Estimating the Intensive Margin Effect in Panel Data Settings

Javier Viviens

European University Institute
Department of Economics

Joint Research Centre, Ispra
February 10, 2026

Motivation

- Many policies have an effect through two distinct channels: the **intensive margin** and the **extensive margin**

  Examples:

  - Job training policies affect wages, but also labor market participation
  - Education policies affect performance, but also enrollment
  - Carbon taxes affect emissions, but also firms' entry and exit
  - Childcare provision affects parents' labor market outcomes, but also fertility decisions
  - Education affects earnings, but also occupational choices

  In these examples, the intensive margin must be estimated in a selected sample

- **This paper** studies how to disentangle these two effects and correctly estimate the **intensive margin** effect in **panel data** settings (DiD)

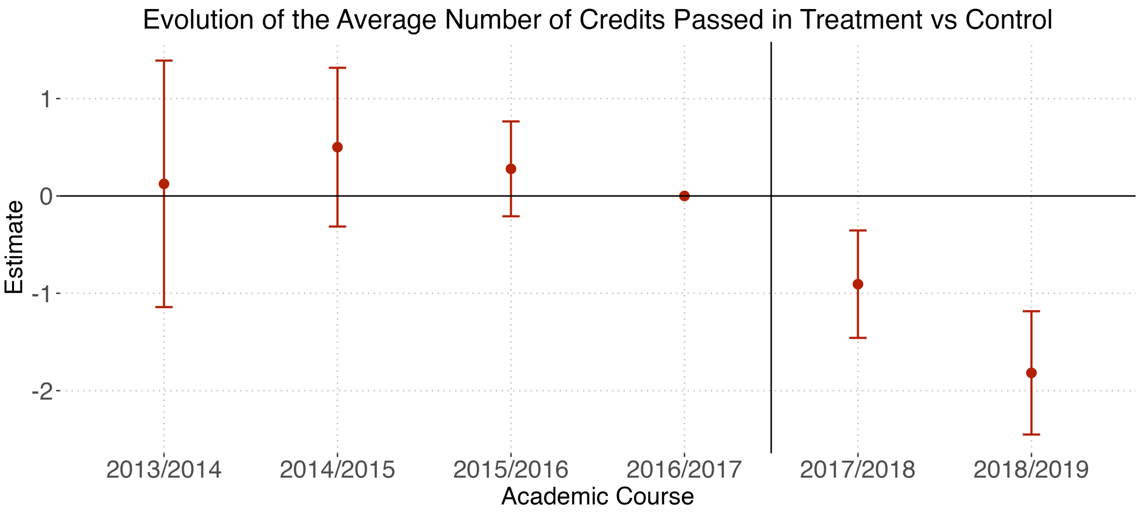
A running (real-world) example

- In 2017, the Andalusian government (Spain) introduced tuition fee waivers based on previous academic performance
- The policy affected all students in Andalusian public universities ($\sim 20\%$ of the total in Spain)
- The main goal of the waiver was to increase academic performance
- **Research Question:**

  How did this policy affect the academic performance of treated students?

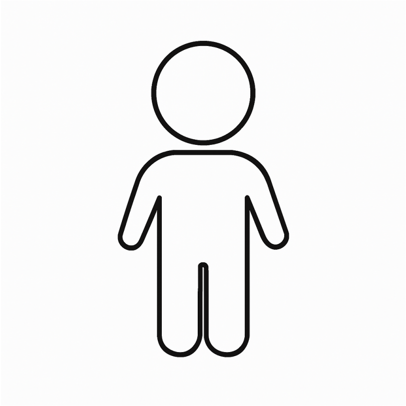
- Available data:
  - Individual performance (i.e., # Credits Passed) before and after the policy
  - Two groups of universities: Andalusian (treated) and the rest of Spain (control)

A descriptive event study



The extensive margin effect

Observed (e.g. Enrolled) **Not observed (e.g. drop out)**



S = 1 S = 0

Enrollment as a post-treatment variable

Control

$S_i(0) = 0$

Treatment

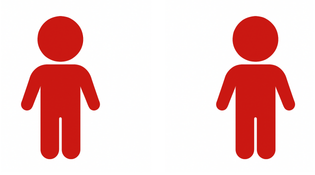
$S_i(1) = 1$

This student would be **observed** in the sample **only if treated**

Principal Strata (Frangakis & Rubin, 2002)

Always Observed

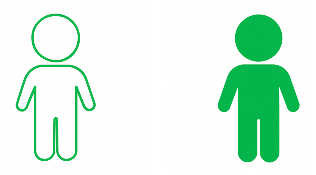
Control Treatment



$S_i(0) = 1 \quad S_i(1) = 1$

Observed-if-Treated

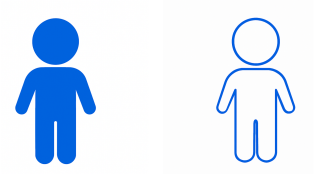
Control Treatment



$S_i(0) = 0 \quad S_i(1) = 1$

Observed-if-Control

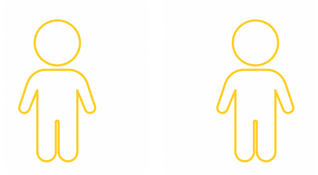
Control Treatment



$S_i(0) = 1 \quad S_i(1) = 0$

Never Observed

Control Treatment



$S_i(0) = 0 \quad S_i(1) = 0$

Drawbacks of the complete case analysis

Suppose you discard observations with missing outcomes and perform your preferred analysis using only the units without missing outcomes

1. **Conditioning on a post-treatment variable** (e.g., being enrolled), your estimand loses its causal interpretation
   Known as 'bad control' (Angrist & Pischke, 2009; Rosenbaum, 1984)

2. Your estimand will have limited interpretation.
   It does not allow you to **distinguish between the extensive and intensive margin**

This paper:

Estimate causal effects for the Always-Observed principal stratum ($S(0) = S(1) = 1$)

The proposed estimands can be interpreted as the effect at the intensive margin

Literature

- One particular, well-studied form of this problem is sample selection (Heckman, 1979; Manski, 1989)
- Seminal paper by Zhang and Rubin (2003) formally introduced the use of Principal Stratification in sample selection settings (Imai, 2008; Zhang et al., 2009)
- Lee (2009) popularized the 'Lee bounds': a trimming approach to partially identify the ATE on the Always-Observed stratum
- This literature has been extended in multiple directions: including covariates (Ding et al., 2011; Grilli & Mealli, 2008; Long & Hudgens, 2013; Samii et al., 2025; Semenova, 2025), instruments (Mattei et al., 2014), post-censoring outcomes (Yang & Small, 2016), structural model for selection (Honoré & Hu, 2020), longitudinal data (Comment et al., 2025; Grossi et al., 2023), noncompliance (Blanco et al., 2020; Chen & Flores, 2015)
- However, **all these papers require the treatment to be unconfounded**

This paper as an extension of the Horowitz-Manski-Lee bounds

The **original Horowitz-Manski-Lee bounds** rely on the assumptions of:

1. Treatment unconfoundedness
2. Monotonicity of selection

This paper :

1. Allows treatment to be confounded with unobservable unit-level characteristics
2. Relaxes monotonicity when multiple sources of selection are available

Some notation

Consider the 2 periods ($t \in \{1, 2\}$), 2 groups case,

- Y_{it}: outcome of interest (e.g., # Credits Passed)
- G_i: treatment group indicator (1 treatment, 0 control)
- S_{it}: binary selection indicator (1 if unit i's outcome is observed at period t, 0 otherwise)
- V_i: Principal Stratum of unit i

$\boldsymbol{S_{i2}(0)}$	$\boldsymbol{S_{i2}(1)}$	**Description**	$\boldsymbol{V_i}$
1	1	Always Observed	AO
0	0	Never Observed	NO
1	0	Observed only in Control	OC
0	1	Observed only in Treatment	OT

Principal Causal Effects

1. Average Treatment Effect on the Treated **Always-Observed** units

$$\mathrm{ATT}_{\mathrm{AO}} = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO]$$

2. Quantile Treatment Effect on the Treated **Always-Observed** units

$$\mathrm{QTT}_{\mathrm{AO}}(q) = Q_{Y_2(1)|G=1,V=AO}(q) - Q_{Y_2(0)|G=1,V=AO}(q)$$
$$q \in [0,1]$$

where

- $F_Y(y)$ is the c.d.f. of Y
- $Q_Y(q)$ is the quantile function of Y, $Q_Y(q) = \inf\{y : F_Y(y) \geq q\}$

Principal Stratification vs Bias-correction

- This paper:

$$\text{ATT}_{\text{AO}} = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO]$$

What's the effect for the units with observed outcomes under both treatment arms?
(Intensive margin effect)

- Bias-correction literature:

$$ATT = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1]$$

What would be the effect if the outcome for all the units were observed?
(Blend of intensive and extensive margin effect)

This literature implicitly assumes...

  - outcomes are always well defined, even when not directly observed
  - selection is manipulable, as they rely on 'a priori counterfactuals' potential outcomes
  - restrictions on how potential outcomes are distributed across principal strata (exclusion restrictions, structural models)

Some useful readings: Angrist et al. (1996) with discussions and rejoinder, Imbens (2010) and Mealli and Mattei (2012)

Identification of latent distributions

- Observed distributions are mixtures of principal strata:
  - Control group: Always-Observed and Observed-if-Control

$$F_{Y_2|G=0,S_2=1}(y) = \pi_0 F_{Y_2|G=0,V=AO}(y) + (1-\pi_0) F_{Y_2|G=0,V=OC}(y)$$

  - Treatment group: Always-Observed and Observed-if-Treated

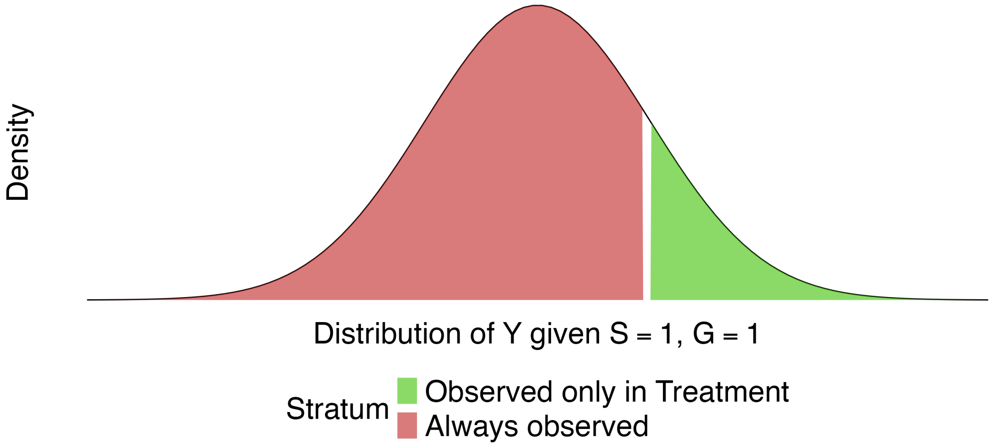
$$F_{Y_2|G=1,S_2=1}(y) = \pi_1 F_{Y_2|G=1,V=AO}(y) + (1-\pi_1) F_{Y_2|G=1,V=OT}(y)$$

where $\pi_g = Pr(V_i = AO \mid G_i = g, S_2 = 1)$

- Given the mixture proportions, I follow Horowitz and Manski (1995) to partially identify the distributions of the observed outcome for the AO units

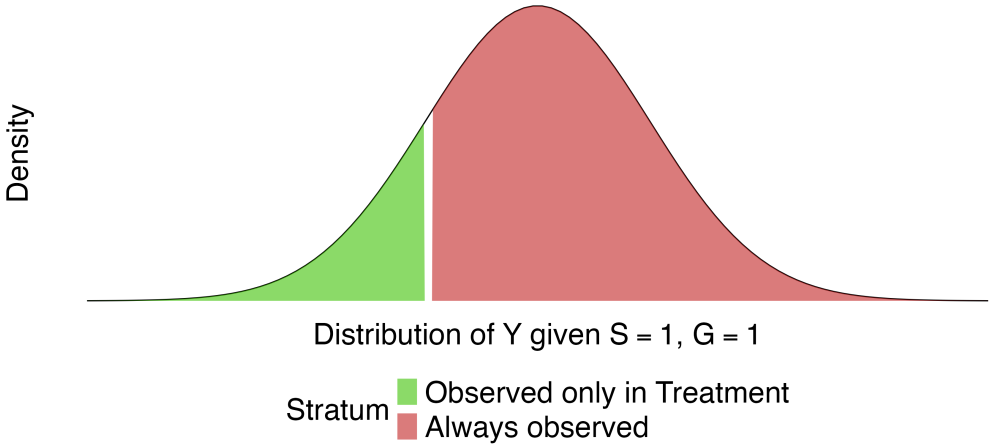
Lower bound

Figure: Intuition behind Horowitz and Manski (1995)

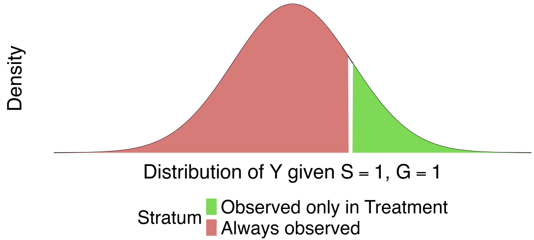


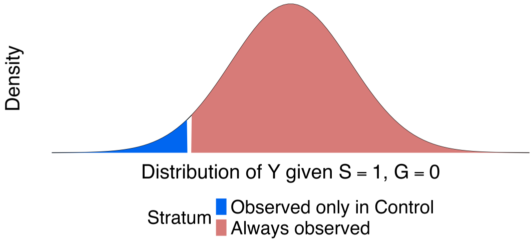
Upper bound

Figure: Intuition behind Horowitz and Manski (1995)



Using the trimmed distributions





Identification assumption: Changes-in-Changes (Athey & Imbens, 2006)

Assumption: Changes-in-changes for the **Always-Observed** **Outcome**:

$$(Y_{it}(0) \mid V_i = AO) = m(\mathcal{U}_{it}, t)$$

where

- $m(\cdot)$ is strictly increasing in $\mathcal{U}$
- the distribution of $\mathcal{U}_{it} \mid G_i, V_i = AO$ is stable over time
- and $\mathbb{Y}_1 \subseteq \mathbb{Y}_0$

Example: $Y_{it}(0) = m(\mathcal{U}_{it}, t) = \lambda_t + \underbrace{\alpha_i + \varepsilon_{it}}_{\mathcal{U}_{it}}$

Intuition behind Changes-in-Changes I

Suppose unit i (treated) and j (control) have the same value of Y in the pre-treatment period:

$$Y_{i1} = Y_{j1}$$

Then, $Y_{it}(0) = m(\mathcal{U}_{it}, t)$, with $m(\cdot)$ is strictly increasing in $\mathcal{U}$ implies

$$\mathcal{U}_{i1} = \mathcal{U}_{i2}$$

If $\mathcal{U}$ were constant over time, the missing potential outcome for unit i would be given by unit j's outcome:

$$Y_{i2}(0) = m(\mathcal{U}_{i2}, 2) = m(\mathcal{U}_{j2}, 2) = Y_{j2}$$

The assumption is not constant $\mathcal{U}$ but constant distribution within groups

Intuition behind Changes-in-Changes II

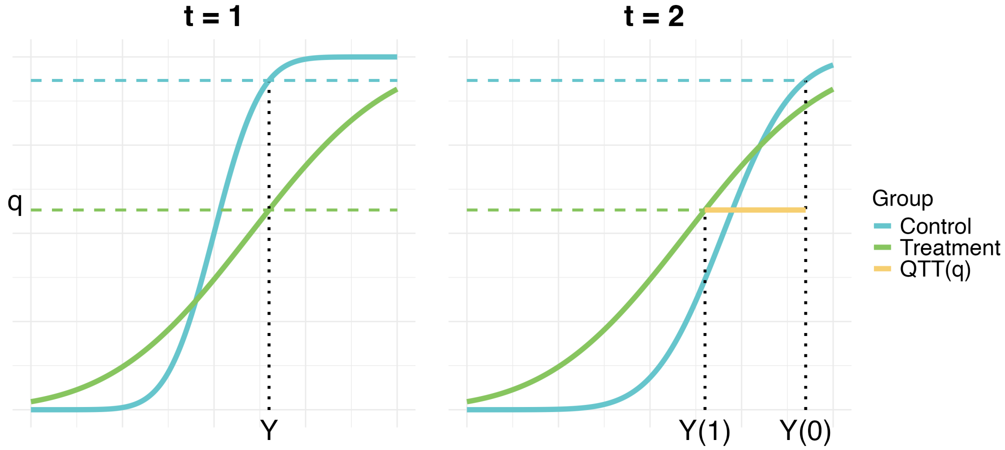
We want to impute the quantile q of the distribution of the missing potential outcome $Y_2(0)$ for treated units: $Q_{Y_2(0)|G=1}(q)$

$$\text{If } Y_{i1} = Y_{j1} \text{ and} \begin{cases} F_{Y_1|G=1}(Y_{i1}) = q \\ F_{Y_1|G=0}(Y_{j1}) = \tilde{q} \end{cases} \implies Q_{Y_1|G=1}(q) = Q_{Y_1|G=0}(\tilde{q})$$

Under the CiC assumptions, we can reconstruct the missing distribution of $Y_{i2}(0)$ for treated units as follows:

$$Q_{Y_2(0)|G=1}(q) = Q_{Y_2|G=0,S_2=1}\left(\underbrace{F_{Y_1|G=0,S_2=1}\left(\underbrace{Q_{Y_1|G=1,S=1}(q)}_{Y_{i1}=Y_{j2}}\right)}_{\tilde{q}}\right)$$

Intuition behind Changes-in-Changes III



Identification result: CiC

Identification result: $\Lambda^{LB}(q)$ and $\Lambda^{UB}(q)$ are sharp lower and upper bounds for the Quantile Treatment Effect on the Treated Always-Observed units $(\text{QTT}_{\text{AO}}(q))$, where:

$$\begin{aligned}
\Lambda^{LB}(q) &= Q_{Y_2|G=1,S_2=1}(q\pi_1) \\
&\quad - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0\right) \\
\Lambda^{UB}(q) &= Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) \\
&\quad - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S=1}(q\pi_1)\right) - (1 - \pi_0)\right) \\
\pi_1 &= Pr(V_i = AO \mid G_i = 1, S_{i2}(1) = 1) \\
\pi_0 &= Pr(V_i = AO \mid G_i = 0, S_{i2}(0) = 1) \\
F_Y(y) &:= Pr(Y \leq y) \\
Q_Y(q) &:= \inf\{y : F_Y(y) \geq q\}
\end{aligned}$$

$$\int_0^1 \Lambda^{LB}(q)dq \quad \leq \quad \text{ATT}_{\text{AO}} \quad \leq \quad \int_0^1 \Lambda^{UB}(q)dq$$

Full Identification result

Identification result: DiD

Assumption: **Principal** Parallel Trends

$$\mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid G_i = 1, V_i = AO] = \mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid G_i = 0, V_i = AO]$$

Identification result: Δ^{LB} and Δ^{UB} are sharp lower and upper bounds for the Average Treatment Effect on the Treated Always-Observed units (ATT_{AO}), where

$$\Delta^{LB} = \mathbb{E}[Y_{i2} - Y_{i1} \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^1_{\pi_1}] - \mathbb{E}[Y_{i2} - Y_{i1} \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \geq \ddot{y}^0_{1-\pi_0}]$$

$$\Delta^{UB} = \mathbb{E}[Y_{i2} - Y_{i1} \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \geq \ddot{y}^1_{1-\pi_1}] - \mathbb{E}[Y_{i2} - Y_{i1} \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^0_{\pi_0}]$$

$$\ddot{Y}_i = Y_{i2} - Y_{i1}$$

$\ddot{y}^g_q = \inf\{\ddot{y} : F(\ddot{y}) \geq q\}$, with F the c.d.f. of $\ddot{Y}$ conditional on $S_2 = 1$ and $G = g$

$$\pi_1 = Pr(V_i = AO \mid G_i = 1, S_{i2}(1) = 1)$$

$$\pi_0 = Pr(V_i = AO \mid G_i = 0, S_{i2}(0) = 1)$$

Identification of the principal strata proportions

Two assumptions needed:

1. Monotonicity:
$$S_{i2}(1) \geq S_{i2}(0) \quad \forall i$$

2. Changes-in-Changes (Athey & Imbens, 2006) for **Selection**
$$S_{it}(0) = h^0(U_{it}, t)$$

   where
   - $h^0(\cdot)$ is non-decreasing in U
   - the distribution of $U_{it} \mid G_i$ is constant over time
   - and $\mathbb{U}_1 \subseteq \mathbb{U}_0$

Expression

Identification of the principal strata proportions

Or conversely:

1. Monotonicity:
$$S_{i2}(1) \leq S_{i2}(0) \quad \forall i$$

2. Changes-in-Changes (Athey & Imbens, 2006) for **Selection**
$$S_{it}(1) = h^1(U_{it}, t)$$

   - $h^1(\cdot)$ is non-decreasing in U
   - the distribution of $U_{it} \mid G_i$ is constant over time
   - and $\mathbb{U}_0 \subseteq \mathbb{U}_1$

▸ Expression

Full Identification Result

Identification result: $\Lambda^{LB}(q)$ and $\Lambda^{UB}(q)$ are lower and upper bounds for the Quantile Treatment Effect on the Treated Always-Observed units $(\text{QTT}_{\text{AO}}(q))$, where:

$$
\begin{aligned}
\Lambda^{LB}(q) &= Q_{Y_2|G=1,S_2=1}(q\pi_1) \\
&\quad - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0\right) \\
\Lambda^{UB}(q) &= Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) \\
&\quad - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S=1}(q\pi_1)\right) - (1 - \pi_0)\right) \\
\pi_1 &= 1 \\
\pi_0 &= \frac{\mathbb{E}[S_{i2}(1) \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} = \frac{\mathbb{E}[S_{i1} \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} \frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} \in [0, 1] \\
F_Y(y) &:= Pr(Y \leq y) \\
Q_Y(q) &:= \inf\{y : F_Y(y) \geq q\}
\end{aligned}
$$

$$
\int_0^1 \Lambda^{LB}(q)dq \quad \leq \quad \text{ATT}_{\text{AO}} \quad \leq \quad \int_0^1 \Lambda^{UB}(q)dq
$$

Relaxing the Monotonicity Assumption

If different unobservables affect selection in different directions, and they translate into different sources of selection, we can leverage them to relax monotonicity

Example:

- tuition waivers may help financially constrained students $\implies$ decrease in dropout

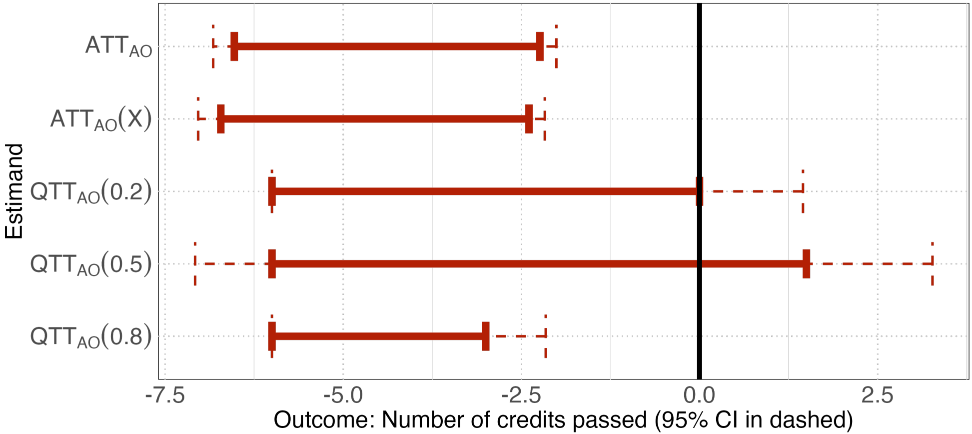
$$s^{drop}(1) \geq s^{drop}(0)$$

- tuition waivers may increase performance $\implies$ increase in graduation

$$s^{grad}(1) \leq s^{grad}(0)$$

▸ Expression

Back to the real-world example



$\mathrm{ATT}_{\mathrm{AO}}(X)$ refers to the $\mathrm{ATT}_{\mathrm{AO}}$ including the following covariates in the estimation: parental education and highschool grades

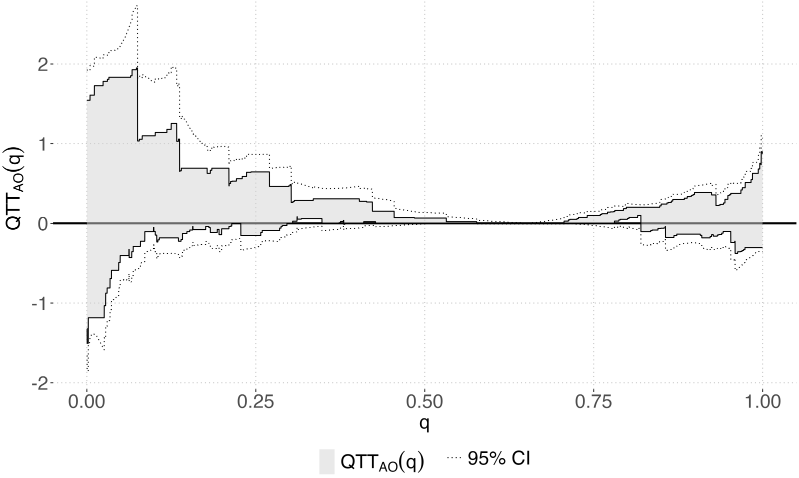
Revisiting Attanasio et al. (2011)

- Attanasio et al. (2011) estimate the effect of a job training program on wages

Outcome	Log of salaried earnings			
Estimand	ATT_{AO}		Complete Case	
Estimate	[-0.11 , 0.429]	[-0.095 , 0.319]	0.129	0.123
95% CI	(-0.169 , 0.527)	(-0.158 , 0.416)	(0.037 , 0.22)	(0.031 , 0.215)
Covariates	No	Yes	No	Yes
N	888	888	888	888

Table: Estimates of the bounds for the ATT_{AO} and naive approach (using complete case analysis)

Distributional effects in Attanasio et al. (2011)



In the paper you will find...

- Asymptotically Normal estimators
- Valid inference (Imbens & Manski, 2004)
- Some extensions
  - Sample selection in the pre-treatment period
  - Relax monotonicity assumption
  - Discrete Outcomes
  - Covariates
  - Repeated Cross Sections
- An application (Attanasio et al., 2011)

\+ R package (coming soon!)

Working Paper:

arxiv.org/abs/2502.08614

Thanks!
Javier.Viviens@eui.eu

Full identification result CiC I

Identification result: $\Lambda^{LB}(q)$ and $\Lambda^{UB}(q)$ are sharp lower and upper bounds for the Quantile Treatment Effect on the Treated Always-Observed units $(\text{QTT}_{\text{AO}}(q))$, where:

$$
\begin{aligned}
\Lambda^{LB}(q) &= Q_{Y_2|G=1,S_2=1}(q\pi_1) \\
&\quad - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1+1-\pi_1)\right)+1-\pi_0\right) \\
\Lambda^{UB}(q) &= Q_{Y_2|G=1,S_2=1}(q\pi_1+1-\pi_1) \\
&\quad - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S=1}(q\pi_1)\right)-(1-\pi_0)\right) \\
\pi_1 &= Pr(V_i = AO \mid G_i = 1, S_{i2}(1) = 1) \\
\pi_0 &= Pr(V_i = AO \mid G_i = 0, S_{i2}(0) = 1) \\
F_Y(y) &:= Pr(Y \leq y) \\
Q_Y(q) &:= \inf\{y : F_Y(y) \geq q\}
\end{aligned}
$$

provided that

$$
\begin{aligned}
F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1+1-\pi_1)\right) &\leq \pi_0 \\
F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\pi_1)\right) &\geq 1-\pi_0
\end{aligned}
$$

Full identification result CiC II

If $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) > \pi_0$, then

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}(1)$$

If $F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=0}(q\pi_1)\right) < 1 - \pi_0$, then

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) - Q_{Y_2|G=0,S_2=1}(0)$$

◂ Return to presentation

Identification PS proportions

Under Positive Monotonicity,

$$S_{i2}(1) \geq S_{i2}(0) \quad \forall i:$$

$$\pi_0 = 1$$

$$\pi_1 = \frac{\mathbb{E}[S_{i2}(0) \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} = \frac{\mathbb{E}[S_{i1} \mid G_i = 1]}{\mathbb{E}[S_{i2} \mid G_i = 1]} \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]} \in [0, 1]$$

Return to presentation

Identification PS proportions

Under Negative Monotonicity,

$$S_{i2}(1) \leq S_{i2}(0) \quad \forall i:$$

$$\pi_0 = \frac{\mathbb{E}[S_{i2}(1) \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} = \frac{\mathbb{E}[S_{i1} \mid G_i = 0]}{\mathbb{E}[S_{i2} \mid G_i = 0]} \frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} \in [0,1]$$

$$\pi_1 = 1$$

Return to presentation

Identification PS proportions with multiple sources

$$\pi_0 = \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 0]} \left(1 - \sum_{j \in J^+} \Big(1 - \mathbb{E}[s^j_{i2} \mid G_i = 0] \Big) - \sum_{j \in J^-} \left(1 - \mathbb{E}[s^j_{i1} \mid G_i = 0] \frac{\mathbb{E}[s^j_{i2} \mid G_i = 1]}{\mathbb{E}[s^j_{i1} \mid G_i = 1]} \right) \right)$$

$$\pi_1 = \frac{1}{\mathbb{E}[S_{i2} \mid G_i = 1]} \left(1 - \sum_{j \in J^+} \left(1 - \mathbb{E}[s^j_{i1} \mid G_i = 1] \frac{\mathbb{E}[s^j_{i2} \mid G_i = 0]}{\mathbb{E}[s^j_{i1} \mid G_i = 0]} \right) - \sum_{j \in J^-} \Big(1 - \mathbb{E}[s^j_{i2} \mid G_i = 1] \Big) \right)$$

◂ Return to presentation

Asymptotically Normal Estimators

$$\hat{\pi}_0 = \frac{1}{\frac{\sum_i S_{i2}(1-G_i)}{N_0}} \left(1 - \left(1 - \frac{\sum_i s^1_{i2}(1-G_i)}{N_0}\right) - \left(1 - \frac{\sum_i s^2_{i1}(1-G_i)}{N_0} \frac{\sum_i s^2_{i2}G_i}{\sum_i s^2_{i1}G_i}\right)\right) \tag{1}$$

$$\hat{\pi}_1 = \frac{1}{\frac{\sum_i S_{i2}G_i}{N_1}} \left(1 - \left(1 - \frac{\sum_i s^1_{i1}G_i}{N_1} \frac{\sum_i s^1_{i2}(1-G_i)}{\sum_i s^1_{i1}(1-G_i)}\right) - \left(1 - \frac{\sum_i s^2_{i2}G_i}{N_1}\right)\right) \tag{2}$$

$$\widehat{\Lambda^{LB}}(q) = \widehat{Q}_{Y_2|G=1,S_2=1}(q\hat{\pi}_1) - \widehat{Q}_{Y_2|G=0,S_2=1}\left(\hat{q}^*_{LB}\right) \tag{3}$$

$$\hat{q}^*_{LB} = \min\left\{\widehat{F}_{Y_1|G=0,S_2=1}\left(\widehat{Q}_{Y_1|G=1,S_2=1}(q\hat{\pi}_1 + 1 - \hat{\pi}_1)\right) + 1 - \hat{\pi}_0, 1\right\} \tag{4}$$

$$\widehat{\Lambda^{UB}}(q) = \widehat{Q}_{Y_2|G=1,S_2=1}(q\hat{\pi}_1 + 1 - \hat{\pi}_1) - \widehat{Q}_{Y_2|G=0,S_2=1}\left(\hat{q}^*_{UB}\right) \tag{5}$$

$$\hat{q}^*_{UB} = \max\left\{\widehat{F}_{Y_1|G=0,S_2=1}\left(\widehat{Q}_{Y_1|G=1,S_2=1}(q\hat{\pi}_1)\right) - (1 - \hat{\pi}_0), 0\right\} \tag{6}$$

where $\widehat{F}_Y$ denotes the empirical distribution of Y Return to presentation

Empirical Distributions

For instance

$$\widehat{F}_{Y_1|G=0,S_2=1}(y) = \frac{\sum_i S_{i2}(1-G_i)\mathbb{1}(Y_{i1} \leq y)}{\sum_i S_{i2}(1-G_i)}$$

$$\widehat{Q}_{Y_2|G=0,S_2=1}(q) = \inf\left\{y : \widehat{F}_{Y_2|G=0,S_2=1}(y) \geq q\right\}$$

Return to presentation

Discrete Outcomes: CiC

$$\Lambda^{LB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1) - Q_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(Q_{Y_1|G=1,S_2=1}(q\pi_1 + 1 - \pi_1)\right) + 1 - \pi_0\right)$$

$$\Lambda^{UB}(q) = Q_{Y_2|G=1,S_2=1}(q\pi_1 + 1 - \pi_1) - \widetilde{Q}_{Y_2|G=0,S_2=1}\left(F_{Y_1|G=0,S_2=1}\left(\widetilde{Q}_{Y_1|G=1,S=1}(q\pi_1)\right) - (1 - \pi_0)\right)$$

$$\pi_1 = Pr(S_{i2}(0) = 1 \mid G_i = 1, S_{i2}(1) = 1)$$
$$\pi_0 = Pr(S_{i2}(1) = 1 \mid G_i = 0, S_{i2}(0) = 1)$$

$$F_Y(y) := Pr(Y \leq y)$$

$$Q_Y(q) := \inf\{y : F_Y(y) \geq q\}$$

$$\widetilde{Q}_Y(q) := \sup\{y \cup \{-\infty\} : F_Y(y) \leq q\}$$

◂ Return to presentation

Discrete Outcomes: DiD (reweighting)

$$\underline{\mathcal{P}}^{\pi}_{Y}(y) = \begin{cases} \frac{Pr(Y=y)}{\pi} & \text{if} \quad y < Q_Y(\pi) \\ \frac{\pi - Pr(y < Q_Y(\pi))}{\pi} & \text{if} \quad y = Q_Y(\pi) \\ 0 & \text{if} \quad y > Q_Y(\pi) \end{cases} \tag{7}$$

$$\overline{\mathcal{P}}^{\pi}_{Y}(y) = \begin{cases} 0 & \text{if} \quad y < Q_Y(1-\pi) \\ \frac{\pi - Pr(Y > Q_Y(1-\pi))}{\pi} & \text{if} \quad y = Q_Y(1-\pi) \\ \frac{Pr(Y=y)}{\pi} & \text{if} \quad y > Q_Y(1-\pi) \end{cases} \tag{8}$$

$$\Delta^{LB} = \sum_{i:G_i=1,S_{i2}=1} \ddot{Y}_i \underline{\mathcal{P}}^{\pi_1}_{\ddot{Y}_i|G_i=1,S_{i2}=1}(y) - \sum_{i:G_i=0,S_{i2}=1} \ddot{Y}_i \overline{\mathcal{P}}^{\pi_0}_{\ddot{Y}_i|G_i=0,S_{i2}=1}(y)$$

$$\Delta^{UB} = \sum_{i:G_i=1,S_{i2}=1} \ddot{Y}_i \overline{\mathcal{P}}^{\pi_1}_{\ddot{Y}_i|G_i=1,S_{i2}=1}(y) - \sum_{i:G_i=0,S_{i2}=1} \ddot{Y}_i \underline{\mathcal{P}}^{\pi_0}_{\ddot{Y}_i|G_i=0,S_{i2}=1}(y)$$

◂ Return to presentation

Covariates I

1. Target Conditional Estimands: e.g., Conditional Average Treatment Effect on the Treated

$$CATT_{\text{AO}}(x) = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 1, V_i = AO, X_i = x]$$

2. Make Conditional Assumptions: e.g., Conditional Positive Monotonicity

$$S_{i2}(1) \geq S_{i2}(0) \quad \forall i : X_i = x$$

3. Identification and estimation in each cell of the covariates:

$$\Delta^{LB}(x) = \mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^1_{\pi_1}, X_i = x] - \mathbb{E}[\ddot{Y}_i \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \geq \ddot{y}^0_{1-\pi_0}, X_i = x]$$

$$\Delta^{UB}(x) = \mathbb{E}[\ddot{Y}_i \mid G_i = 1, S_{i2} = 1, \ddot{Y}_i \geq \ddot{y}^1_{1-\pi_1}, X_i = x] - \mathbb{E}[\ddot{Y}_i \mid G_i = 0, S_{i2} = 1, \ddot{Y}_i \leq \ddot{y}^0_{\pi_0}, X_i = x]$$

$$\pi_0 = 1$$

$$\pi_1 = \frac{\mathbb{E}[S_{i1} \mid G_i = 1, X_i = x]}{\mathbb{E}[S_{i2} \mid G_i = 1, X_i = x]} \frac{\mathbb{E}[S_{i2} \mid G_i = 0, X_i = x]}{\mathbb{E}[S_{i1} \mid G_i = 0, X_i = x]}$$

Covariates II

1. Aggregate:

$$\overline{\text{ATT}_{\text{AO}}} \in \left[\overline{\Delta^{LB}}, \overline{\Delta^{UB}}\right]$$

where

$$\overline{\Delta^{LB}} = \sum_{x \in X} \Delta^{LB}(x)p(x) \quad ; \quad \overline{\Delta^{UB}} = \sum_{x \in X} \Delta^{UB}(x)p(x)$$

$$p(x) = Pr(X_i = x \mid G_i = 1, V_i = AO)$$

which are estimated as:

$$\widehat{\Delta^{LB}} = \sum_{x \in X} \omega_x \widehat{\Delta^{LB}}(x) \quad ; \quad \widehat{\Delta^{UB}} = \sum_{x \in X} \omega_x \widehat{\Delta^{UB}}(x)$$

$$\omega_x = \frac{\widehat{\pi}_1^x}{\Pi_1} \frac{\sum_i \mathbb{1}(G_i = 1, S_{i2} = 1, X_i = x)}{N_1} \quad ; \quad \Pi_1 = \sum_{x \in X} \widehat{\pi}_1^x$$

Return to Graph   Return to presentation

Repeated Cross-Sections

The proportions of AO units in the pre-treatment period are given by:

$$\varpi_1 = \pi_1 \frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]} \qquad ; \qquad \varpi_0 = \pi_0 \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]}$$

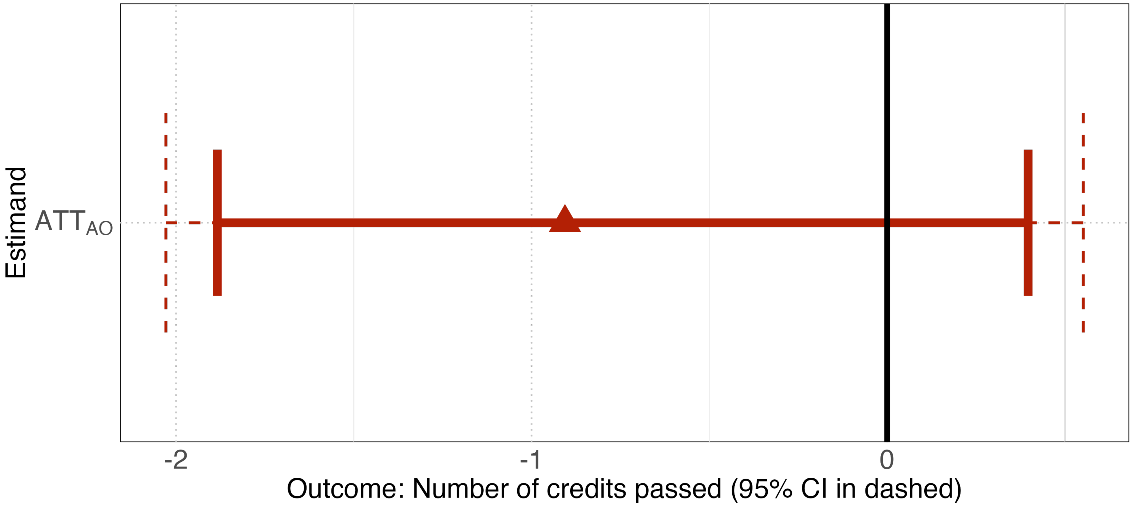
or, if missigness is not an absorbing state, they are bounded by

$$\max\left\{0, \frac{\pi_1\mathbb{E}[S_{i2} \mid G_i = 1] - (1 - \mathbb{E}[S_{i1} \mid G_i = 1])}{\mathbb{E}[S_{i1} \mid G_i = 1]}\right\} \leq \varpi_1 \leq \min\left\{1, \pi_1 \frac{\mathbb{E}[S_{i2} \mid G_i = 1]}{\mathbb{E}[S_{i1} \mid G_i = 1]}\right\}$$

$$max\left\{0, \frac{\pi_0\mathbb{E}[S_{i2} \mid G_i = 0] - (1 - \mathbb{E}[S_{i1} \mid G_i = 0])}{\mathbb{E}[S_{i1} \mid G_i = 0]}\right\} \leq \varpi_0 \leq \min\left\{1, \pi_0 \frac{\mathbb{E}[S_{i2} \mid G_i = 0]}{\mathbb{E}[S_{i1} \mid G_i = 0]}\right\}$$

Return to presentation

Bounds with DiD



References I

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996).Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.2307/2291634

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. https://doi.org/10.2307/j.ctvcm4j72

Athey, S., & Imbens, G. W. (2006).Identification and inference in nonlinear difference-in-differences models. *Econometrica*, *74*(2), 431–497. https://doi.org/10.1111/j.1468-0262.2006.00668.x

Attanasio, O., Kugler, A., & Meghir, C. (2011).Subsidizing vocational training for disadvantaged youth in colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, *3*(3), 188–220. https://doi.org/10.1257/app.3.3.188

Blanco, G., Chen, X., Flores, C. A., & Flores-Lagunes, A. (2020).Bounds on average and quantile treatment effects on duration outcomes under censoring, selection, and noncompliance. *Journal of Business & Economic Statistics*, *38*(4), 901–920. https://doi.org/10.1080/07350015.2019.1609975

References II

Chen, X., & Flores, C. A. (2015).Bounds on treatment effects in the presence of sample selection and noncompliance: The wage effects of job corps. *Journal of Business & Economic Statistics*.

Comment, L., Mealli, F., Haneuse, S., & Zigler, C. M. (2025).Survivor average causal effects for continuous time: A principal stratification approach to causal inference with semicompeting risks. *Biometrical Journal*, *67*(2), e70041. https://doi.org/10.1002/bimj.70041

Ding, P., Geng, Z., Yan, W., & Zhou, X.-H. (2011).Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association*, *106*(496), 1578–1591. https://www.jstor.org/stable/23239560

Frangakis, C. E., & Rubin, D. B. (2002).Principal stratification in causal inference. *Biometrics*, *58*(1), 21–29. https://doi.org/10.1111/j.0006-341x.2002.00021.x

Grilli, L., & Mealli, F. (2008).Nonparametric bounds on the causal effect of university studies on job opportunities using principal stratification. *Journal of Educational and Behavioral Statistics*, *33*(1), 111–130. Retrieved May 22, 2024, from https://www.jstor.org/stable/20172106

References III

Grossi, G., Mariani, M., Mattei, A., & Mealli, F. (2023, December 30). Bayesian principal stratification with longitudinal data and truncation by death. https://doi.org/10.48550/arXiv.2401.00196

Heckman, J. J. (1979).Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161. https://doi.org/10.2307/1912352

Honoré, B. E., & Hu, L. (2020).Selection without exclusion. *Econometrica*, *88*(3), 1007–1029. https://doi.org/10.3982/ECTA16481

Horowitz, J. L., & Manski, C. F. (1995).Identification and robustness with contaminated and corrupted data. *Econometrica*, *63*(2), 281–302. https://doi.org/10.2307/2951627

Imai, K. (2008).Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". *Statistics & Probability Letters*, *78*(2), 144–149. https://doi.org/10.1016/j.spl.2007.05.015

Imbens, G. W. (2010).Better late than nothing: Some comments on deaton (2009) and heckman and urzua (2009). *Journal of Economic Literature*, *48*(2), 399–423. https://doi.org/10.1257/jel.48.2.399

References IV

Imbens, G. W., & Manski, C. F. (2004).Confidence intervals for partially identified parameters. *Econometrica*, *72*(6), 1845–1857. https://doi.org/10.1111/j.1468-0262.2004.00555.x

Lee, D. S. (2009).Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, *76*(3), 1071–1102. https://doi.org/10.1111/j.1467-937X.2009.00536.x

Long, D. M., & Hudgens, M. G. (2013).Sharpening bounds on principal effects with covariates. *Biometrics*, *69*(4), 812–819. https://doi.org/10.1111/biom.12103

Manski, C. F. (1989).Anatomy of the selection problem. *The Journal of Human Resources*, *24*(3), 343–360. https://doi.org/10.2307/145818

Mattei, A., Mealli, F., & Pacini, B. (2014).Identification of causal effects in the presence of nonignorable missing outcome values. *Biometrics*, *70*(2), 278–288.

Mealli, F., & Mattei, A. (2012).A refreshing account of principal stratification. *The International Journal of Biostatistics*, *8*(1). https://doi.org/10.1515/1557-4679.1380

References V

Rosenbaum, P. R. (1984).The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Royal Statistical Society. Journal. Series A: General*, *147*(5), 656–666. https://doi.org/10.2307/2981697

Samii, C., Wang, Y., & Zhou, J. A. (2025).Generalizing trimming bounds for endogenously missing outcome data using random forests. *Political Analysis*, 1–15. https://doi.org/10.1017/pan.2025.10001

Semenova, V. (2025).Generalized lee bounds. *Journal of Econometrics*, *251*, 106055. https://doi.org/10.1016/j.jeconom.2025.106055

Yang, F., & Small, D. S. (2016).Using post-outcome measurement information in censoring-by-death problems. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *78*(1), 299–318.

Zhang, J. L., & Rubin, D. B. (2003).Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, *28*(4), 353–368. https://www.jstor.org.eui.idm.oclc.org/stable/3701340

References VI

Zhang, J. L., Rubin, D. B., & Mealli, F. (2009).Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, *104*(485), 166–176.