

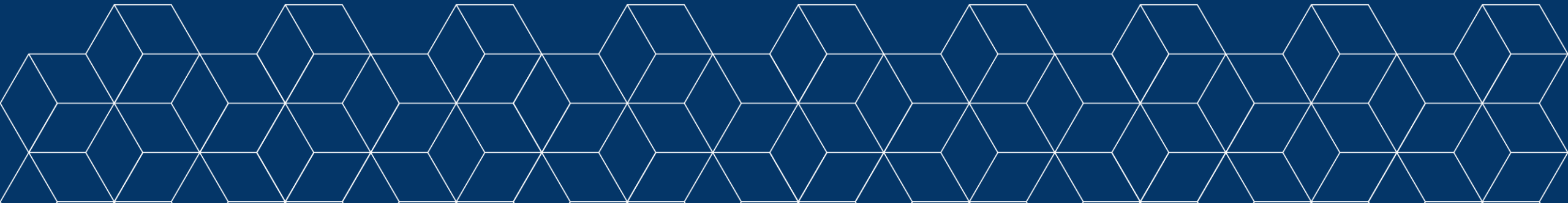
M. BIAGETTI*
G. SIMONE*

PUBLIC PROCUREMENT AND FIRM PERFORMANCE: EVIDENCE FROM ITALIAN MUNICIPALITIES

Event-Study, a Relative-Spike Design and Their Limits

XXXVIII SIEP Conference, Modena Italy September 10th, 2026

***INAPP, Rome Italy**



SUMMARY

MOTIVATION

LITERATURE, METHODOLOGY

DEFINITIONS

DATA SOURCES

DESCRIPTIVE STATS

RESULTS

ROBUSTNESS

CONCLUSIONS AND FURTHER RESEARCH

ANNEX

A relevant tool for economies: an appealing, open question

Public procurement is around **12.7% of GDP** across OECD countries (2023).

A growing body of economic research asks whether it raises the firms' profitability and performance when they are located where contracts are awarded.

Firm-level designs

Winners vs. losers of the same tender. Explicit counterfactual, credible identification: positive effects on firm growth, survival, credit

Aggregate designs

Exposure at the municipality-year level. No losing-bidder counterfactual: this is where our setting originally lay.



Why not a Continuous Dose nor a TWFE?

As a continuous, non-absorbing dose in a TWFE regression, the aggregate design cannot identify these effects.

−0.41

first-difference autocorrelation of the annual flow — close to the −0.5 of pure white noise

Around 41%

of treated TWFE cells carry a negative weight: the coefficient is uninterpretable (our computations)

$ResDos$

$= Dose$

$- \text{mean of single ob.} - \text{mean of years}$

$+ \text{overall mean}$

$$w = \frac{ResDos}{\sum ResDos}$$

No Timing

the dose rises and falls every year: there cannot be any adoption date, or stable control group (in other words, the dose – i.e. the amount of money awarded might go up and down and might not be continuous)

We change the question: not a dose, but an event. Therefore, the case becomes binary and absorbing once again (once treated, always treated. Control group == not-yet-treated obs.)



The First Drawbacks: dCDH vs CS

dCDH

Non-negative weights

no cell enters with a flipped sign

Valid controls only

already-treated units never used as controls

Continuous switching (reversible) dose

Continuous doses (i.e. public funding) rise and fall...

No profile over time (static dCdH)

Only a single average response of the outcome to a change in dose, pooling all year-on-year switches when the dose moves.

Problems when dose change is near zero

CS

Needs binary adoption

A continuous dose needs to be dichotomized

51.3% up · 44.0% down

Needs absorbing treatment

adopters fall back

80.3% revert

Needs stable controls

almost no stayers

22 of 7,548

Three strands of literature

Procurement & Firms

Winners vs. losers of a tender – tantamount to an explicit counterfactual.

Ferraz et al. (2015); Cappelletti et al. (2024); Gabriel (2024); Hebous–Zimmermann (2021)

Local Fiscal Multipliers

Public demand at the territorial level. Mixed evidence, no losing-bidder counterfactual.

Acconcia et al. (2014); Auricchio et al. (2020); Pieri et al. (2026)

DiD with Heterogeneity

TWFE breaks under staggered, heterogeneous effects. Robust estimators and sensitivity tools.

Goodman-Bacon (2021); Callaway–Sant'Anna (2021); de Chaisemartin–D'Haultfoeille (2020); Borusyak et al. (2024)

We bring the diagnostics of the third strand to a setting in the tradition of the second.



Five estimators that fail in different ways

TWFE	The workhorse of the aggregate tradition. de Chaisemartin–D'Haultfœuille (2020) and Goodman-Bacon (2021): a weighted sum of group-time effects with possibly negative weights.	Benchmark
Callaway–Sant'Anna	reference: clean group-time ATTs, doubly robust	
Borusyak et al.	imputation: most efficient, most trend-sensitive to parallel trends failures	
de Chaisemartin–D'H.	permissive on heterogeneity	
Wooldridge	regression-adjusted, flexible	
Stacked (Cengiz et al.)	the most legible: stacked mini-experiments	



Agreement across methods with different weaknesses limits the chance that a number is dependent uniquely on the method used

Treatment as a relative investment spike

The first year a municipality's per-capita works procurement exceeds 3× its own historical median.

$$G_m = \min \{ t : \text{euro_wrk}(m, t) > 3 \times \text{median}_m, \text{median}_m > 0 \}$$

Relative, not absolute

each municipality judged against its own history — keeps treated and controls comparable

Lumpy-investment roots

transposes the 'relative spike' of Power (1998) to municipal budgets

Staggered adoption

restores testable pre-trends and a not-yet-treated comparison group



Six Treatment Definitions

Label	Authority perimeter	Assignment	Sector	Amount
C_comm_lav	Municipalities (BDAP)	Contracting municipality	Works	Tender base
C_comm	Municipalities (BDAP)	Contracting municipality	All sectors	Tender base
C_comm_aw_lav	Municipalities (BDAP)	Contracting municipality	Works	Awarded
C_luogo_lav	Municipalities (BDAP)	Place of execution	Works	Tender base
PR	Pieri-Ricci perimeter	Contracting municipality	All sectors	Awarded
PR_lav	Pieri-Ricci perimeter	Contracting municipality	Works	Awarded



A Municipal Panel: Procurement Meets Firm Accounts

ANAC contracts define the treatment; AIDA balance sheets provide the outcomes. 2015–2023.

ANAC

11m lot-level contracts, classified to municipalities via the BDAP register. It's where we get the treatment from

AIDA

Incorporated-firm balance sheets (legal seat): productivity, labour share, EBITDA, employment

71,389

municipality-years

7,942

municipalities

5,708

treated

1,998

never treated

Other sources for population (2015, Istat) and deflator (2020, Eurostat)



The Main Four AIDA Outcomes

Municipality-year panel, 2015–2023. Log outcomes; labor share as a ratio. Medians reported.

Outcome	N	Mean	SD	Median
Employment $\ln(1+\text{workers})$	68,103	4.97	2.11	5.12
Productivity $\log \text{VA per worker}$	65,742	10.68	0.59	10.75
Labor share	66,871	0.66	0.26	0.66
EBITDA per worker (natural log)	63,153	9.57	0.92	9.65

A municipal economy of small firms: strong right-skew, so medians rather than means are exhibited. Sample sizes differ by outcome (they have their own missing patterns).

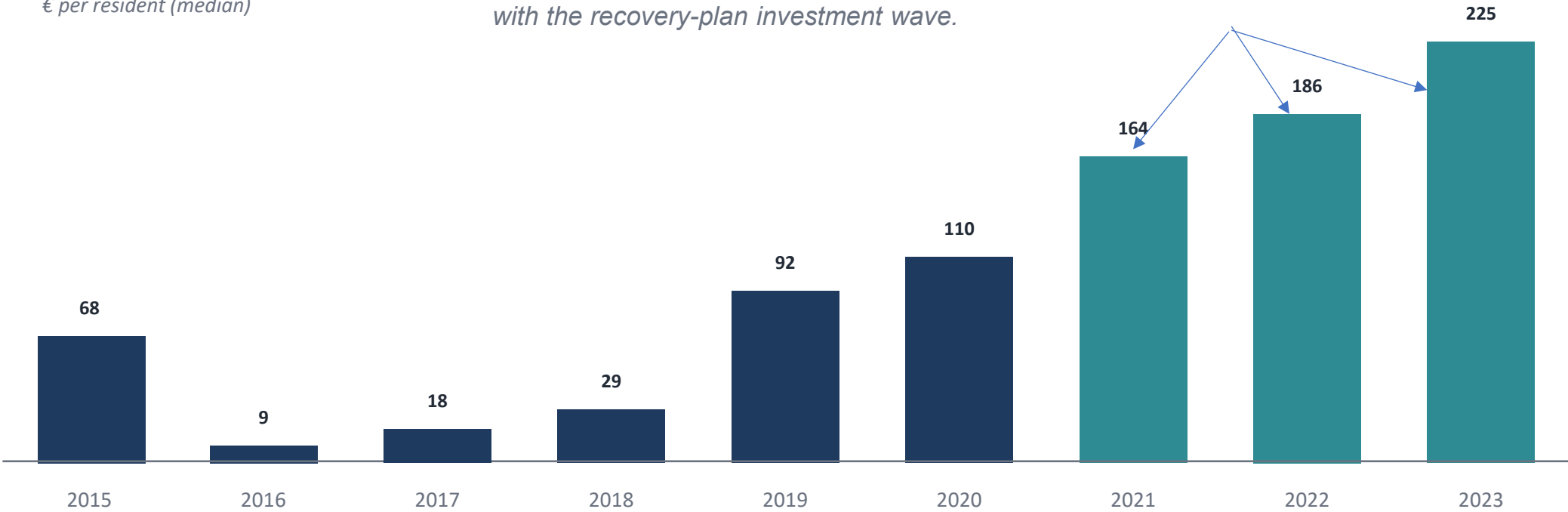


Works procurement rises over the decade

ANAC works tendered by municipalities, real euros per resident (2015 population). Median per year.

€ per resident (median)

The median municipality tenders three times more works per resident in 2023 than in 2015; the post-2020 acceleration coincides with the recovery-plan investment wave.





Main specification

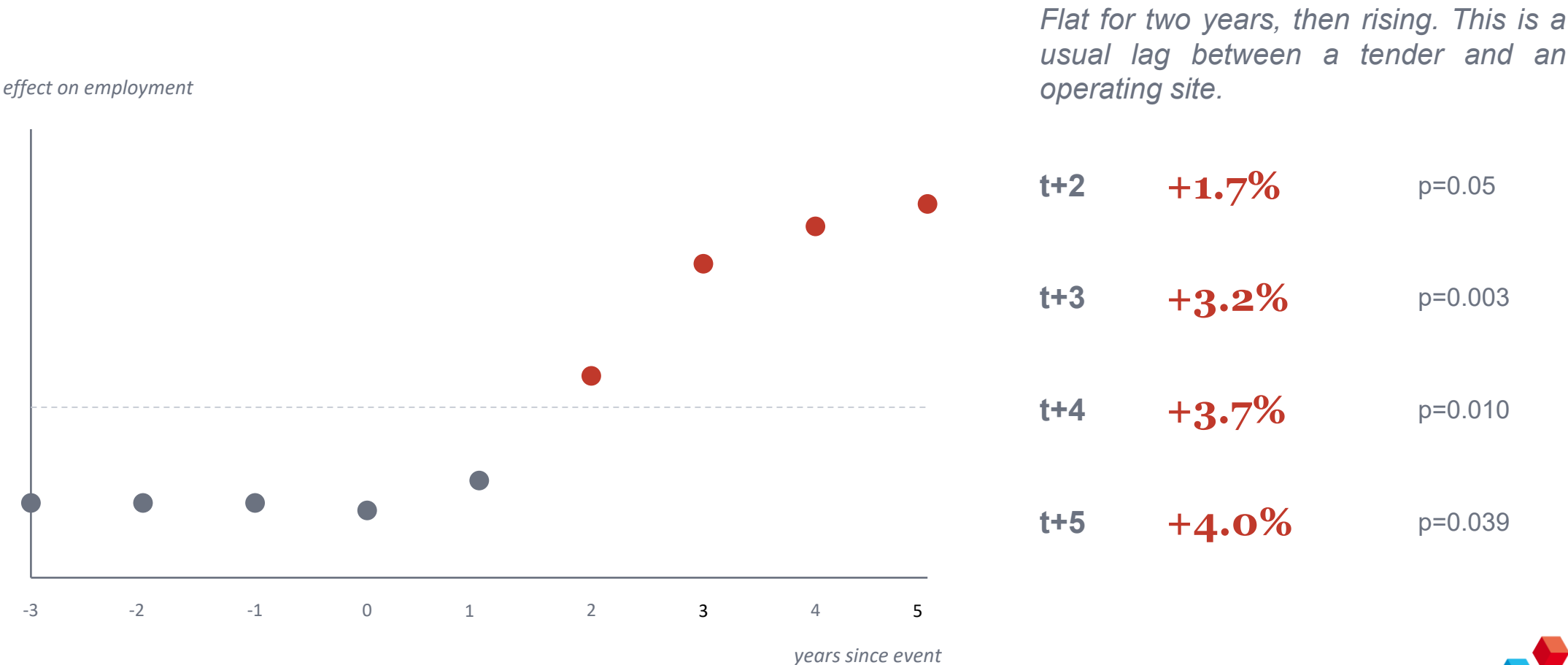
CS ATT, AIDA outcomes. Log outcomes ×100 percent.

Outcome	ATT	p	Pre-trend p
Productivity (VA per worker)	+1.8%	0.009	0.74 ✓
Total Value Added	+1.5%	0.015	0.03 x
Employment (headcount)	+1.4%	0.037	0.98 ✓
Labor share	−1.0 pp	0.020	0.52 ✓
EBITDA per worker	+2.7%	0.024	0.21 ✓
Labor cost per worker	+0.8%	0.155	0.73
Sales per worker	+0.2%	0.734	0.79

Lab prod grows, wages flat consequently labor share falls, margins rise.



Timing: Employment Arrives Late



Placebos: the Same Rule on Spending that Opens no Site

Treatment	Productivity	EBITDA p.w.	Labor share	Employment
Works (main)	+1.8%**	+2.7%**	-1.0pp**	+1.4%**
Services	+0.9%	-0.4%	-0.7pp*	+0.4% x
Supplies	+0.6% x	+1.1%	-0.4pp	+0.3% x
All public bodies	+0.4%	+1.2%	+0.0pp	-0.0%

If the effect were generic fiscal expansion, non-works procurement would show it too, but it doesn't

Two caveats: placebos have very low statistical power (services MDE = 0.018 > public works estimate), and a formal works-vs-services test can't be rejected. But these weaken alternatives and cannot be considered a decision rule written in stone.



Stable Across Design and Specification Choices

Check	Productivity	EBITDA p.w.	Labor share	Employment
Threshold 2×	+2.2%**	+4.2%***	−1.2pp***	+0.4%
Threshold 5×	+1.8%**	+2.0%	−0.6pp	+2.3%***
Never-treated only	+1.5%**	+2.9%**	−1.1pp**	+1.7%**
Place of execution	+1.4%**	+2.3%**	−0.6pp	0.9% (pre-trend rej. p=0.027)
Center–North	+1.9%**	+2.2%*	−0.9pp*	+1.9%**
Reference t−2	+2.3%***	+3.2%***	−1.2pp***	+1.7%**

Employment shows a dose-response (+0.4%, +1.4%, +2.3% at 2×/3×/5×), i.e. the larger shocks, the more are the hirings.

T-2 is the anticipation test. Normally the reference period (the basis the effect is measured against) is the year BEFORE the event, t-1. This check shifts the reference two years before. The logic is the following: were municipalities to anticipate the event – i.e. they would move before the construction site begins – shifting behind the reference should weaken the estimate. Yet, it strengthens it. We conclude that we are not capturing any anticipation, therefore the effect comes with the event not before.



Inference is not the Weak Link but we Stay Cautious

Clustering, permutation and multiple-testing corrections, on a frozen 60-cell grid.

Is the P-value of the main effects credible

Inference holds

999 replications to see if the coefficient wobbles

Wild bootstrap (999)

p within 3% of analytical

Permutation (500)

prod p=0.006, share p=0.002

Province clusters

prod 0.020, share 0.007

500 draws, 500 fake effects then compare the true effect to the fake distribution

Multiple testing

Something turns out significant by chance

15 / 60

significant at 5% (raw)

11 / 24

Robust to Holm-Bonferroni within family correction

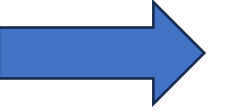
0 / 60

Survive Benjamini-Hochberg even though designed for independent discoveries

**Main works definition survives Holm-Bonferroni for all four outcomes.
No cell clears the grid-wide BH correction (min p=0.17). We interpret this result as suggestive, not decisive as it is conceived**

It allows municipalities within the same province to be correlated





The electoral cycle: the objection reverses

Mid-mandate events

+3.3%

productivity, $p=0.002$ — where the political cycle cannot explain it (EBITDA +6.4%)

Pre-electoral events

+1.3%

productivity, $p=0.139$ — not significant where political timing would predict one

Mayors concentrate tenders before elections. If timing were political, the effect should live there.

The productivity and profitability effects are largest where the political motive is absent (Employment is the exception).

Anyway, this tool can be only considered as diagnostic, not an IV yet.



CONCLUSIONS AND FURTHER RESEARCH

What we offer, and what comes next



A design that makes the question estimable

the relative-spike event restores testable pre-trends where the aggregate dose cannot



A coherent, robust pattern

productivity, delayed employment, falling labor share, rising margins — across 5 estimators & many checks



An honest interpretation

suggestive descriptive evidence; no cell clears the grid-wide correction, trends untestable post-event



Next steps

IV on the electoral calendar; extend outcomes to the universe of firms (local-unit entry)

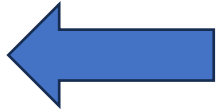




THANKS FOR YOUR ATTENTION



www.inapp.org

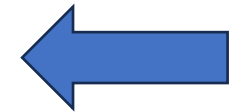


Five estimators agree on the four main outcomes

Outcome	CS	BJS	dCDH	Wooldridge	Stacked
Productivity	+1.8%	+2.1%	+1.7%	+2.0%	+2.1%
Labor share	−1.0pp	−1.3pp	−1.0pp	−1.2pp	−1.0pp
Employment	+1.4%	+1.7%	+1.2%	+1.6%	+1.3%
EBITDA/worker	+2.7%	+3.3%	+2.6%	+3.1%	+3.7%

The are five ways of computing the same contrast, not five independent experiments. Agreement rules out the lilikelihood that results depend on methods.





Electoral cycle, outcome by outcome

Outcome	Mid-mandate	Pre-electoral
Productivity	+3.3% (0.002)	+1.3% (0.139)
EBITDA per worker	+6.4% (0.001)	+1.6% (0.309)
Labor share	−1.4pp (0.055)	−1.1pp (0.056)
Employment	+1.2% (0.247)	+1.8% (0.051)

Productivity, margins: concentrated mid-mandate (cycle can't explain). Labor share: indifferent to the cycle which a natural feature of public works. Employment instead proves to be larger in pre-electoral times.

