Staggered Synthetic Difference-in-Differences (SDID)

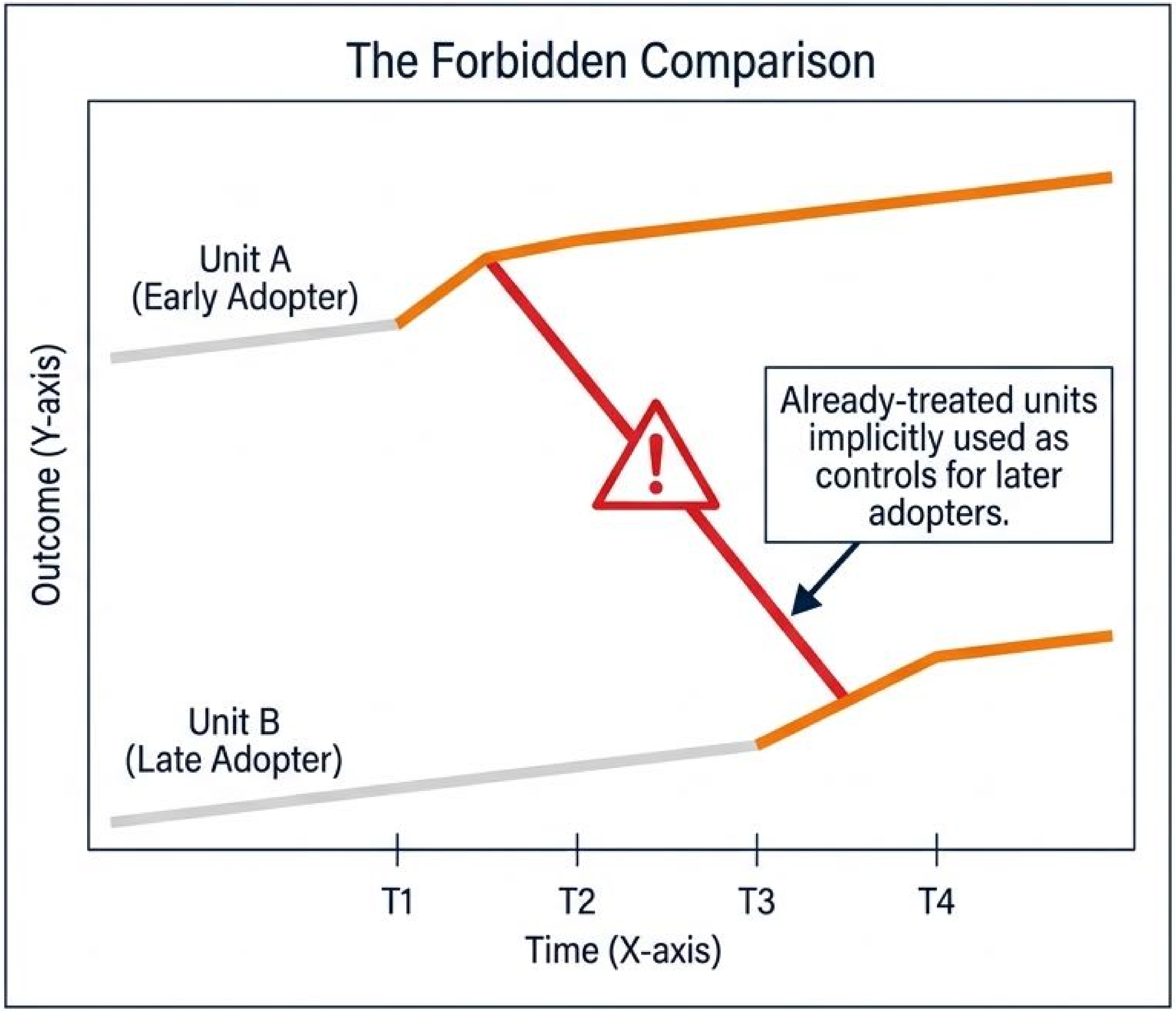
Methodology & Implementation: A Study Guide on Theory, Staggered Extensions, and Stata Application

The Econometric Masterclass Series

The Methodological Problem: Why TWFE Fails

Core Concept: Staggered adoption (different units treated in different years) combined with heterogeneous effects breaks standard Two-Way Fixed Effects.

The Consequence: TWFE yields a variance-weighted average of every 2x2 comparison. Because it uses already-treated units as controls, these weights can turn negative, pulling the estimate away from the true effect or flipping its sign.



Empirical Context & The Estimand

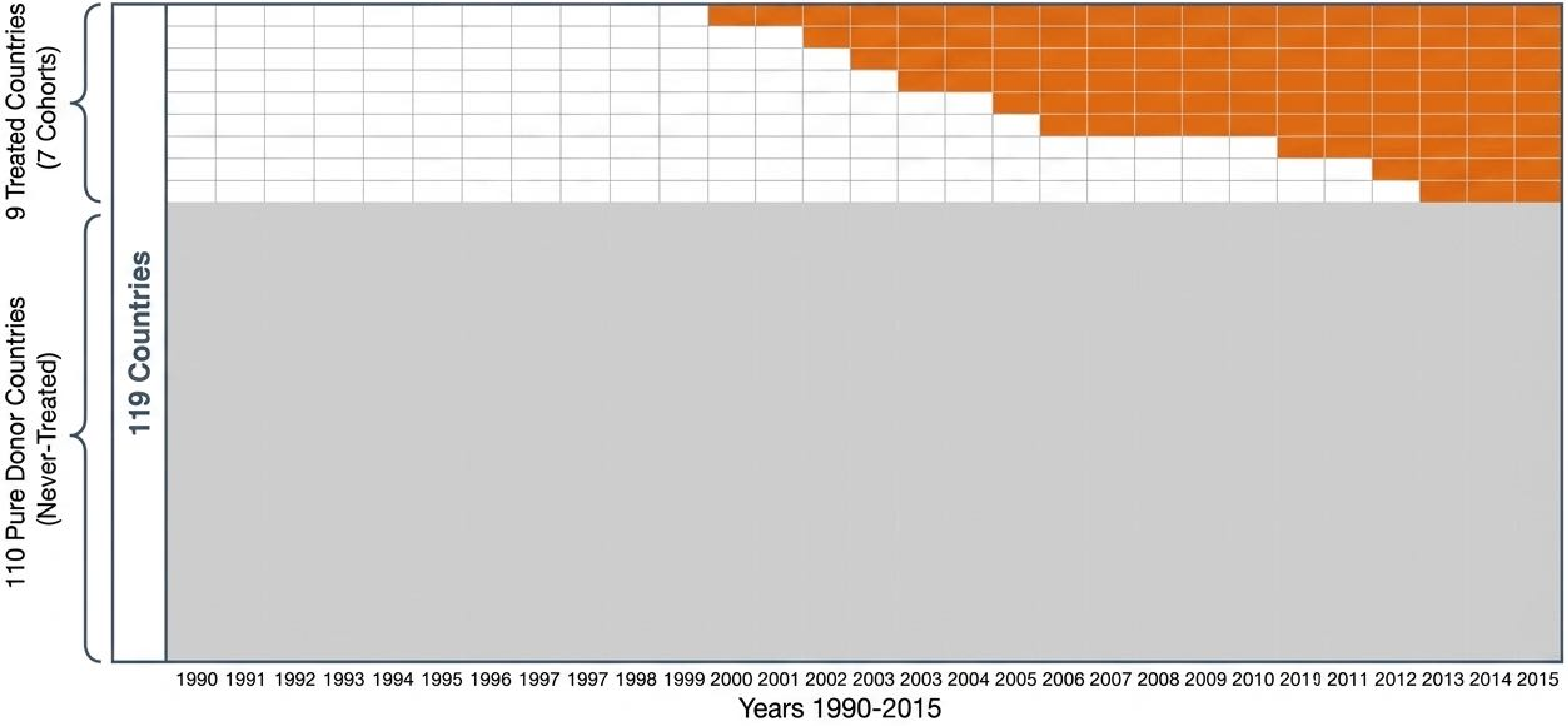
Math Variable	Role	Dataset Definition
i	Unit	119 countries (9 ever-treated, 110 never-treated donors)
t	Time	1990-2015 (26 years)
Y_{it}	Outcome	% women in national parliament
W_{it}	Treatment	1 once quota adopted (absorbing), 0 otherwise
X_{it}	Covariate	Log GDP per capita

The Target Estimand (Average Treatment Effect on the Treated)

$$\tau = \frac{1}{N_{tr} \times T_{post}} \sum_{i:W_i=1} \sum_{t>T_{pre}} \left[Y_{it}(1) - Y_{it}(0) \right]$$

The unobserved counterfactual that SDID must impute.

Visualizing the Staggered Structure



Key Insight: Treatment is an absorbing state but arrives on 7 different clocks.

Conclusion: Naive comparison of group means is confounded by varying baselines. We require cohort-specific counterfactuals, cleanly built from the pure donor countries.

First Principles: The SDID Objective Function

The Machinery: SDID is fundamentally a weighted TWFE regression.

$$(\hat{\tau}, \hat{\mu}, \hat{\alpha}, \hat{\beta}) = \underset{\tau,\mu,\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \mu - \hat{\alpha}_i - \hat{\beta}_t - W_{it} * \hat{\tau})^2 \, \hat{\omega}_i \cdot \hat{\lambda}_t$$

Standard Unit and Time Fixed Effects.

The Twist: The weights (ω, λ) are not free parameters; each solves its own distinct optimization problem to construct the optimal counterfactual.

Deconstructing ωmega: The Unit Weights

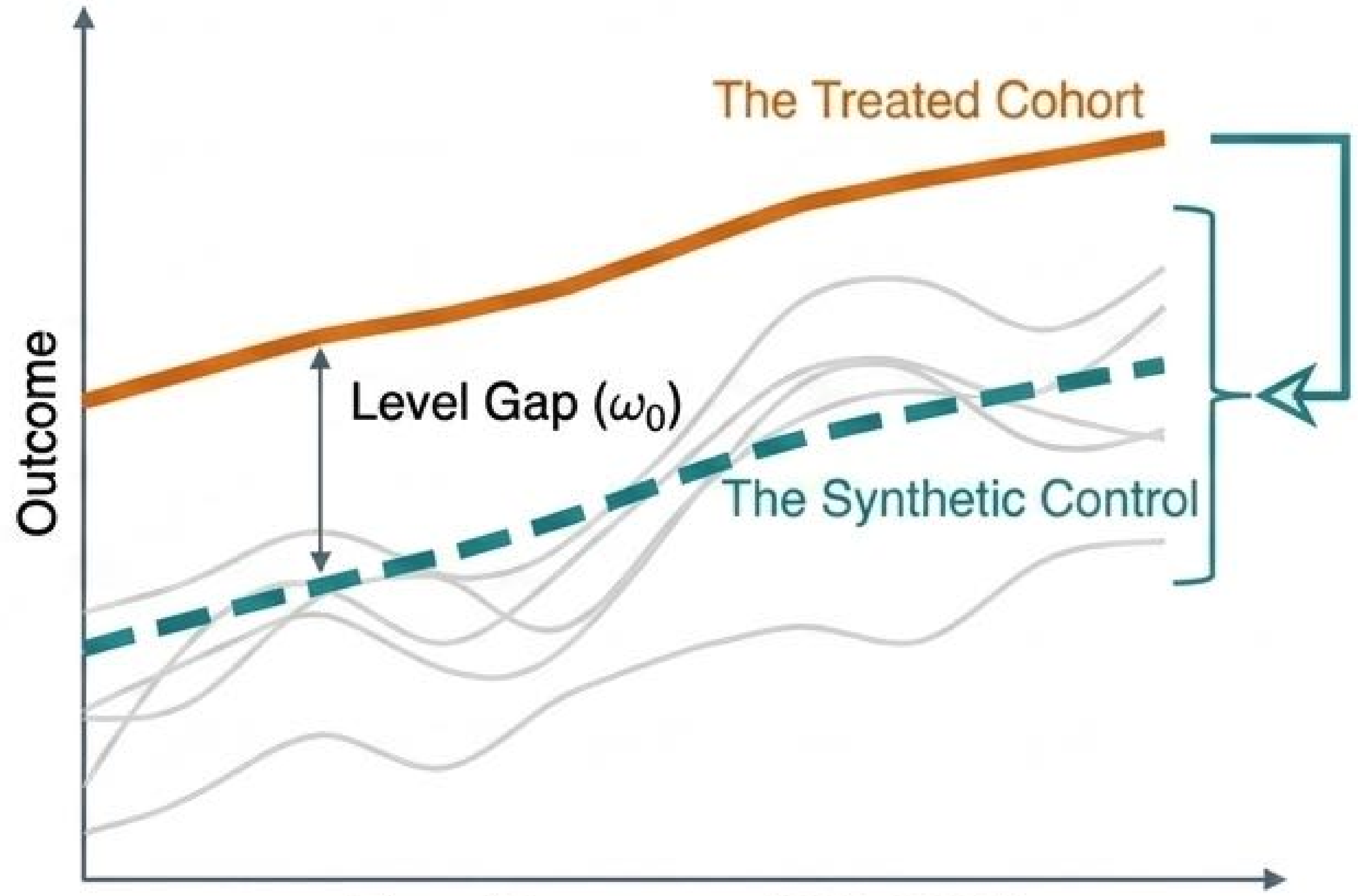
The Optimization Problem

$$\hat{\omega} = \underset{\omega_0, \omega \geq 0}{\operatorname{argmin}} \sum_{t=1}^{T_{pre}} \left(\omega_0 + \sum_{i=1}^{N_{co}} \omega_i Y_{it} - \frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} Y_{it} \right)^2 + \zeta^2 T_{pre} \|\omega\|^2$$

The **Intercept** (ω_0):
Allows the synthetic control to match the trend while ignoring baseline level gaps (absorbed later by unit fixed effects).

Ridge Penalty (ζ):
Spreads weights across multiple donors rather than concentrating on a few, ensuring stability.

Visual Metaphor: Trend Matching with Level Gap



Deconstructing lambda: The Time Weights

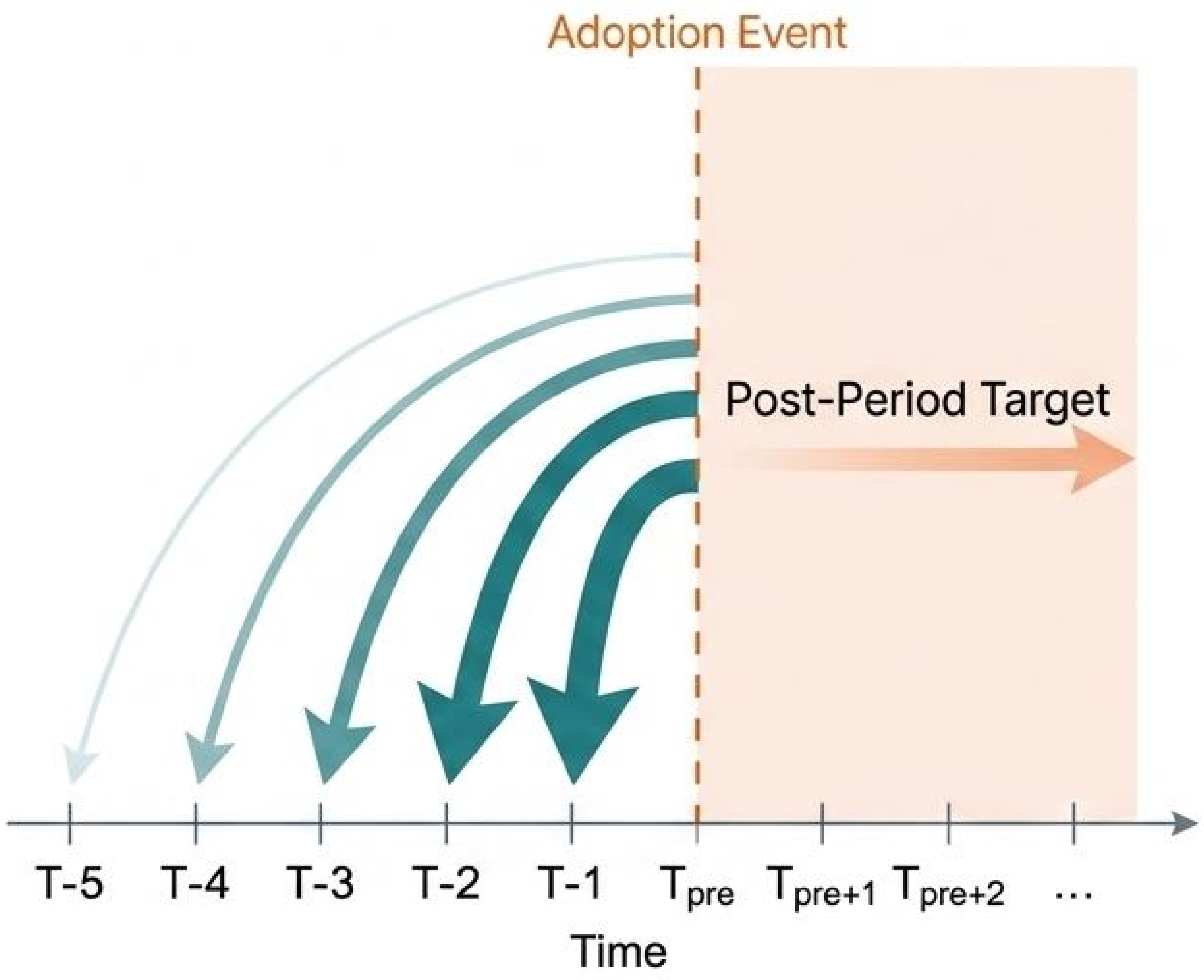
The Optimization Problem

$$\hat{\lambda} = \underset{\lambda_0, \lambda \geq 0}{\operatorname{argmin}} \sum_{i=1}^{N_{co}} \left(\lambda_0 + \sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^{T} Y_{it} \right)^2 + \zeta \lambda^2 N_{co} \|\boldsymbol{\lambda}\|^2$$

Predictive Baseline: Abandons uniform averaging. Up-weights the specific pre-adoption years that best predict the post-period trajectory.

The 'Before' Anchor: The counterfactual baseline is built from the most relevant history, not a flat, potentially noisy average of early years.

Visual Metaphor: Predictive Baseline



The Methodological Comparison Matrix

Method	Unit weights (ω)	Time weights (λ)	Unit FE (α_i)	Must match
DiD	Uniform	Uniform	✅ Yes	Trend on all controls
Synthetic Control	Optimized	Uniform	❌ No	Level and trend
SDID	Optimized	Optimized	✅ Yes	Trend (level gap allowed)

Takeaway: SDID bridges the gap. It optimizes unit weights like Synthetic Control but retains unit fixed effects like DiD, freeing it from having to match exact baseline levels.

The Staggered Extension: Cohort-by-Cohort Aggregation

Assembly Line

Extract Cohort A + Pure Never-Treated Donors → Solve SDID → $\hat{\tau}_A$ → Weight by Size/Duration: $(N_{trA} * T_{postA})$ → $\sum$ → Overall $AT\hat{T}_a$

$$AT\hat{T} = \sum_{a \in A} \left[\frac{N_{\text{tr}}^{a} \cdot T_{\text{post}}^{a}}{\sum_{b \in A} (N_{\text{tr}}^{b} \cdot T_{\text{post}}^{b})} \right] \hat{\tau}_a$$

The Mechanism:
Estimate a separate, mothemnictated self-
self-contained SDID effect for each cohort using only never-treated controls.

The Advantage:
By aggregating with weights strictly proportional to treated country-years, the method mathematically guarantees non-negative weights. The TWFE “forbidden comparison” is eliminated.

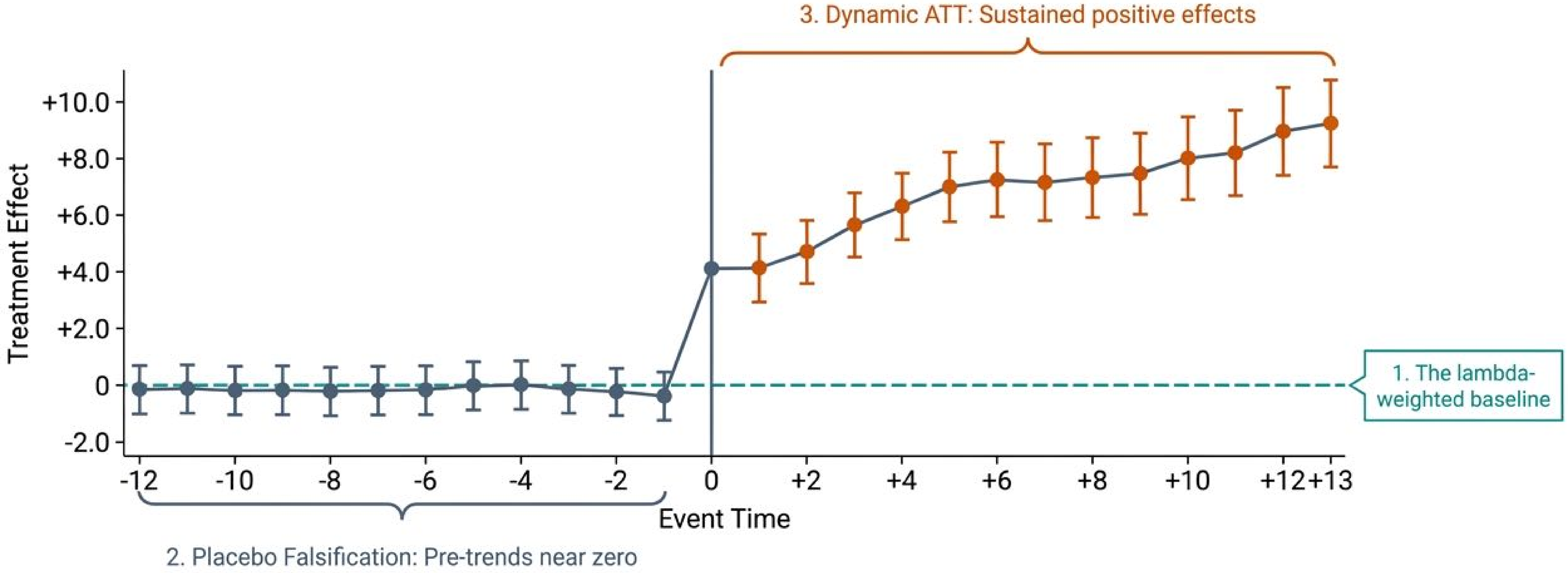
Application Results: The Aggregate vs. The Heterogeneous

Headline Result: Overall **ATT = +8.03 percentage points** (SE 3.74, p=0.032)

Interpretation: Gender quotas raise female parliament share by ~8 points on average.

The Nuance: The aggregate is a time/size-weighted average, not a simple mean. It masks extreme heterogeneity. Some quotas were transformative (+21.8); others had no measurable effect (-3.5).

Reading the SDID Event Study



1. The Baseline: Measured against a lambda-weighted optimal average of the pre-period, not just t=-1.

2. Placebo Falsification (Left): Pre-trend coefficients near zero confirm that parallel synthetic trends hold before adoption.

3. Dynamic Effects (Right): Tracks the evolution of treatment. For quotas, the effect appears immediately (+4.1 points) and sustains a higher level for over a decade.

Covariate Adjustment Strategies

The Confounder Check: Does the quota effect merely reflect economic development (Log GDP)?

Strategy 1: Optimized Method

Folds covariate adjustment directly into the SDID optimization. Flexible but computationally heavy.

ATT: +8.05

Strategy 2: Projected Method

A. Regress Y on X for untreated → B. Extract Residuals → C. Run SDID on Residuals

Regresses outcome on covariate among untreated observations, then runs SDID on residuals. Numerically stable and fast.

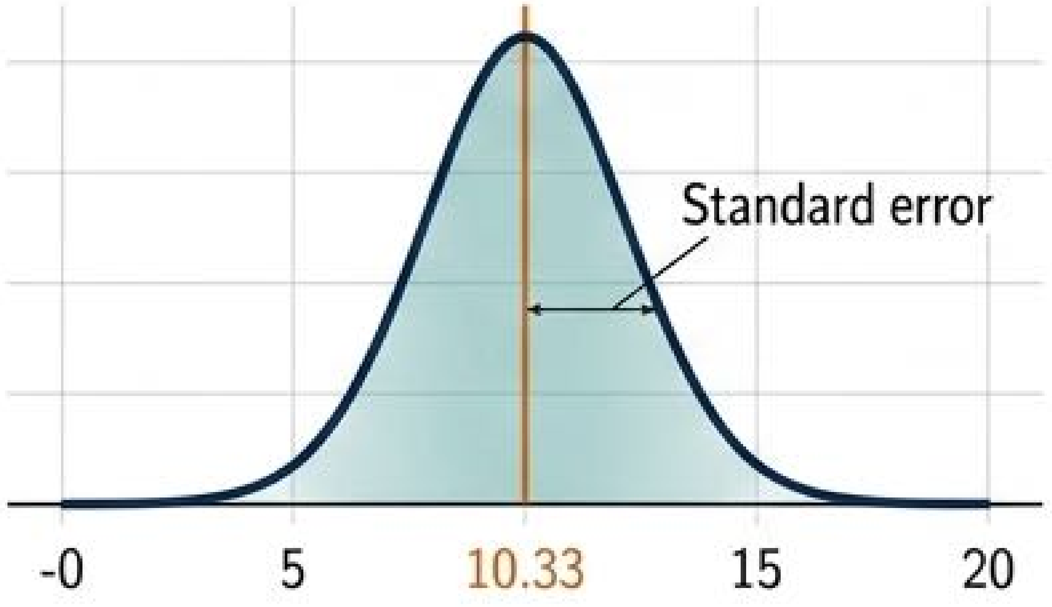
ATT: +8.06

Result Robustness: Both yield virtually identical estimates, proving the the +8.03 effect is fully robust to income differences.

Inference Diagnostics

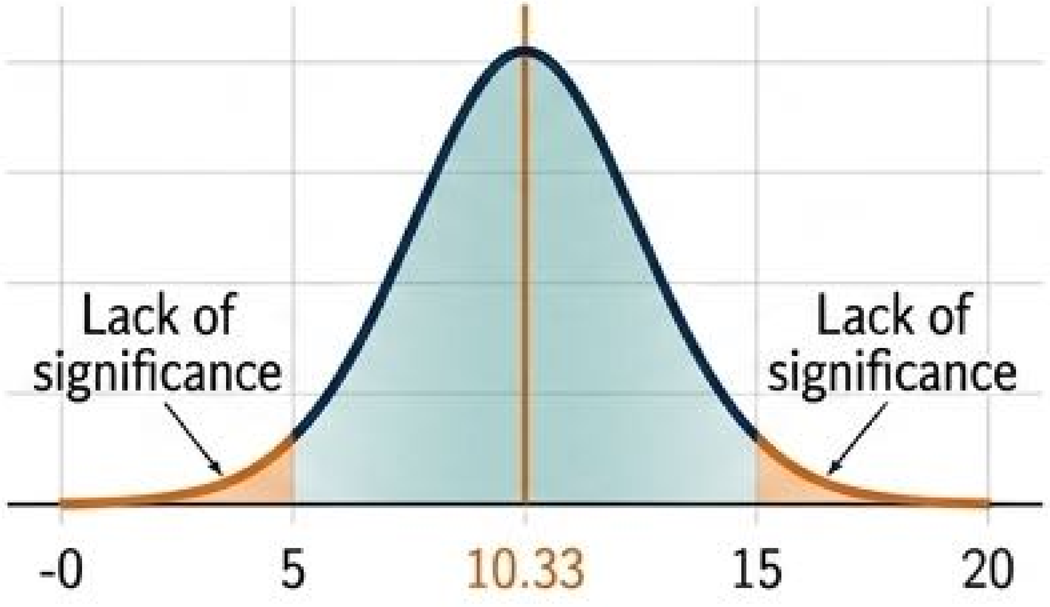
Context: Tested on the 2002/2003 cohort sub-sample (N>1 treated units per period).
Point estimate is identical (10.33), but standard errors vary drastically.

Bootstrap (Default)



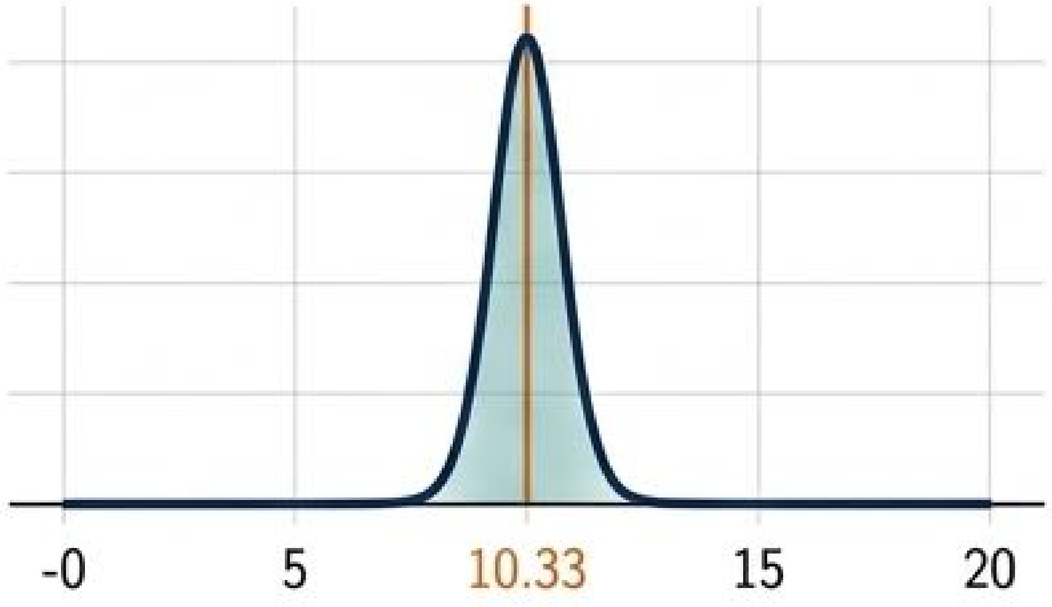
- Resamples units.
- SE: 4.73
- The recommended baseline standard error.

Jackknife (Conservative)



- Leave-one-out sampling.
- SE: 6.01 (Crosses zero)
- Most cautious approach, requires >1 treated unit per cohort.

Placebo (Aggressive)



- Assigns fake treatment to controls.
- SE: 2.34
- Tightest bounds, assumes homoskedasticity.

Takeaway: With few treated units, cross-check standard bootstrap results against jackknife or placebo bounds to ensure robustness.

Implementation Guide (Stata)

Core SDID Estimation

```
* Estimate overall ATT with standard bootstrap inference
sdid womparl country year quota, vce(bootstrap) seed(1234)
```

e(tau) stores the underlying heterogeneous cohort-specific effects.

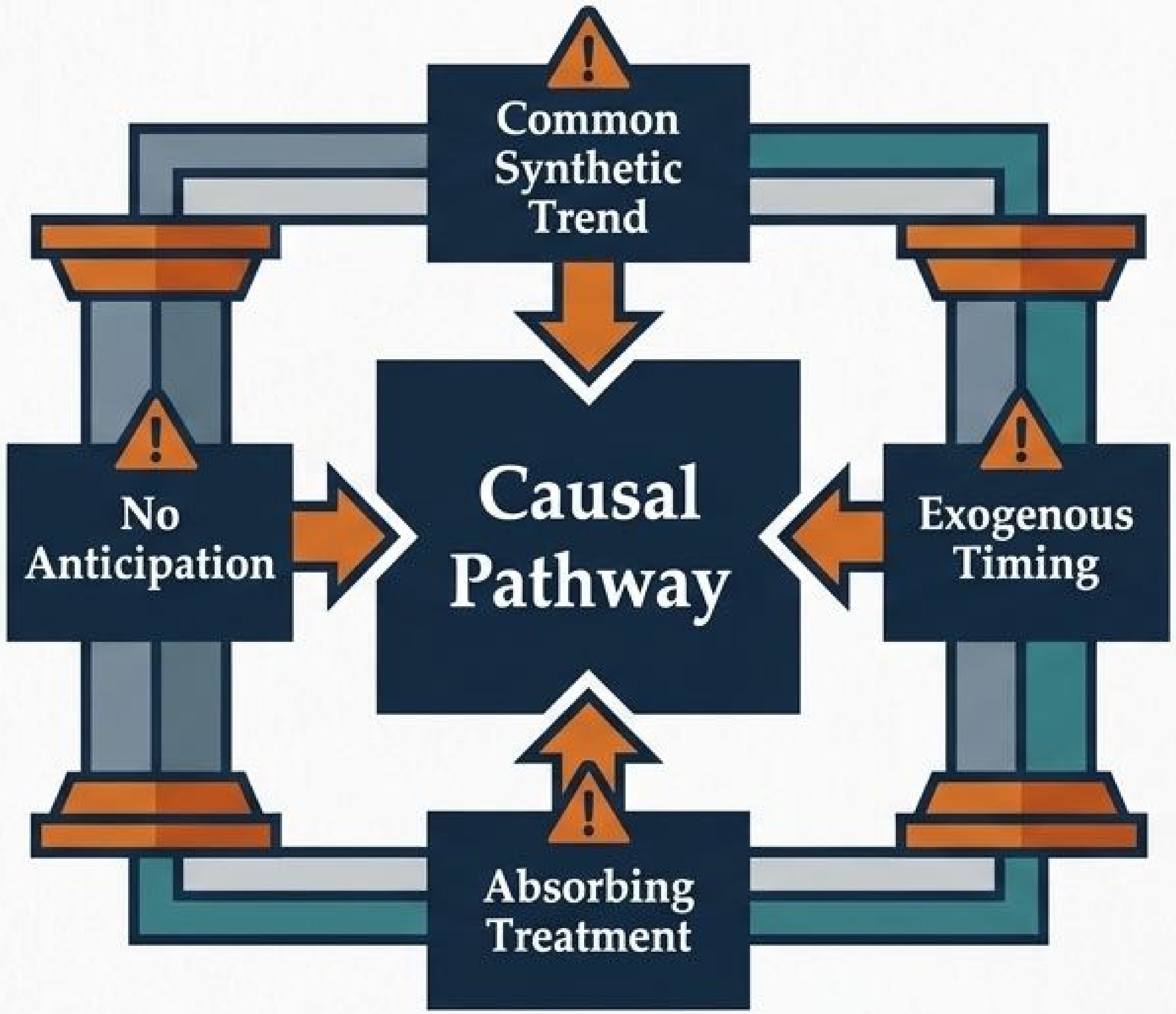
Event Study Execution

```
* Generate dynamic effects and placebo tests
sdid_event womparl country year quota, vce(bootstrap) seed(1234)
```

Covariate Adjustment (Projected)

```
* Robustness check conditioning on log GDP per capita
sdid womparl country year quota, covariates(lngdp, projected)
```

Synthesis & Identifying Assumptions



- **1. Common Synthetic Trend:** The optimally weighted donors accurately represent the counterfactual trajectory (supported by flat event-study placebos).
- **2. No Anticipation:** Units do not change behavior before formal adoption.
- **3. Absorbing Treatment:** Once treated, units do not revert to untreated status.
- **4. Exogenous Timing:** Adoption timing is not a mechanical response to the outcome's future path.

Final Takeaway: Staggered SDID isolates clean variations free from forbidden comparisons, unmasking the true structural heterogeneity hidden within panel data.