Y_i(t, m)$$
- NPSEM:
$$\begin{aligned} Y &= f_Y(M, T, \epsilon) \\ M &= f_M(T, Z, \eta) \end{aligned} \quad \text{where} \quad \epsilon \not\perp\!\!\!\perp \eta$$




- If $M$ and $Z$ are continuous, we can use the control function approach (Imbens and Newey, 2009)
  1. Independence: $Z \perp\!\!\!\perp (\epsilon, \eta)$
  2. Monotonicity: $\eta$ is a continuous scalar variable with its CDF and $f_M(\cdot, \cdot, \eta)$ being strictly monotonic in $\eta$
- Then, $(M, T) \perp\!\!\!\perp \epsilon \mid C$ where $C = F_{M|T,Z}(T, Z) = F_\eta(\eta)$
  - Recall the control function approach to 2SLS
  - Regress $Y$ on $M$, $T$ and the first stage residual $\hat{\eta}$
- Extension: an additional instrument for $T$ (Florich and Huber, 2017)

# Gender Bias Application: IV Analysis

- Instrument $Z$: twinning at the second birth

$$M_i = \alpha_M + \beta_M T_i + \zeta_M Z_i + \lambda_M T_i Z_i + \boldsymbol{\xi}_M^\top \boldsymbol{X}_i + \eta_i$$

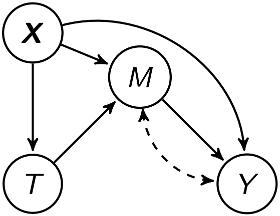
- Assumptions:
  - exogenous instrument: twinning is random conditional on $\boldsymbol{X}$
  - exclusion restriction: twinning affects $Y$ only through $M$

- Findings:

| | **Standard analysis** | **IV analysis** |
|---|---|---|
| $\widehat{\text{ATE}}$ | 0.0020 (0.0013) | 0.0021 (0.0013) |
| $\widehat{\overline{\text{CDE}}(\overline{M})}$ | −0.0010 (0.0014) | −0.0092 (0.0061) |
| $\widehat{\overline{\text{NDE}}(1)}$ | −0.0001 (0.0014) | −0.0203 (0.0106) |
| $\overline{\text{NIE}}(0)$ | 0.0022 (0.0005) | 0.0224 (0.0105) |

# Complete Mediation Analysis (Kwon and Roth 2024)

- Complete mediation: $Y_i(t, m) = Y_i(m)$
- Assumption: No unobserved confounding between $T$ and $M$ and between $T$ and $Y$
- Possible unobserved confounding between $M$ and $Y$




- Under monotonicity $M_i(1) \geq M_i(0)$ (in the binary mediator case), we can use the following test of instrumental validity

$$P(Y, M = 0 \mid T = 0, \boldsymbol{X}) \geq P(Y, M = 0 \mid T = 1, \boldsymbol{X})$$
$$P(Y, M = 1 \mid T = 1, \boldsymbol{X}) \geq P(Y, M = 1 \mid T = 0, \boldsymbol{X})$$

- Randomized experiment: test of complete mediation
- Observational study: unobserved confounding between $T$ and $Y$ can also lead to the rejection of the null hypothesis

* *This slide has been revised to more accurately represent the method proposed by Kwon and Roth (2024).*

# Implicit Mediation

- What if we want to avoid the untestable assumptions at all costs?
- What can we infer from $\text{ATE}_M$ and $\text{ATE}_Y$ that are identifiable without such assumptions?

**Table 1.** Possible Implicit-Mediation Findings

| Result | Inference | Rationale |
|---|---|---|
| $X$ affects $M$ and $Y$ | $M$ may be a mediator. | $X$ appears to influence $M$, and this effect seems to coincide with a change in $Y$, as would be expected if $M$ were a mediator. |
| $X$ affects $M$ but not $Y$ | $M$ appears not to be a mediator. | Although $X$ affects $M$, this effect seems not to have any consequences for $Y$. |
| $X$ affects $Y$ but not $M$ | Some variable other than $M$ may be a mediator. | $X$ appears to have no effect on $M$, which means that $X$'s apparent effect on $Y$ is not due to changes in $M$. |
| $X$ affects neither $M$ nor $Y$ | There seem to be no indirect pathways from $X$ to $Y$ through $M$ or other mediators. | $X$ seems not to set in motion any causal effects. |

Bullock and Green (2021)

# Identification Analysis of Implicit Mediation

- Questions:
  1. Does $\text{ATE}_M = 0$ imply $\overline{\text{NIE}} = 0$ and/or $\overline{\text{NDE}} \neq 0$?
  2. Does $\text{ATE}_M > 0$ and $\text{ATE}_Y > 0$ imply $\overline{\text{NIE}} > 0$?
- No! Recall even the no-assumption bounds from the parallel experiment design always contain zero
- The decomposition under a binary mediator:

$$\overline{\text{NIE}}(t) = \underbrace{\mathbb{E}[Y_i(t,1) - Y_i(t,0) \mid M(1) = 1, M(0) = 0]}_{\text{ATE of } M \text{ on } Y \text{ for compliers}} \cdot p_{10} - \underbrace{\mathbb{E}[Y_i(t,1) - Y_i(t,0) \mid M(1) = 0, M(0) = 1]}_{\text{ATE of } M \text{ on } Y \text{ for defiers}} \cdot p_{01}$$

where $p_{m_1 m_0} = \Pr(M(1) = m_1, M(0) = m_0)$

- Cross-world assumption or homogeneity assumption leads to the usual product estimator

$$\overline{\text{NIE}}(t) = \underbrace{\mathbb{E}[Y_i(t,1) - Y_i(t,0)]}_{=\text{ATE of } M \text{ on } Y} \times \underbrace{(p_{10} - p_{01})}_{=\text{ATE}_M}$$

# Identification under Monotonicity

(Blackwell et al. 2024; Kwon and Roth 2024)

- Monotonicity assumption (no defier) yields:

$$\overline{\mathrm{NIE}}(t) = \mathbb{E}[Y_i(t,1) - Y_i(t,0) \mid M(1) = 1, M(0) = 0] \cdot p_{10}$$

- Sharp bounds

$$\max\{-\mathrm{ATE}_M, -q_{1-t,t|t}\} \leq \overline{\mathrm{NIE}}(t) \leq \min\{\mathrm{ATE}_M, q_{tt|t}\}$$

  where $q_{ym|t} = \Pr(Y = y, M = m \mid T = t)$

- Two fundamental difficulties remain:
  1. effect heterogeneity
  2. endogeneity of mediator
- Even under an additional assumption of $\mathbb{E}[Y(t,1) - Y(t,0)] > 0$, the sharp bounds still contain zero

# Take-aways III

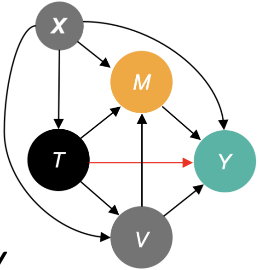
- Instrumental variable approach
  - addressing the endogeneity problem
  - the instrument must be exogeneous
  - exclusion restriction needs to be satisfied
  - nonparametric estimation is possible

- Complete mediation
  - hypothesis testing approach
  - no need to assume the exogeneity of mediator
  - no unobserved confounding between $T$ and $Y$ (satisfied in RCT)

- Implicit mediation
  - an attempt to sidestep assumptions
  - not informative even about the signs of NIE/NDE
  - monotonicity is not sufficient

# Part IV. Mediation Analysis under Posttreatment Confounding

# Identification of CDE with Posttreatment Confounding




- Replace the following assumption

$$\{Y_i(t,m)\}_m \perp\!\!\!\perp M_i \mid T_i = t, \boldsymbol{X}_i = \boldsymbol{x},$$

with

$$\{Y_i(t,m)\}_m \perp\!\!\!\perp M_i \mid \boldsymbol{V}_i = \boldsymbol{v}, T_i = t, \boldsymbol{X}_i = \boldsymbol{x}$$

- Post-treatment bias: cannot simply control for $\boldsymbol{V}$

$$\overline{\text{CDE}}(m) \neq \sum_{\boldsymbol{X},\boldsymbol{V}} (\mathbb{E}[Y \mid T=1, M=m, \boldsymbol{X}, \boldsymbol{V}] - \mathbb{E}[Y \mid T=0, M=m, \boldsymbol{X}, \boldsymbol{V}])P(\boldsymbol{X}, \boldsymbol{V})$$

- Identification: model $\boldsymbol{V}$ given $T$ and $\boldsymbol{X}$

$$\overline{\text{CDE}}(m) = \sum_{\boldsymbol{X},\boldsymbol{V}} \{\mathbb{E}[Y \mid T=1, M=m, \boldsymbol{X}, \boldsymbol{V}]P(\boldsymbol{V} \mid T=1, \boldsymbol{X}) - \mathbb{E}[Y \mid T=0, M=m, \boldsymbol{X}, \boldsymbol{V}]P(\boldsymbol{V} \mid T=0, \boldsymbol{X})\} P(\boldsymbol{X})$$

# Estimation of Controlled Direct Effects

1. Directly use the identification formula

$$\bar{\xi}(m) = \sum_{\boldsymbol{X},\boldsymbol{V}} \{\mathbb{E}[Y \mid T=1, M=m, \boldsymbol{X}, \boldsymbol{V}] P(\boldsymbol{V} \mid T=1, \boldsymbol{X}) \\ -\mathbb{E}(Y \mid T=0, M=m, \boldsymbol{X}, \boldsymbol{V}) P(\boldsymbol{V} \mid T=0, \boldsymbol{X})\} \, P(\boldsymbol{X})$$

   - regression of $Y$ on $T, M, \boldsymbol{X}, \boldsymbol{V}$
   - model $\boldsymbol{V}$ given $T$ and $\boldsymbol{X}$ $\rightsquigarrow$ difficult if $\boldsymbol{V}$ is high-dimensional

2. Marginal structural models (Robins et al. 2000)

$$\mathbb{E}[Y(t,m)] = \mathbb{E}\left[\underbrace{\frac{1\{T=t, M=m\}}{\Pr(T=t \mid \boldsymbol{X})}}_{\text{IPW for treatment}} \cdot \underbrace{\frac{1}{\Pr(M=m \mid T=t, \boldsymbol{X}, \boldsymbol{V})}}_{\text{IPW for mediator given treatment}} \times Y\right]$$

   - no need to model $\boldsymbol{V}$
   - covariate balancing methods are also available (Imai and Ratkovic, 2015)

# Identification of NDE/NIE with Posttreatment Confounding

- Identification is impossible with *observed* posttreatment confounding
- Consider the following NPSEM

$$\begin{aligned} Y &= f_Y(M, \boldsymbol{V}, T, \epsilon) \\ M &= f_M(\boldsymbol{V}, T, \eta) \\ \boldsymbol{V} &= f_{\boldsymbol{V}}(T, \xi) \end{aligned}$$

- Cross-world independence cannot hold

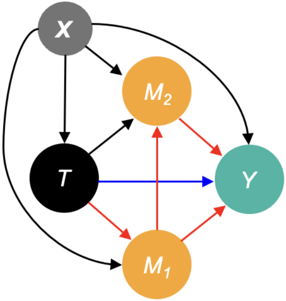
$$\underbrace{\boldsymbol{V}(1)}_{=f_{\boldsymbol{V}}(1,\xi)} \not\perp\!\!\!\perp \underbrace{\boldsymbol{V}(0)}_{=f_{\boldsymbol{V}}(0,\xi)} \quad \Longrightarrow \quad Y(t', m, \boldsymbol{V}(t'), \epsilon) \not\perp\!\!\!\perp M(t, \boldsymbol{V}(t), \eta)$$

- Conditioning on $T$ and $\boldsymbol{V}$ does not solve this problem

# Multiple Causally Related Mediators

- Same as the posttreatment confounding setting
- Path specific effects
  1. $T \to Y$
  2. $T \to M_1 \to Y$
  3. $T \to M_2 \to Y$
  4. $T \to M_1 \to M_2 \to Y$
- Combined effect:

$$\begin{aligned} & T \to M_1 \rightsquigarrow Y \\ = & (T \to M_1 \to Y) + (T \to M_1 \to M_2 \to Y) \end{aligned}$$




- Generalized cross-world independence assumptions:
  1. $\{M_{1i}(t), M_{2i}(t, m_1), Y_i(t, m_1, m_2)\}_{t, m_1, m_2} \perp\!\!\!\perp T_i \mid \boldsymbol{X}_i = \boldsymbol{x}$
  2. $\{M_{2i}(t', m_1), Y_i(t', m_1, m_2)\}_{t', m_1, m_2} \perp\!\!\!\perp M_{1i}(t) \mid T_i = t, \boldsymbol{X}_i = \boldsymbol{x}$
  3. $\{Y_i(t', m_1, m_2)\}_{t', m_2} \perp\!\!\!\perp M_{2i}(t, m_1) \mid M_{1i} = m_1, T_i = t, \boldsymbol{X}_i = \boldsymbol{x}$
- Identifiable decomposition:

$$\text{ATE} = (T \to Y) + (T \to M_2 \to Y) + (T \to M_1 \rightsquigarrow Y)$$

# Interventional Direct and Indirect Effects (IDE and IIE)

- $\mathcal{P}_{M(t)}$: interventional distribution that independently generates $M(t)$
- Definition (Geneletti, 2007; Lok, 2016)

$$\text{Individual:} \begin{cases} \text{IIE}_i(t) &= Y_i(t, \mathcal{P}_{M(1)}) - Y_i(t, \mathcal{P}_{M(0)}) \\ \text{IDE}_i(t) &= Y_i(1, \mathcal{P}_{M(t)}) - Y_i(0, \mathcal{P}_{M(t)}) \end{cases}$$

$$\text{Average:} \begin{cases} \overline{\text{IIE}}(t) &= \mathbb{E}[Y(t, \mathcal{P}_{M(1)}) - Y(t, \mathcal{P}_{M(0)})] \\ \overline{\text{IDE}}(t) &= \mathbb{E}[Y(t, \mathcal{P}_{M(1)}) - Y(t, \mathcal{P}_{M(0)})] \end{cases}$$

- Interpretation
  - similar to NIE and NDE
  - IDE is a function of CDE:

$$\text{IDE}_i(t) = \sum_m \text{CDE}_i(m) \times P(M(t) = m)$$

  - no mediation: zero treatment effect on $M$ implies zero IIE
- Effect decomposition

$$\underbrace{Y_i(1, \mathcal{P}_{M(1)}) - Y_i(0, \mathcal{P}_{M(0)})}_{\text{Interventional Total Effect (ITE)} \neq \text{TE}} = \text{IIE}_i(t) + \text{IDE}_i(1-t)$$

# Identification of IDE and IIE

- Once CDE is identified, we can identify IDE:

$$\overline{\text{IDE}}(t) = \sum_m \overline{\text{CDE}}(m) P(M(t) = m)$$

- IIE is also identifiable:

$$\overline{\text{IIE}}(t) = \sum_m \mathbb{E}[Y(t,m)] \{P(M(1) = m) - P(M(0) = m)\}$$

- Effect decomposition

$$\underbrace{\mathbb{E}[Y(1, \mathcal{P}_{M(1)}) - Y(0, \mathcal{P}_{M(0)})]}_{\neq \mathbb{E}[Y(1, M(1)) - Y(0, M(0))]} = \overline{\text{IDE}}(t) + \overline{\text{IIE}}(1-t)$$

- Complete mediation: $\overline{\text{IDE}} = 0$
- Identification is possible with observed pretreatment and posttreatment confounding
- Experimental identification via parallel design is also possible

# Take-aways IV

- Posttreatment confounding
  - CDE can be identified under exogeneity
  - estimation of CDE requires marginalizing posttreatment confounders
  - NIE/NDE are not identifiable under exogeneity
  - Different decomposition is identifiable under cross-world independence

- Alternative estimands
  - interventional direct and indirect effects (IDE/IIE)
  - interventional distribution on $M$
  - enables decomposition of alternative total effect
  - identification of CDE implies that of IDE/IIE

# Conclusion, Resources, and References

# Concluding Remarks on Causal Mechanisms

- Study of causal mechanisms is essential but challenging
- Triangulation of evidence is necessary
  - causal quantities
    - CDE
    - NDE/NIE, path specific effects
    - IDE/IIE
  - causal identification strategies
    - selection on observables
    - instrumental variables
    - experimental designs
    - partial identification
  - statistical methodologies
    - weighting and regression
    - sensitivity analysis
    - nonparametric modeling and machine learning

# Resources

- Statistical software:
  - mediation (R and Stata)
  - Valeri and VanderWeele macros (SPSS, SAS, Stata)

- Review article by an economist:

Huber, Martin (2020). "Mediation Analysis". Handbook of Labor, Human Resources and Population Economics. Ed. by Klaus F. Zimmermann. Cham: Springer.

- Monographs:

VanderWeele, Tyler J. (2015). Explanation in Causal Inference: Methods for Mediation and Interaction. New York: Oxford University Press.
Wodtke, Geoffrey T. and Xiang Zhou (Forthcoming). Causal Mediation Analysis. Cambridge University Press.

# Works Cited I


Blackwell, Matthew, Ruofan Ma, and Aleksei Opacic (2024). "Assumption Smuggling in Intermediate Outcome Tests of Causal Mechanisms". arXiv preprint arXiv:2407.07072.

Bullock, John G. and Donald P. Green (2021). "The Failings of Conventional Mediation Analysis and a Design-Based Alternative". Advances in Methods and Practices in Psychological Science 4.4, pp. 1–14.

Chen, Stacey H., YenâChien Chen, and JinâTan Liu (2019). "The Impact of Family Composition on Educational Achievement". Journal of Human Resources 54.1, pp. 122–170.

Farbmacher, Helmut et al. (Jan. 2022). "Causal mediation analysis with double machine learning". The Econometrics Journal 25.2, pp. 277–300.

# Works Cited II


Frölich, Markus and Martin Huber (2017). "Direct and Indirect Treatment Effects–Causal Chains and Mediation Analysis with Instrumental Variables". Journal of the Royal Statistical Society Series B: Statistical Methodology 79.5, pp. 1645–1666.

Geneletti, Sara (2007). "Identifying direct and indirect effects in a non-counterfactual framework". Journal of the Royal Statistical Society, Series B (Statistical Methodology) 69.2, pp. 199–215.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto (2010). "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects". Statistical Science 25.1, pp. 51–71.

# Works Cited III


Imai, Kosuke and Marc Ratkovic (2015). "Robust Estimation of Inverse Probability Weights for Marginal Structural Models". Journal of the American Statistical Association 110.511, pp. 1013–1023.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto (2013). "Experimental Designs for Identifying Causal Mechanisms (with discussions)". Journal of the Royal Statistical Society, Series A (Statistics in Society) 176.1, pp. 5–51.

Imai, Kosuke and Teppei Yamamoto (Spring 2013). "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments". Political Analysis 21.2, pp. 141–171.

Imbens, Guido W. and Whitney K. Newey (2009). "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity". Econometrica 77.5, pp. 1481–1512.

# Works Cited IV


Kwon, Soonwoo and Jonathan Roth (Apr. 2024). "Testing Mechanisms". arXiv preprint arXiv:2404.11739.

Lok, Judith J. (2016). "Defining and Estimating Causal Direct and Indirect Effects When Setting the Mediator to Specific Values Is Not Feasible". Statistics in Medicine 35.22, pp. 4008–4020.

Pearl, Judea (2000). Causality: Models, Reasoning, and Inference. New York: Cambridge University Press.

— (2001). "Direct and Indirect Effects". Proc. of the 17th Conference on Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, pp. 411–420.

Robins, James M. and Sander Greenland (Mar. 1992). "Identifiability and exchangeability for direct and indirect effects". Epidemiology 3.2, pp. 143–155.

# Works Cited V

Robins, James M., Miguel Ángel Hernán, and Babette Brumback (2000). "Marginal Structural Models and Causal Inference in Epidemiology". Epidemiology 11.5, pp. 550–560.

Tchetgen, Eric J. and Ilya Shpitser (2012). "Semiparametric Theory for Causal Mediation Analysis: Efficiency Bounds, Multiple Robustness, and Sensitivity Analysis". Annals of Statistics 40.3, pp. 1816–1845.

# Part 2: Surrogate Indices

## Identifying Treatment Effects on Primary Outcomes Using Mediators

# Identifying Long-Term Treatment Effects Using Proxies

- In this part of the lecture, focus on three methods that formalize when and how we can use proxies to predict treatment effects on a primary outcome of interest

1. Surrogates: predict long-term outcomes using intermediate outcomes
   [e.g., Prentice 1989, VanderWeele 2013, Athey, Chetty, Imbens, Kang 2025]
2. Remotely Sensed Variables: predict outcomes using post-treatment proxies
   [Rambachan, Singh, Viviano 2025]
3. Experimental Selection Correction: control for selection using intermediate outcomes
   [Athey, Chetty, Imbens 2025]

# Surrogates

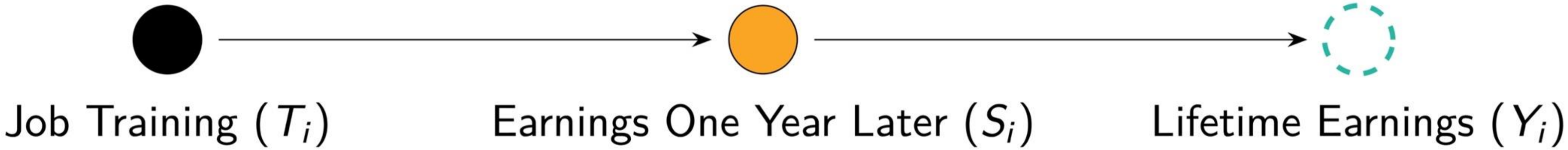
## Predicting Treatment Effects Using Intermediate Outcomes

# Estimating Long-Term Impacts of Interventions

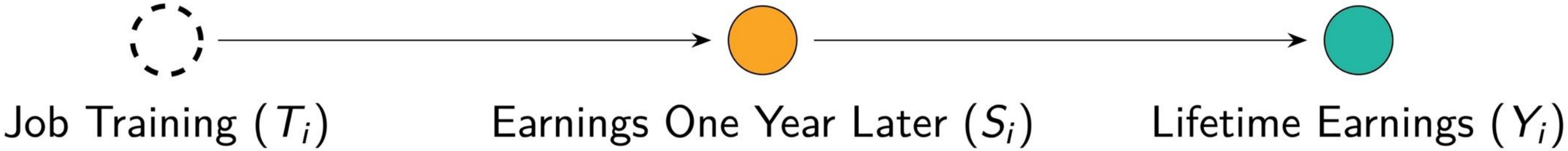

- How do job training programs affect career trajectories?
- Would take decades to measure full impact on lifetime earnings
- Intuitive, widely used heuristic: use short-term proxies to predict long-term outcomes
  - If job training improves earnings one year later, and earnings are highly serially correlated over time, can we extrapolate to predict long-term earnings gains?

# Surrogate Model: Setup

**Experimental Data (e.g., Job Training Experiment)**
$G_i = E$




**Observational Data (e.g., Administrative Tax Records)**
$G_i = O$



# Identification Using Surrogates: Key Assumptions

**Assumption 1.** Unconfounded Experiment (or Quasi-Experiment)

$$T_i \perp\!\!\!\perp (Y_i(0), Y_i(1), S_i(0), S_i(1)) \mid G_i = E$$

**Assumption 2.** Comparability of Experimental and Observational Samples

$$Y_i \mid S_i, G_i = E \sim Y_i \mid S_i, G_i = O$$

# Comparability Assumption in Linear Models

$$Y_i \mid S_i, G_i = E \sim Y_i \mid S_i, G_i = O$$

- Suppose the conditional expectation of $Y_i \mid S_i$ is linear in both samples:

**Experimental Data:** $\mathrm{E}[Y_i \mid S_i, G_i = E] = \alpha_E + \beta^E \cdot S_i$

**Observational Data:** $\mathrm{E}[Y_i \mid S_i, G_i = O] = \alpha_O + \beta^O \cdot S_i$

- In this case, comparability requires that the slopes are the same in both samples:

$$\beta^E = \beta^O$$

# Identification Using Surrogates: Key Assumptions

**Assumption 1.** Unconfounded Experiment (or Quasi-Experiment)

$$T_i \perp\!\!\!\perp (Y_i(0), Y_i(1), S_i(0), S_i(1)) \mid G_i = E$$

**Assumption 2.** Comparability of Experimental and Observational Samples

$$Y_i \mid S_i, G_i = E \sim Y_i \mid S_i, G_i = O$$

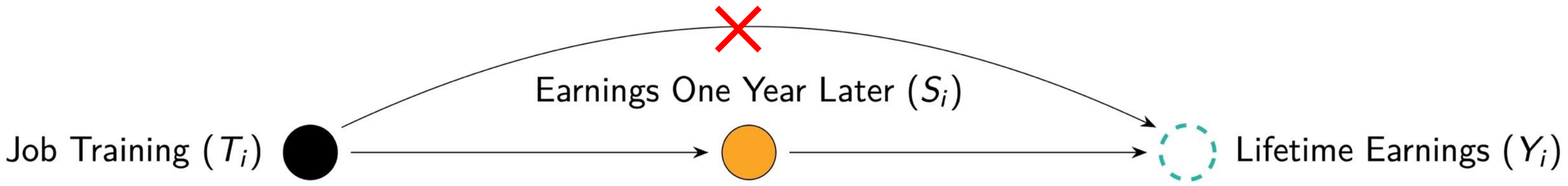
**Assumption 3.** Surrogacy: $T_i$ has no effect on long-term outcome $Y_i$ once we control for $S_i$

$$T_i \perp\!\!\!\perp Y_i \mid S_i, G_i = E$$

# Violations of Surrogacy Assumption

$$T_i \perp\!\!\!\perp Y_i \mid S_i, G_i = E$$

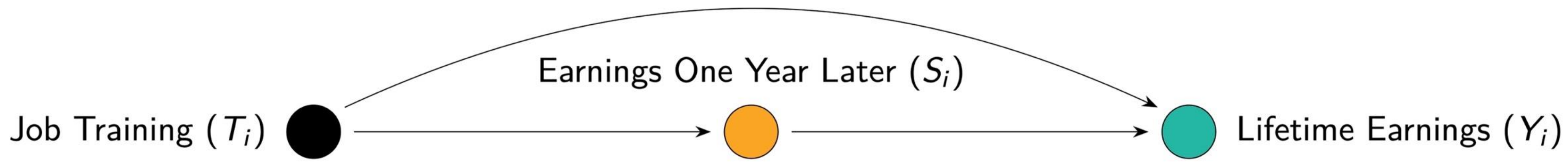
**Experimental Data**



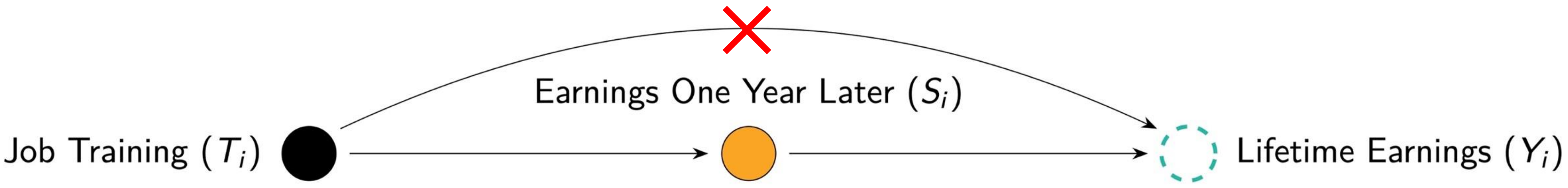

**Source:** VanderWeele (2013); Athey, Chetty, Imbens, Kang (2025)

# Surrogacy vs. Mediation

**Mediation:** Allow treatment to have direct effect on $Y_i$ and estimate direct vs. indirect effects using data on $Y_i$



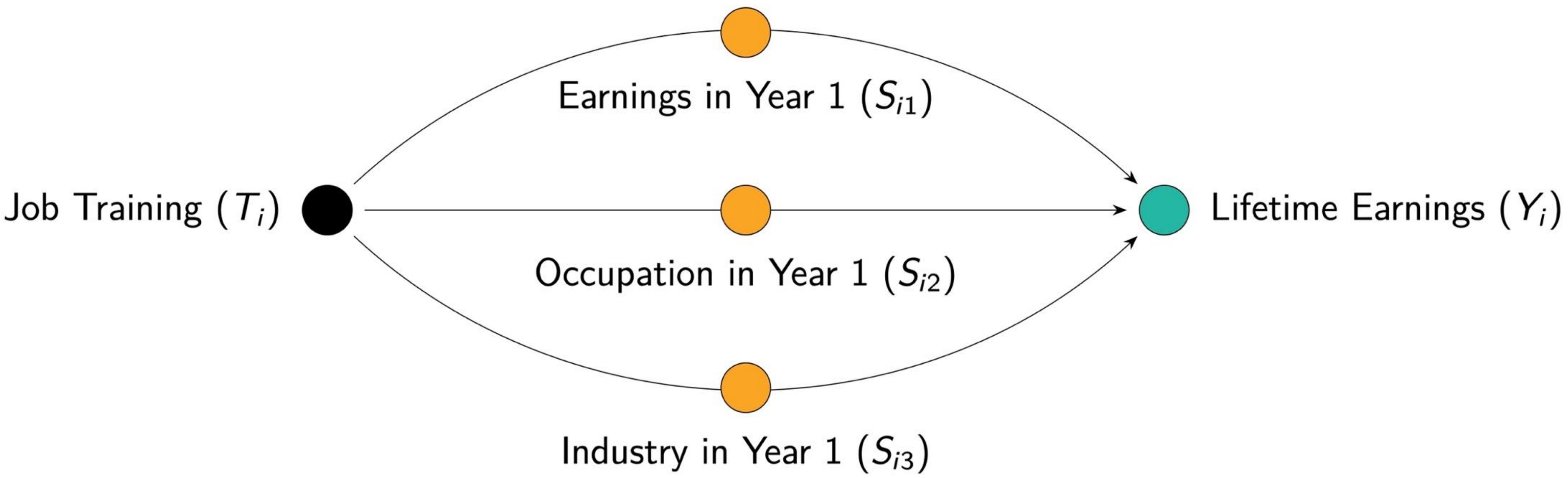

**Surrogacy:** Assume no direct effect of treatment on $Y_i$ → indirect effect = total effect; do not observe $Y_i$ in experimental sample, so use observational data to predict it



# Surrogacy with Multiple Mediators

In many social science applications, unlikely to have a single mediator that captures causal pathway from treatment to outcome (unlike specific biomarkers in biology)

More plausible that causal pathways are spanned by multiple mediators that together satisfy surrogacy assumption



# Surrogate Index: Combining Multiple Mediators

- Combine multiple mediators by constructing a surrogate index
- Surrogate index is conditional expectation of long-term outcome (e.g., total lifetime earnings) given the surrogates (e.g., early-career earnings) in the observational data

$$\mu(s, O) = E[Y_i \mid S_i = s, G_i = O]$$

- In a linear model, surrogate index is the predicted value from a regression of long-term outcome on surrogates $S_{in}$:

$$Y_i = \beta_0 + \sum_{n=1}^{N} \beta_n S_{in} + \epsilon_i$$

$$\mu(S_i, O) = \beta_0 + \sum_{n=1}^{N} \hat{\beta}_n S_{in}$$

# Identification Using Surrogate Index

Under Assumptions 1-3 (and regularity conditions), treatment effect on the surrogate index in the experimental sample is the average treatment effect on long-term outcome:

$$\mathrm{ATE_Y} = \mathrm{E}[\mu(S_i, O) \mid T_i = 1, G_i = E] - \mathrm{E}[\mu(S_i, O) \mid T_i = 0, G_i = E]$$

Two steps to estimate treatment effect on long-term outcome:

1. Regress long-term outcome on surrogates to construct surrogate index in observational data

```
reg lifetime_earnings earnings_yr1 earnings_yr2 earnings_yr3 if obs == 1
predict lifetime_earnings_pred
```

2. Regress surrogate index on treatment in experimental data

```
reg lifetime_earnings_pred treatment if exp == 1
```

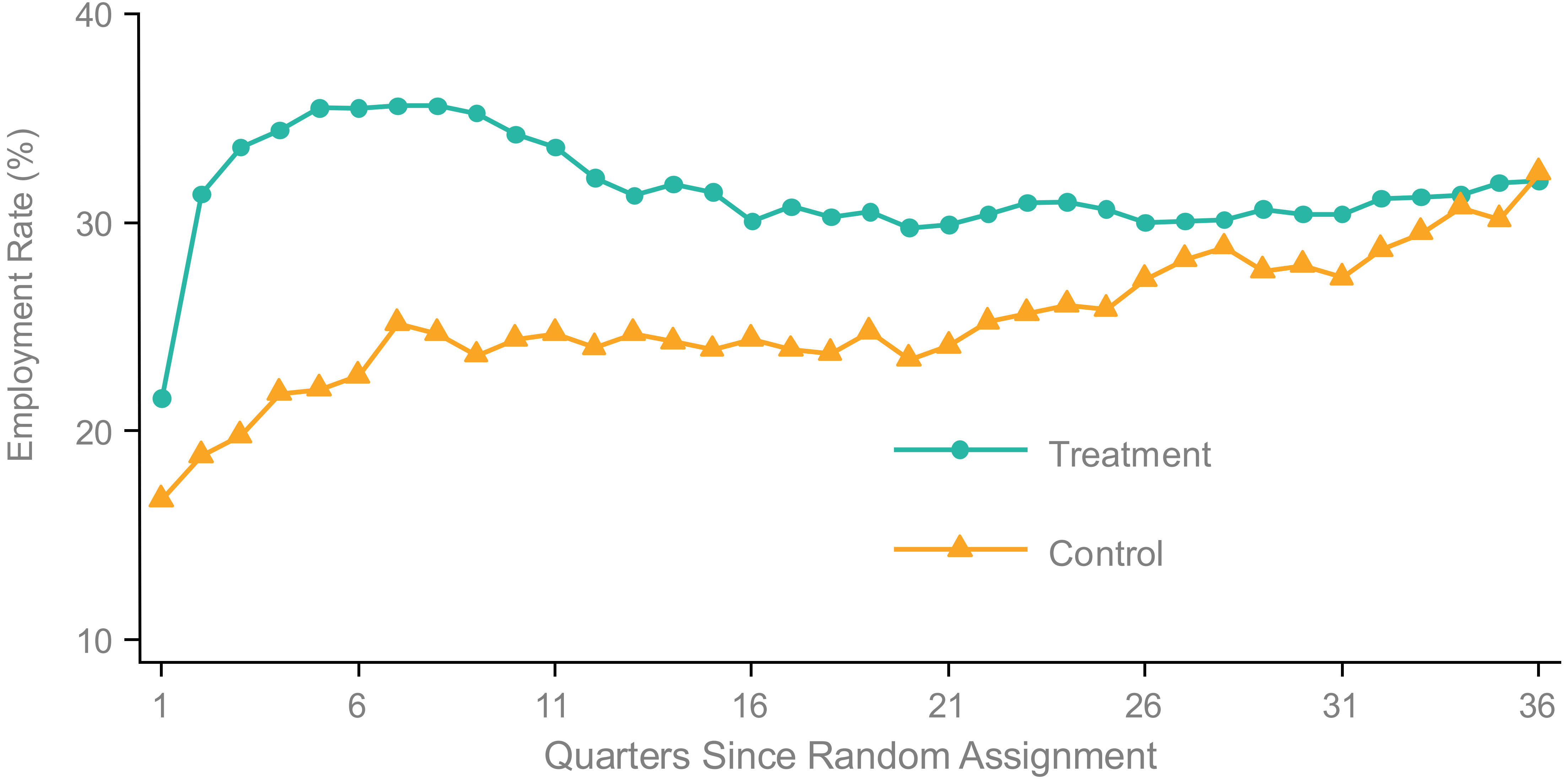
Note: standard errors must be adjusted to account for noise in estimating surrogate index; can be estimated using bootstrap

# Application: California GAIN Training Program

- California Greater Avenues to Independence (GAIN) program: job training program implemented in late 1980s to help welfare recipients find work
- Evaluated using randomized trials in Oakland, Los Angeles, Riverside, and San Diego
- Focus first on Riverside program, which had the largest impacts, perhaps because of its “jobs first” (rather than training-based) approach
- Then return to the other sites, which we use for out-of-sample validation

# Employment Rates in Treatment vs. Control Group

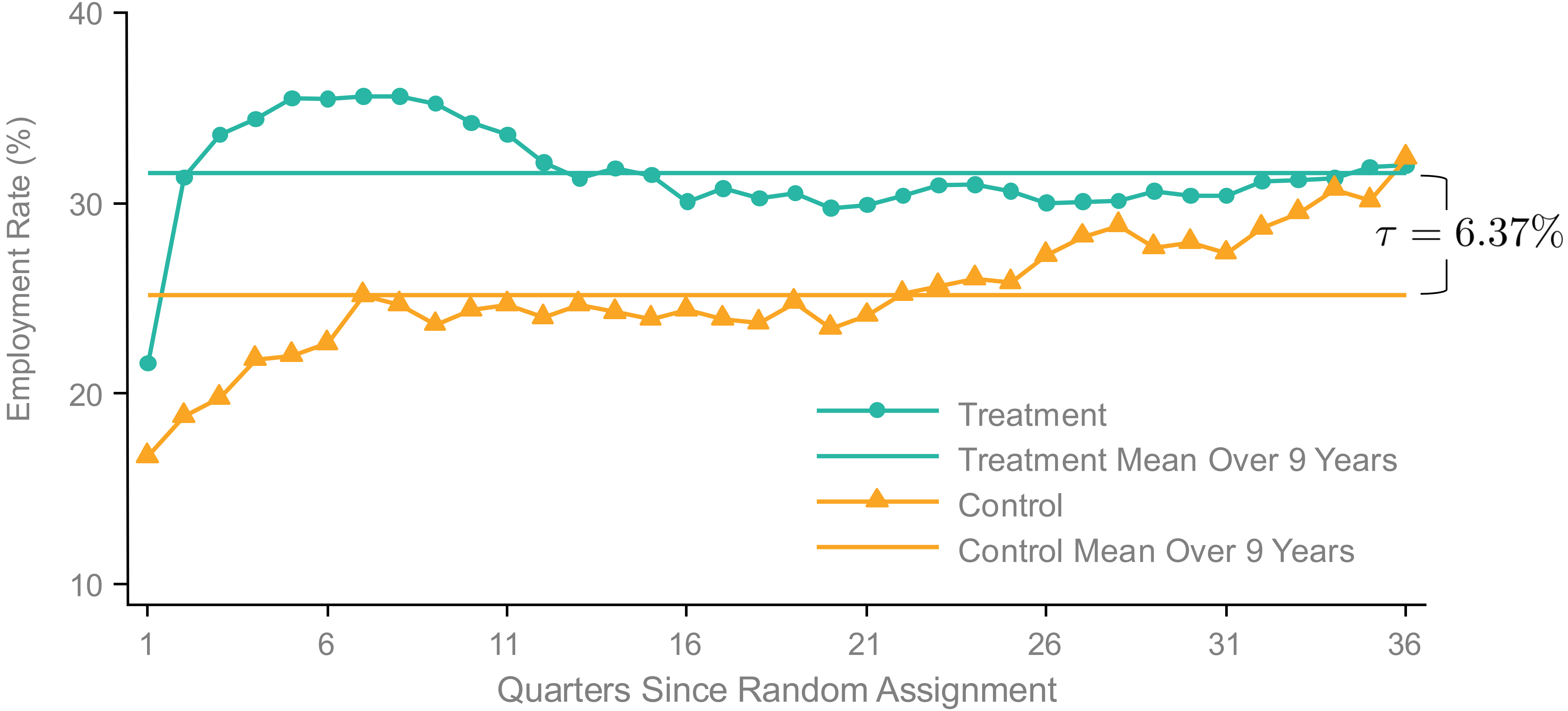
By Quarter in Riverside




**Source:** Hotz, Imbens, Klerman (2006); Athey, Chetty, Imbens, Kang (2025)

# Employment Rates in Treatment vs. Control Group

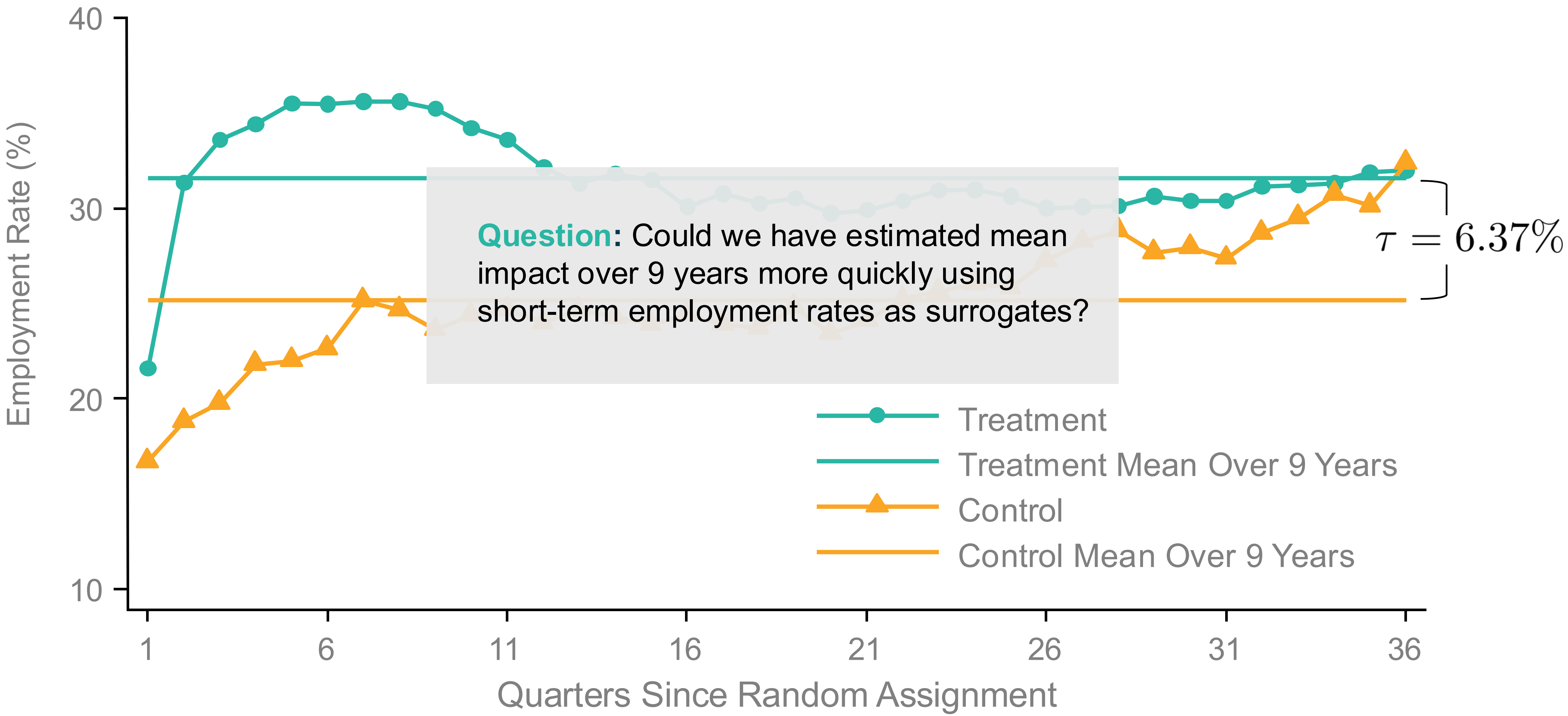
By Quarter in Riverside




**Source:** Hotz, Imbens, Klerman (2006); Athey, Chetty, Imbens, Kang (2025)

# Employment Rates in Treatment vs. Control Group

By Quarter in Riverside




**Source:** Hotz, Imbens, Klerman (2006); Athey, Chetty, Imbens, Kang (2025)

# Construction of the Surrogate Index

- Construct surrogate index using data on quarterly employment rates immediately after job training program

- Regress mean employment rate over 36 quarters on employment indicators from quarter 1 to quarter $N$ :

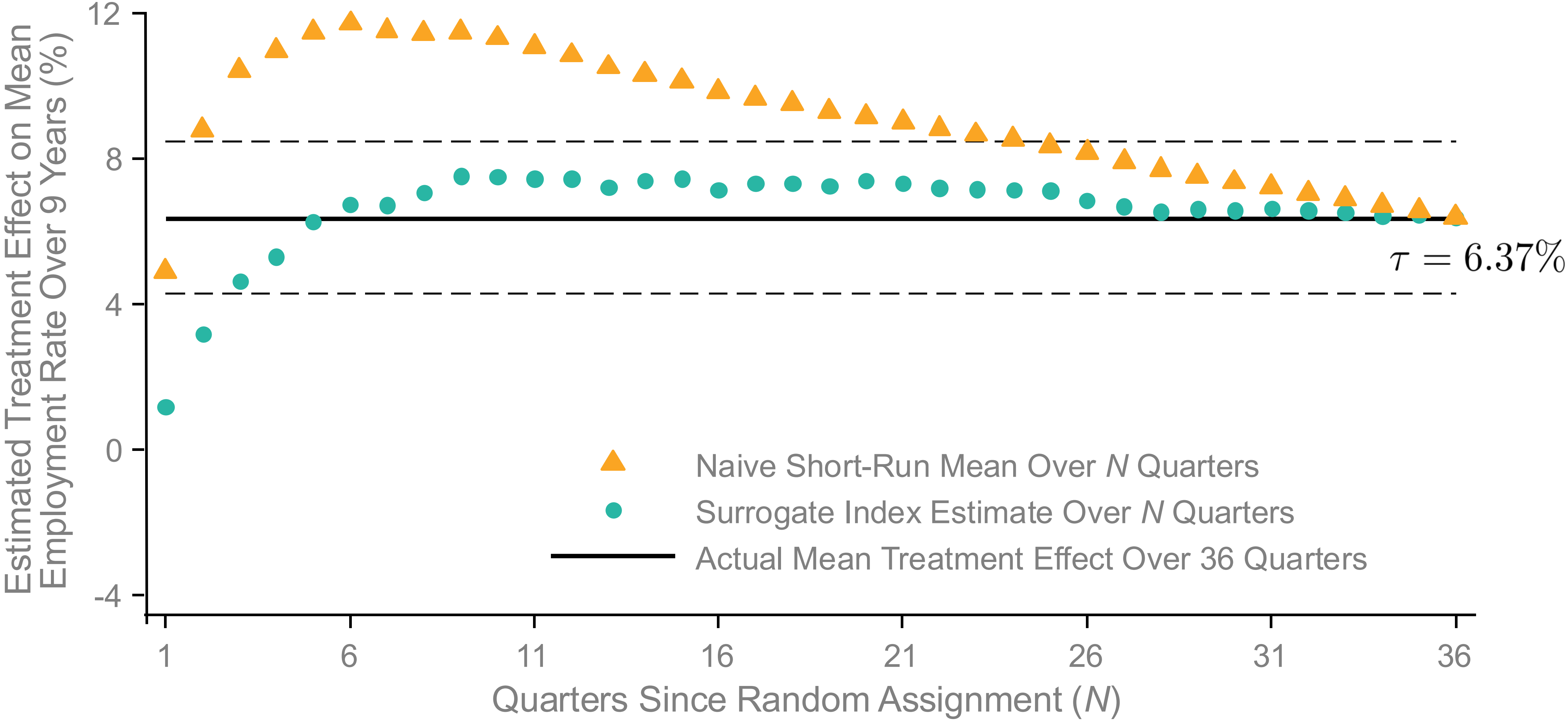
$$\overline{Y}_{iT} = \beta_0 + \sum_{n=1}^{N} \beta_n Y_{in} + \epsilon_i$$

- Then estimate treatment effect on surrogate index based on employment rates up to quarter $N$

- Evaluate how quickly (at what value of $N$) we can estimate nine-year mean impact accurately

# Estimate of Treatment Effect on Mean Employment Rate Over Nine Years

Varying Quarters of Data used to Construct Estimate



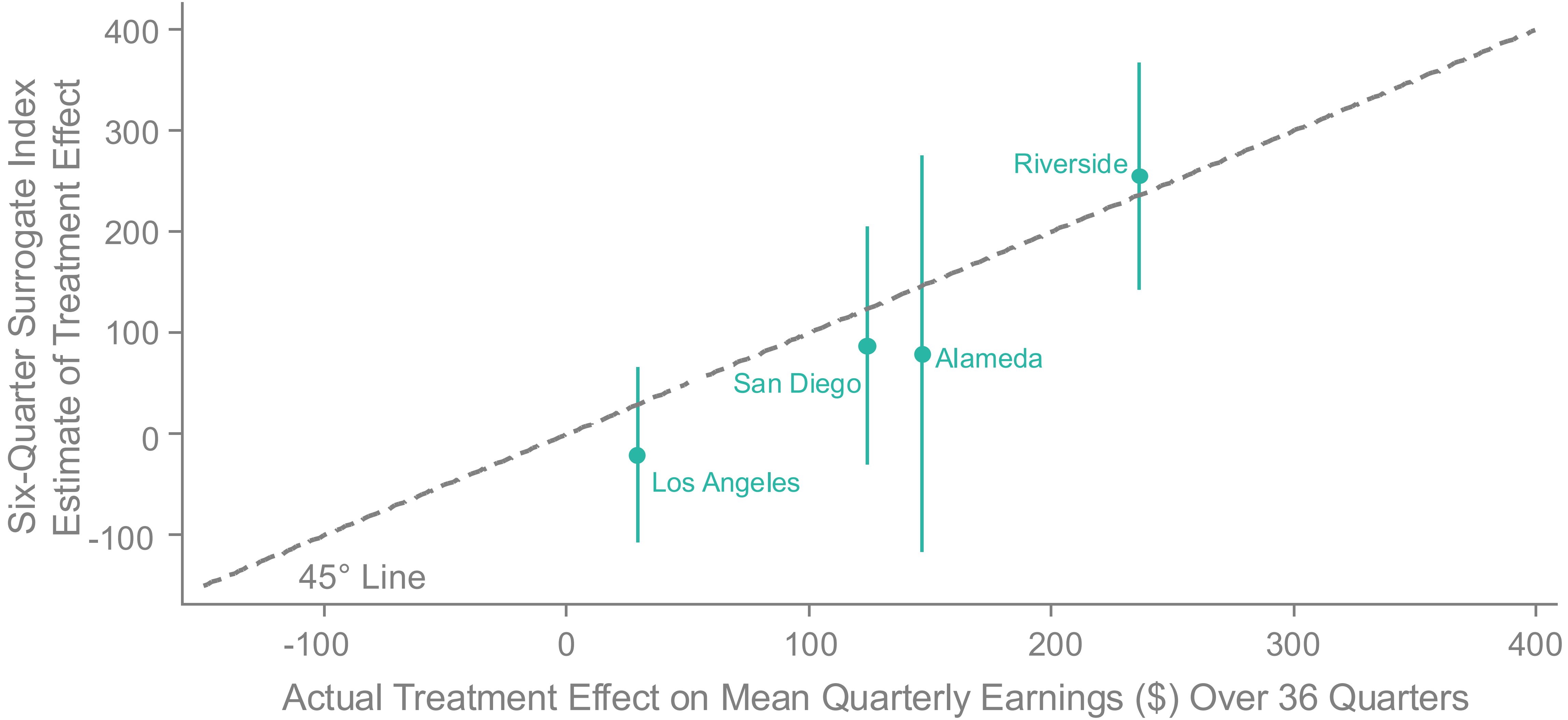

**Source:** Athey, Chetty, Imbens, Kang (2025)

# Out of Sample Validation: Predicting Cross-Site Heterogeneity

- Now use six-quarter surrogate index estimated in Riverside and ask how well it performs in predicting heterogeneity in treatment effects across other sites
  - Out-of-sample validation that jointly tests surrogacy and comparability assumptions

## Surrogate Index Estimates vs. Actual Experimental Estimates, by Site

Mean Quarterly Earnings ($) Over Nine Years

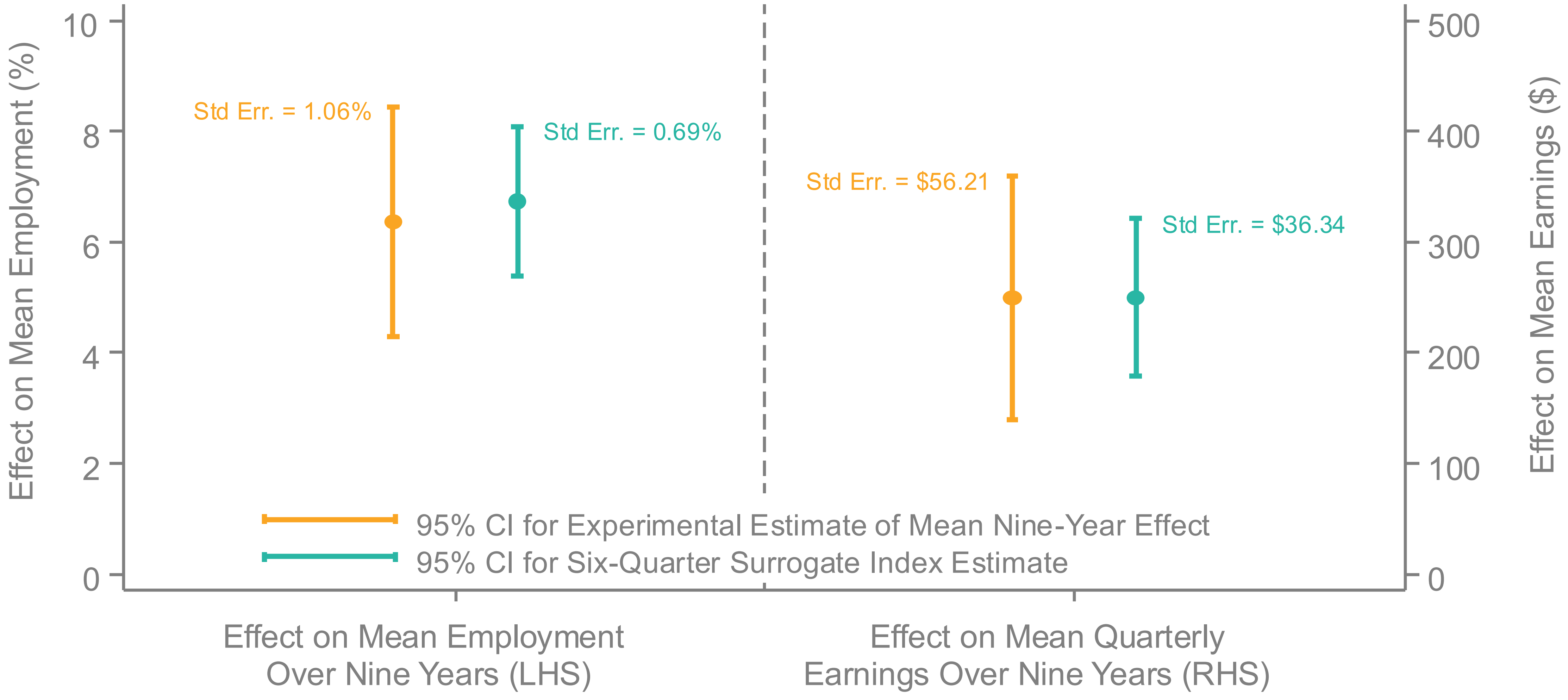



**Source:** Athey, Chetty, Imbens, Kang (2025)

# Secondary Benefit of Surrogate Indices: Gains in Precision

- Even when evaluating impacts with nine-years of data, may be preferable to use surrogate index based on first six quarters to gain precision
  - Intuition: under surrogacy assumption, any residual variation in long-term outcome conditional on surrogate index is noise orthogonal to treatment
  - → Reduce standard errors by estimating treatment effect on surrogate index instead of long-term outcome itself

# Gains in Precision from Using Surrogate Index




**Source:** Athey, Chetty, Imbens, Kang (2025)

# Further Reading on Surrogate Indices

### Identification with Alternative Data Configurations

- Battocchi et al. 2021 [dynamically adjusted treatments]
- Bibaut et al. 2023 [combining many weak experiments]
- Rambachan, Singh, Viviano 2025 [post-treatment outcomes]

### Estimation

- Chen and Ritzwoller 2022 [efficient influence function]
- Kallus and Mao 2024 [role of surrogacy assumption in efficiency]

### Applications

- Chetty, Deming, Friedman 2023
- Carlana, Miglino, Tincani 2024
- Athey, Castillo, Chandar 2025

# Remotely Sensed Variables

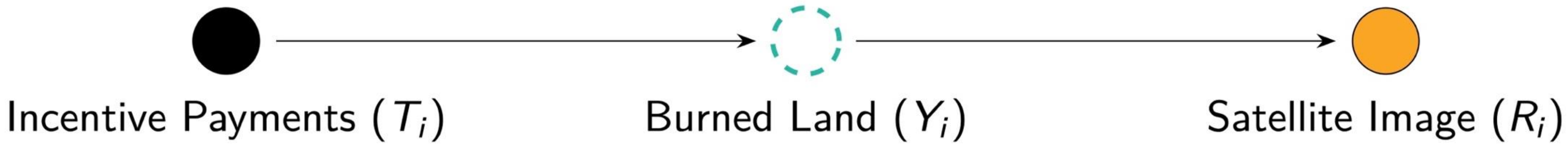
## Predicting Treatment Effects Using Post-Treatment Outcomes

# Post-Outcome Remotely Sensed Variables

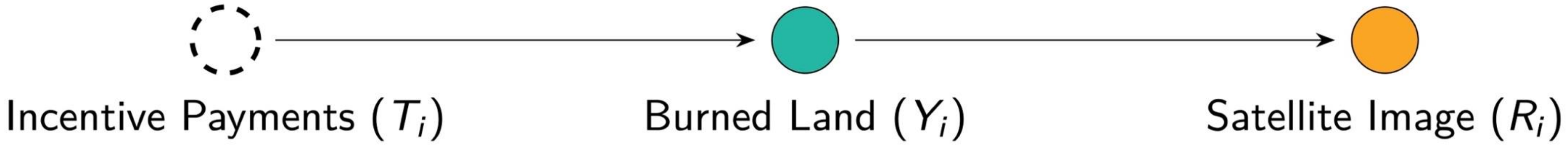
- In some settings, variables used as proxies for primary outcomes are not intermediate outcomes but instead measured **after** the primary outcome
- Example: Jack, Jayachandran, Kala, and Pande (2025) randomize incentive payments ($T_i$) to farmers in India to stop burning crop residue from land ($Y_i$)
  - Since objective measures of land burning are costly to obtain, use satellite images (remotely sensed variable $R_i$) to predict burning

# Post-Outcome Remotely Sensed Variables: Setup

**Experimental Data**




**Observational Data**



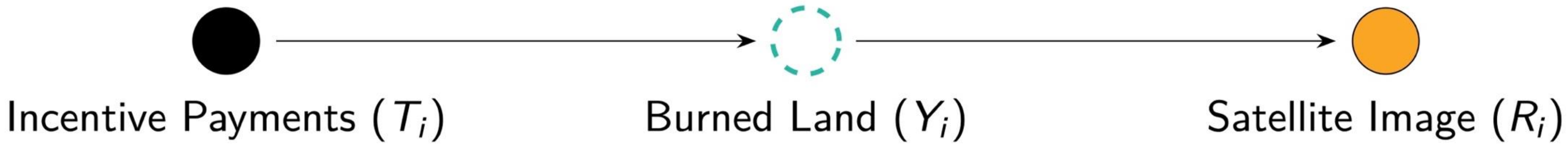
# Post-Outcome Remotely Sensed Variables

- Common, intuitive practice is to use surrogate approach in such settings: predict primary outcome $Y_i$ using $R_i$ and then estimate treatment effect on predicted outcome

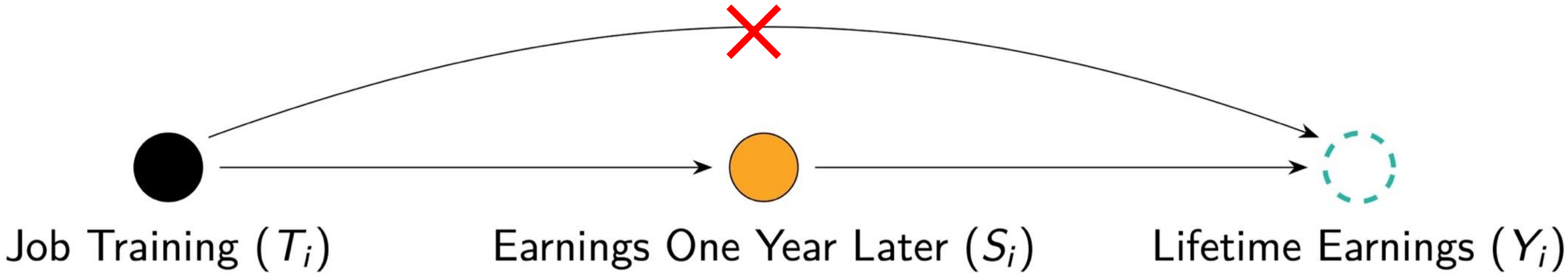
- Rambachan, Singh, Viviano (2025) show that this generally yields biased estimates, because surrogacy assumption is violated: $T_i$ directly affects $Y_i$ independent of $R_i$.

# Remote Sensing vs. Surrogate Model

**Experimental Data in Remote Sensing Model**




**Experimental Data in Surrogate Model**



# Post-Outcome Remotely Sensed Variables

- Common, intuitive practice in such settings is to use surrogate approach: predict primary outcome $Y_i$ using $R_i$ and then estimate treatment effect on predicted outcome

- Rambachan, Singh, Viviano (2025) show that this generally yields biased estimates, because surrogacy assumption is violated: $T_i$ directly affects $Y_i$ independent of $R_i$

- They show how one can identify treatment effect on $Y_i$ in general setting without parametric restrictions

- Focus here on a linear example that captures key difference relative to surrogate model

# Identification Using Remotely Sensed Variables: Key Assumptions

**Assumption 1.** Unconfounded Treatment in Experimental Sample

$$T_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid G_i = E$$

**Assumption 2.** Stability Across Experimental and Observational Samples

$$R_i \mid Y_i, T_i, G_i = E \;\sim\; R_i \mid Y_i, T_i, G_i = O$$

**Assumption 3.** No Direct Effect of $T_i$ on post-treatment outcome $R_i$ conditional on $Y_i$

$$T_i \perp\!\!\!\perp R_i \mid Y_i$$

**Source:** Rambachan, Singh, Viviano (2025)

# Stability Assumption in Linear Case

- Suppose the conditional expectation of $R_i \mid Y_i$ is linear in both samples:

**Experimental Data:** $$\mathrm{E}[R_i \mid Y_i, G_i = E] = \alpha_{\mathrm{E}} + \beta^{\mathrm{E}} \cdot Y_i$$

**Observational Data:** $$\mathrm{E}[R_i \mid Y_i, G_i = O] = \alpha_{\mathrm{O}} + \beta^{\mathrm{O}} \cdot Y_i$$

- In this case, stability requires that the slopes are the same in both samples:

$$\beta^{\mathrm{E}} = \beta^{\mathrm{O}} = \beta$$

# Identification Using Remotely Sensed Variables

Under Assumptions 1-3 (and regularity conditions), treatment effect on $Y_i$ is:

$$\mathrm{ATE_Y} = \{\mathrm{E}[R_i \mid T_i = 1, G_i = E] - \mathrm{E}[R_i \mid T_i = 0, G_i = E]\} \cdot \frac{1}{\beta}$$

$$\mathrm{ATE_Y} = \mathrm{ATE_R}/\beta$$

Three steps to estimate treatment effect on $Y_i$ :

1. Regress the RSV on treatment in experimental data to identify $\mathrm{ATE_R}$

```
reg satellite_prediction T if exp == 1
```

2. Regress the RSV on outcomes in observational data to estimate $\beta$

```
reg satellite_prediction burned_land if obs == 1
```

3. Divide $\mathrm{ATE_R}$ from (1) by $\beta$ from (2)

# Application of Remote Sensing Estimator

$$\mathsf{ATE_Y} = \mathsf{ATE_R}/\beta$$

| | (1) | (2) | (3) |
|---|---|---|---|
| Estimand | $\mathrm{ATE_R}$ | $d\mathrm{R}$ / $d\mathrm{Y}$ ($\beta$) | $\mathrm{ATE_Y} = \mathrm{ATE_R}$ / $\beta$ |
| Estimate | $0.079^{*}$ | $0.530^{***}$ | $0.148^{*}$ |
| | (0.041) | (0.072) | (0.084) |

*Note: $R_i$ is a prediction that a farmer has not burned their field, based on satellite imagery in Jack et al. (2025).*

Intuition: satellite prediction captures 53% of actual non-burned plots → inflate treatment effect on satellite prediction by 1/.53 to get effect on *Y*

**Source:** Rambachan, Singh, Viviano (2025)

# Experimental Selection Correction

Controlling for Selection Using Intermediate Outcomes

# Failures of Surrogacy Assumption

- Many applications where surrogacy assumptions may not hold
- Test scores often used as a surrogate for long-term outcomes (e.g., high school graduation or earnings)
- But early interventions affect long-term outcomes through many channels besides test scores [e.g., Heckman et al. 2006, Deming 2009, Chetty et al. 2011]

# How Can We Make Progress When Surrogacy Assumption Fails?

- How can we make progress when we have data on intermediate outcomes such as test scores but they do not satisfy surrogacy assumptions?

- One approach: experimental selection correction [Athey, Chetty, Imbens 2025]
  - Key requirement: must observe treatment ($T_i$) along with intermediate and long-term outcomes ($S_i, Y_i$) in observational data
  - With this additional information, we can strictly weaken surrogacy assumptions

- Goal: causal inference when we have observational data on long-term outcomes for interventions implemented in the past (e.g., class size reduction)
  - Not applicable to new interventions for which we do not yet have long-term outcome data (unlike surrogate estimator)

# Treatment Effects of Class Size: Experimental vs. Observational Data

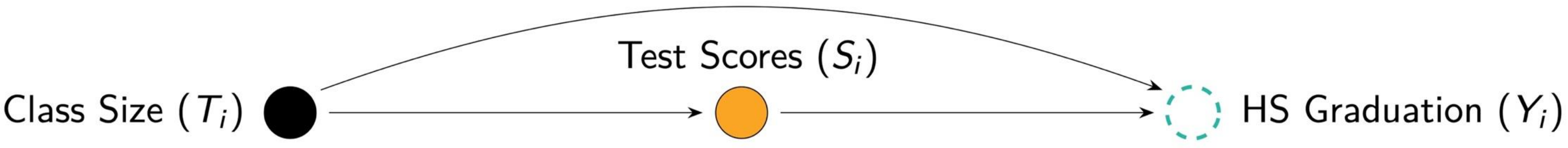
| | $G_i = E$<br>Experimental<br>(Tennessee STAR) | $G_i = O$<br>Observational<br>(NYC Admin Data) |
|---|---|---|
| Treatment Effect of Assignment to Small 3rd Grade Class on: | | |
| 3rd Grade Test Scores ($S_i$) | 0.19 SD<br>(0.04) | -0.12 SD<br>(0.01) |
| HS Graduation Rate ($Y_i$) | ? | -1.76 pp<br>(0.29) |

*Note: All estimates based on OLS regressions, controlling for school x cohort fixed effects*

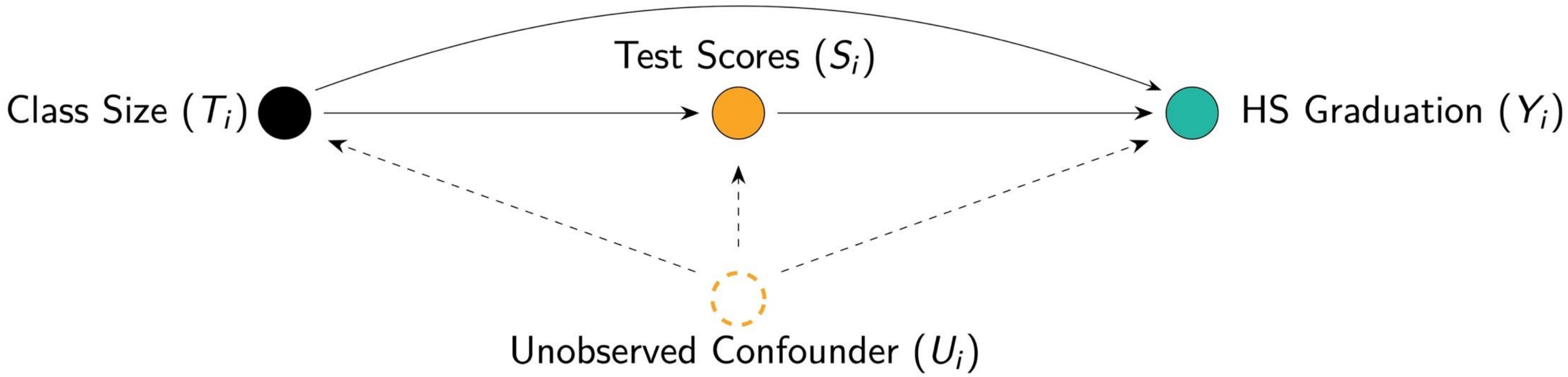
**Source:** Athey, Chetty, Imbens (2025)

# Experimental Selection Correction Model: Setup

**Experimental Data**




**Observational Data**



# Experimental Selection Correction: Key Identification Assumptions

**Assumption 1.** Unconfounded Experiment

$$T_i \perp\!\!\!\perp (Y_i(0), Y_i(1), S_i(0), S_i(1)) \mid G_i = E$$

**Assumption 2.** External Validity of Experimental Sample for Observational Sample

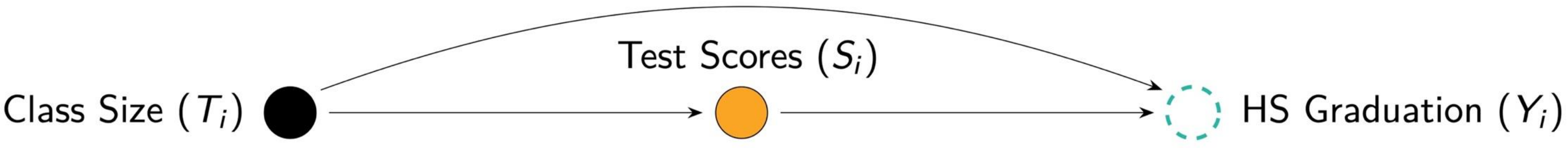
$$Y_i(t), S_i(t) \mid G_i = E \sim Y_i(t), S_i(t) \mid G_i = O$$

**Assumption 3.** Latent Unconfoundedness: Confounds that affect $S_i$ and $Y_i$ are the same

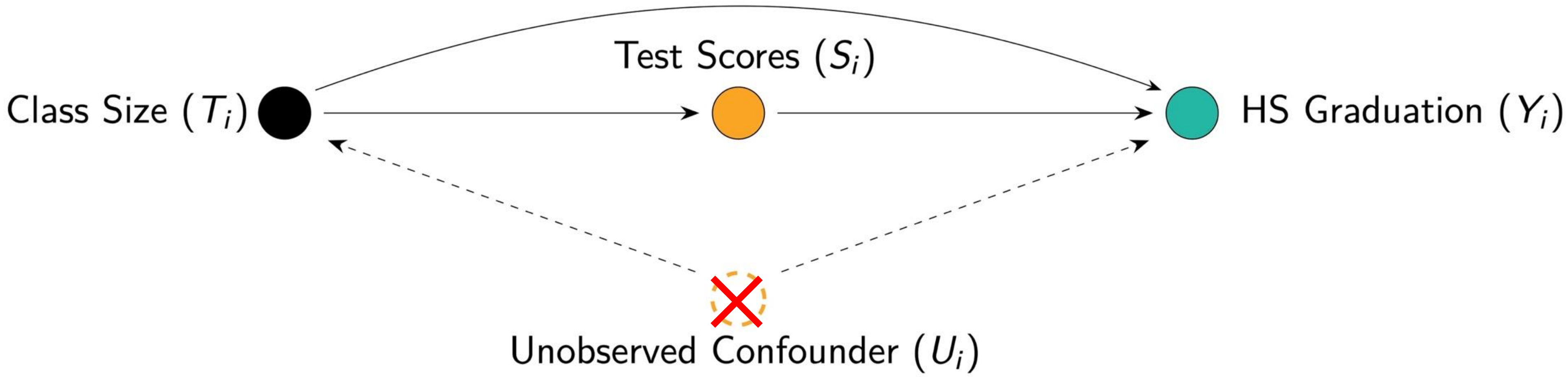
$$T_i \perp\!\!\!\perp Y_i(t) \mid S_i(t), G_i = O \;\; \forall t$$

# Violations of Latent Unconfoundedness Assumption

**Experimental Data**




**Observational Data**



# Linear Case

- Linear models for intermediate and long-term outcomes

$$\underbrace{S_i}_{\text{score}} = \tau^S \underbrace{T_i}_{\text{class size}} + \underbrace{\alpha_i^S}_{\text{parental input}} \qquad \underbrace{Y_i}_{\text{HS graduation}} = \tau^Y \underbrace{T_i}_{\text{class size}} + \underbrace{\alpha_i^Y}_{\text{residual variation in } Y}$$

- $(\alpha_i^Y, \alpha_i^S) \perp\!\!\!\perp T_i$ in experimental sample, but not in observational sample

- Here, latent unconfoundedness assumption requires:

$$\alpha_i^Y = \delta \alpha_i^S + \epsilon_i \quad \text{with} \quad T_i \perp\!\!\!\perp \epsilon_i \mid \alpha_i^S, G_i = O$$

  - Once we control for parental input, residual variation in graduation rates ($\epsilon_i$) must be independent of class size ($T_i$) in observational sample

# Identification in Linear Case

- If we observed $\alpha_i^S$ (parent input), we could identify treatment effect by controlling for $\alpha_i^S$:

$$Y_i = \tau^Y T_i + \gamma \, \alpha_i^S + \epsilon_i$$

- We do not observe $\alpha_i^S$, but can infer it by comparing scores in observational sample to **predicted** scores based on the experimental estimate:

$$\alpha_i^S = S_i - \hat{\tau}^S T_i$$

  where $\hat{\tau}^S$ is estimated in the experimental sample.

- In application, we can infer that there is negative selection into smaller classrooms in NYC schools because children in smaller classes have **lower** test scores

# Identification with Experimental Selection Correction

Under Assumptions 1-3 (and regularity conditions), treatment effect on $Y_i$ is:

$$\mathsf{ATE_Y} = \mathsf{E}[Y_i \mid T_i = 1, \alpha_i^S, G_i = O] - \mathsf{E}[Y_i \mid T_i = 0, \alpha_i^S, G_i = O]$$

Three steps to estimate treatment effect on $Y_i$:

1. Regress $S_i$ on $T_i$ in experimental sample to estimate $\tau^S$

```
reg score treatment if exp == 1
```

2. Construct residuals of $S_i$ in the observational sample based on predicted values

```
predict score_pred

gen selection = score - score_pred if obs == 1
```

3. Regress $Y_i$ on $T_i$ in observational sample, controlling for residuals from Step 2

```
reg graduation treatment selection if obs == 1
```

# Experimental Selection Correction: Connections to Existing Estimators

- Analogous to control function methods with instrumental variables
  - In IV models, can identify by controlling for first-stage residuals (which captures endogenous variation due to selection) in second-stage estimating equation
  - Here, control for difference between observed short-term outcome and prediction based on experimental estimate (which again captures variation due to selection)

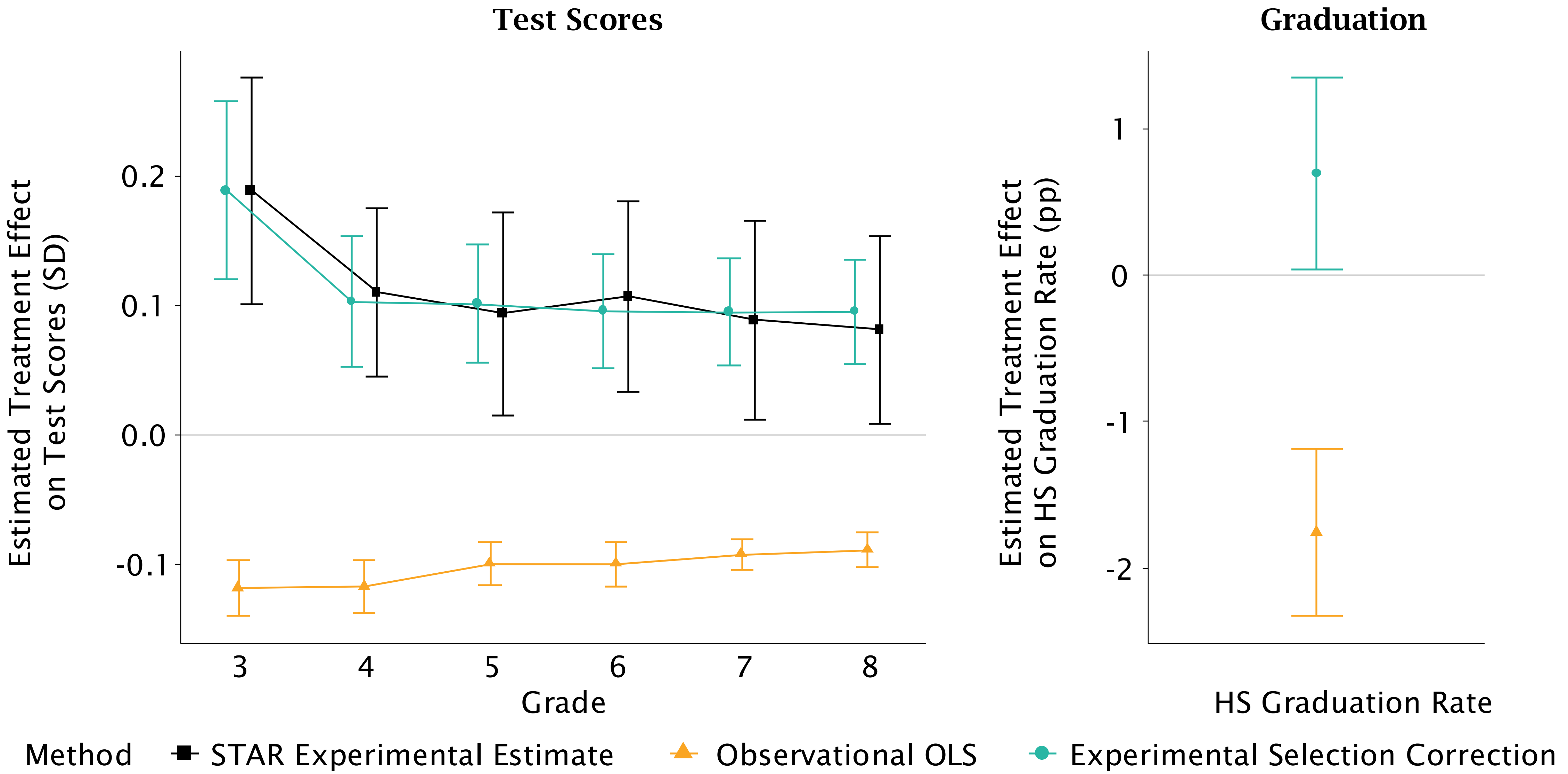
- Also connected to traditional selection correction based on parametric assumptions and exclusion restrictions [e.g., Heckman 1979]
  - Here, selection correction is based on an experimental estimate of treatment effect on intermediate outcome

# Application: Causal Effects of Class Size on High School Graduation

- Observational data from New York City public schools with information on 3rd grade class size, subsequent test scores, and high school graduation
  [Chetty, Friedman, Rockoff 2014, Mariano et al. 2024]

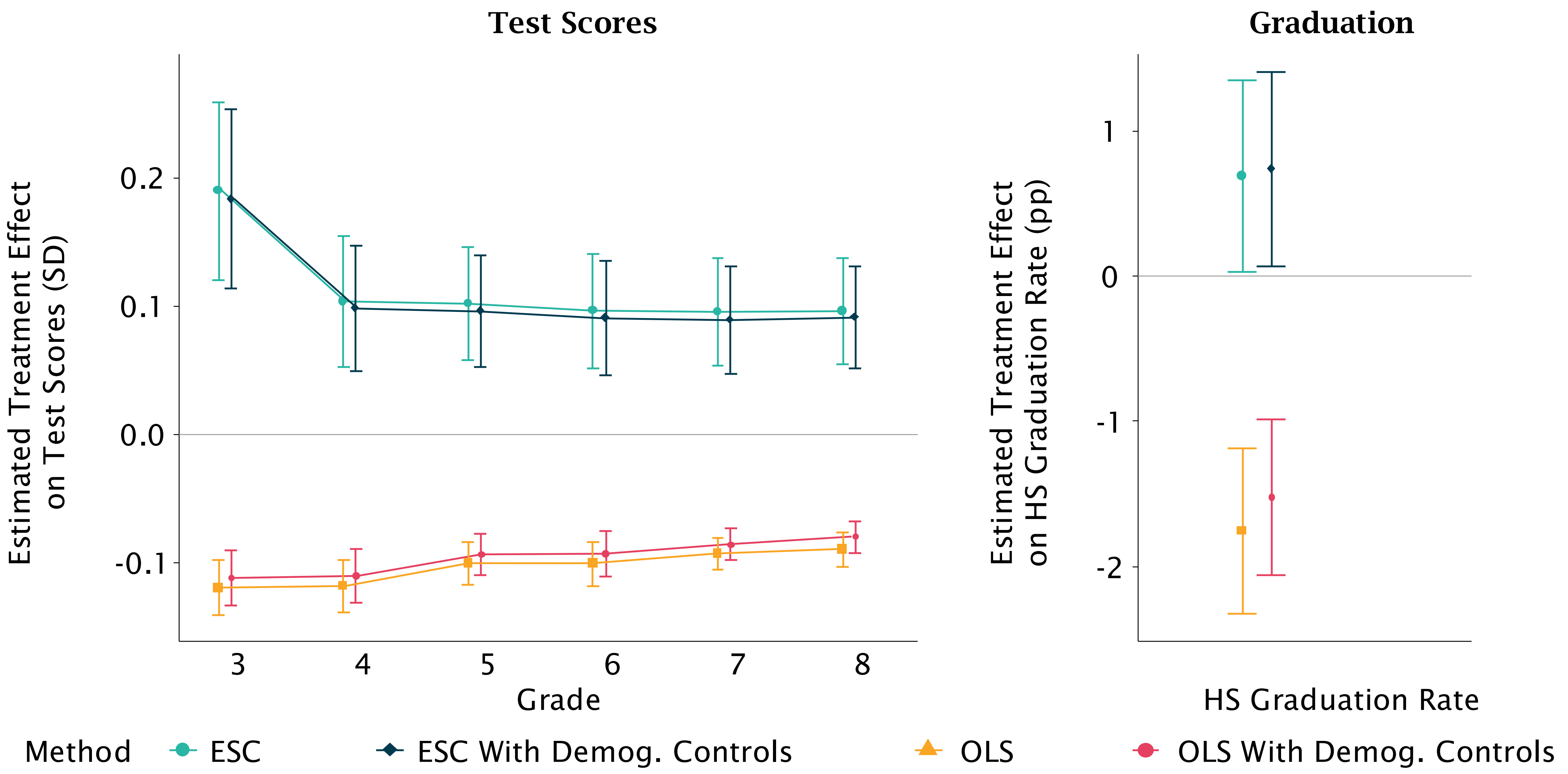
- Experimental data from Tennessee STAR, which randomly assigned 12,000 children to small vs. large classrooms [see e.g., Krueger 1999]
  - Many papers on impacts of STAR on short-term outcomes, but much less evidence on long-term impacts; no estimates of impacts on graduation rates in particular

# Treatment Effects of Assignment to Small Class in 3rd Grade




**Source:** Athey, Chetty, Imbens (2025)

# Effect of Controlling for Observables on Treatment Effect Estimates




**Source:** Athey, Chetty, Imbens (2025)

# Further Reading on Experimental Selection Correction

## Identification with Alternative Assumptions

- Heckman and Pinto 2015, Garcia et al. 2020 [assumptions based on economic theory]
- Ghassami et al. 2022 [parallel trends]
- Park and Sasaki 2024a,b [partial identification]

## Identification with Alternative Data Configurations

- Obradovic 2024 [partial identification with additional instruments]
- Imbens et al. 2025 [identification with sequential intermediate outcomes]

## Estimation

- Meza and Singh 2024 [nested nonparametric IV]

## Applications

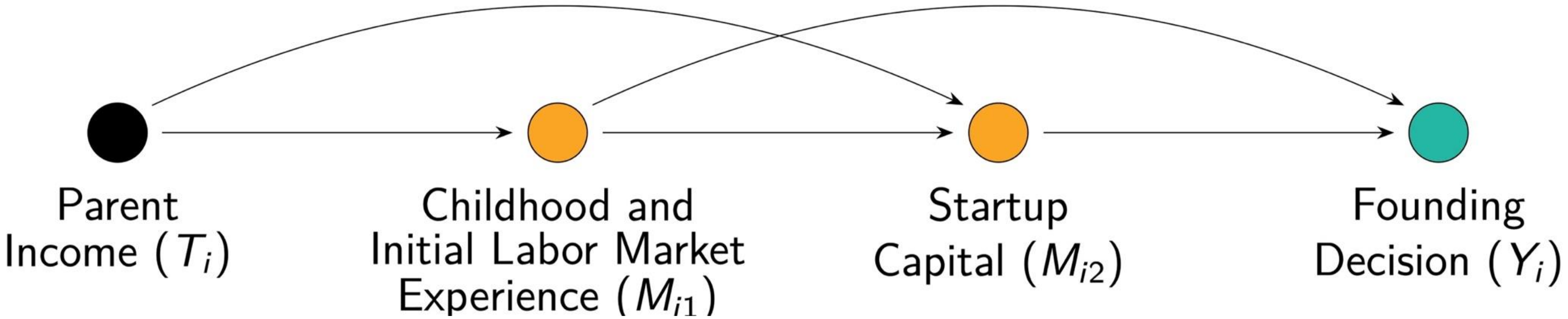
- Aizer, Early, Eli, Imbens, Lee, Lleras-Muney, Strand 2024

# Application and Recommendations for Empirical Practice

# Application: Determinants of and Returns to Entrepreneurship

- Now apply these methods to return to motivating example of determinants and returns to entrepreneurship [Chetty, Dossi, Smith, van Reenen, Zidar, Zwick 2025]

# The Entrepreneurial Pipeline: Causal Mechanisms



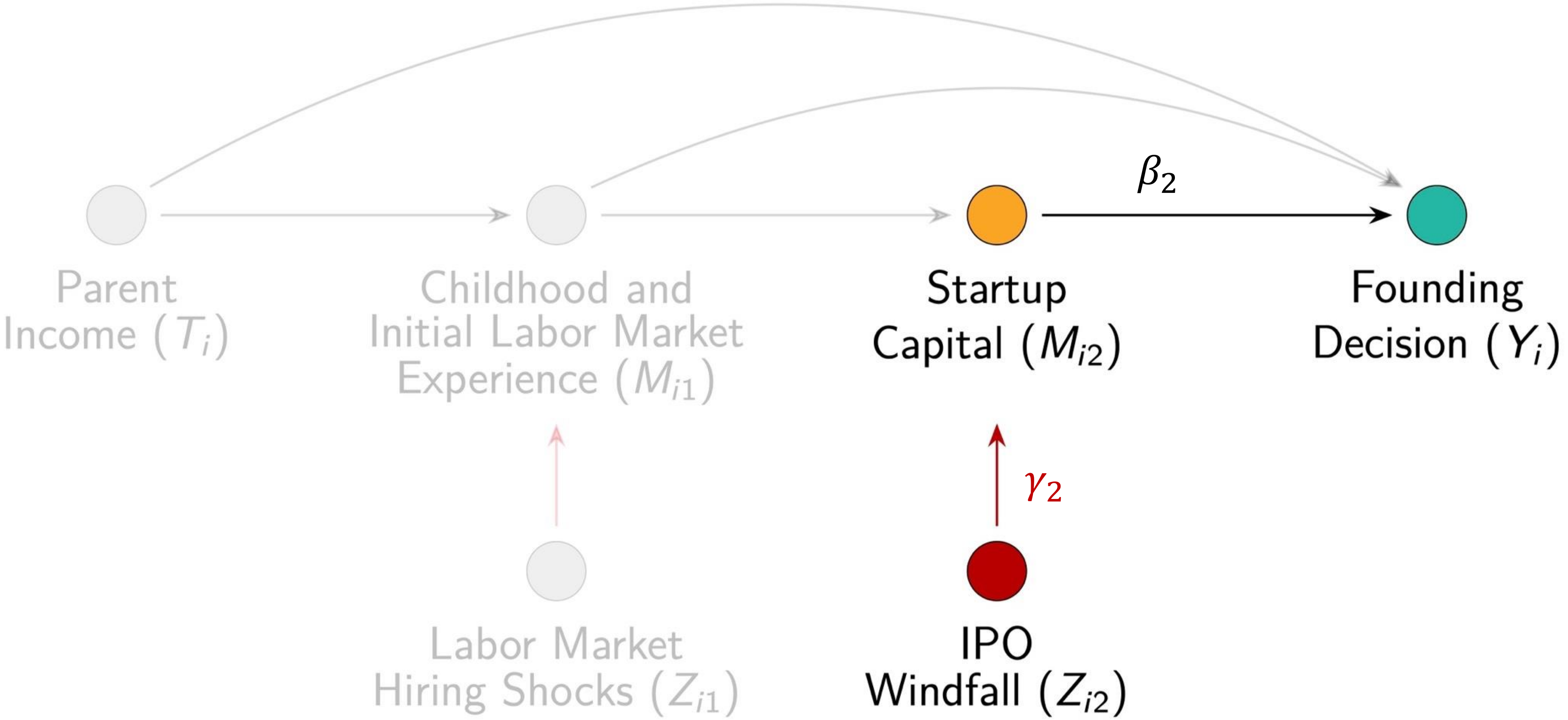
# Identifying Mediating Mechanisms

- Use multiple instruments to obtain exogenous variation in mediators and quantify role of each mechanism

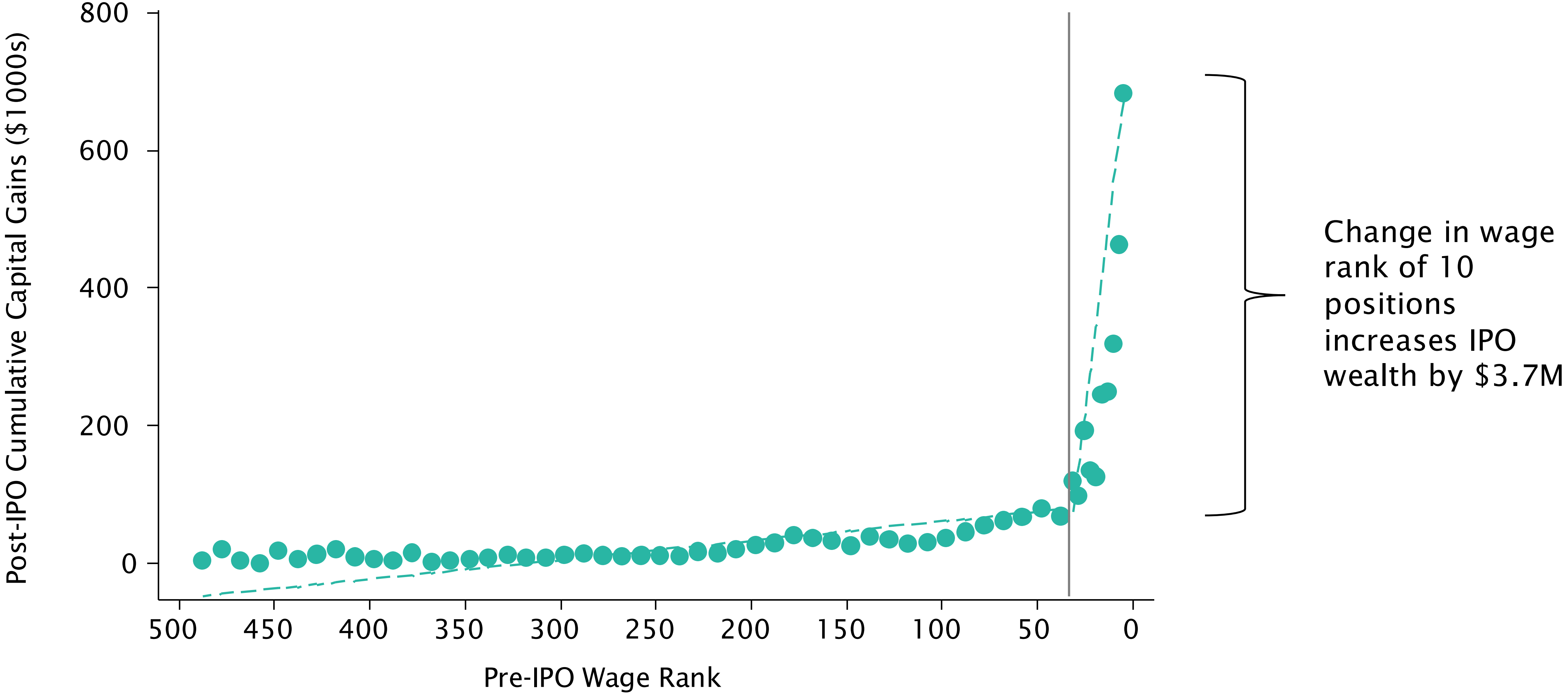
- Start by assuming homogeneous treatment effects and additivity of treatment effects

- Then revisit validity of these assumptions

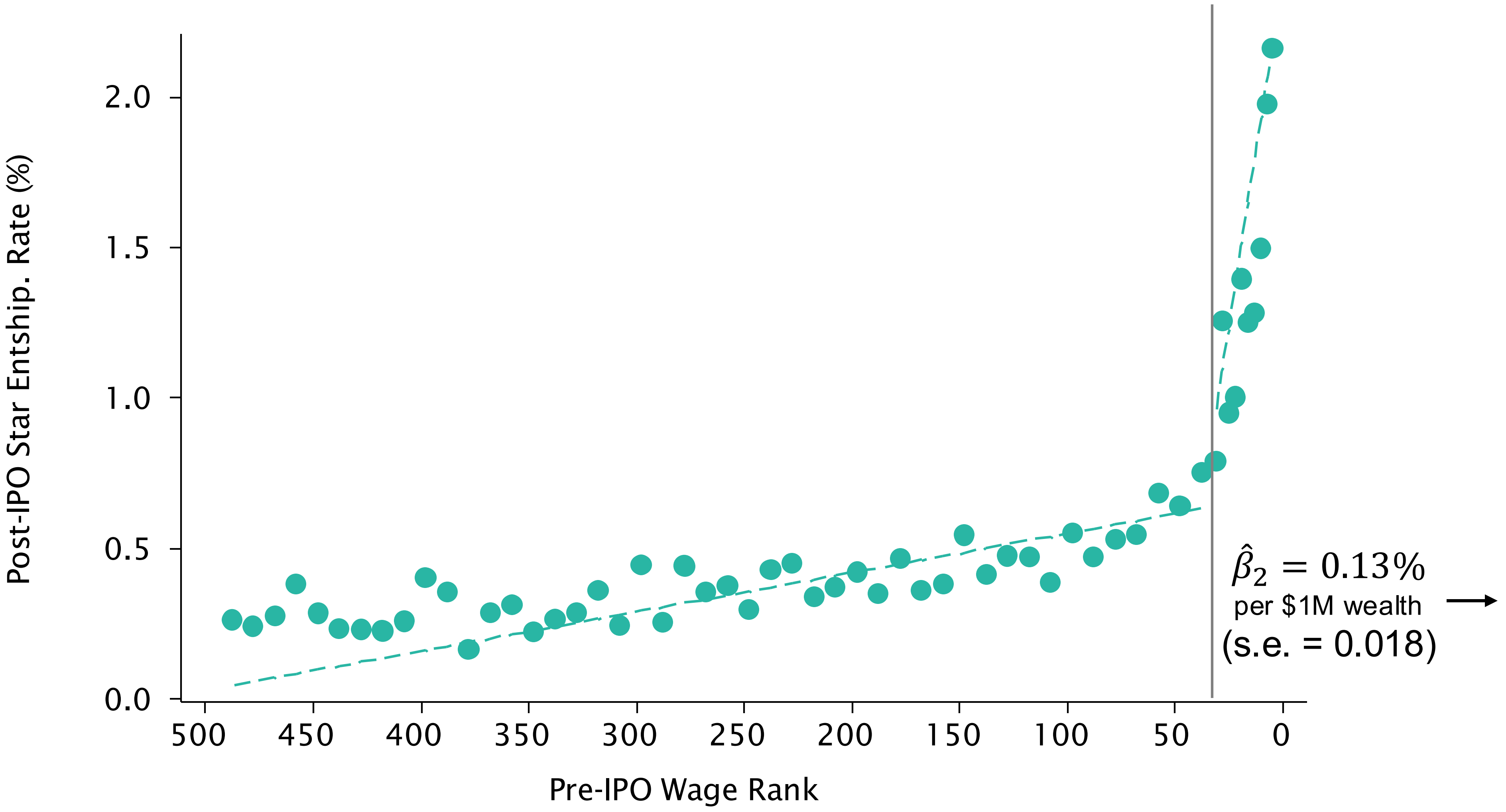
# The Entrepreneurial Pipeline: Causal Mechanisms



# Wealth Windfall by Wage Rank in Initial Public Offerings

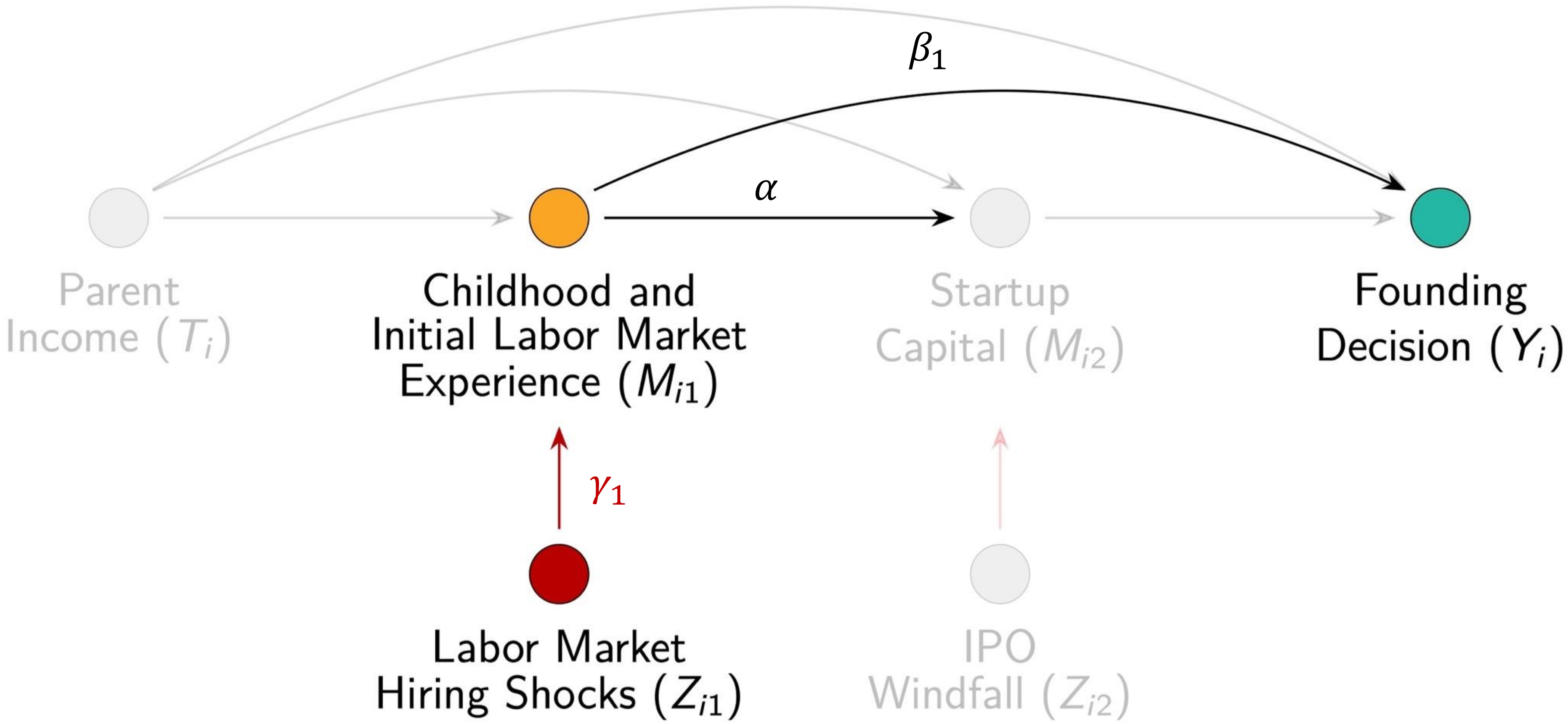


# Impact of IPO Cash Windfalls on Subsequent Entrepreneurship Rates




1.4% of gap in entrepreneurship rates by parent income is driven by wealth

# The Entrepreneurial Pipeline: Causal Mechanisms



# Addressing Potentially Heterogeneous Treatment Effects

- Key implication of mediation analysis: Wealth at startup relatively unimportant in explaining gaps compared to upstream pipeline (e.g., sector of first job)

- Conclusion hinges on assuming homogeneous treatment effects. Two concerns:

1. Samples used to estimate treatment effects in each design differ on observables (e.g. age, cohort)
    - Solution: condition on observables by reweighting on cohort and age
    - Yields wealth effect that is similar to baseline estimate

# Addressing Potentially Heterogeneous Treatment Effects

- Key implication of mediation analysis: Wealth at startup relatively unimportant in explaining gaps compared to upstream pipeline (e.g., sector of first job)

- Conclusion hinges on assuming homogeneous treatment effects. Two concerns:

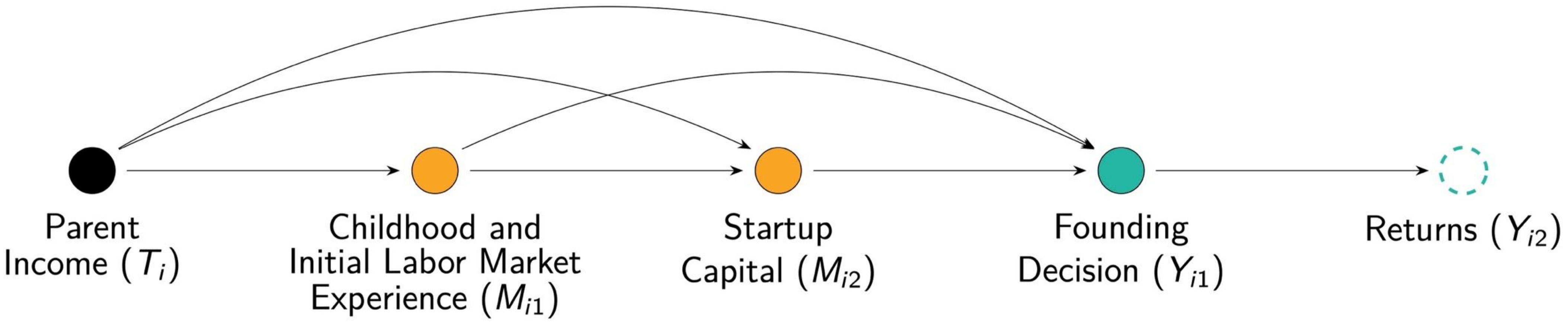
  1. Samples used to estimate treatment effects in each design differ on observables (e.g. age, cohort)

  2. LATE estimated by instruments may differ due to unobserved heterogeneity: those involved in IPOs may have good access to capital markets already

     - Solution: use instruments that generate different LATEs

     - Similarly small effects on business startup estimated using lottery winnings [Golosov et al. 2024]

# The Entrepreneurial Pipeline: Predicting Returns




What are the marginal returns to new businesses startups induced by a shock to labor market experience? Are they as successful as existing businesses?

# Predicting Returns Using a Surrogate Index

- Long-term returns (e.g., 8-10 years after startup) observed for few firms due to censoring

- Use surrogate index to predict outcomes using historical data where we observe both short-run (2 year after founding) and long-run (8-year) outcomes

  - Regress eight-year returns on revenues, wages, and number of employees two years after founding

- Compare predicted returns of marginal entrepreneurs induced to enter by labor market experience shock to returns of average entrepreneur

  - Key finding: returns of marginal entrepreneur are very similar to those of average entrepreneur

# Takeaways from Empirical Application

1. IV-based mediation analysis shows that low-income children are less likely to start firms because of a lack of exposure to entrepreneurial industries, not lack of wealth
   - Opposite of conclusion obtained from controlling for wealth
   - Sequential ignorability assumption fails here because wealth is correlated with many other unobservable factors that directly affect entrepreneurship

2. Surrogate indices reveal that increasing labor market experience induces startups whose returns are similar to current mean

→ Early-career incubator or apprenticeship programs targeted to certain subgroups could increase business startups and growth significantly

# Recommendations for Empirical Practice

- Start by writing down causal graphs (DAGs) underlying empirical analysis

# Recommendations for Empirical Practice

- Mediation analysis:

1. Avoid common practice of controlling for endogenous regressors unless sequential ignorability (conditionally exogenous mediators) can be justified
2. Generally need to find **instruments or experiments** that vary mediators, as we do for treatment itself in design-based paradigm
3. Start by assuming additivity + no heterogeneity, then evaluate robustness by reweighting, estimating different LATEs, and estimating interaction effects

# Recommendations for Empirical Practice

- Predicting long-term/primary outcomes using intermediates or proxies:

  1. When intermediate outcomes are likely to span set of causal pathways to later outcomes, use **surrogate index** to maximize efficiency

     - Often works well when using early labor market or firm outcomes as surrogates for later outcomes, perhaps because there are certain paths to later success

     - When long-term outcome is observed, surrogate index may still be helpful for analyzing heterogeneity or additional tests that demand more power

# Recommendations for Empirical Practice

- Predicting long-term/primary outcomes using intermediates or proxies:

    1. When intermediate outcomes are likely to span set of causal pathways to later outcomes, use **surrogate index** to maximize efficiency

    2. If surrogacy assumptions are debated and treatment is observed alongside long-term outcome in observational data, use **experimental selection correction**

        - Ex: when forecasting long-term impacts of early childhood interventions, where there are many causal pathways beyond those that run through test scores

# Recommendations for Empirical Practice

- Predicting long-term/primary outcomes using intermediates or proxies:

  1. When intermediate outcomes are likely to span set of causal pathways to later outcomes, use **surrogate index** to maximize efficiency
  2. If surrogacy assumptions are debated and treatment is observed alongside long-term outcome in observational data, use **experimental selection correction**
  3. If proxies are observed post-treatment, use **remote sensing estimator**

- In general: validate any surrogate/proxy method using direct experimental or quasi-experimental estimates on holdout outcomes

  - Both in a given paper and more generally as a field → will yield better understanding of best proxies and causal pathways

# Potential Directions for Further Methodological Research

1. Alternative assumptions and data configurations
   - Ex: Identification when surrogacy fails and treatment unobserved in observational data
2. Incorporating common issues in applications
   - Ex: Measurement error in surrogates or short-term outcomes; noncompliance in experiments; dynamic selection; spillover effects
3. Target parameters beyond average treatment effect
   - Ex: quantile treatment effects; marginal treatment effects; heterogeneity
4. Connections to traditional econometric methods
   - Ex: data combination; panel data [e.g., Ridder and Moffitt 2007, Arkhangelsky and Imbens 2024]

# References


- Aizer, A., Early, N., Eli, S., Imbens, G., Lee, K., Lleras-Muney, A., & Strand, A. (2024). “The lifetime impacts of the New Deal’s youth employment program.” *The Quarterly Journal of Economics*, *139*(4), 2579–2635. https://doi.org/10.1093/qje/qjae016
- Arkhangelsky, D., & Imbens, G. W. (2024). “Causal models for longitudinal and panel data: A survey.” arXiv preprint, arXiv:2311.15458. https://doi.org/10.48550/arXiv.2311.15458
- Athey, S., Chetty, R., Imbens, G. W., & Kang, H. (2024). “The surrogate index: combining short-term proxies to estimate long-term treatment effects more rapidly and precisely.” *Review of Economic Studies* (forthcoming). https://doi.org/10.3386/w26463
- Athey, S., Chetty, R., & Imbens, G. W. (2025). “The experimental selection correction estimator: Using experiments to remove biases in observational estimates.” NBER Working Paper No. 33817. https://www.nber.org/papers/w33817
- Athey, S., Castillo, J. C., & Chandar, B. (2024). “Service quality on online platforms: Empirical evidence about driving quality at Uber.” NBER Working Paper No. 33087. https://www.nber.org/papers/w33087
- Battocchi, K., Dillon, E., Hei, M., Lewis, G., Oprescu, M., & Syrgkanis, V. (2022). “Estimating the long-term effects of novel treatments.” arXiv preprint, arXiv:2103.08390. https://doi.org/10.48550/arXiv.2103.08390

# References


- Bibaut, A., Kallus, N., Ejdemyr, S., Zhao, M. "Long-term causal inference with imperfect surrogates using many weak experiments, proxies, and cross-fold moments." arXiv preprint arXiv:2311.04657. https://doi.org/10.48550/arXiv.2311.04657
- Carlana, M., Miglino, E., & Tincani, M. M. (2024). "How far can inclusion go? The long-term impacts of preferential college admissions." NBER Working Paper No. 32525. https://www.nber.org/papers/w32525
- Cattan, S., Salvanes, K. G., & Tominey, E. (2025). "First generation elite: The role of school networks." IZA Discussion Paper No. 15560. Institute of Labor Economics.
- Chen, J., & Ritzwoller, D. M. (2023). "Semiparametric estimation of long-term treatment effects". *Journal of Econometrics*, *237*(2), Part A, 105545. https://doi.org/10.1016/j.jeconom.2023.105545
- Chetty, R., Dossi, G., Smith, M., Van Reenen, J., Zidar, O. M., & Zwick, E. (2025). "Creating new businesses in America: The determinants of and returns to entrepreneurship." *Harvard University manuscript*.
- Chetty, R., Deming, D. J., & Friedman, J. N. (2023). "Diversifying society's leaders? The determinants and causal effects of admission to highly selective private colleges." NBER Working Paper No. 31492. https://www.nber.org/papers/31492

# References


- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). “How does your kindergarten classroom affect your earnings? Evidence from Project Star.” *The Quarterly Journal of Economics*, *126*(4), 1593–1660. 10.1093/qje/qjr041
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates.” *American Economic Review*, *104*(9), 2593–2632. 10.1257/aer.104.9.2593
- Garcia, L. J., Heckman, J. J., Leaf, D. E., & Prados, M. J. (2020). “Quantifying the life-cycle benefits of an influential early-childhood program.” *Journal of Political Economy*, *128*(7), 2502–2541. https://doi.org/10.1086/706926
- Ghassami, A., Yang, A., Richardson, D., Shpitser, I., & Tchetgen Tchetgen, E. (2022). “Combining experimental and observational data for identification and estimation of long-term causal effects.” arXiv preprint, arXiv:2201.10743. https://doi.org/10.48550/arXiv.2201.10743
- Golosov, M., Graber, M., Mogstad, M., & Novgorodsky, D. (2024). “How Americans respond to idiosyncratic and exogenous changes in household wealth and unearned income.” *The Quarterly Journal of Economics*, *139*(2), 1321–1395. https://doi.org/10.1093/qje/qjad053
- Heckman, J.J. (1979). “Sample selection bias as a specification error.” *Econometrica* 47(1), 153-161. https://doi.org/10.2307/1912352

# References


- Heckman, J.J., Pinto, R. (2015). "Causal analysis after Haavelmo" *Econometric Theory* 31(1), 115-151. https://doi.org/10.1017/S026646661400022X
- Heckman, J. J., Stixrud, J., & Urzúa, S. (2006). "The effects of cognitive and noncognitive abilities on labor-market outcomes and social behavior." *Journal of Labor Economics*, 24(3), 411–482. https://doi.org/10.1086/504455
- Hotz, V. J., Imbens, G. W., & Klerman, J. A. (2006). "Evaluating the differential effects of alternative welfare-to-work training components: A re-analysis of the California GAIN program." *Journal of Labor Economics*, *24*(3), 521–566. https://doi.org/10.1086/505050
- Imbens, G., Kallus, N., Mao, X., & Wang, Y. (2024). "Long-term causal inference under persistent confounding via data combination." arXiv preprint, arXiv:2202.07234. https://doi.org/10.48550/arXiv.2202.07234
- Jack, B.K., Jayachandran, S., Kala N., Pande, R. (2025). "Money (not) to burn: payments for ecosystem services to reduce crop residue burning." *American Economic Review* 7(1) : 39–55. https://doi.org/10.1257/aeri.20230431
- Kallus, N., & Mao, X. (2024). "On the role of surrogates in the efficient estimation of treatment effects with limited outcome data." arXiv preprint, arXiv:2003.12408. https://doi.org/10.48550/arXiv.2003.12408

# References


- Krueger, A. B. (1999). "Experimental estimates of education production functions." *Quarterly Journal of Economics, 114*(2), 497–532. https://doi.org/10.1162/003355399556052
- Mariano, T.L., Martorell, P., Berglund, T. (2024). "The effects of grade retention on high school outcomes: evidence from New York City schools." *Journal of Educational Effectiveness*, 1-31. https://doi.org/10.1080/19345747.2023.2287607
- Meza, I., & Singh, R. (2025). "Nested nonparametric instrumental variable regression." arXiv preprint, arXiv:2112.14249. https://doi.org/10.48550/arXiv.2112.14249
- Obradović, F. (2024). "Identification of long-term treatment effect via temporal links, observational, and experimental data." arXiv preprint, arXiv:2411.04380. https://doi.org/10.48550/arXiv.2411.04380
- Park, Y., & Sasaki, Y. (2024a). "A bracketing relationship for long-term policy evaluation with combined experimental and observational data." arXiv preprint, arXiv:2401.12050. https://doi.org/10.48550/arXiv.2401.12050
- Park, Y., & Sasaki, Y. (2024b). "The informativeness of combined experimental and observational data under dynamic selection. arXiv preprint arXiv:2403.16177. https://doi.org/10.48550/arXiv.2403.16177.
- Pearl, J. (2000). "Causality: Models, reasoning, and inference." New York, NY: Cambridge University Press.

# References


- Prentice, R.L. (1989). "Surrogate endpoints in clinical trials: definition and operational criteria." *Statistics in medicine*, 8(4): 431-440. https://doi.org/10.1002/sim.4780080407
- Rambachan, A., Singh, R., & Viviano, D. (2024). "Program evaluation with remotely sensed outcomes." arXiv preprint, arXiv:2411.10959. https://doi.org/10.48550/arXiv.2411.10959
- Ridder, G., & Moffitt, R. (2007). "The econometrics of data combination." Handbook of Econometrics 6, 5469-5547. https://doi.org/10.1016/S1573-4412(07)06075-8
- VanderWeele, T.J.. (2013). "Surrogate measures and consistent surrogates." *Biometrics* 69(3). https://doi.org/10.1111/biom.12071