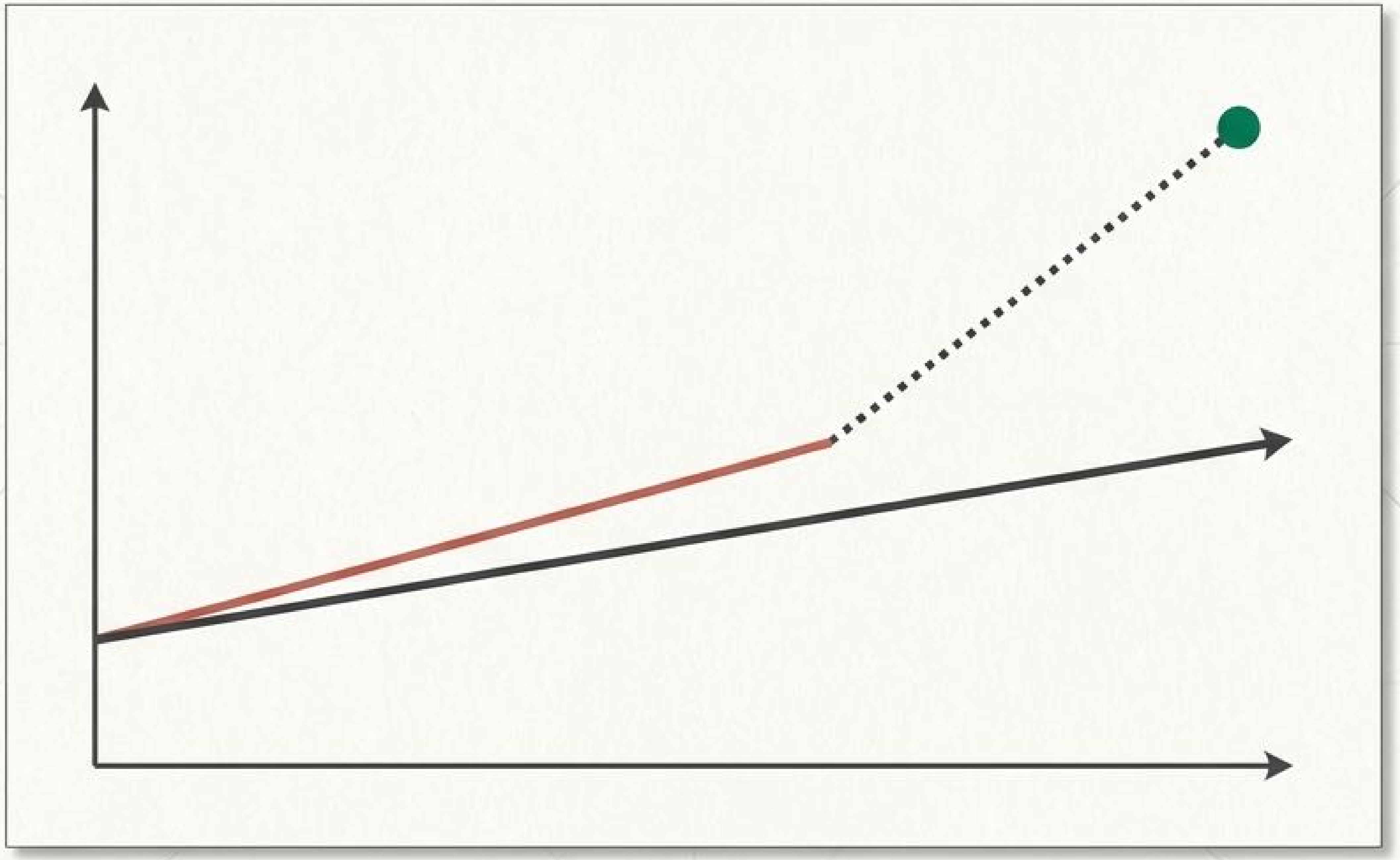
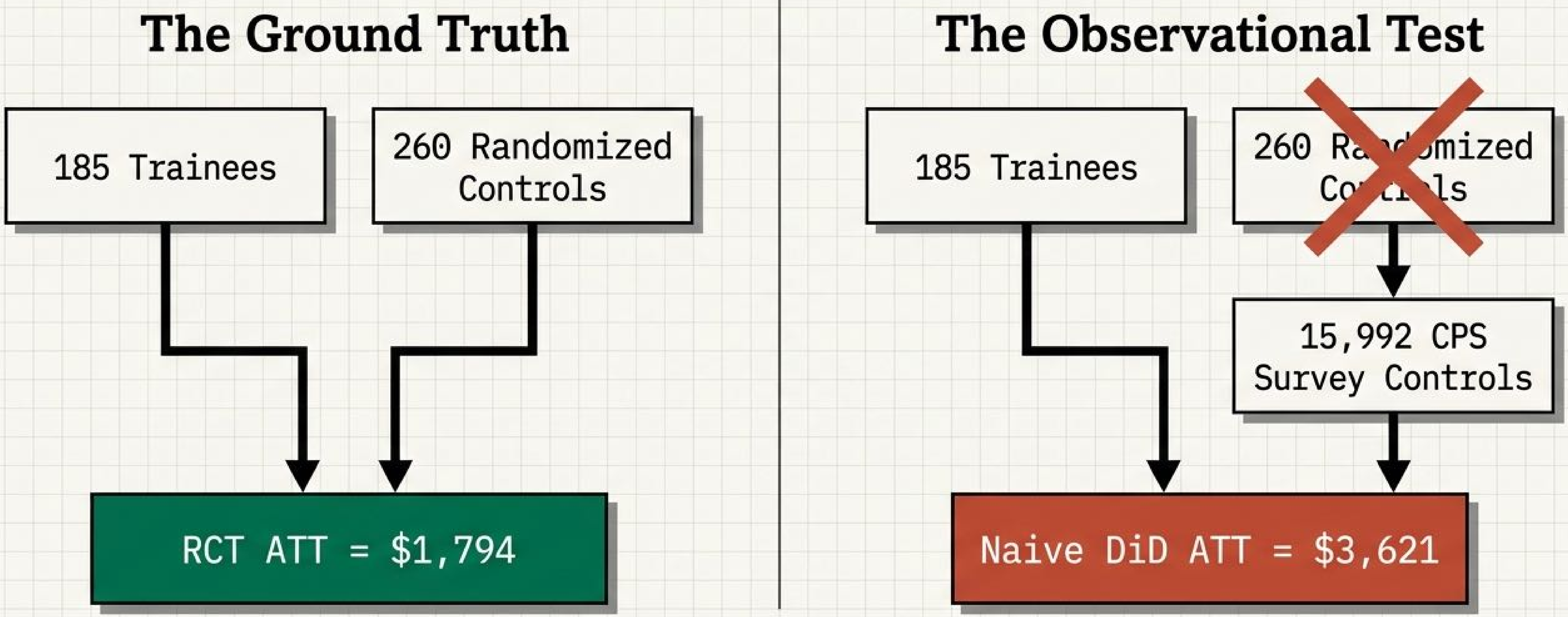
Covariates in Difference-in-Differences

A Methodological Study Guide on Trend, Level, and Treatment Effects



Audience: Graduate Economics | **Focus:** Methodology & Implementation

The LaLonde Test Grades the Estimator, Not the Data



Analogy: Grading a thermometer against a calibrated one. If it reads 40°C in boiling water, you learn about the thermometer, not the water.
The question: Can a DiD equipped with covariates recover the **$1,794?**

Foundational Concepts: Estimands and Imbalance

Estimand (ATT vs. ATE)

Definition: The focus is strictly on the Average Treatment on the Treated (**ATT**). Replacing the experimental control with a survey control changes who the comparison represents, but leaves the treated group intact.

Analogy: Measuring how much your team improved after coaching. Swapping which rival you compare against does not alter your team's gain. Target remains **$1,794**.

Covariate Imbalance

Definition: Systematic differences in traits (**X**). The **CPS** controls differ from trainees by a Standardized Mean Difference (**SMD**) of **+2.3** on race and **-1.6** on prior earnings.

Analogy: Comparing marathon times of teenagers and retirees. The age gap distorts comparisons only because age simultaneously drives the trend.

Foundational Concepts: Parallel Trends and Model Placement

Conditional Parallel Trends

Definition: With highly imbalanced groups, raw parallel trends fail. Parallelism is only theoretically credible conditional on covariates X.

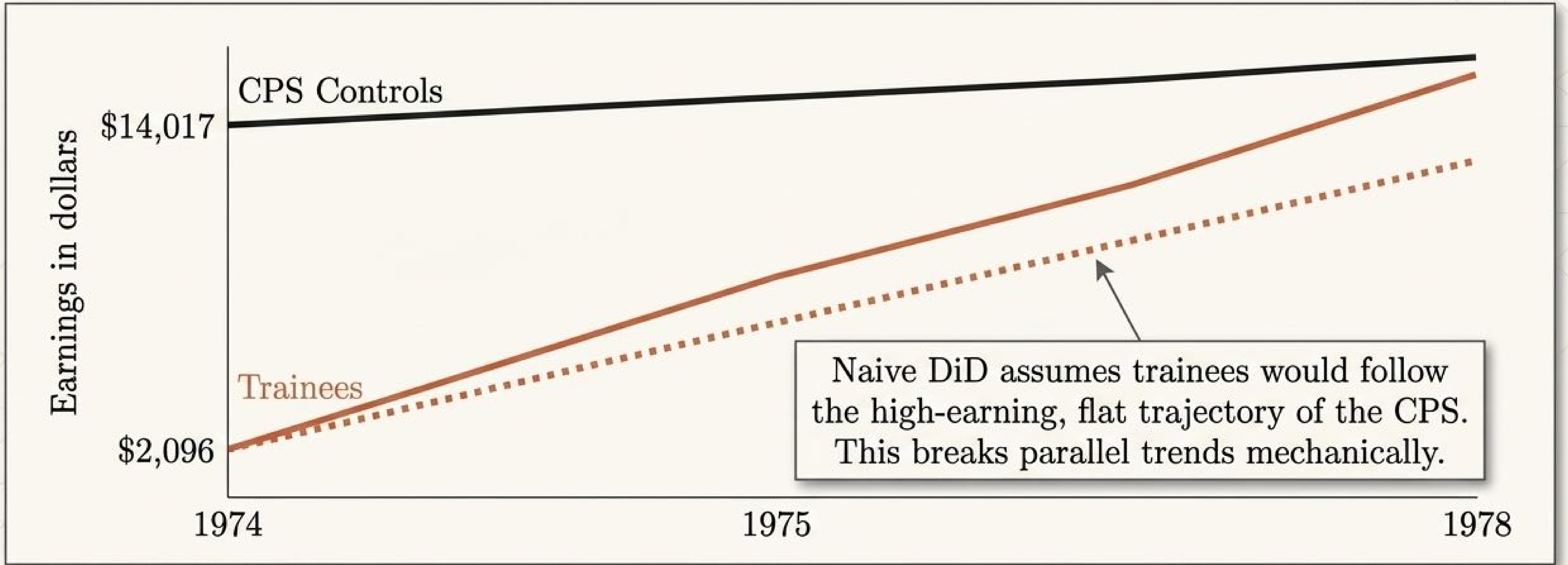
Analogy: Two runners on different hills. They only move in parallel once you mathematically account for the slope each is standing on.

Trend vs. Level vs. Effect

Definition: A covariate in the level (additive) or the treatment effect ($X \times D$) leaves the counterfactual trend untouched. A covariate in the trend ($X \times post$) **bends the counterfactual and corrects bias.**

Analogy: A corrective lens. Placed over the eye, it fixes vision (trend); placed in your pocket, it does nothing (level).

Why Naive TWFE Fails: Divergent Raw Trends

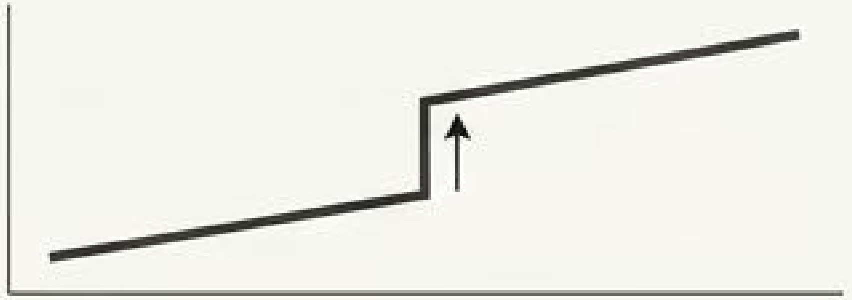


Methodological Conclusion: Covariates must model the divergent trajectories across groups, not merely adjust the starting baseline.

The Anatomy of Covariate Placement

Level (Additive X)

$$y = \ldots + \beta X + \ldots$$



Counterfactual trend untouched. Inert.

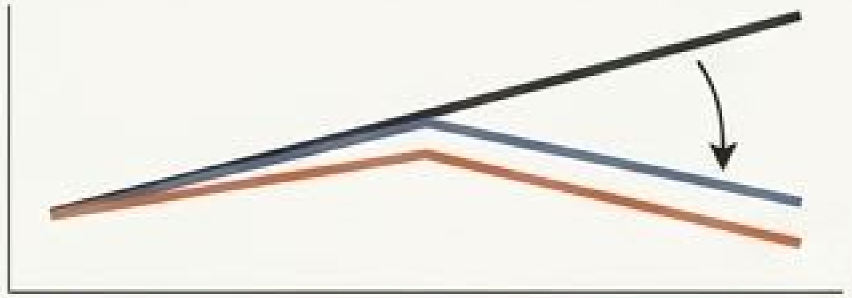
Effect ($X \times D$)

$$y = \ldots + \beta(X \times D) + \ldots$$

Counterfactual trend untouched. Inert.

Trend ($X \times post$)

$$y = \ldots + \beta(X \times post) + \ldots$$



Counterfactual corrected. Success.

The Inert Specifications: Naive TWFE and Additive Controls

Spec 0: Naive TWFE

- Method: No covariates.
- Result: **$3,621**
- Diagnosis: Precise, plausible, wildly wrong.

Spec A: Additive X

- Method: TWFE with additive baseline controls (`pyfixest` standard implementation).
- Result: **$3,621**

Methodological Autopsy

Because the baseline covariates are time-invariant (race, 1974 earnings), the ‘within’ transformation in a TWFE model mechanically sweeps them out entirely. The DiD coefficient never depended on them. Adding additive controls is not a robustness check; it literally does nothing to the point estimate.

Spec BT: Heterogeneous Effects Do Not Fix Divergent Trends

The Approach

Interact covariates with the treatment switch ($T = post \times D$), recover ATT via g-computation over the treated-post cell.

Result: $3,621

Methodological Autopsy

- **What it does:** Saturates the model in levels, allowing the treatment effect to vary based on traits.
- **Why it fails:** The empirical problem was never that treatment effects varied across groups; the problem was the control group's counterfactual trend. Modifying the treatment gap leaves the underlying divergent trajectories mismodeled.

Spec B: The Pivot — Placing Covariates in the Trend

The Approach: Interact each baseline covariate with the post indicator (`X × post`).

Result: $1,711 | Target Truth: $1,794

Methodological Autopsy

By multiplying the time-invariant characteristic by `post`, the model now mathematically allows workers with different baseline traits to follow different earnings trajectories over time.

Key Insight

Covariates rescue a Difference-in-Differences estimate only when they are permitted to bend the control group's counterfactual trend.

Spec C: Saturated First Differences (HIT 1997)

The Mechanics

Result: $1,770

The Approach:

First-difference the outcome: $\Delta y = y_{1978} - y_{1975}$

Regress on covariates, treatment indicator, and full interactions.

Recover ATT via g-computation.

Methodological Autopsy

Because the outcome itself is a change, a covariate's coefficient is now entirely free to act as its effect on the trend. This single regression simultaneously bends the control's counterfactual trend and relaxes constant treatment effects.

Interpretation

Oaxaca-Blinder logic applied to a change. Learn how normal workers' pay evolves, then impute exactly how much more the trainees gained than their statistical twins would have.

Alternative Paradigm: Modeling Treatment (IPW)

The Approach

Estimate the probability of being a trainee. Reweight the CPS controls to resemble the treated using Abadie's (2005) inverse-propensity weights applied to the first-differenced outcome.

$$w_i = \left[\frac{D_i - p(X_i)}{1 - p(X_i)}\right] * \frac{1}{\Pr(D=1)}$$

The Outcome

Result: $1,861

Interpretation

Two completely independent modeling philosophies (modeling the outcome trend vs. modeling treatment assignment) converge near the $1,794 experimental benchmark. Convergence builds confidence.

The Ultimate Safety Net: Doubly Robust DiD

Doubly Robust (`Sant'Anna-Zhao 2020`) — Consistent if either the trend model or the propensity model is correctly specified.

Outcome Regression
(Trend Model)

Result: $1,993

`Cross-checked via diff-diff package at $1,979.`

Inverse-Propensity Reweighting
(Treatment Model)

Methodological Autopsy

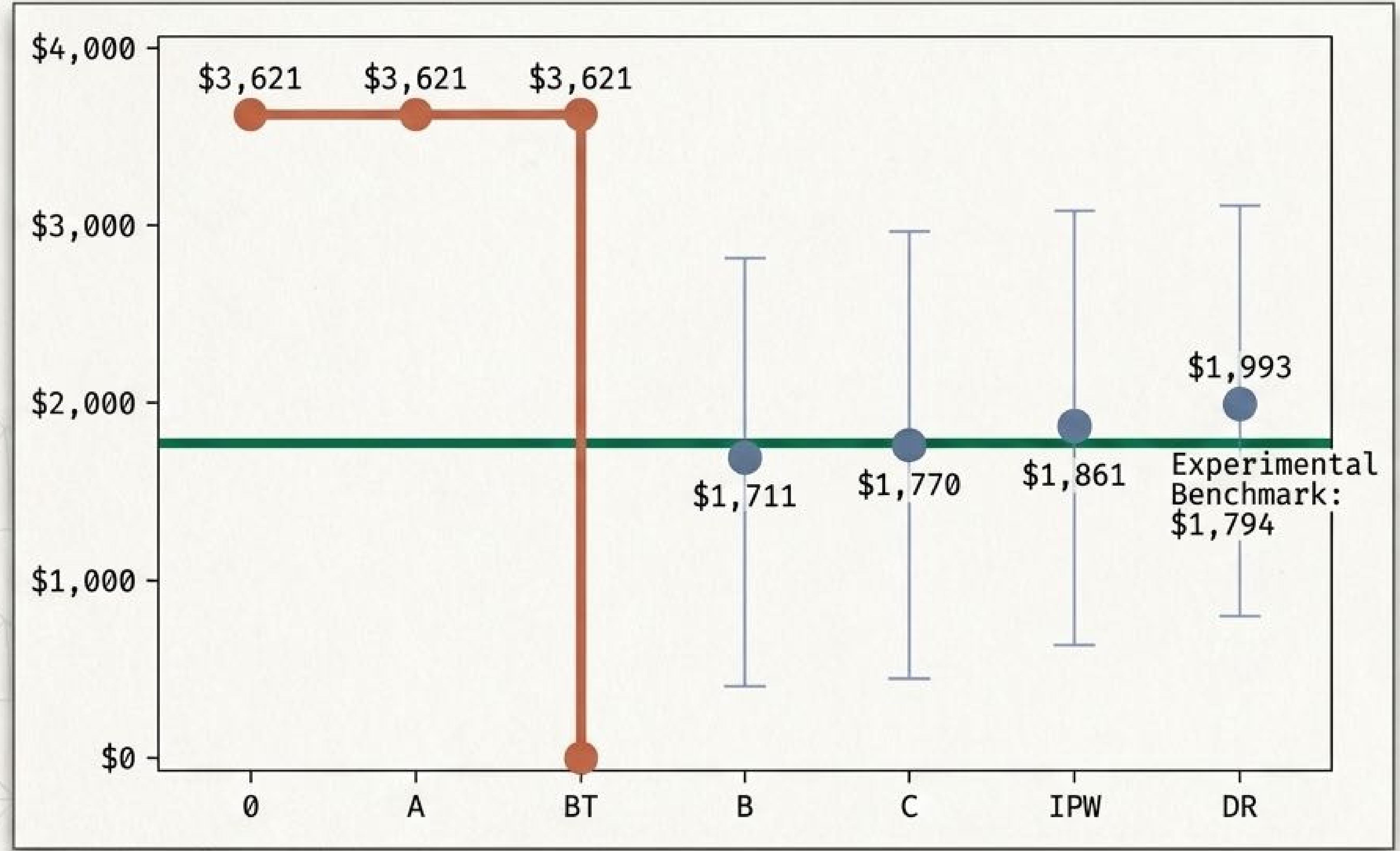
Two independent safety nets. You fall through to a biased estimate only if both models fail simultaneously. In applied work where true models are unknown, this insurance is critical.

Synthesis: Covariate Placement Dictates the Bias

The Covariate Arc

Spec	Estimator	Where X enters	ATT Estimate
0	Naive TWFE	nowhere	$3,621
A	Additive X	level	$3,621
BT	X × treatment	effect	$3,621
B	X × post	trend	$1,711
C	Saturated FD (HIT)	trend + effect	$1,770
--	IPW (Abadie)	propensity	$1,861
--	DR (Sant'Anna)	propensity	$1,993
Truth	**RCT Benchmark**	**ground truth**	**$1,794**

The “Snap” to the Experimental Benchmark



Trust the Pattern, Not the Decimal

The benchmark’s own Standard Error is ~$671. Given this noise, do not rank the corrected estimates ($1,770 is not inherently “better” than $1,711). The rigorous finding is the clean split between trend-ignoring and trend-modeling specifications.

Implementation Blueprint for Applied Panel Work

1 Step 1: Diagnose Imbalance. Always compute Standardized Mean Differences (SMDs) between groups. If SMD > 0.1 or > 0.25, raw parallel trends are fundamentally suspect.

2 Step 2: Ditch Additive Controls. Stop treating 'TWFE + baseline controls' as a robustness check. It mechanically sweeps out time-invariant variables and ignores the trend.

3 Step 3: Model the Trend. Explicitly use $X \times \texttt{post}$ or first-differencing in regression software (e.g., `pyfixest`) to bend the control trajectory.

4 Step 4: Use Modern Estimators. Default to Doubly-Robust methods using dedicated causal packages (e.g., `diff-diff` or `Callaway-Sant'Anna` implementations in Python/R) to leverage built-in safety nets against model misspecification.