Difference-in-Differences using Mixed-Integer Programming Matching Approach

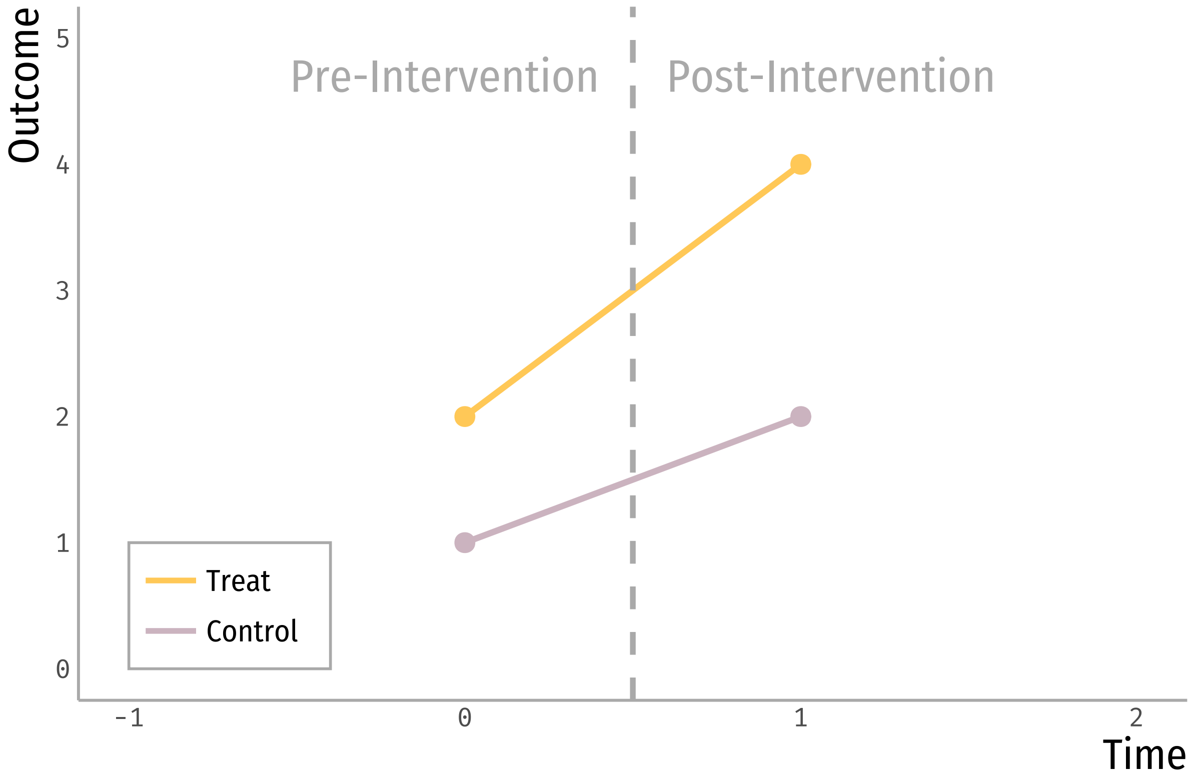
Magdalena Bennett

McCombs School of Business, The University of Texas at Austin

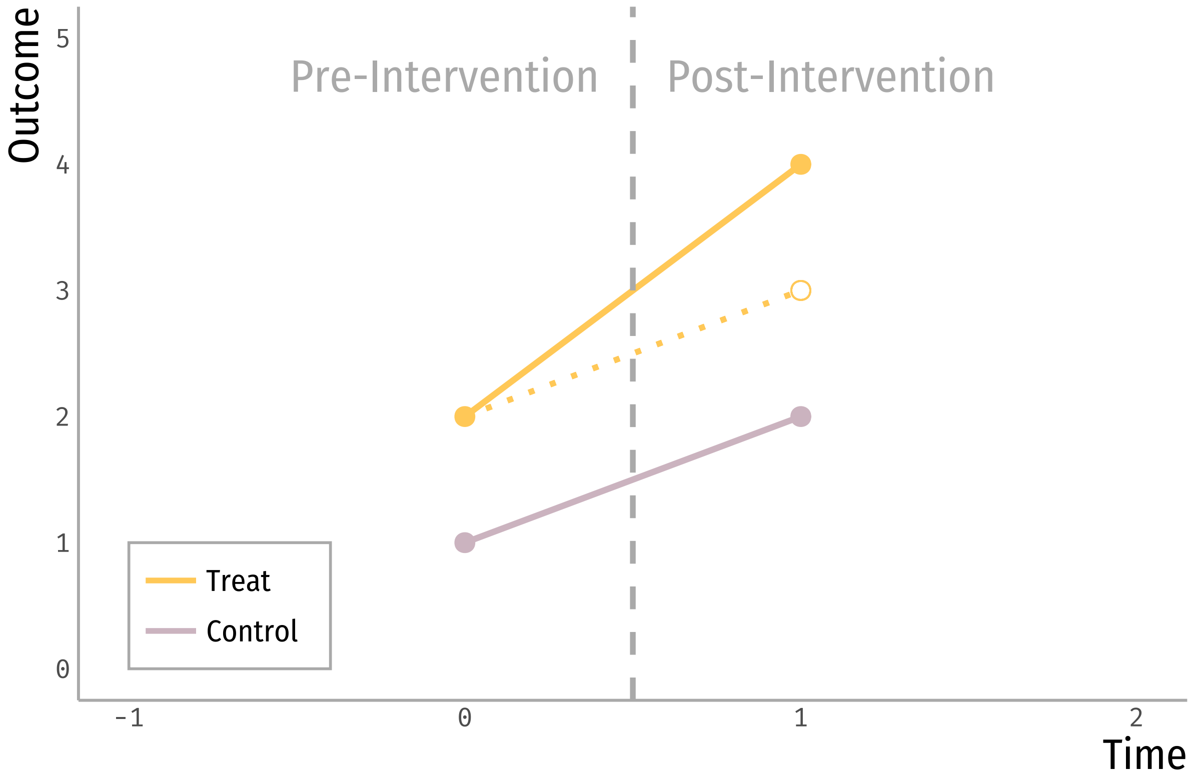
AEFP 50th Conference, Washington DC

March 13th, 2025

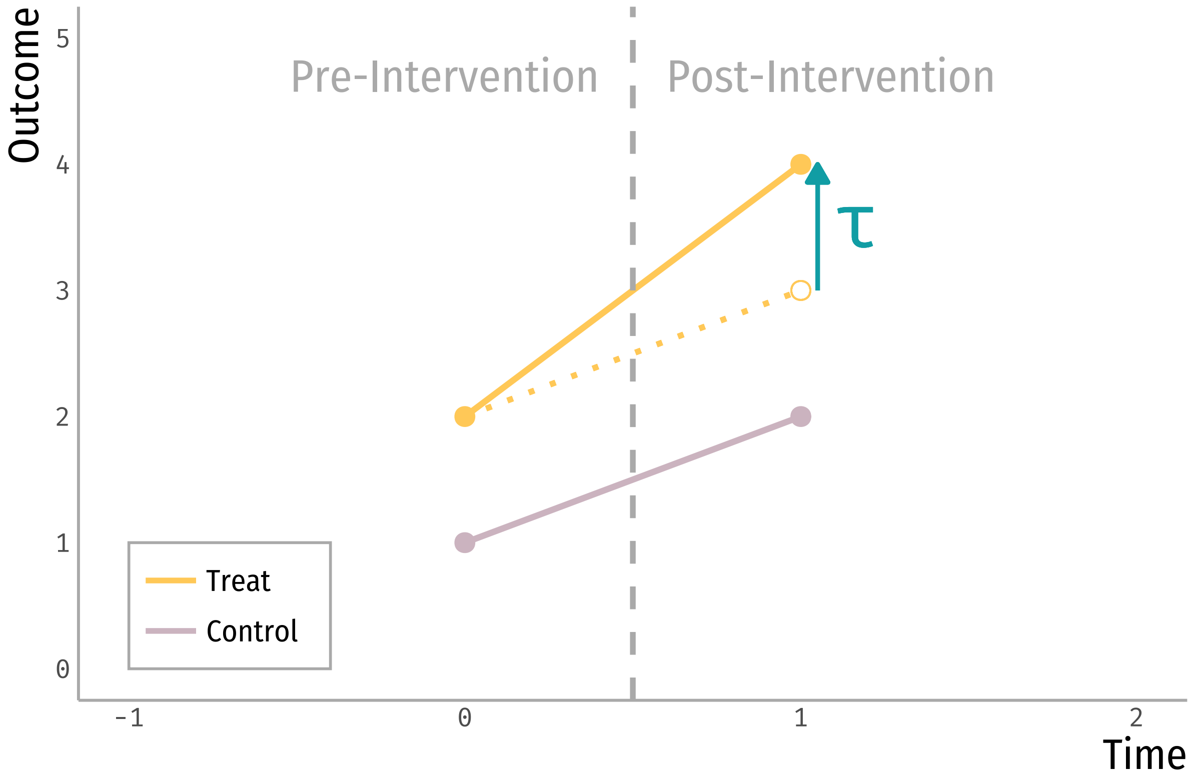
Diff-in-Diff as an identification strategy



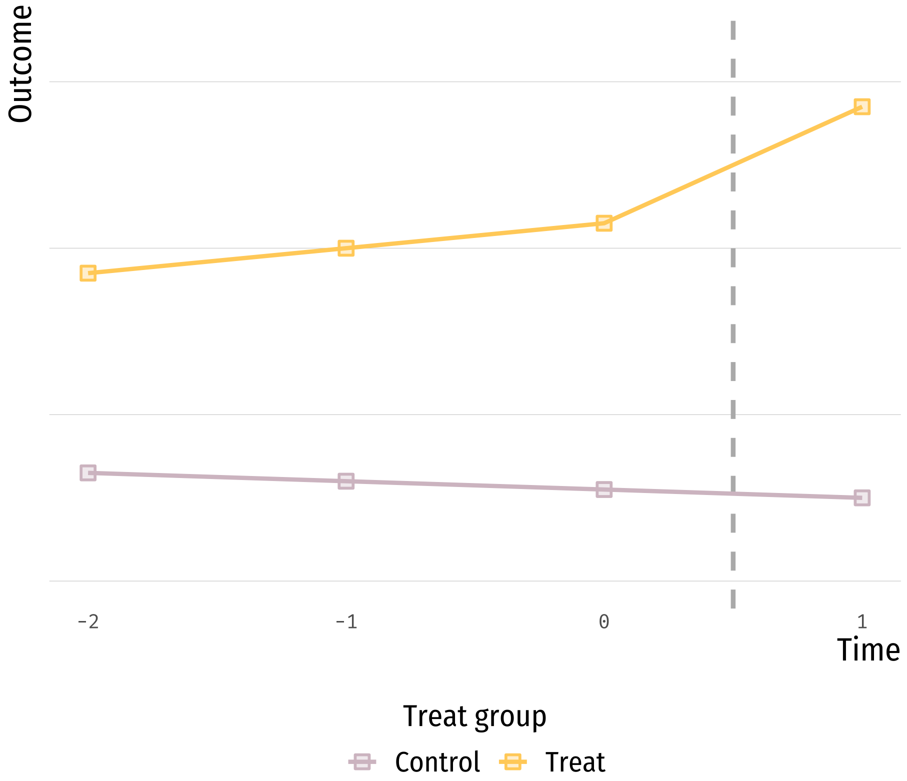
Parallel trend assumption (PTA)



Estimate Average Treatment Effect on the Treated (ATT)

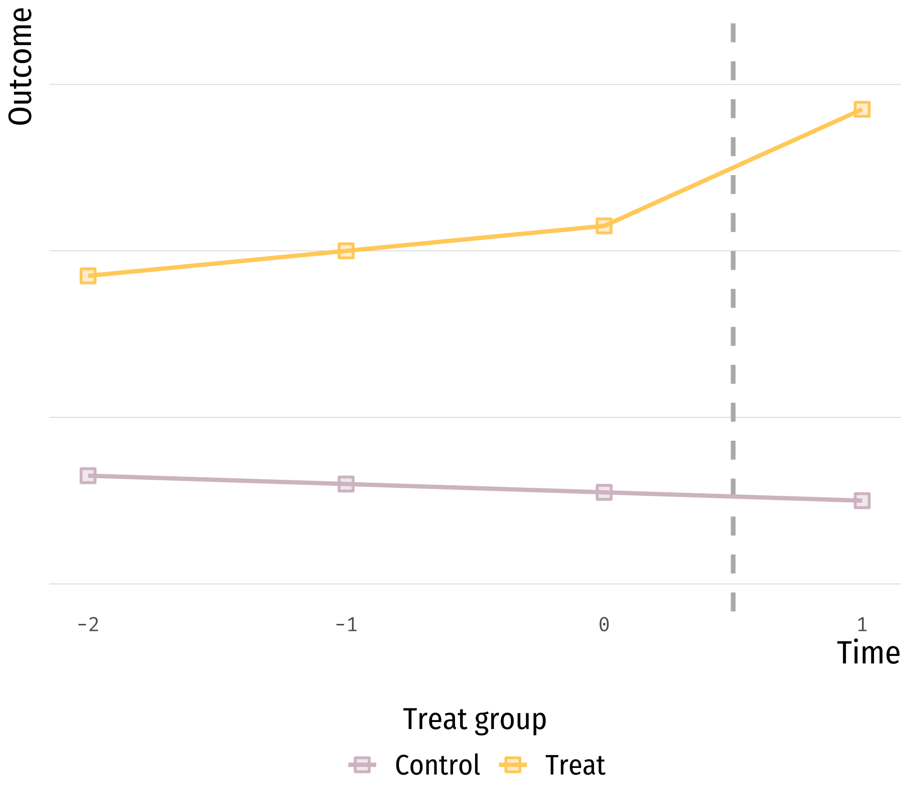


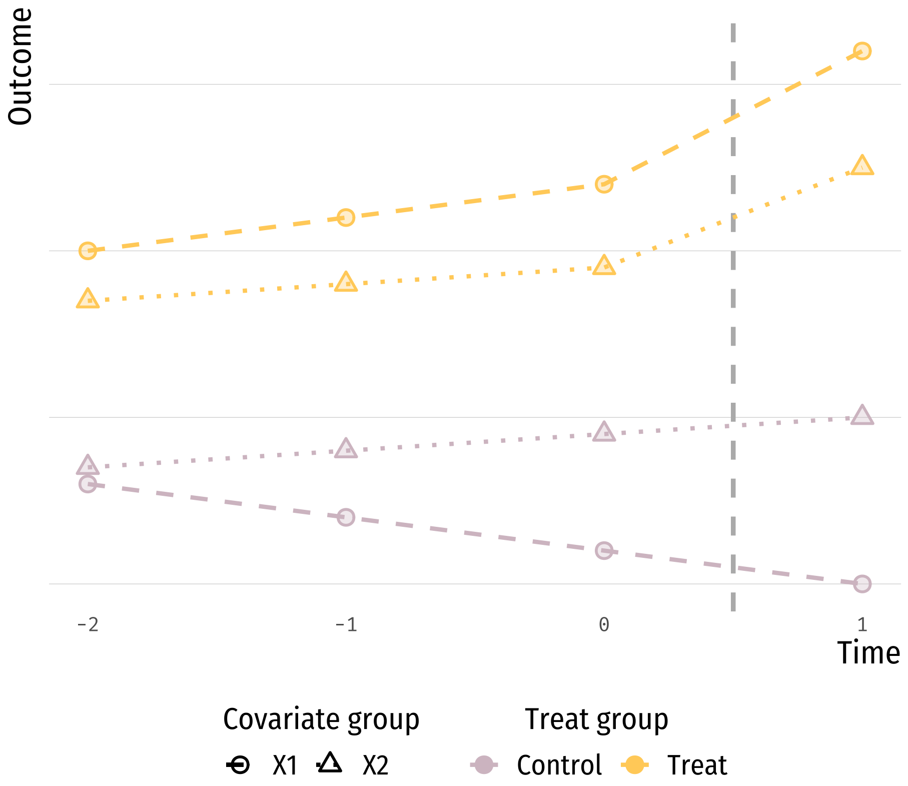
But what if the PTA doesn't hold?



But what if the PTA doesn't hold?

We can potentially remove [part of] the bias by matching on $X^s{}_{it}=X_i$





This paper

- Identify contexts when matching can recover causal estimates under **certain violations of the parallel trend assumption**.
  - Overall bias reduction and increase in robustness for sensitivity analysis.
- Use **mixed-integer programming matching (MIP)** to balance covariates directly.

Simulations:
Different DGP scenarios

Application:
School segregation & vouchers

Let's set up the problem

DD Setup

- Let $Y_{it}(z)$ be the potential outcome for unit i in period t under treatment z.
- Intervention implemented in $T_0 \rightarrow$ No units are treated in $t \leq T_0$
- Difference-in-Differences (DD) focuses on ATT for $t > T_0$:

$$ATT(t) = E[Y_{it}(1) - Y_{it}(0)|Z = 1]$$

DD Setup

- Let $Y_{it}(z)$ be the potential outcome for unit i in period t under treatment z.
- Intervention implemented in $T_0 \to$ No units are treated in $t \leq T_0$
- Difference-in-Differences (DD) focuses on ATT for $t > T_0$:

$$ATT(t) = E[Y_{it}(1) - Y_{it}(0)|Z = 1]$$

- Under the PTA:

$$\hat{\tau}^{DD} = \overbrace{E[Y_{i1}|Z = 1] - E[Y_{i1}|Z = 0]}^{\Delta_{post}} - \underbrace{(E[Y_{i0}|Z = 1] - E[Y_{i0}|Z = 0])}_{\Delta_{pre}}$$

Bias in a DD setting

Bias can be introduced to DD in different ways:

1) **Time-invariant covariates with time-varying effects**: *Obs. Bias*

- e.g. Effect of gender on salaries.

2) **Differential time-varying effects**: *Obs. Diff. Bias*

- e.g. Effect of race on salaries evolve differently over time by group.

3) **Observed or unobserved time-varying covariates**: *Unobs. Bias*

- e.g. Test scores

If the PTA holds...

$$\overbrace{(\bar{\gamma}_1(X^1,t') - \bar{\gamma}_1(X^0,t')) - (\bar{\gamma}_1(X^1,t) - \bar{\gamma}_1(X^0,t))}^{Obs.Bias} + \underbrace{(\bar{\gamma}_2(X^1,t') - \bar{\gamma}_2(X^1,t))}_{Obs.Diff.Bias} + \underbrace{(\lambda_{t'1} - \lambda_{t'0}) - (\lambda_{t1} - \lambda_{t0})}_{Unobs.Bias} = 0$$

One of the two conditions need to hold:

1) No effect or constant effect of X on Y over time: $\mathbb{E}[\gamma_1(X,t)] = \mathbb{E}[\gamma_1(X)]$

2) Equal distribution of observed covariates between groups: $X_i|Z=1 \overset{d}{=} X_i|Z=0$

in addition to:

3) No differential time effect of X on Y by treatment group: $\mathbb{E}[\gamma_2(X,t)] = 0$

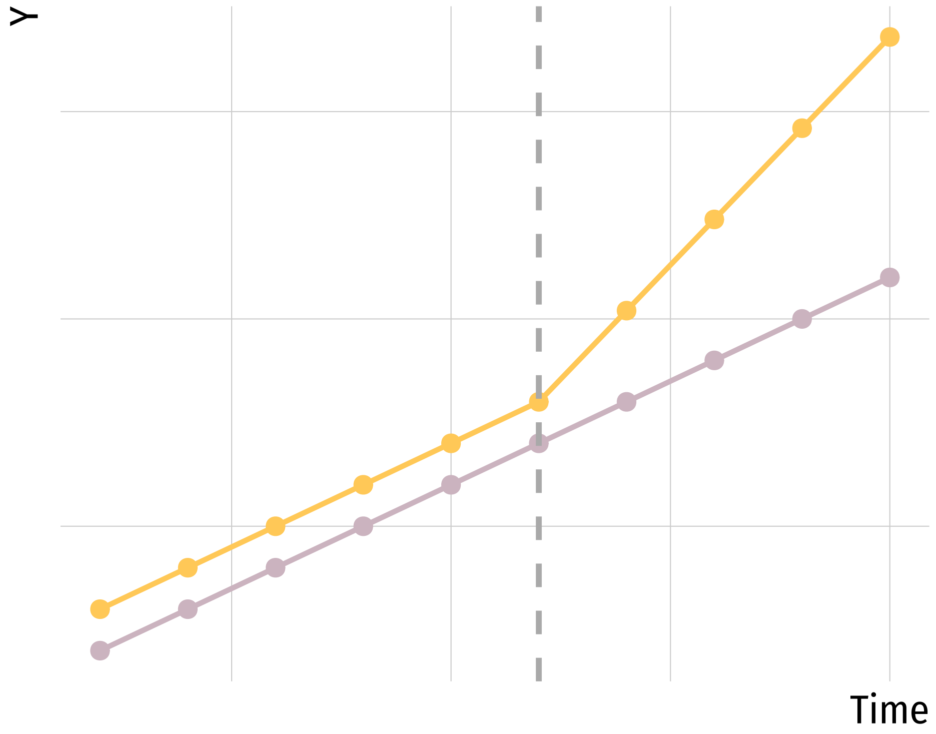
4) No unobserved time-varying effects: $\lambda_{t1} = \lambda_{t0}$

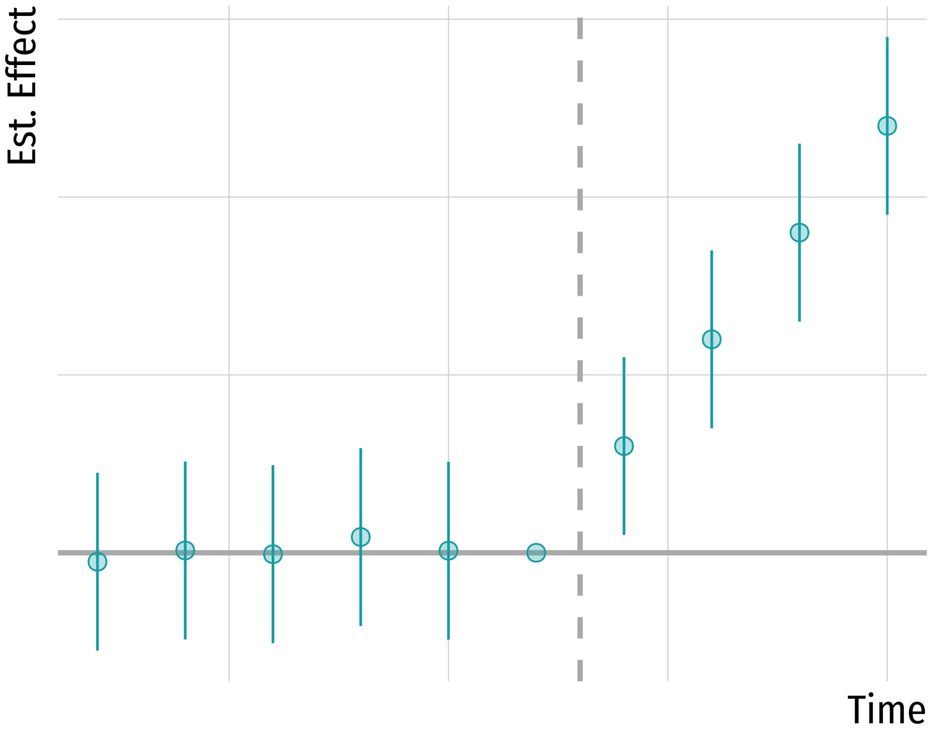
Cond. 2 can hold through **matching**

Cond. 3 and 4 can be tested with **sensitivity analysis**

Sensitivity analysis for Diff-in-Diff

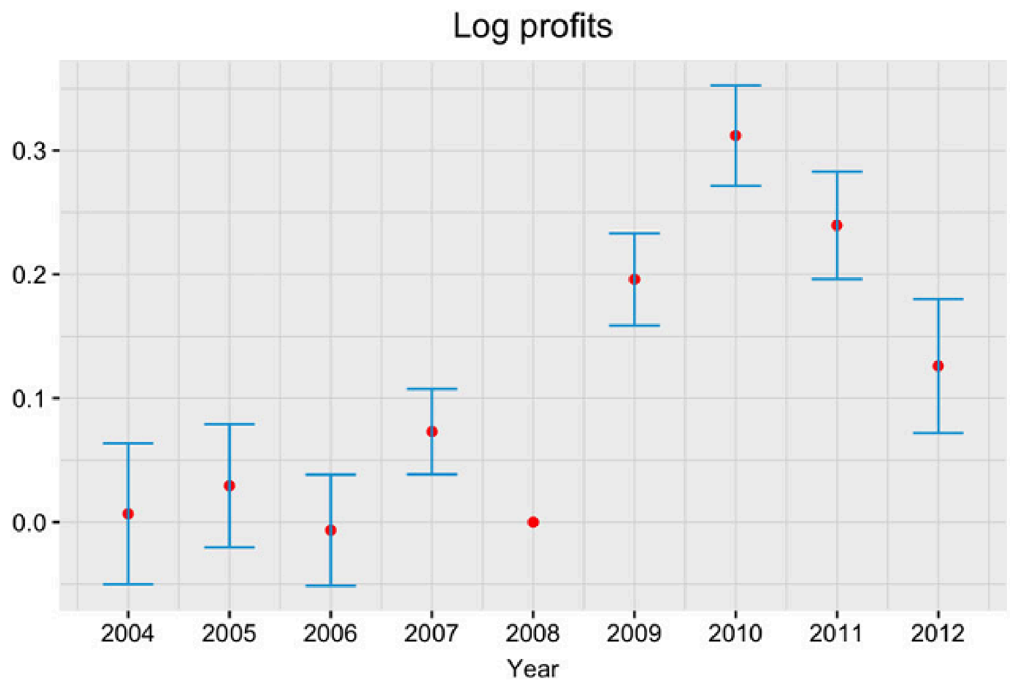
- In an event study $\rightarrow$ null effects prior to the intervention:

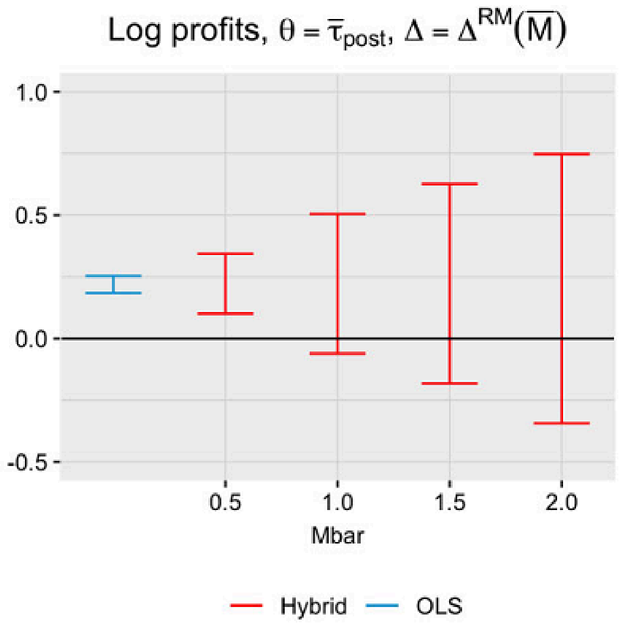




Honest approach to test pretrends

- One main issue with the previous test $\rightarrow$ **Underpowered**
- Rambachan & Roth (2023) propose **sensitivity bounds** to allow pre-trends violations:
  - E.g. Violations in the post-intervention period can be *at most* M times the max violation in the pre-intervention period.





Simulations

Different scenarios

For linear and quadratic functions:

S1: No interaction between X and t

S2: Equal interaction between X and t

S3: Differential interaction between X and t

S4: S3 + Bias cancellation

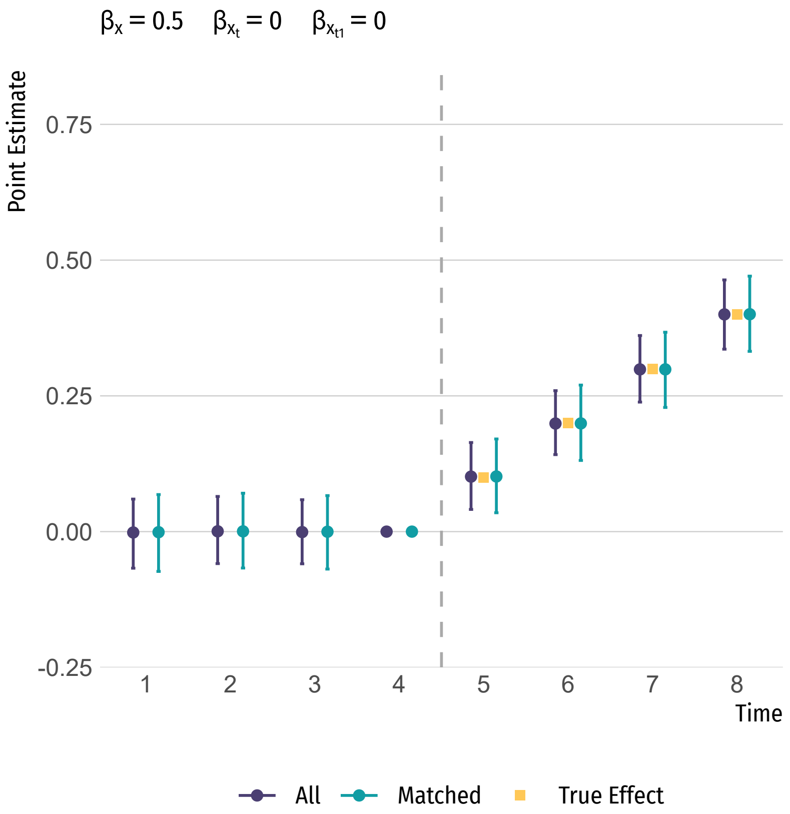
- For all scenarios, differential distribution of covariates X between groups

Parameters:

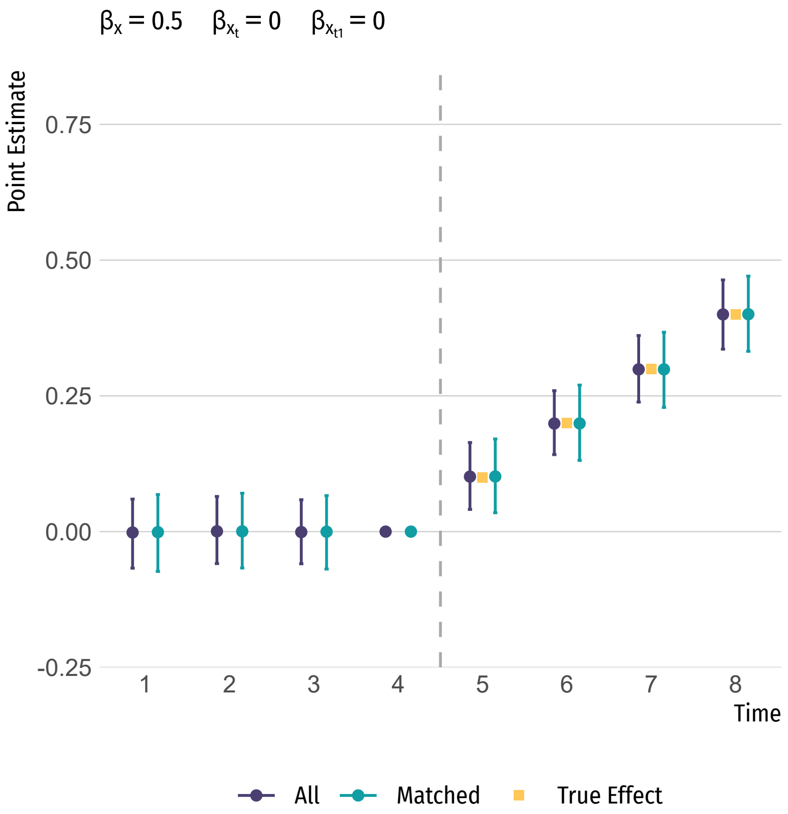
Parameter	Value
Number of obs (N)	1,000
`Pr(Z=1)`	0.5
Time periods (T)	8
Last pre-intervention period (T_0)	4
Matching PS	Nearest neighbor (using calipers)
MIP Matching tolerance	.01 SD
Number of simulations	1,000

- Estimate compared to sample ATT (*can be different for matching*)

S1 - No interaction between X and t

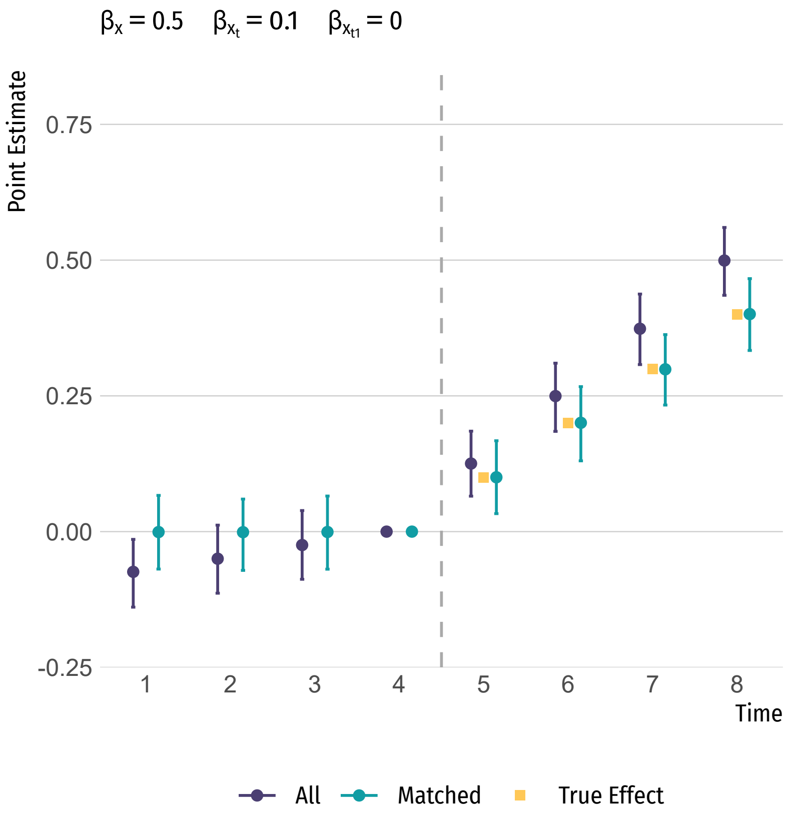


(a) Linear

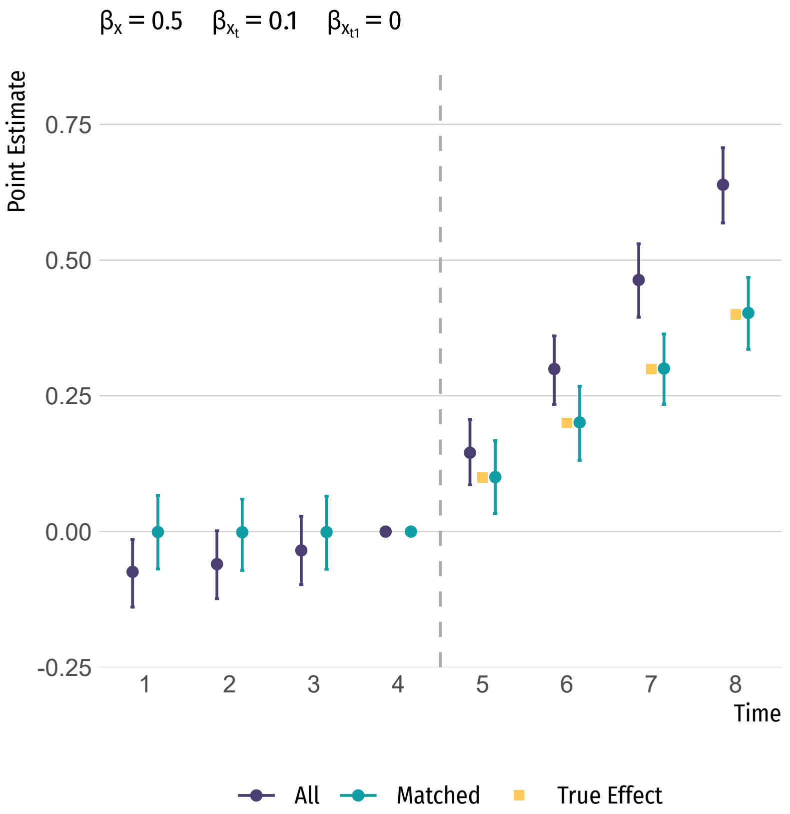


(b) Quadratic

S2 - Equal interaction between X and t by treatment

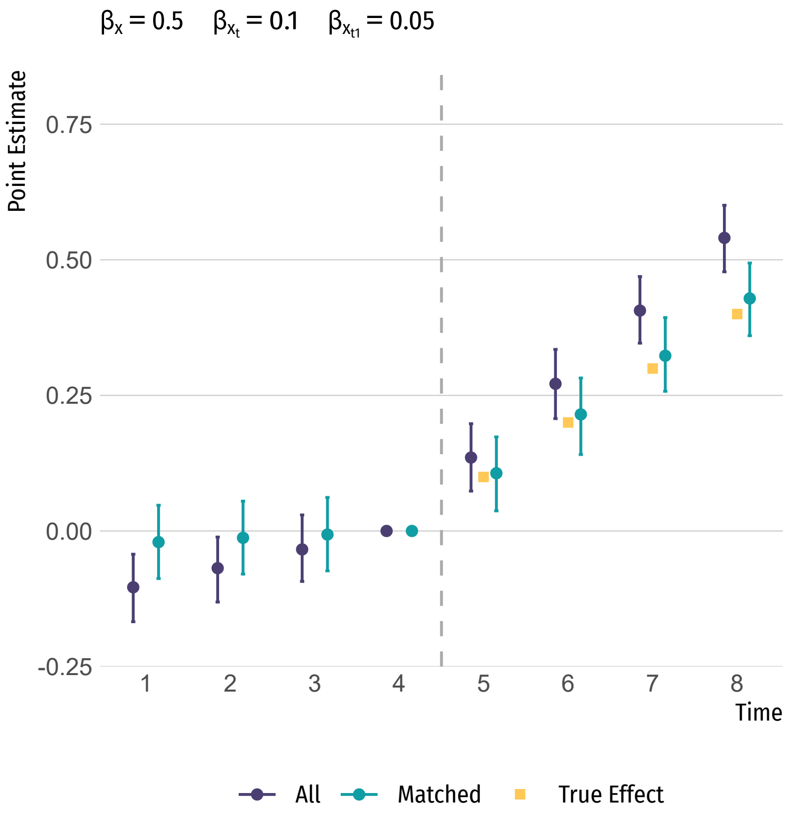


(a) Linear

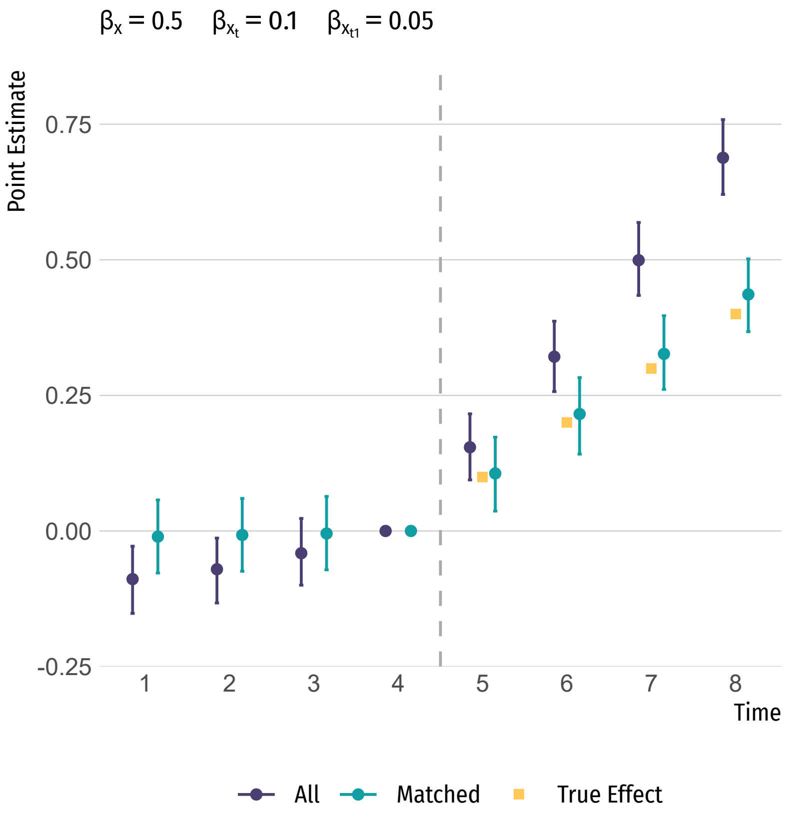


(b) Quadratic

S3 - Differential interaction between X and t by treatment



(a) Linear

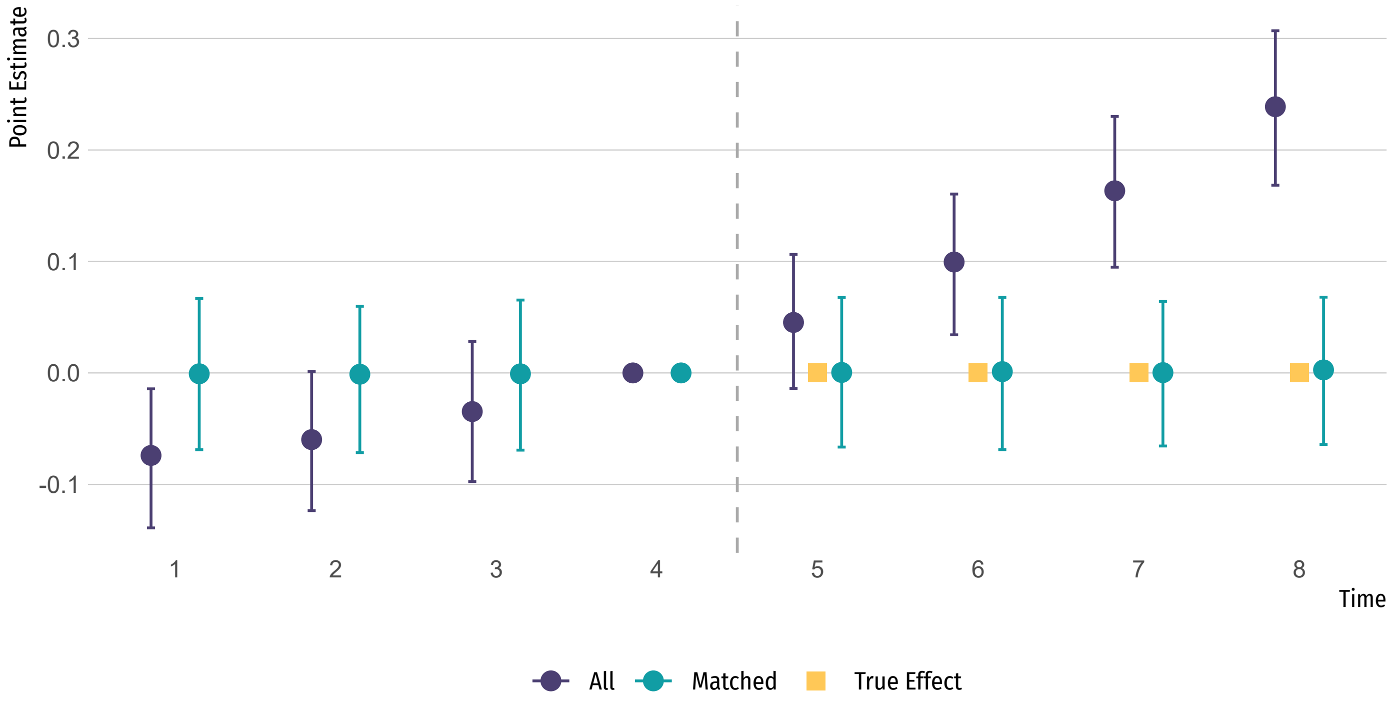


(b) Quadratic

Why is this bias reduction important?

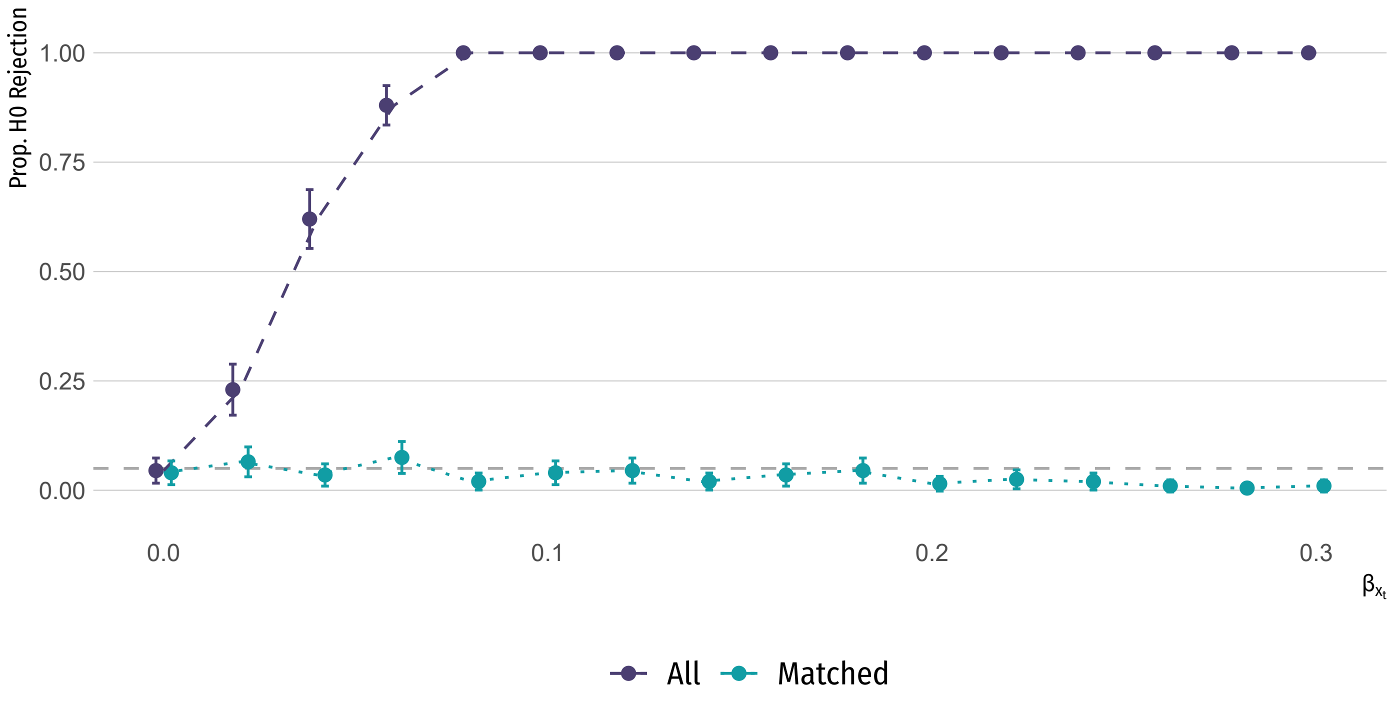
- Example of S2 (Quadratic) with no true effect:

$\beta_X = 0.5 \quad \beta_{X_t} = 0.1 \quad \beta_{X_{t1}} = 0$



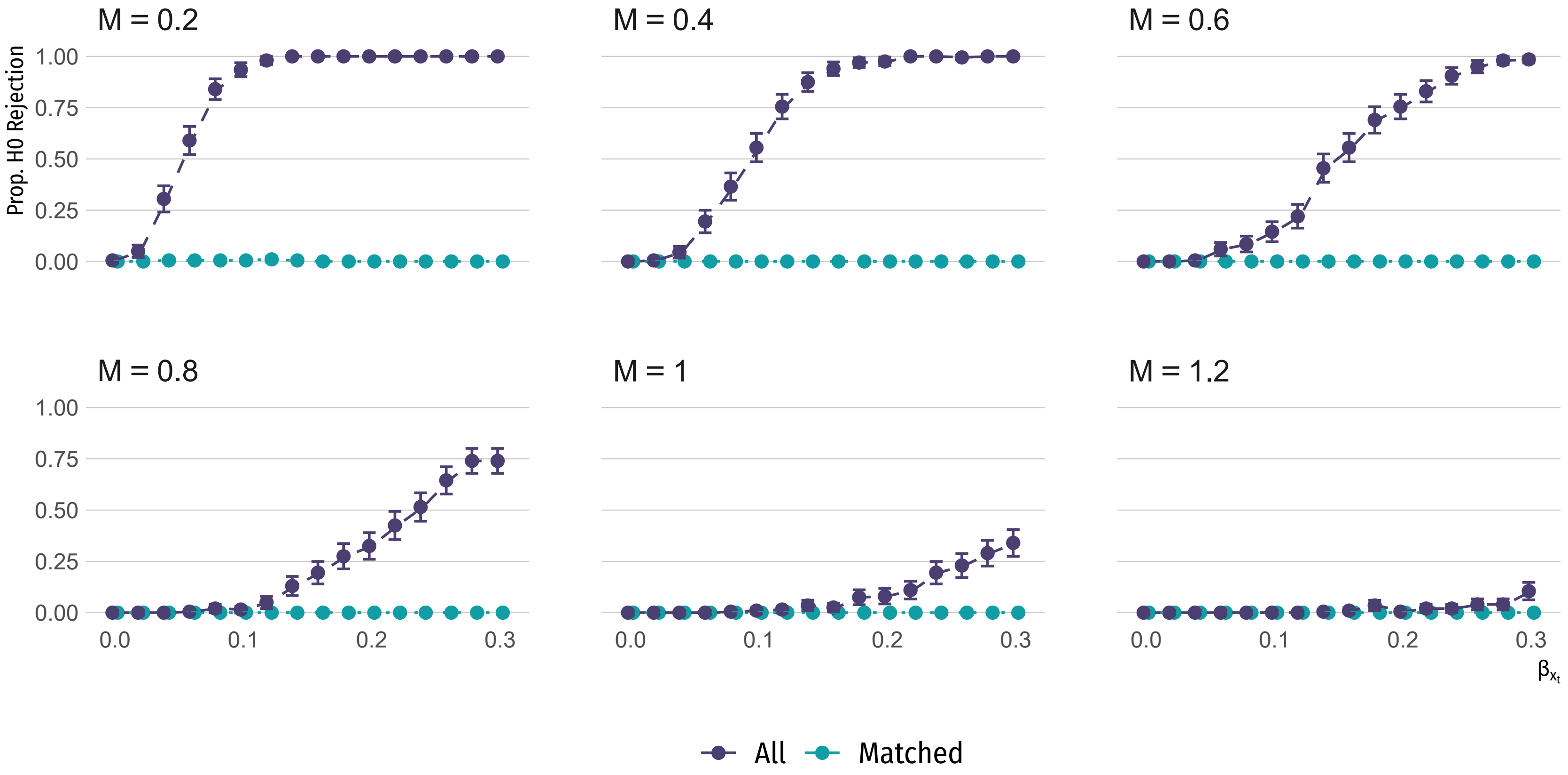
Why is this bias reduction important?

- Even under modest bias, we would incorrectly reject the null 20% of the time.

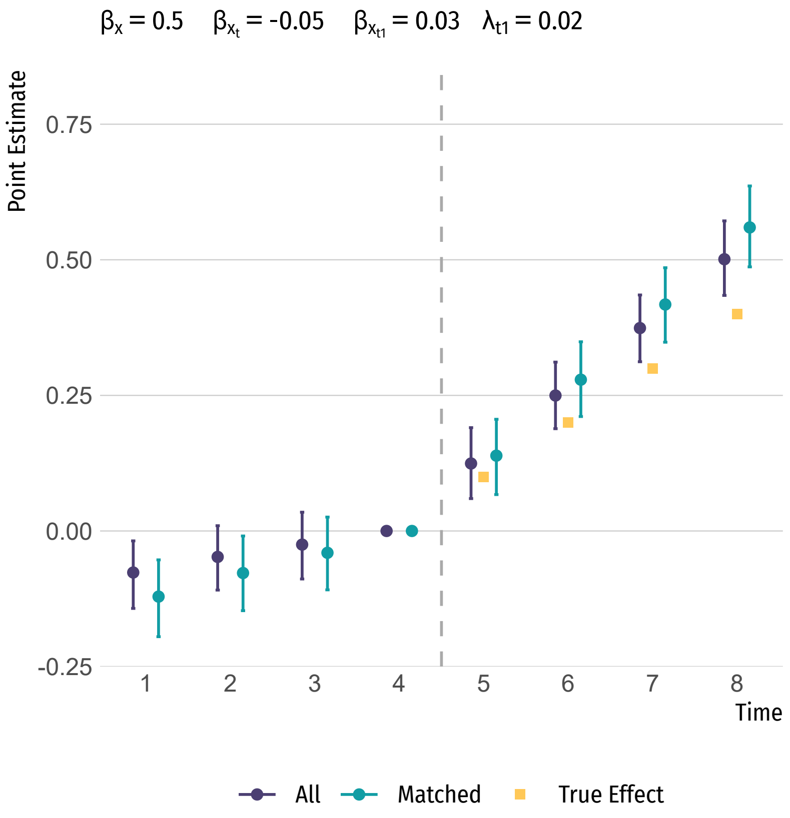


Why is this bias reduction important?

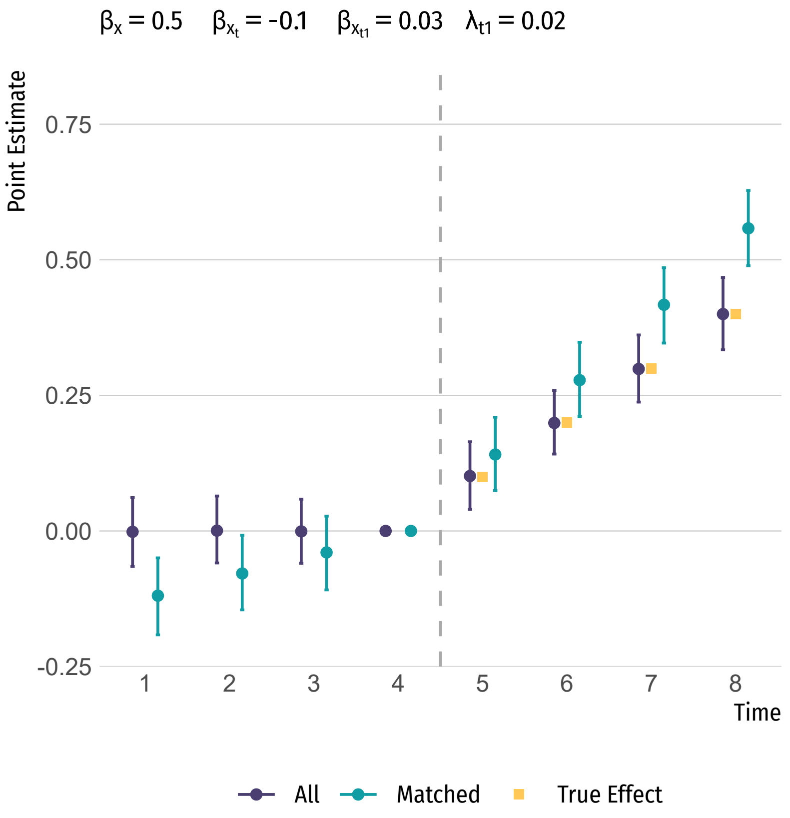
- Sensitivity analysis results are skewed by the magnitude of the bias.



S4: Bias cancellation



(a) Larger bias in matched DD vs Smaller bias in unmatched DD



(a) Larger bias in matched DD vs No bias in unmatched DD

Application

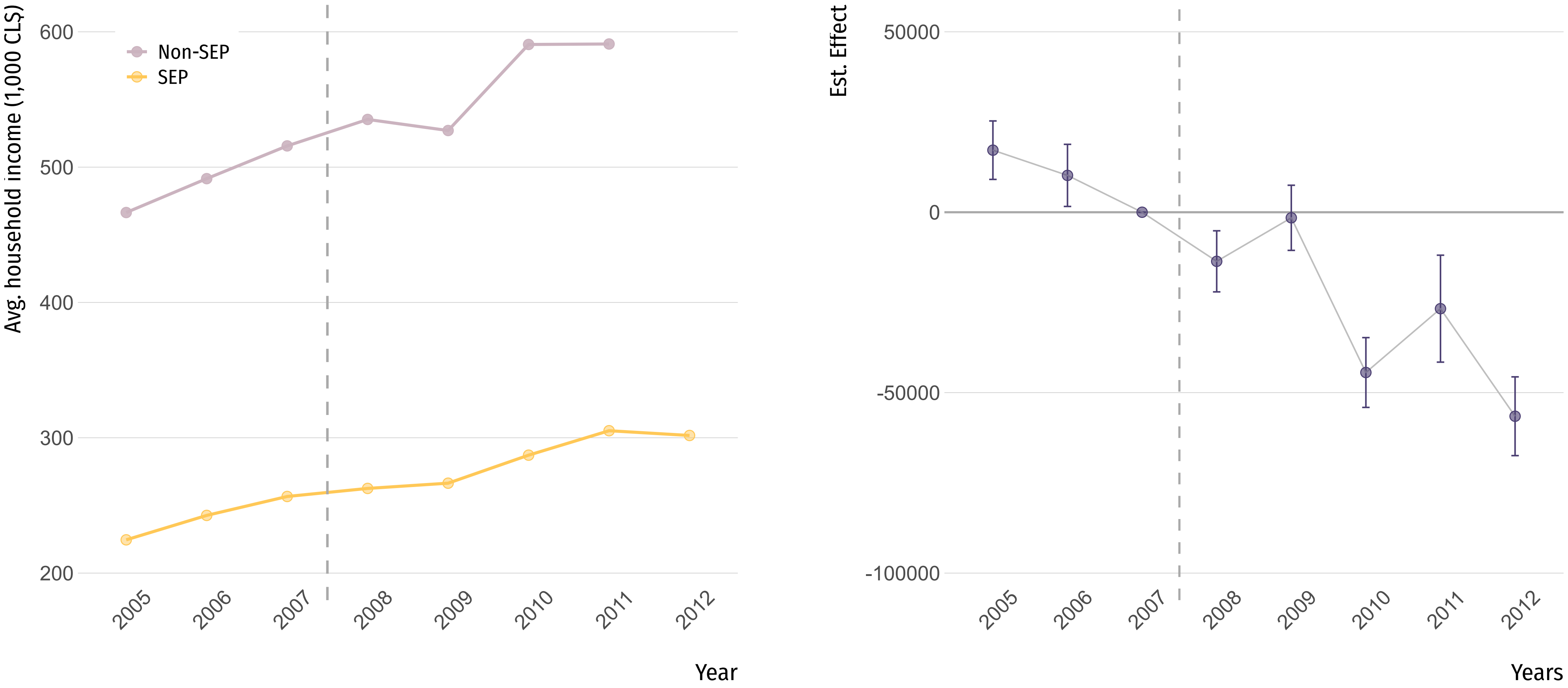
Preferential Voucher Scheme in Chile

- Universal flat voucher scheme $\xrightarrow{\mathbf{2008}}$ Universal + preferential voucher scheme
- Preferential voucher scheme:
  - Targeted to bottom 40% of vulnerable students
  - Additional 50% of voucher per student
  - Additional money for concentration of SEP students.

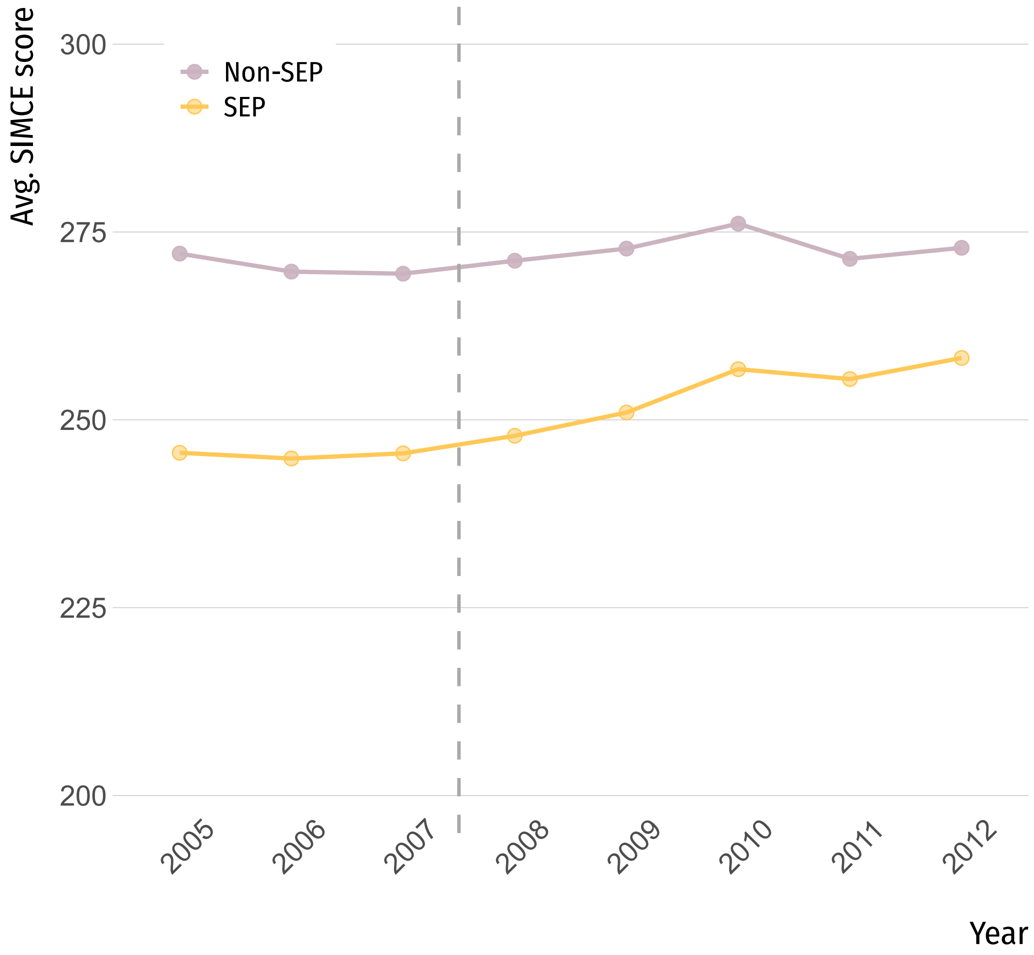
Students:
- Verify SEP status
- Attend a SEP school

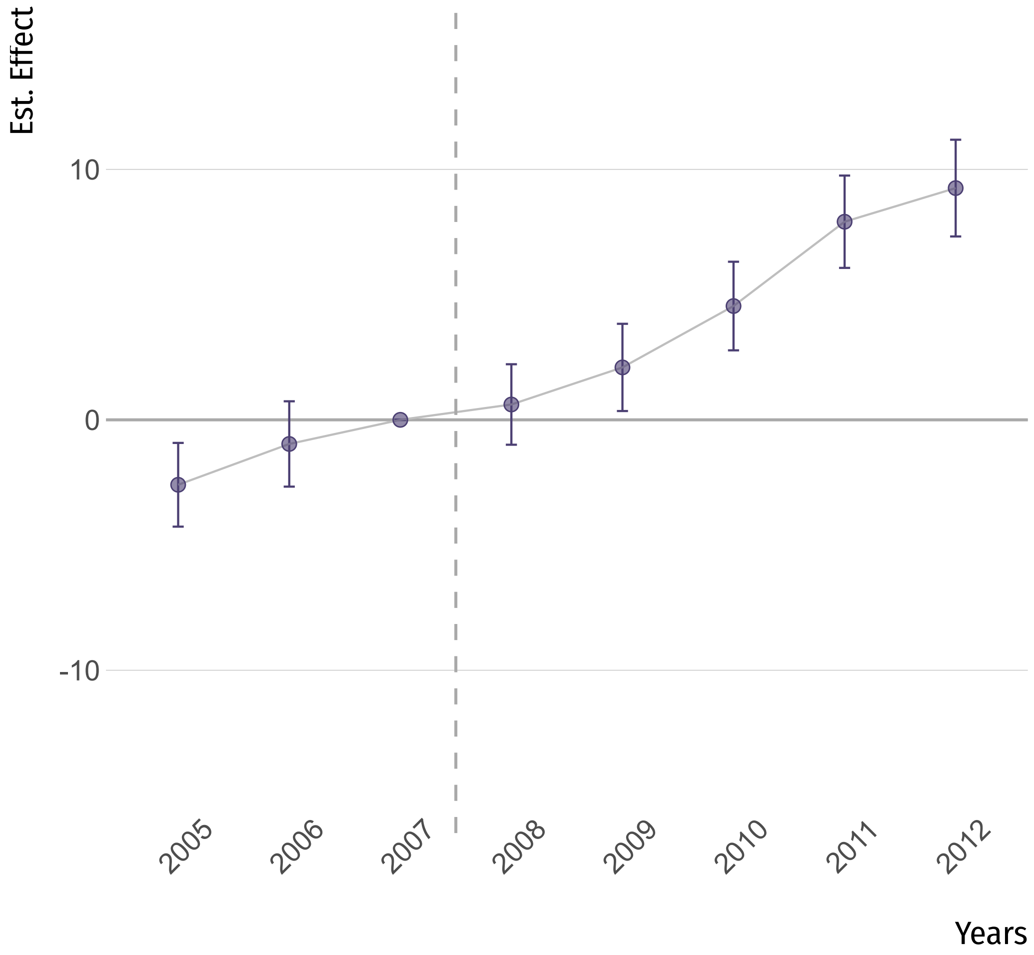
Schools:
- Opt-into the policy
- No selection, no fees
- Resources ~ performance

Before matching: Household income



Before matching: Average SIMCE

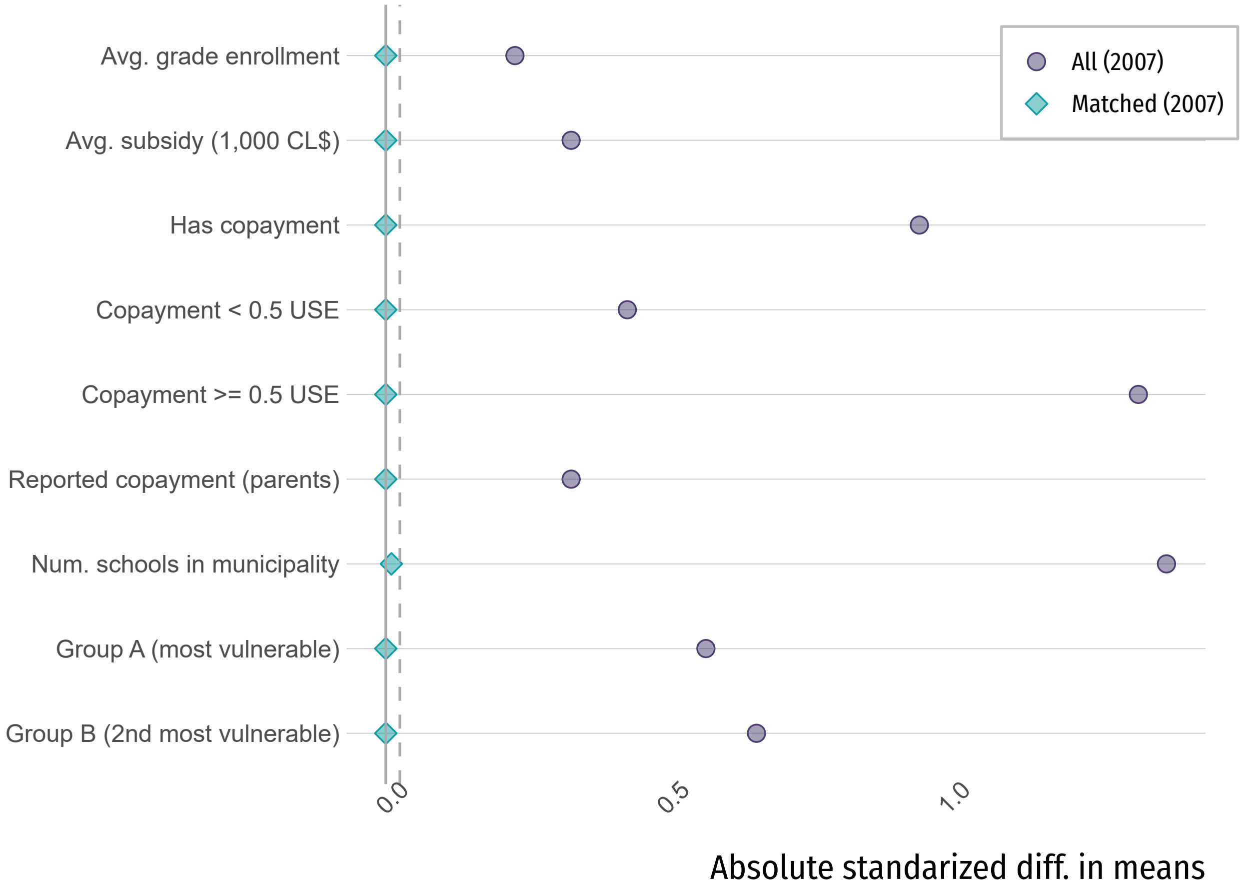




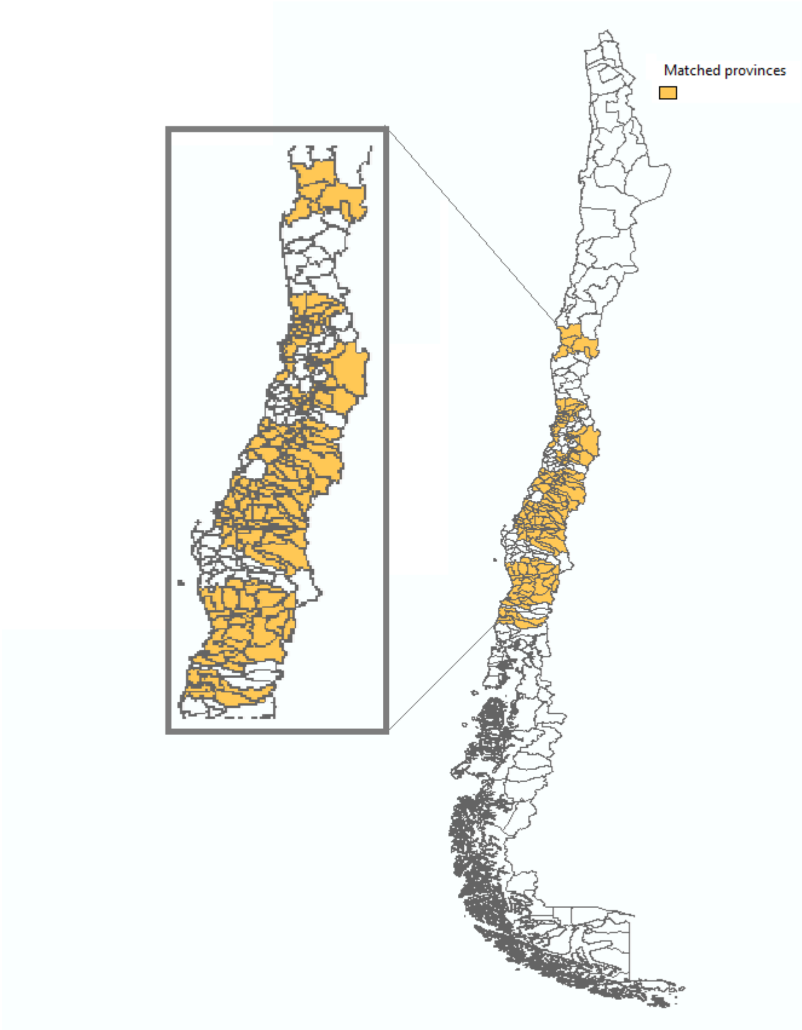
Matching + DD

- Prior to matching: No parallel pre-trend
- Different types of schools:
  - Schools that charge high co-payment fees.
  - Schools with low number of SEP student enrolled.
- MIP Matching using constant or "sticky" covariates:
  - Mean balance (0.025 SD): Enrollment, average yearly subsidy, number of voucher schools in county, charges add-on fees
  - Exact balance: Geographic province

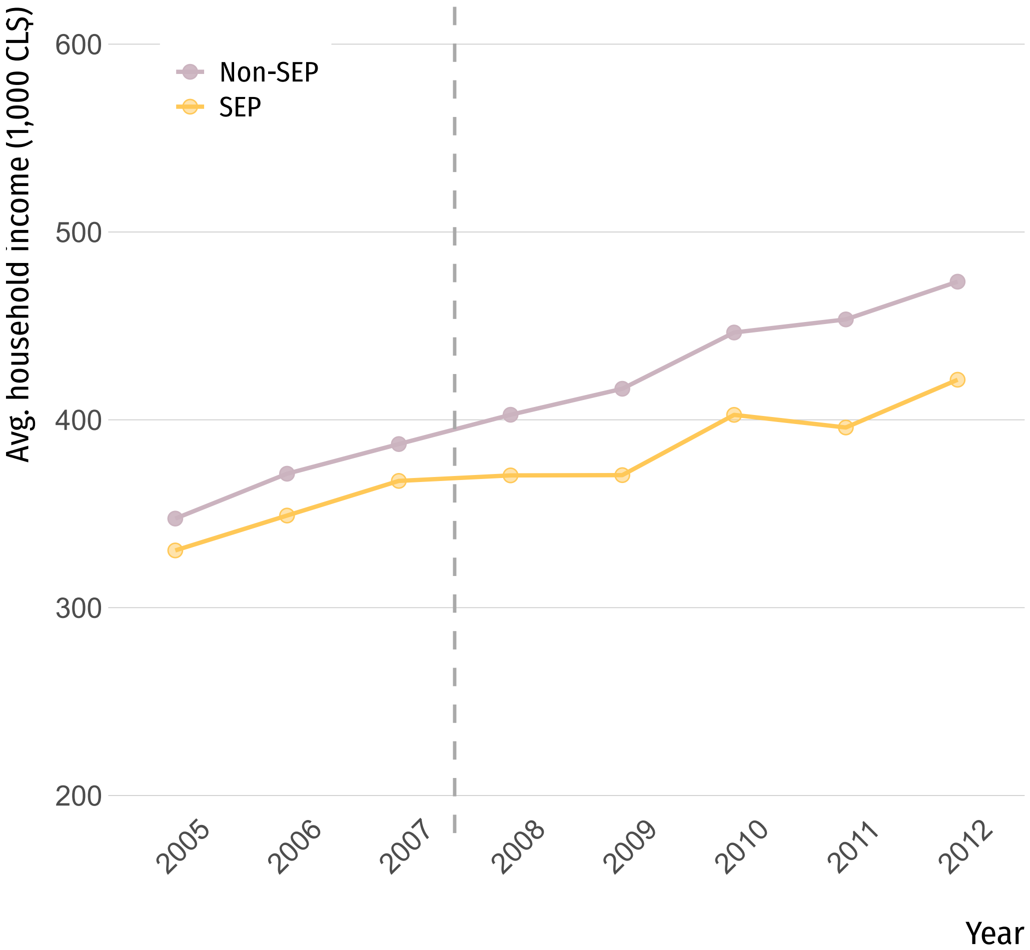
Groups are balanced in specific characteristics

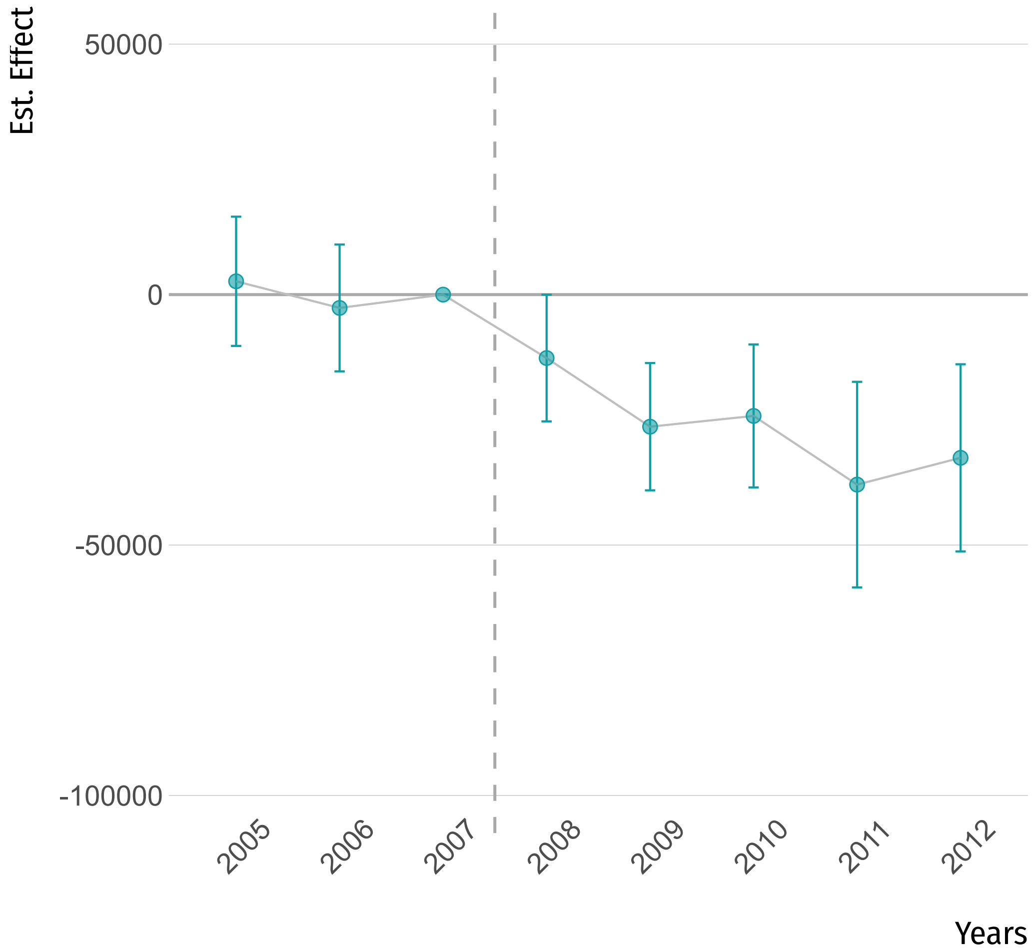


Matching in 16 out of 53 provinces

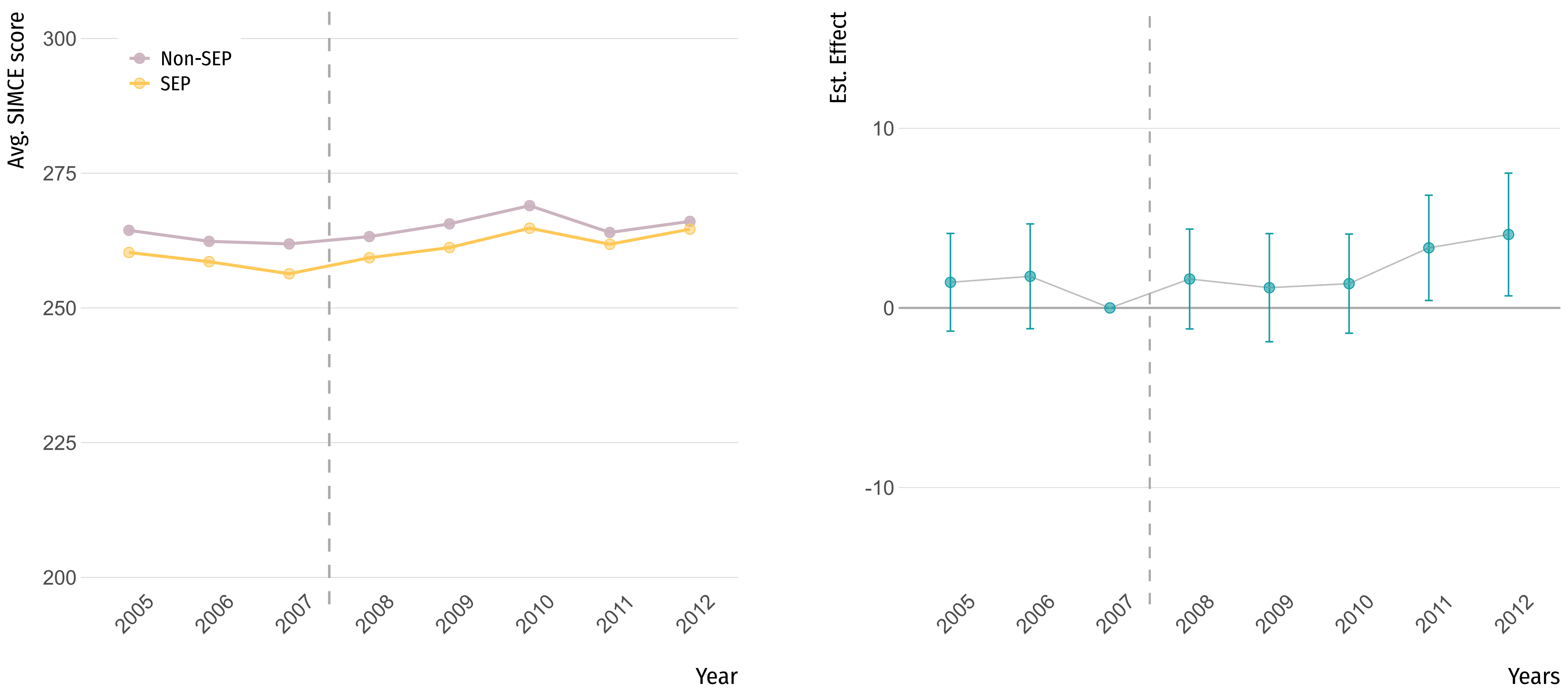


After matching: Household income





After matching: Average SIMCE

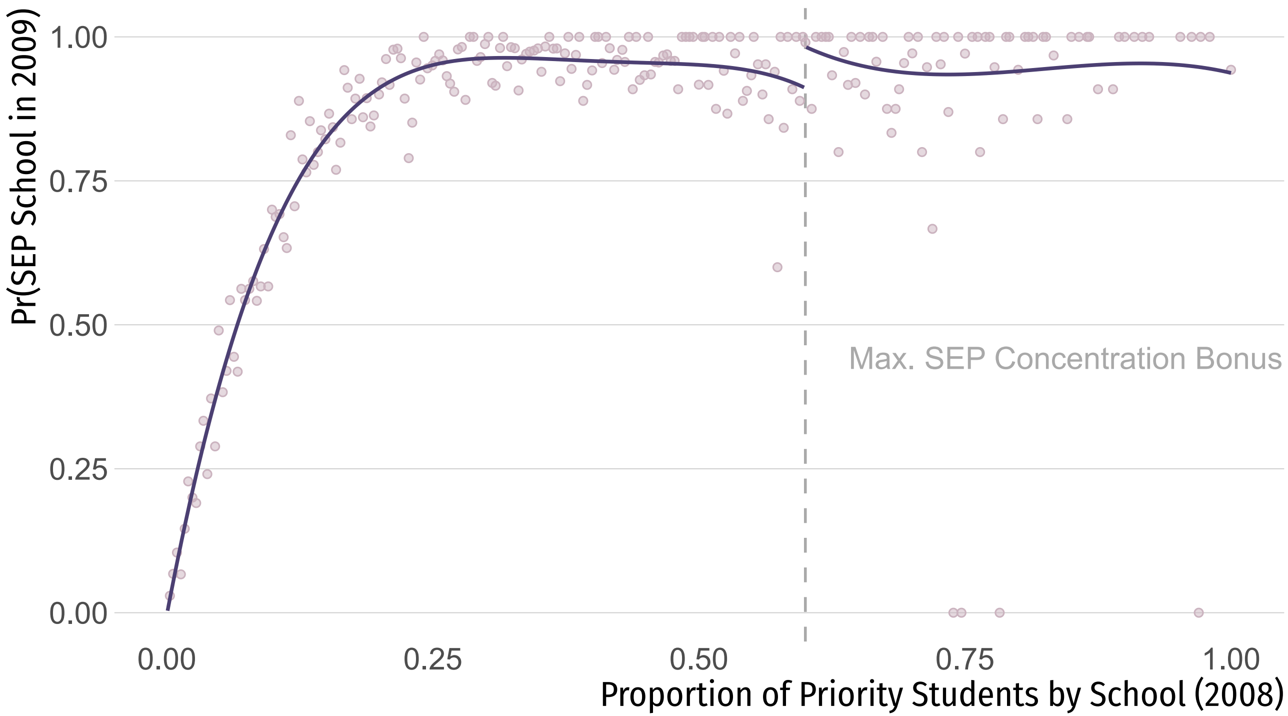


Results

- Matched schools:
  - More vulnerable and lower test scores than the population mean.
- 9pp increase in the income gap between SEP and non-SEP schools in matched DD:
  - SEP schools attracted even more vulnerable students.
  - Non-SEP schools increased their average family income.
- No evidence of increase in SIMCE score:
  - Could be a longer-term outcome.
- Findings in segregation are moderately robust to hidden bias (Keele et al., 2019):
  - $\Gamma_c = 1.76 \rightarrow$ Unobserved confounder would have to change the probability of assignment from 50% vs 50% to 32.7% vs 67.3%.
  - Allows up to 70% of the maximum deviation in the pre-intervention period (*M = 0.7*) vs 50% without matching (Rambachan & Roth, 2023)

Potential reasons?

- Increase in probability of becoming SEP in 2009 jumps discontinuously at 60% of SEP student concentration in 2008 (4.7 pp; SE = 0.024)



Let's wrap it up

Conclusions and Next Steps

- Matching can be an important tool to address violations in PTA.
- Bias reduction is very important for sensitivity analysis.
- Serial correlation also plays an important role: Don't match on random noise.
- Next steps: Partial identification using time-varying covariates

Difference-in-Differences using Mixed-Integer Programming Matching Approach

Magdalena Bennett

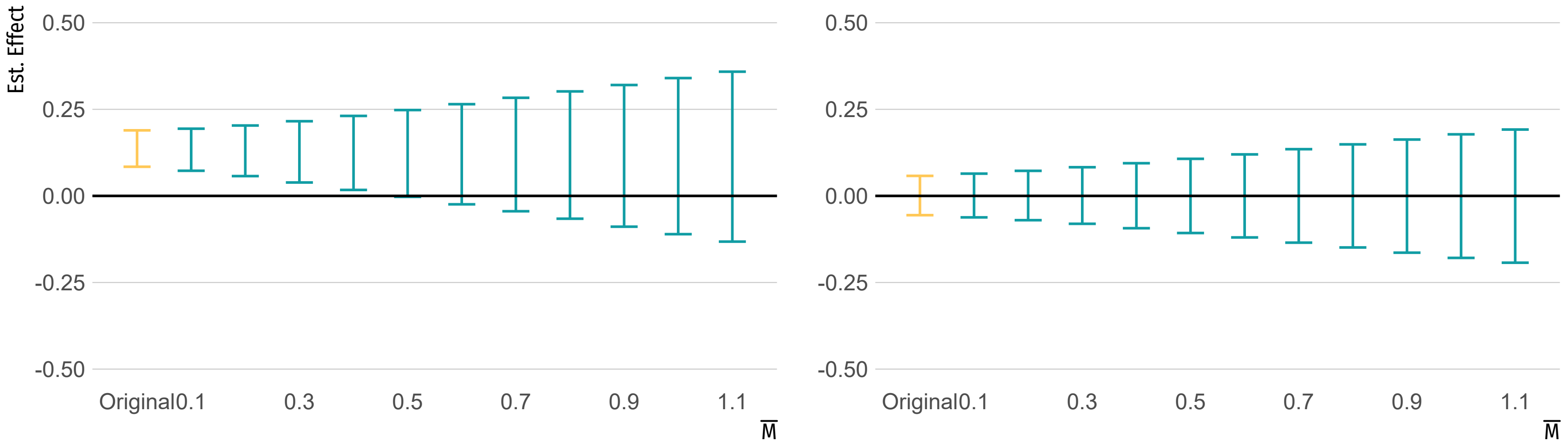
McCombs School of Business, The University of Texas at Austin

AEFP 50th Conference, Washington DC

March 13th, 2025

Honest approach to test pretrends

- One drawback of the previous method is that it can **overstate** (or understate) the robustness of findings if the point estimate is biased.
  - Honest CIs depend on the **magnitude of the point estimate** as well as the **pre-trend violations**.
- Matching can **reduce the overall bias** of the point estimate



(a) Biased estimate

(b) Unbiased estimate

How do we match?

- Match on covariates or outcomes? Levels or trends?
- Propensity score matching? Optimal matching? etc.

This paper:

- Match on time-invariant covariates that could make groups behave differently.
    - Use distribution of covariates to match on a template.
- Use of Mixed-Integer Programming (MIP) Matching (Zubizarreta, 2015; Bennett, Zubizarreta, & Vielma, 2020):
    - Balance covariates directly
    - Yield largest matched sample under balancing constraints (cardinality matching)
    - Works fast with large samples

Data Generating Processes

Scenarios	**Functions**		
Linear			
(1) No interaction between X and t	$\gamma_0(X) = \beta_x \cdot X$	$\gamma_1 = \gamma_2 = 0$	
(2) Equal interaction between X and t by treatment	$\gamma_0(X) = \beta_x \cdot X$	$\gamma_1(X,t) = \beta_{x_t} \cdot X \cdot \frac{t}{2}$	$\gamma_2(X,t) = 0$
(3) Different interaction between X and t by treatment	$\gamma_0(X) = \beta_x \cdot X$	$\gamma_1(X,t) = \beta_{x_t} \cdot X \cdot \frac{t}{2}$	$\gamma_2(X,t) = \beta_{x_{t1}} \cdot X \cdot \frac{t}{5} \cdot Z$
Quadratic			
(1) No interaction between X and t	$\gamma_0(X) = \beta_x \cdot X + \beta_x \cdot \frac{X^2}{10}$	$\gamma_1 = \gamma_2 = 0$	
(2) Equal interaction between X and t by treatment	$\gamma_0(X) = \beta_x \cdot X + \beta_x \cdot \frac{X^2}{10}$	$\gamma_1(X,t) = \beta_{x_t} \cdot X \cdot \frac{t^2}{10}$	$\gamma_2(X,t) = 0$
(3) Different interaction between X and t by treatment	$\gamma_0(X) = \beta_x \cdot X + \beta_x \cdot \frac{X^2}{10}$	$\gamma_1(X,t) = \beta_{x_t} \cdot X \cdot \frac{t^2}{10}$	$\gamma_2(X,t) = \beta_{x_{t1}} \cdot X \cdot \frac{t^2}{50} \cdot Z$

SEP adoption over time