

# diff - diff

v2.4

**Your variance estimator  
is lying to you.**

Gardner (2022) Two-Stage DiD

GMM sandwich variance that tells the truth

Per-observation treatment effects

# What is diff-diff?

- + Complete DiD toolkit for Python
- + sklearn-like API, statsmodels-style output
- + 10 methods, 12 tutorials, validated vs R
- + The most complete DiD toolkit in any language

Now with the Two-Stage DiD estimator.

# The TWFE

## Problem

	t=1	t=2	t=3	t=4	t=5	t=6
Unit 1	Untreated	Untreated	Treated	Treated	Treated	Treated
Unit 2	Untreated	Untreated	Untreated	Treated	Treated	Treated
Unit 3	Untreated	Untreated	Untreated	Untreated	Untreated	Untreated
Unit 4	Untreated	Untreated	Untreated	Untreated	Untreated	Untreated

*TWFE estimates FEs using ALL data, including treated outcomes*



Untreated



Treated

**Treated outcomes contaminate the counterfactual**

Heterogeneous effects create negative weights

**Solution: estimate the model on untreated data only.**

diff-diff

# Two-Stage DiD

*Gardner (2022) | Butts & Gardner (R Journal, 2022)*

1

Estimate unit + time FEs on untreated observations only



2

Residualize ALL outcomes, regress on treatment

Clean counterfactual from untreated data.

Unbiased treatment effects from the residuals.

Requires parallel trends + no anticipation + absorbing treatment.

**diff-diff**

# The Math

Two stages, three equations

Stage 1

$$Y_{it} = \alpha_i + \delta_t + \varepsilon_{it}$$

*(on  $D_{it} = 0$  only)*

Stage 2

$$\tilde{Y}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\delta}_t$$

$$\tilde{Y}_{it} = \tau \cdot D_{it} + u_{it}$$

*(on ALL observations)*

Variance

$$V = (D'D)^{-1} \left[ \sum_c S_c S_c' \right] (D'D)^{-1}$$

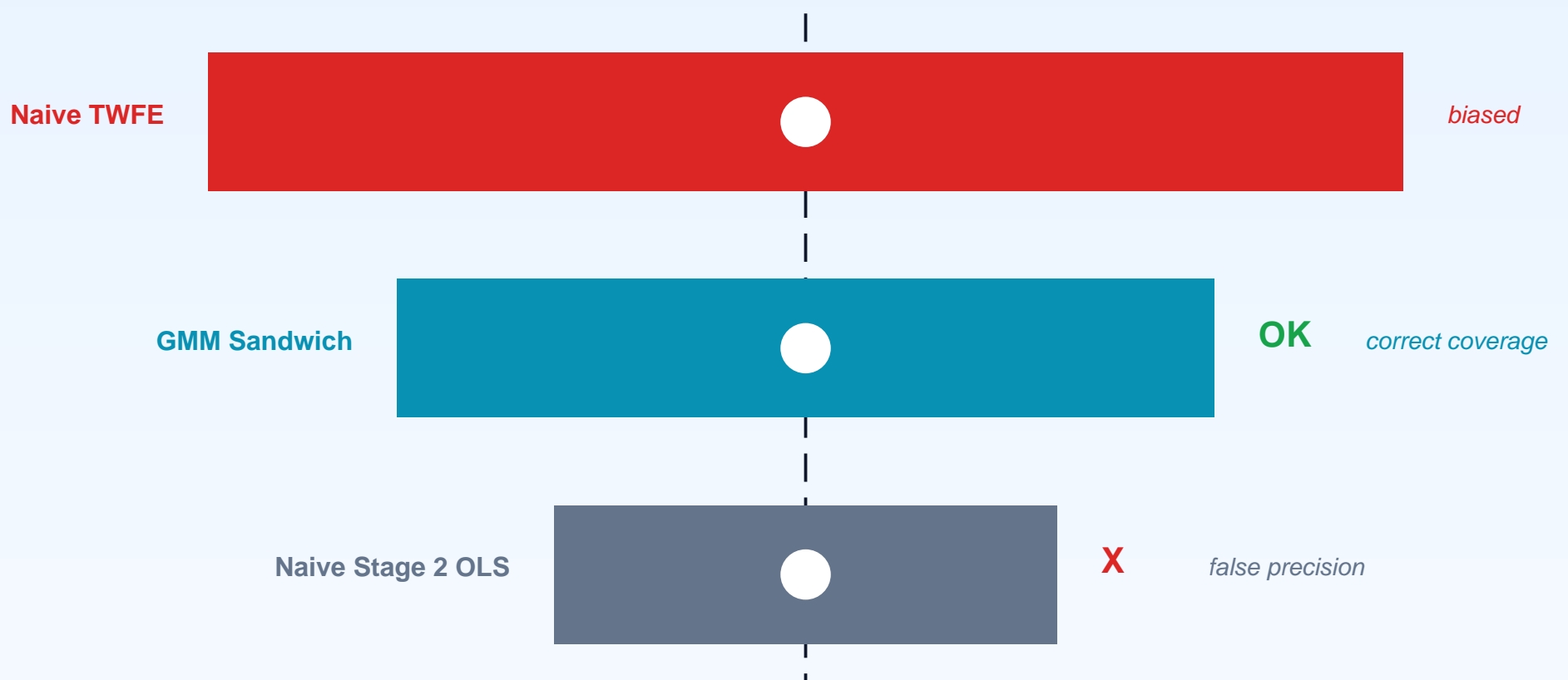
**GMM sandwich corrects for Stage 1 uncertainty**

diff-diff

# Honest

## Standard Errors

Not all confidence intervals tell the truth



Naive OLS ignores that  $\alpha$ -hat and  $\delta$ -hat are estimated.

The GMM correction accounts for first-stage uncertainty.

**Narrower isn't better if it's wrong. GMM gets it right.**

Under homogeneous effects. Compare with ImputationDiD for robustness.

diff-diff

# Per-Observation Treatment Effects

Every treated unit-period gets its own tau-hat

unit	time	tau_hat	weight
firm_3	2019	2.14	0.0033
firm_3	2020	1.87	0.0033
firm_7	2020	3.21	0.0033
firm_7	2021	2.95	0.0033
...	...	...	...

**Aggregate to: static ATT, event study, or by cohort**

Or analyze individual treatment effect heterogeneity

diff-diff

# Drop-in

## Replacement

```
# Switch in one line

from diff_diff import TwoStageDiD

est = TwoStageDiD()

results = est.fit(

    data,

    outcome='sales',

    unit='firm_id',

    time='year',

    first_treat='first_treat',

    aggregate='event_study'

)

# Per-observation effects

results.treatment_effects.head()
```

Same fit() API as CallawaySantAnna and ImputationDiD.

Identical point estimates to ImputationDiD.

**diff-diff**



# Every Method

## You Need

### Basic DiD / TWFE

Classic 2x2 and panel

### Sun-Abraham

Interaction-weighted (2021)

### Two-Stage DiD [NEW]

Gardner (2022)

### Triple Difference

DDD with proper covariates

### Honest DiD

Rambachan-Roth sensitivity

### Callaway-Sant'Anna

Staggered adoption (2021)

### Imputation DiD

Borusyak et al. (2024)

### Synthetic DiD

Arkhangelsky et al. (2021)

### TROP

Factor-adjusted DiD (2025)

### Bacon Decomposition

TWFE diagnostic weights

The most complete DiD toolkit in any language.

diff-diff

# Upgrade to

## v2.4

```
$ pip install --upgrade diff-diff
```

[github.com/igerber/diff-diff](https://github.com/igerber/diff-diff)

Full documentation & 12 tutorials included

MIT Licensed | Open Source

## diff - diff

Difference-in-Differences for Python