

Geographies of Discontent, A Reanalysis and Discussion

Jesse Frederik

De Correspondent

jesse@decorrespondent.nl

Word count: ~3,150

Abstract

[Cremaschi et al. \(2024\)](#) report that a 2010 Italian municipal reform increased far-right vote share by 1.5 percentage points, using difference-in-differences around a population threshold. I show that this estimate reflects a time-varying size gradient—far-right support has grown faster in small municipalities—rather than a policy effect. The estimate attenuates to zero in narrow bands around the threshold, appears before the reform, arises at placebo population thresholds where no reform exists, and replicates in France where no comparable reform took place; adding flexible size-by-year controls eliminates it entirely. The proposed mechanism is also unsupported on its own terms: the reform was never systematically enforced and was abolished in 2024, compliance with the joint-service-delivery mandate was partial rather than universal, and where compliance did occur it produced no detectable shift in either service quality or far-right voting.

1 Introduction

Cremaschi et al. (2024) study a 2010 Italian reform that required municipalities below 5,000 inhabitants (below 3,000 for designated mountain municipalities) to jointly manage public services with neighboring communes.¹ Their identification strategy compares municipalities below these population cutoffs (treated) to those above (controls), before and after the reform, using difference-in-differences. They conclude that the reform increased far-right vote share by roughly 1.5 percentage points in the 2013 and 2018 general elections, relative to 2001–2008.

Because treatment is assigned by a population cutoff, treated municipalities are systematically smaller than controls. Far-right support has grown faster in small municipalities since the early 2000s—a differential trend that a standard DID specification, even with municipality and year fixed effects, cannot absorb.

Four anomalies confirm that this size gradient, not the reform, drives the estimate. The estimate attenuates to zero near the population threshold (Section 2.1). A temporal placebo using only pre-reform data is significant (Section 2.2). The same specification produces significant estimates at placebo population thresholds where no reform exists (Section 2.3). The identical pattern appears in France, where no comparable reform took place (Section 2.4). Controlling for size-specific time trends eliminates the treatment effect entirely (Section 3.1). A direct test of the proposed mechanism—comparing municipalities that actually complied with the mandate to similar ones that did not—finds no compliance effect either (Section 4).

¹**AI disclosure.** This manuscript used Claude Opus 4.6 and 4.7 (Anthropic). The research questions, research design, choice of analyses, argument structure, discussion and conclusions are the author’s own. Claude was used to (i) generate R code specified by the author; (ii) render the bias decomposition in Appendix B into formal notation, following the mechanism specified by the author; (iii) provide suggestions to improve clarity of author-drafted prose. The author checked, rewrote and refined all generated code and text suggestions. All code and data collection were reviewed by the author and outputs verified against original replication materials where applicable; all derivations and final text were reviewed and approved by the author, who retains full responsibility for the manuscript.

1.1 Data and methods

All analyses use the paper’s electoral panel dataset (7,964 municipalities \times 5 national elections: 2001, 2006, 2008, 2013, 2018), available through the Harvard Dataverse replication package, and the paper’s own specifications. I extend the panel to include the September 2022 general election (7,506 matched municipalities; 458 dropped because their 2008 ISTAT codes have been retired since 2008).² I also construct a parallel French panel (34,710 communes \times 5 presidential elections: 2002, 2007, 2012, 2017, 2022) as a placebo country where no comparable reform exists (Appendix A). Standard errors are clustered at the municipality level throughout.

TWFE (two-way fixed effects) regresses far-right vote share on a treatment indicator (treated \times post) with municipality and year fixed effects (see Roth et al. 2023, for a recent overview of DID methods). This is the paper’s baseline estimator. The authors note that pre-trends differ between treated and control groups and therefore turn to two alternative estimators as their preferred specifications. *MTWFE* (Ho et al. 2007) first matches treated municipalities to controls using Mahalanobis distance on pre-treatment covariates (population, demographics, income, education, altitude), then runs *TWFE* on the matched sample. *SDID* (Arkhangelsky et al. 2021) reweights control units and pre-treatment periods to balance the treated group’s pre-reform outcome trajectory. I replicate all three estimators and additionally use regression discontinuity (Calonico et al. 2014) to test for a jump in far-right support at the population threshold.

²Source: Italian Ministry of the Interior via the onData project’s processed Eligendo files (<https://github.com/ondata/elezioni-politiche-2022>). Party classification follows Cremaschi et al. (2024): far-right = Fratelli d’Italia + Lega.

2 Four Anomalies

2.1 The estimate attenuates near the population threshold

If the reform caused the far-right shift, it should appear as a jump at the population threshold. The logic of a regression-discontinuity design is that municipalities just below 5,000 are nearly identical to those just above except for the treatment: comparing only them isolates the policy effect from confounding by size and any size-correlated characteristic. A genuine discontinuity at the cutoff should therefore become *sharper*, not weaker, as the comparison narrows.

Figure 1 shows the opposite. The estimate sits near 0.015 when most of the country is included in the control group, begins to fall once the bandwidth narrows below $\pm 5,000$, and is indistinguishable from zero at bands of $\pm 2,000$ or below.³ The headline estimate is generated by including municipalities far larger than those near the cutoff in the control group, not by any jump at the cutoff itself.⁴

A more formal test uses regression discontinuity (implemented via `rdrobust`; [Calonico et al. 2014](#)), which fits separate local regressions on each side of the 5,000 threshold and tests whether far-right vote share jumps at the cutoff. The method selects the range of municipalities to include automatically (here $\pm 1,420$ inhabitants, $N = 1,792$). The pooled estimate is $+0.005$ ($p = 0.17$): same direction as the headline but a third of the size and statistically indistinguishable from zero. Running the same test separately on the cross-section in each election year confirms the null in all three post-reform elections: 2013 (-0.006 , $p = 0.52$), 2018 ($+0.008$, $p = 0.54$), and 2022 ($+0.004$, $p = 0.76$).⁵

³The attenuation is not just widening standard errors: at 24 of the 27 bandwidths below $\pm 2,000$, the upper 95% CI lies strictly below the headline.

⁴Running the sweep separately for non-mountain communes (cutoff 5,000) and mountain communes (cutoff 3,000) produces the same attenuation pattern in each group: full-sample estimates of $+0.016$ and $+0.011$ respectively, falling to mean narrow-band (± 200 to $\pm 2,000$) estimates of $+0.006$ and $+0.003$. The pooled-threshold figure is not generating the pattern by averaging two structurally different local comparisons.

⁵A McCrary density test ([Cattaneo et al. 2020](#)), which checks whether municipalities cluster suspiciously on one side of the threshold, finds no manipulation of the running variable ($T = -0.22$, $p = 0.83$).

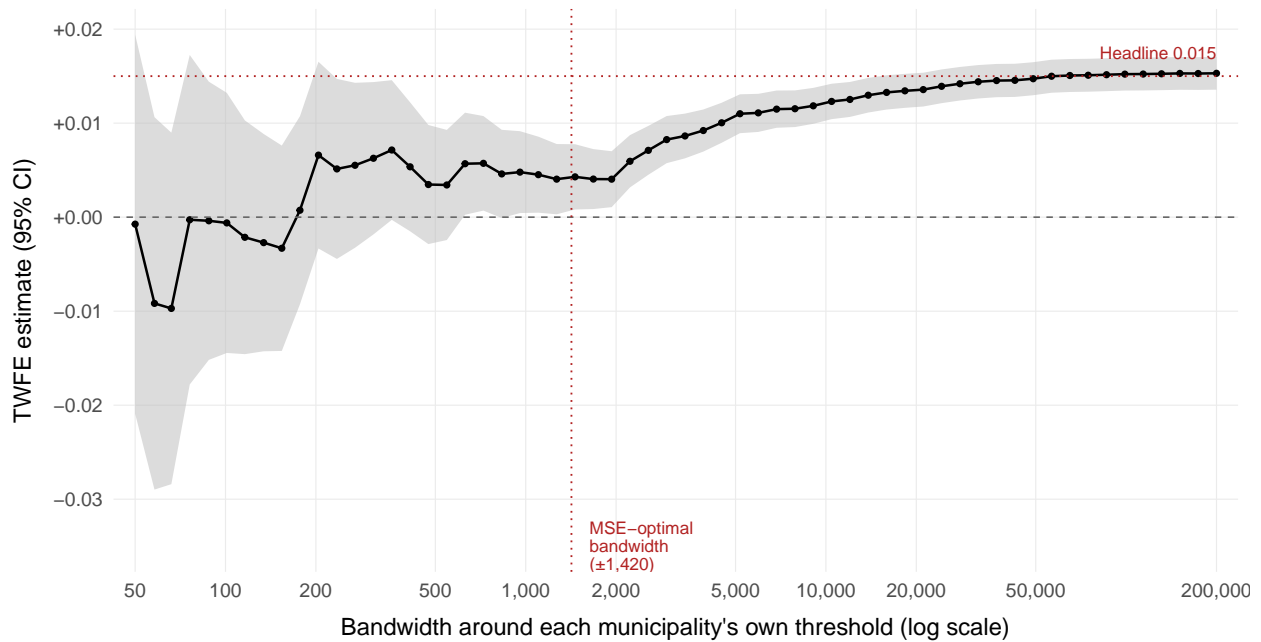


Figure 1: TWFE estimate of the reform effect as a function of the population bandwidth around each municipality's own threshold (5,000 for non-mountain, 3,000 for mountain), on a log scale. Each point is a separate TWFE regression of far-right vote share on treated \times post with municipality and year fixed effects and standard errors clustered by municipality. Shaded region is the 95% confidence interval at each bandwidth. The horizontal red dotted line marks the full-sample (headline) estimate of 0.015; the vertical red dotted line marks the MSE-optimal RD bandwidth ($\pm 1,420$ inhabitants) selected by `rdrobust`. At 24 of the 27 narrow-band bandwidths below $\pm 2,000$, the upper 95% CI lies strictly below the headline.

2.2 The estimate appears before the reform and keeps growing after it

Table 1 applies the paper’s specification using alternative break years, an extended sample window, and a post-reform placebo. Pre-reform placebos are significant: using 2006 as the break year yields 0.009 ($p < 0.001$), and using 2001 yields 0.005 ($p < 0.001$). The “treatment effect” was present before the reform existed. The estimates grow in roughly equal increments across break years (0.005, 0.009, 0.015, 0.020), consistent with a steadily steepening trend rather than a discontinuity at 2010. The authors themselves flag this problem, noting that “a strict causal interpretation may not be warranted given that the pre-trends of the control and treatment groups are different” (p. 1590), and turn to MTWFE and SDID as preferred estimators. Extending the panel to include the 2022 general election increases the estimate by one-third (from 0.015 to 0.020).

Table 1: Temporal placebo and extended-sample estimates

Reform date	Elections used	Estimate	<i>N</i>
2001 (placebo)	2001, 2006, 2008	0.005 (0.001)	23,892
2006 (placebo)	2001, 2006, 2008	0.009 (0.002)	23,892
2010 (actual)	2001–2018	0.015 (0.001)	39,820
2010 (actual)	2001–2022	0.020 (0.001)	47,326
2014 (placebo)	2013, 2018, 2022	0.020 (0.002)	23,434

Notes: Each row uses the paper’s specification (municipality and year FE, clustered SEs) and dual-threshold treatment. The first two rows use only pre-reform elections, with “post” redefined as elections after the alternative break year. The fourth extends the panel through 2022. The last uses only post-reform elections with break at 2014.

2.3 Placebo thresholds produce estimates of similar magnitude

If the reform at 5,000 caused the far-right shift, the estimate should peak at that cutoff—a local maximum at 5,000, with smaller estimates at nearby thresholds where the placebo

dummy splits reform-treated municipalities across both sides. Figure 2 shows nothing of the kind: across twenty thresholds from 1,000 to 50,000, the estimate at 5,000 does not stand out from its neighbors, and the TWFE and MTWFE estimates grow monotonically with the cutoff.

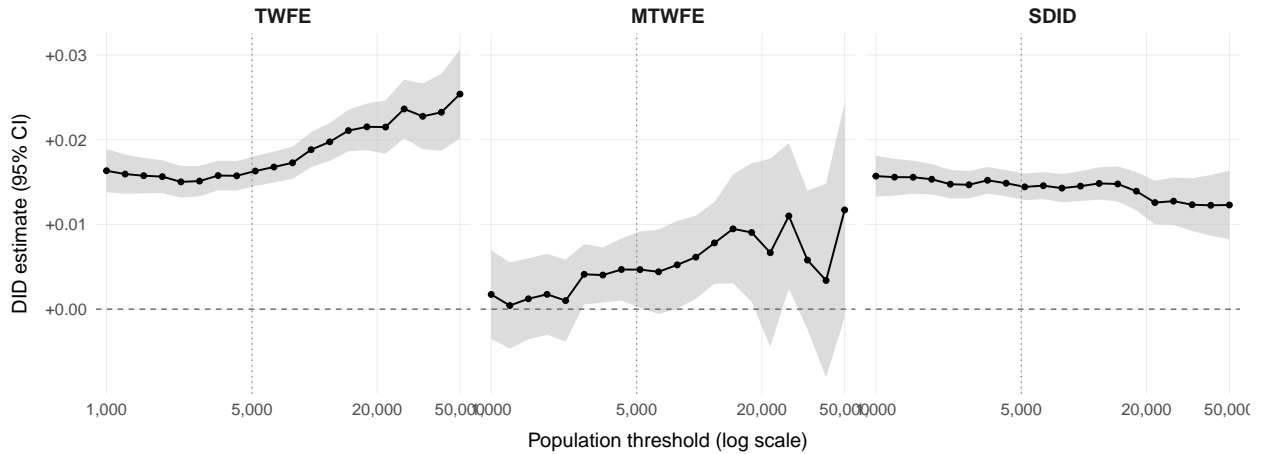


Figure 2: Placebo threshold estimates across all three estimators, on a log-scale population grid. Each point is a separate estimation at the indicated cutoff. TWFE: municipality and year FE, clustered SEs. MTWFE: Mahalanobis matching with replacement on pre-treatment covariates. SDID: [Arkhangelsky et al. \(2021\)](#), jackknife SEs. Shaded regions are 95% confidence intervals. The dotted vertical line marks the actual reform threshold (5,000); the treatment definition uses the dual threshold (5,000/3,000) only at that cutoff.

The pattern survives an even stronger placebo. Dropping every municipality below 5,000 leaves 2,420 untreated municipalities; running the same sweep on this subsample still produces a significant estimate ($p < 0.001$) at every threshold from 7,000 to 30,000. The estimator cannot be picking up a reform effect in a sample that contains no reform-treated units.

2.4 The same pattern appears in France, where no reform exists

If the estimator is picking up a genuine Italian policy effect, applying it in a country with no comparable reform should produce null results. I test this using French presidential elections (2002–2022, 34,710 communes; details in Appendix A, with an institutional comparison of the three relevant French laws in A.1).

At a 5,000-inhabitant cutoff, the French TWFE yields 0.039 ($p < 0.001$)—larger than the Italian estimate. The estimate attenuates in narrow bands around any cutoff and grows with the cutoff itself, just as in Italy (Figures 3 and 4).⁶

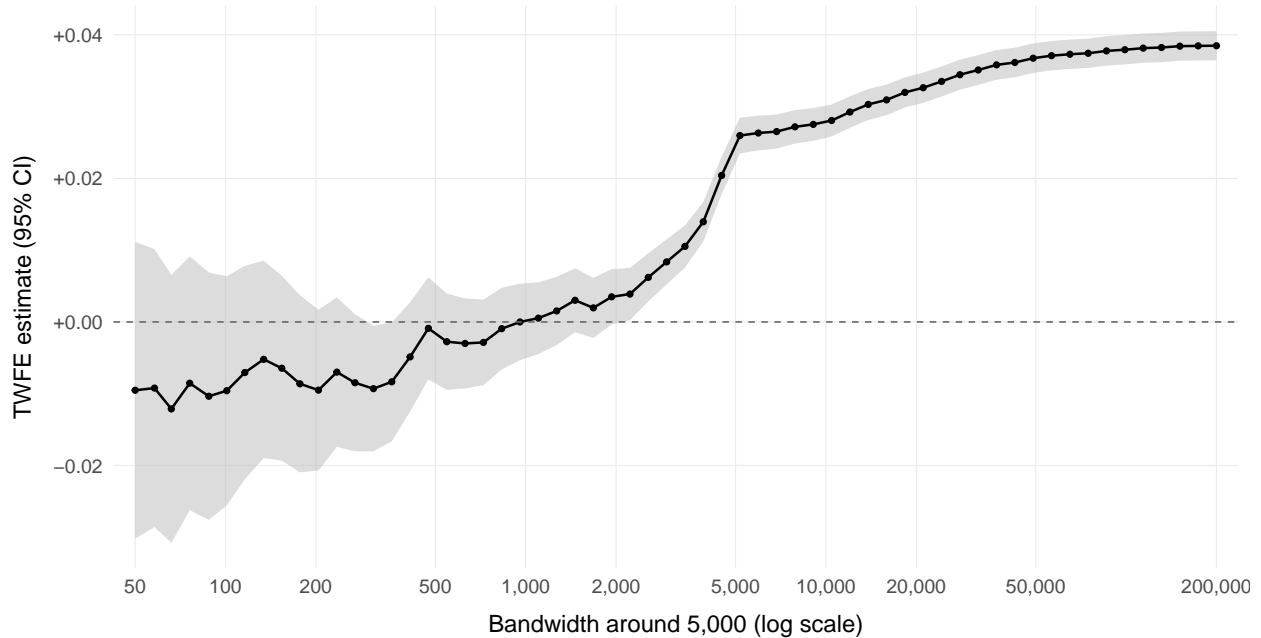


Figure 3: France: TWFE estimate as a function of population bandwidth around the 5,000-inhabitant placebo cutoff, on a log scale. Each point is a separate TWFE regression of far-right vote share on treated \times post with commune and year fixed effects and standard errors clustered by commune. Pre = {2002, 2007}, post = {2012, 2017, 2022}. Shaded region is the 95% confidence interval.

⁶France passed a major local-government reform in December 2010 (Loi RCT) that mandated membership in an *établissement public de coopération intercommunale* (EPCI, an inter-municipal cooperation body grouping neighboring communes) for all communes, with a minimum aggregate population of 5,000 for the EPCI itself. The 5,000 threshold applies to the EPCI as a whole, not the individual commune, so it does not create a sharp treatment at commune-population = 5,000. The sign of any plausible commune-size mechanism also runs the wrong way: smaller communes are *more* likely to be effectively treated by the mandate, since the 5,000-aggregate requirement binds them harder (a 200-inhabitant commune cannot satisfy it alone, a 30,000-inhabitant commune already does). This would predict estimates that grow as the cutoff is lowered—but the data show the opposite (estimates grow *with* the cutoff) and attenuate to zero in narrow bands around every cutoff.

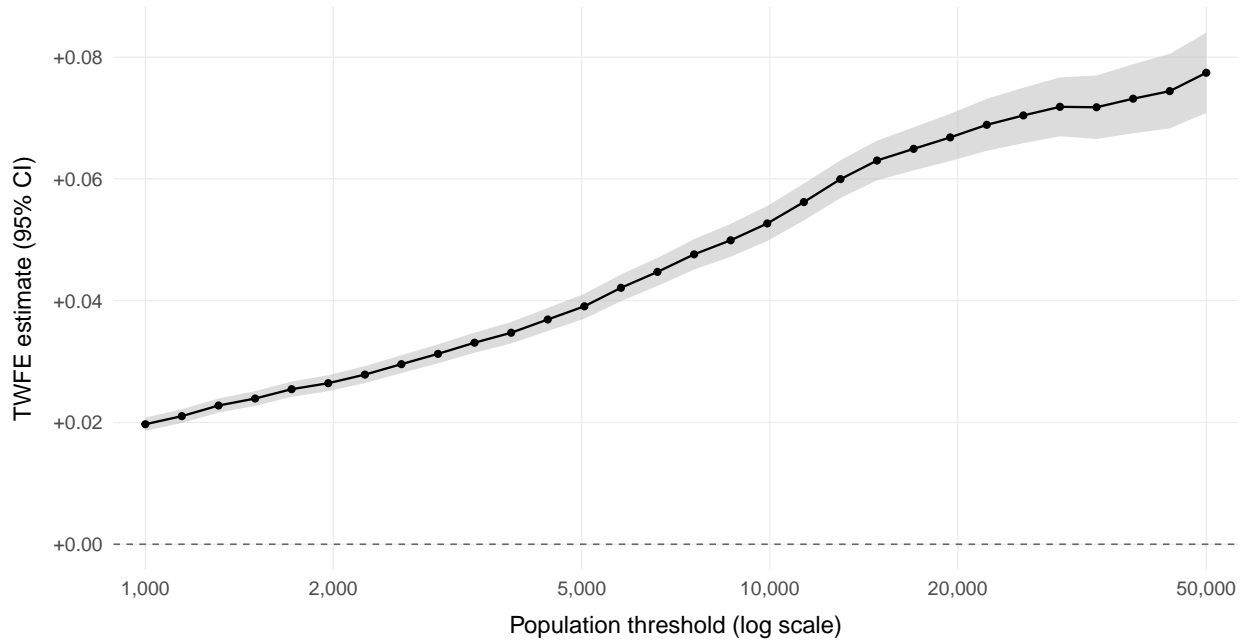


Figure 4: France: TWFE estimate at placebo population thresholds from 1,000 to 50,000, on a log scale. Each point is a separate estimation at the indicated cutoff. Same specification, pre/post window, and clustering as Figure 3. Shaded region is the 95% confidence interval.

3 A Unifying Explanation

These four anomalies are consistent with a single explanation: far-right support has been growing faster in small municipalities than in large ones, and because treatment is assigned by population, the DID cannot distinguish a reform effect from this differential trend. Figures 5 and 6 show this size gradient—the relationship between far-right vote share and municipality population—for Italy and France.

Municipality and year fixed effects absorb permanent place differences and election-wide shocks, but neither absorbs a size gradient that steepens over time. Because treatment is assigned by population, the DID cannot distinguish the reform from this trend; Appendix B shows formally that the spurious component equals the rate at which the gradient steepens times the population gap between “treated” and “control” groups, so the bias grows mechanically with the cutoff, and empirically with the passage of time as the gradient continues to steepen.

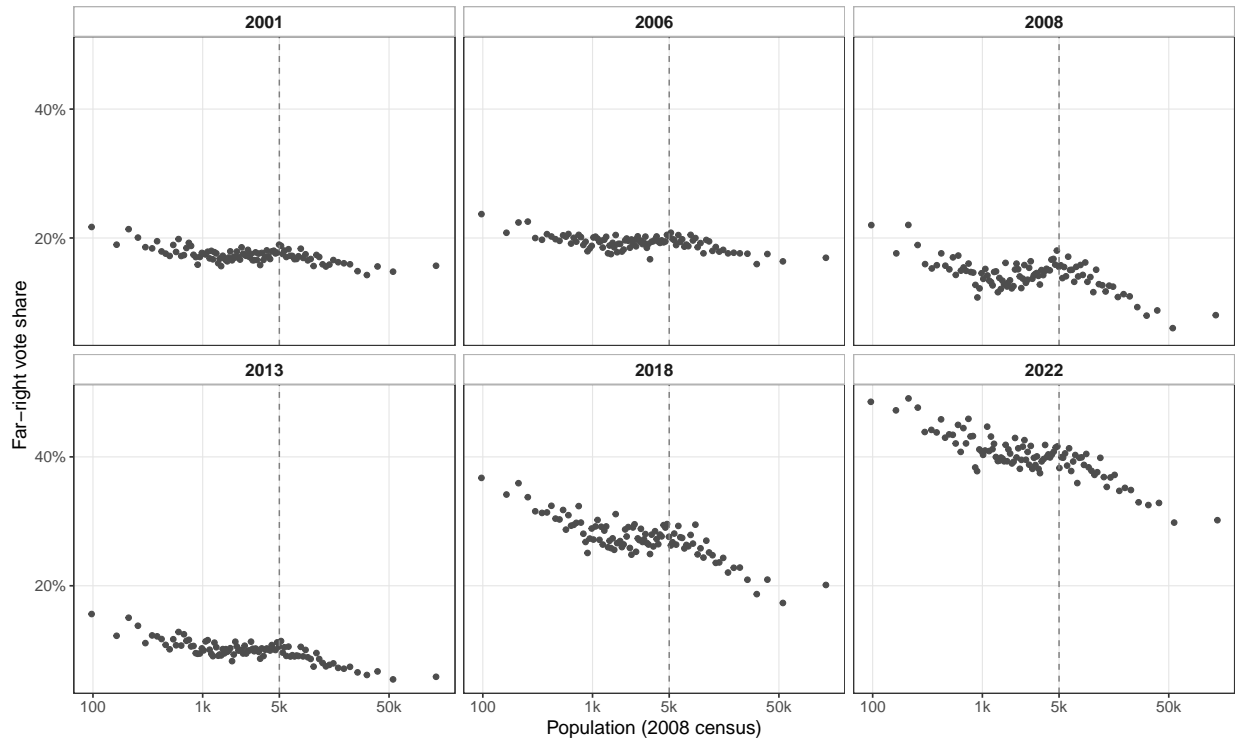


Figure 5: Far-right vote share versus municipality population by election year (Italy). Each point is a population-percentile bin mean. The dashed vertical line marks the 5,000-inhabitant reform threshold.

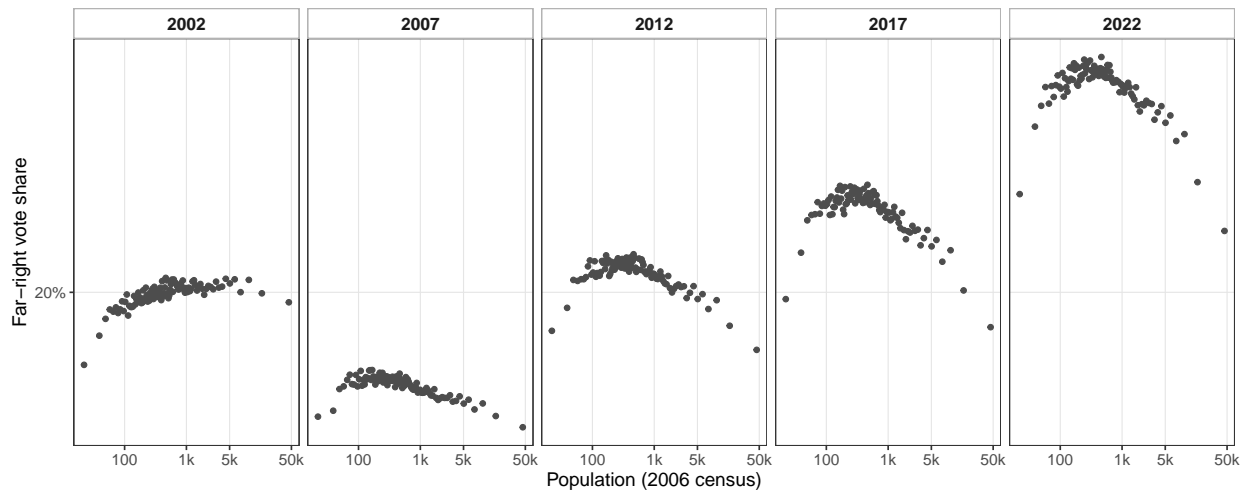


Figure 6: Far-right vote share versus commune population by election year (France, 2002–2022). Each point is a population-percentile bin mean. No French reform comparable to Italy’s 2010 service-consolidation mandate exists.

Figure 7 confirms the formula directly. Across 197 placebo cutoffs the DID estimate is a near-perfect linear function of the log-population gap between control and treated groups ($R^2 = 0.92$). At the actual reform threshold of 5,000, the fitted line predicts a spurious DID of 0.017—within 7% of the observed 0.015. The headline estimate is what the size-gradient mechanism alone would produce.

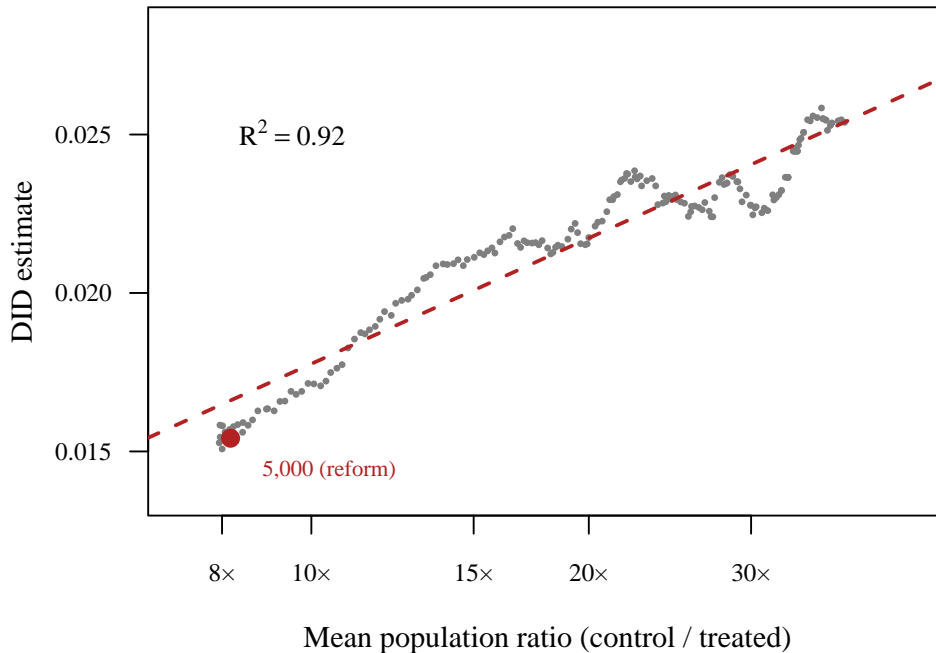


Figure 7: DID estimate as a function of the mean log-population gap between control and treated groups, across 197 placebo cutoffs from 1,000 to 50,000 (in steps of 250 inhabitants). Each grey point is a separate TWFE regression at the indicated cutoff. The red marker is the actual reform threshold (5,000). The dashed line is the OLS fit: $\widehat{DID} = 0.0046 + 0.0057 \times \text{log-pop gap}$, slope SE = 0.0001, $R^2 = 0.92$. The formula predicts a DID of 0.0166 at the 5,000 cutoff (observed: 0.0154).

Matching and synthetic-control approaches do not escape the problem (Appendix B.1). MTWFE (Ho et al. 2007) matches treated municipalities to controls on pre-treatment covariates including population, narrowing the mean population gap from 16,250 to 4,552 inhabitants—but it cannot close it. By construction, every treated unit lies below 5,000 and every control above.

SDID (Arkhangelsky et al. 2021) balances on a different target: pre-period *outcomes*

rather than covariates. It chooses unit and time weights so that the weighted control trajectory matches the treated trajectory before treatment. If the relationship between population and far-right vote share were constant over time, matching pre-period outcomes would also balance population, and SDID would be unbiased. But the relationship is not constant—it steepens—so weights chosen to match the pre-period no longer match the post-period, and balancing population directly is once again ruled out by the cutoff. The bias has two factors: the population gap and the steepening itself. SDID’s unit weights cannot close the population gap—the cutoff forbids it—and empirically close less than 1% of it at every placebo cutoff. Time weights, however, address the other factor: they pick a weighted average of pre-periods whose gradient approximates the post-period’s, shrinking the steepening factor. That is why SDID stays flatter than TWFE across cutoffs without being unbiased: a residual mismatch survives, leaving the +0.014 estimate at the 5,000 cutoff (Appendix B).

3.1 Controlling for the gradient

The simplest test of the two hypotheses is to add a flexible size control to the paper’s TWFE specification and see whether the headline estimate survives. It does not (Table 2): adding $\log(\text{pop}) \times \text{year}$ interactions drops the estimate from +0.015 to -0.000 ($p = 0.76$), and the same control eliminates the estimate at every placebo cutoff from 2,000 to 50,000. It also reverses the matched MTWFE estimate from +0.004 to -0.007 ($p < 0.01$): matching reduces the population gap but cannot close it, so residual size variation still correlates with the post-period trend.

A sharper formulation puts both hypotheses in the same regression: a below-5,000 \times post dummy (the discontinuity) alongside the full set of $\log(\text{pop}) \times \text{year}$ interactions (the gradient). Once the gradient is in the model, the discontinuity dummy is zero ($F = 0.08$, $p = 0.78$)—knowing whether a municipality lies above or below 5,000 adds nothing to what its population and the year already supply. The gradient terms, by contrast, remain

strongly non-zero with the dummy included ($F = 63.5$, $p < 2 \times 10^{-16}$). AIC and BIC both prefer the gradient-only specification.

Table 2: Electoral estimates with and without size-specific time trends

Specification	Estimate
TWFE (paper)	0.015 (0.001)
TWFE + $\log(\text{pop}) \times \text{year}$	0.000 (0.002)
TWFE + $\text{pctile}(\text{pop}) \times \text{year}$	-0.002 (0.002)
MTWFE (paper)	0.004 (0.002)
MTWFE + $\log(\text{pop}) \times \text{year}$	-0.007 (0.002)
MTWFE + $\text{pctile}(\text{pop}) \times \text{year}$	-0.008 (0.002)

Notes: All specifications include municipality and year FE with SEs clustered by municipality. $\text{pctile}(\text{pop})$ is each municipality’s percentile rank in the national population distribution (continuous, 0–1 scale). MTWFE uses Mahalanobis matching on pre-treatment covariates including population (with replacement), which reduces the mean population gap from 16,250 to 4,552.

A natural objection is that $\log(\text{pop}) \times \text{year}$ is bad control: if the reform itself operated through a size-correlated channel—smaller treated municipalities reshaped more by the mandate—then the gradient terms would absorb the true effect along with the confound. The objection requires the gradient to be downstream of the 2010 reform. It is not. The same gradient pattern appears in the pre-reform temporal placebo (Table 1, $p < 0.001$ for both the 2001 and 2006 break years), in France where no comparable reform exists (Section 2.4), and in the never-treated subsample (Section 2.3). A direct triple-interaction test pinpoints the issue: regressing far-right share on the DID treatment, the year-by-year size gradient, and a treated \times post \times $\log(\text{pop})$ interaction yields a coefficient on the triple term of +0.0000 (SE 0.0012, $p = 0.97$). The gradient steepens identically in treated and control communes post-reform; the reform did not differentially steepen it. The bad-control critique would have to explain how a 2010 reform produces variation in samples

that pre-date it, that lie outside it, that exclude every treated unit, and that exhibit no differential steepening among the treated.

4 A direct test of the mechanism

The size-gradient critique above does not directly engage [Cremaschi et al. \(2024\)](#)'s proposed mechanism: that the 2010 reform reduced the local responsiveness of public administration in small municipalities, provoking a far-right backlash. The mechanism requires three things to be true—the reform changed how small municipalities deliver public services, that change degraded the services, and the affected municipalities moved further right than similar non-affected ones. The remainder of this section works through each.

4.1 Was the reform implemented?

Article 14 of [Repubblica Italiana \(2010\)](#) (published 31 May 2010) required municipalities below 5,000 inhabitants (3,000 for designated mountain municipalities) to deliver ten “fundamental functions” jointly with neighbouring municipalities: police, waste, social services, territorial planning, road maintenance, civil registry, tax collection, education, civil defence, and IT. Compliance could take the form of a formal *unione di comuni* or a less binding inter-municipal agreement (*convenzione*). Implementation was statutory rather than voluntary.

In practice, the deadline was postponed at least twelve times between 2013 and 2024, no sanction was ever applied to a non-complying municipality, the Constitutional Court declared the rigid mandate partially unconstitutional ([Corte Costituzionale 2019](#)), and Article 21 of [Repubblica Italiana \(2024\)](#) abolished it. The Conferenza Stato-Città states that “the obligation for associated exercise of all fundamental functions of small municipalities never entered into force” ([Conferenza Stato-Città ed Autonomie Locali 2025](#)); [Di Ielsi et al. \(2022\)](#), writing from inside SOSE, confirm that the mandatory associated exercise “never actually materialized.”

Crevaschi et al. do not engage with any of this. They seem to claim near-universal compliance—“200 (2.51%) of the affected municipalities . . . merged, 1,562 (19.61%) formed a new union, and the rest adopted a convention” (p. 12)—but offer no source for the residual 78% and themselves acknowledge that “systematic data is not available for conventions” (online appendix, p. 5). Section 4.2 examines whether the available data supports the universal-compliance claim.

4.2 Did sub-threshold municipalities use shared service delivery more than above-threshold ones?

The reform allowed three compliance vehicles: municipal mergers (rare), formal *unioni di comuni*, and *convenzioni*—inter-municipal agreements for joint provision (Crevaschi et al. 2024, appendix p. 5). Only *unioni* appear in the Ministry of the Interior’s registry (Ministero dell’Interno 2020); the OpenCivitas database records associated delivery under either vehicle.

For *unioni*, the Ministry registry permits a time-series view of the treatment effect. Figure 8 plots cumulative membership and annual new-formation rates separately for sub- and above-threshold municipalities. Sub-threshold membership rises from 18.6% in 2009 to 42.6% in 2018 ($\Delta = +24.0$ pp). Above-threshold membership rises from 13.3% to 27.0% ($\Delta = +13.7$ pp). The implied DID is +10.2 pp—non-zero, but more than half of the sub-threshold rise would have occurred anyway under the counterfactual of the above-threshold trend.⁷

Membership in an *unione* does not by itself imply joint delivery. Spano and Tagliagambe (2018) examine official cost data for 384 municipalities across 54 *unioni* in three regions, treating disappearance of a municipality’s expenditure on a function—with continued joint delivery—as evidence the function was actually transferred. Only 11 of the 384 (under 3%) met this test for any function in 2013–2015, leading them to conclude that “the transfer

⁷The 2020 registry contains only surviving *unioni*, so pre-2010 *unioni* that dissolved before 2020 are missing, which biases the cumulative series toward more recent formations.

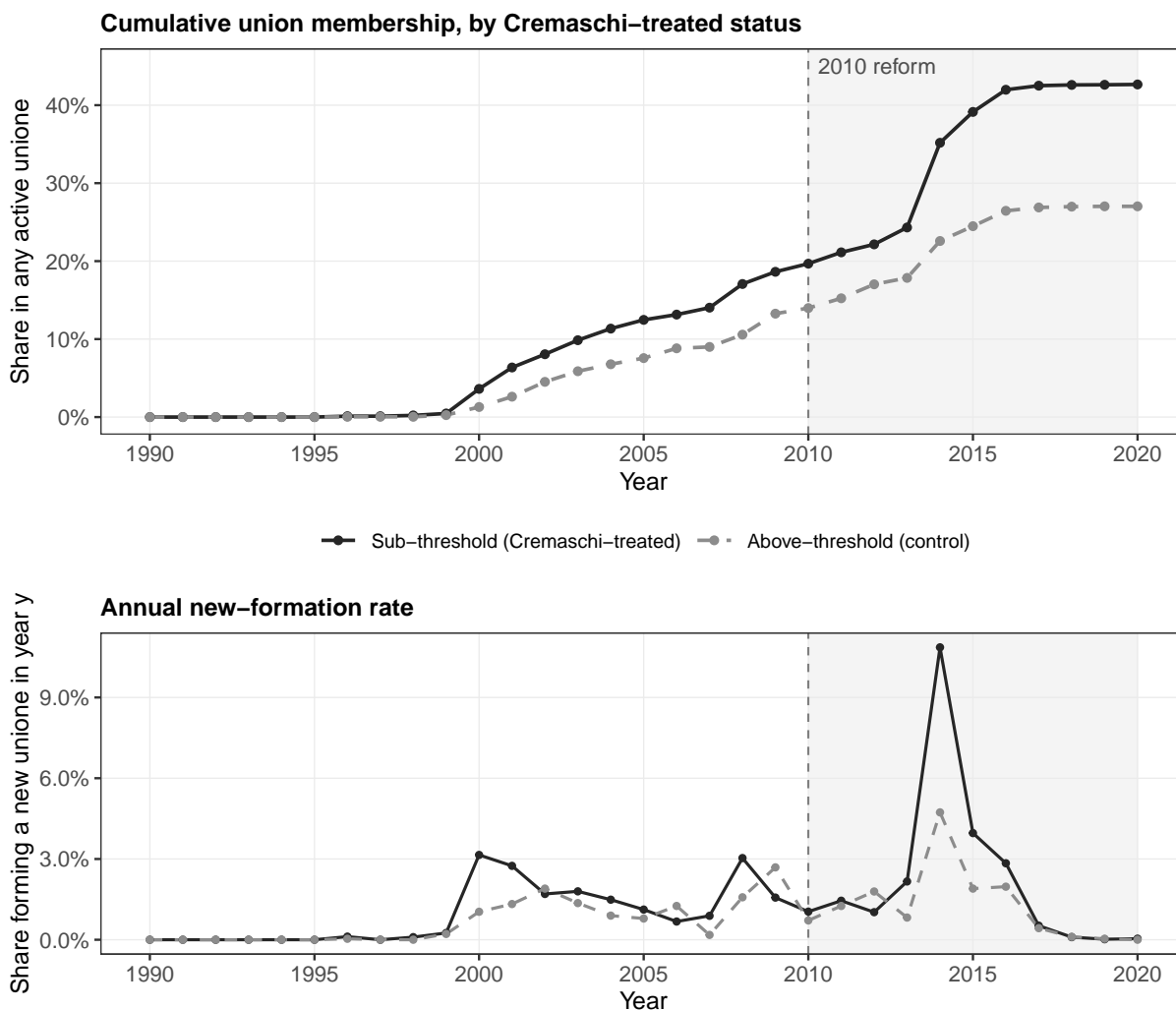


Figure 8: *Unione di comuni* formation over time, sub-threshold (Cremaschi-treated) vs. above-threshold municipalities. Top: cumulative share in any active *unione* as of year y . Bottom: share forming a new *unione* in year y . Source: [Ministero dell’Interno \(2020\)](#).

of functions from Italian municipalities to a *Unione* for a joint service delivery has been irrelevant so far” ([Spano and Tagliagambe 2018](#), p. 9).

The OpenCivitas function-level flags offer a looser, self-reported measure of the same question. The two sources are not directly comparable: [Spano and Tagliagambe \(2018\)](#) apply a strict budgetary test (the municipality’s own expenditure line for a function must disappear), whereas OpenCivitas captures whether the municipality reports delivering a function in associated form—a much weaker bar that does not require the function to have

been formally transferred. The OpenCivitas flags are drawn from the annual *Fabbisogni Standard* questionnaires that municipalities are required to file with SOSE, the Italian government agency that produces standard-needs estimates for municipal financing. For each fundamental function, the questionnaire asks whether the function is delivered in associated form with neighbouring municipalities, capturing both *unione* and *convenzione* arrangements. The variables we use are a 2015 snapshot, so they describe the post-reform state only and cannot identify a pre/post change in joint delivery. Figure 9 plots the share of municipalities with associated delivery against population for nine compliance measures: post-2010 *unione* membership and each of the eight Article 14 fundamental functions delivered in associated form in 2015.

For sub-threshold municipalities, associated delivery reaches roughly 70% for social services, 55% for garbage and education, 35% for local police, 25% for post-2010 *unione* membership and territorial planning, and only about 10% for civil registry, road maintenance, and tax collection.

More strikingly, the bin means trace a continuous pattern through the legal threshold in every panel: there is no visible jump in compliance behaviour at $x = 0$. Where there is a gradient, it runs through the cutoff continuously; where compliance is nearly flat, it remains so on both sides.⁸ Service-sharing follows a continuous gradient through the size distribution, not a sharp policy discontinuity at the legal threshold.

4.3 Did it worsen public service delivery?

The mechanism's second link requires that the reform, where it was applied, degraded service delivery. [Cremaschi et al. \(2024\)](#) report the reform's effect on three services—local police, public registry, garbage collection—noting that for the others “missing data and coding issues constrain the analysis” and that public-transport coverage is missing for

⁸Formal local-linear RD tests at the threshold confirm the visual reading: eight of the nine compliance indicators yield discontinuity estimates indistinguishable from zero, and the lone $p < 0.05$ result (local police, -0.110) does not survive Bonferroni for nine tests. See Appendix C.

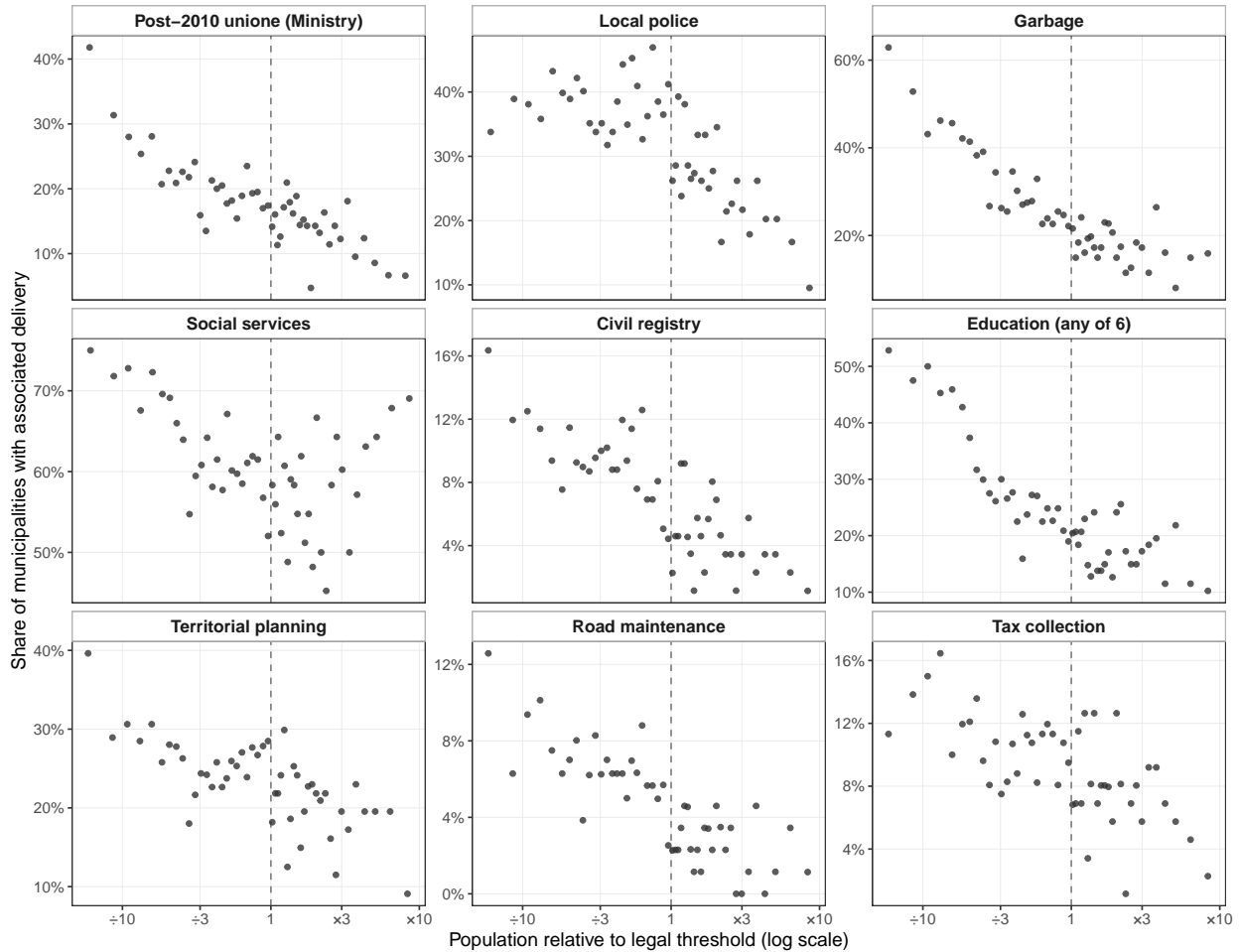


Figure 9: Share of municipalities with associated service delivery by distance to the legal population threshold. Points are percentile-bin means. The horizontal axis is $\log(\text{pop} / \text{threshold})$, pooling 5,000-inhabitant (non-mountain) and 3,000-inhabitant (mountain) thresholds at $x = 0$.

95% of municipalities (online appendix, p. 15). Their main-text Table 3 reports MTWFE estimates of -0.196 ($\text{SE} = 0.047$) for police, -0.157 ($\text{SE} = 0.080$) for registry, and -0.076 ($\text{SE} = 0.036$) for garbage—all in standard-deviation units of the “Services Against Standard Demand” outcome, all statistically significant. The paper reads these as evidence that the reform reduced public-service provision in affected municipalities. Note that Cremaschi et al.’s Table 3 compares 2013 to 2009, but the bulk of post-reform *unione* formation among sub-threshold municipalities did not occur until 2014 (Figure 8). The service-quality decline they detect predates most of the compliance through which their mechanism is supposed

to operate.

These estimates are themselves vulnerable to the size-gradient confound documented in Section 2 for the political outcome. I reproduce [Cremaschi et al. \(2024\)](#)'s Table 3 estimates exactly using their replication data and matching specification, and then plot the muni-level change in their standardized service-output measure between 2009 and 2013 against distance from the legal threshold (Figure 10). The change is a smooth function of population in all three services. Sub-threshold municipalities did experience worse trajectories than above-threshold municipalities for police and registry capacity—but the divergence is operating continuously across the size distribution, with no jump at the legal cutoff. RD estimates of the threshold discontinuity in the change are -0.078 (SE = 0.077) for police, -0.018 (SE = 0.096) for registry, and $+0.091$ (SE = 0.079) for garbage—all statistically indistinguishable from zero.

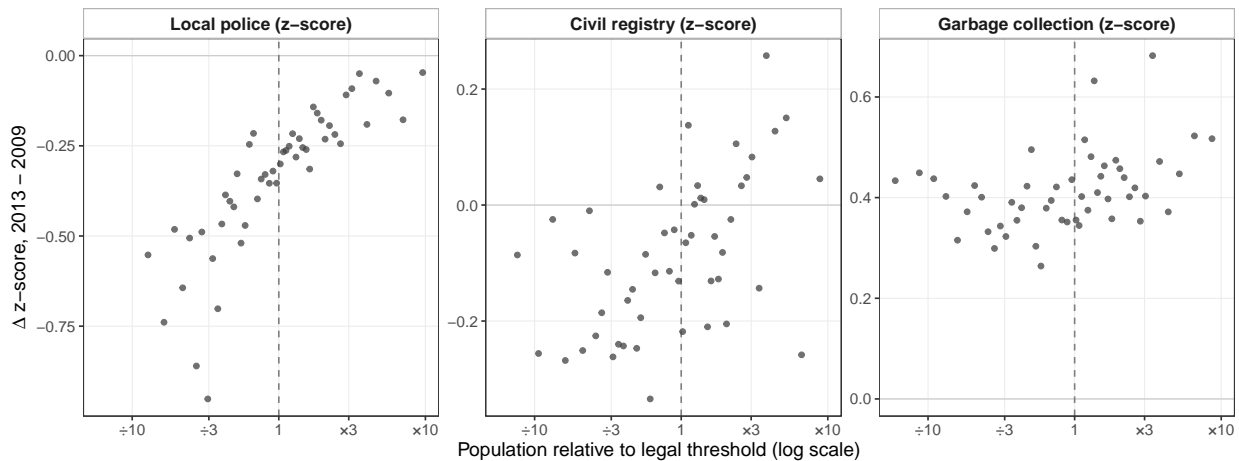


Figure 10: Muni-level change in standardized “Services Against Standard Demand” (2013 minus 2009) by distance to the legal threshold. Each point is a percentile-bin mean. Axis as in Figure 9. Outcome, sample, and standardization mirror [Cremaschi et al. \(2024\)](#)'s Table 3.

It is not obvious that joint service delivery would degrade services in any case—the stated goal of the reform was to improve them by eliminating duplication, and the academic literature on Italian inter-municipal cooperation finds no consistent evidence of service-quality decline ([Ferraresi et al. 2018](#); [Gori et al. 2025](#); [Luca and Modrego 2021](#)).

4.4 From intent-to-treat to treatment-on-the-treated

Sections 4.1–4.3 establish that the 2010 reform was at most a partial treatment. Cremaschi et al.’s 1.5-percentage-point effect is therefore an intent-to-treat estimate; under partial compliance the treatment-on-the-treated parameter should be substantially larger (Angrist et al. 1996). To test this, I match compliers to non-compliers within the Cremaschi-treated stratum—using their Mahalanobis specification on their seven pre-reform covariates—and run TWFE on the matched sample. I do this for three aggregate compliance measures (post-2010 *unione* membership and two OpenCivitas indicators) and for each of the eight Article 14 functions separately.

The matched ATTs are essentially zero (Table 3): all three aggregate estimates sit within a thousandth of zero, and six of the eight single-function estimates do too. Under $ITT = \pi \cdot ATT$, partial compliance implies an ATT *larger* than the 1.5-pp ITT; the matched estimates instead run smaller.

Table 3: Pooled compliance ATT estimates across aggregate and single-function measures

Compliance measure	N treated	MTWFE
<i>Aggregate measures</i>		
Post-2010 <i>unione</i> (Ministry)	1,193	0.000 (0.002)
8-of-8 functions \geq 4 (OpenCivitas)	891	0.001 (0.002)
6-of-8 functions \geq 4 (OpenCivitas)	1,442	0.000 (0.002)
<i>Individual Article 14 functions (OpenCivitas)</i>		
Social services	2,427	0.001 (0.002)
Local police	1,450	0.006 (0.002)
Waste collection	1,418	-0.008 (0.002)
Education (any sub-function)	1,281	0.000 (0.002)
Territorial planning	1,097	0.002 (0.002)
Tax collection	455	-0.001 (0.003)
Civil registry	396	0.000 (0.003)
Road maintenance	278	0.001 (0.003)

Notes: Pooled (Cremaschi-treated) stratum, panel 2001–2018. The post-2010 *unione* flag indicates membership in a consortium constituted after 31 May 2010 (Ministry registry). The OpenCivitas aggregate measures dichotomise the count of Article 14 fundamental functions shared in 2015 at 4-of-8 (8-function) or 4-of-6 (excluding tax and registry). Single-function rows use the OpenCivitas 2015 binary indicator for that function; education pools six sub-functions into “shares at least one.” MTWFE: Mahalanobis matching with replacement on the seven baseline covariates Cremaschi et al. use, then TWFE on the matched sample (Ho et al. 2007), with municipality and year fixed effects and clustered SEs. Mean absolute standardised differences across the seven covariates fall from 0.17–0.31 pre-match to 0.017–0.037 post-match.

5 Conclusion

The 2010 reform did not detectably increase far-right vote share. The headline estimate captures a pre-existing size gradient, not a policy effect: the same pattern appears before the reform, at placebo population thresholds, and in France where no comparable reform exists. The proposed mechanism fails on its own terms too: the mandate was never systematically enforced and was abolished in 2024, compliance was partial at best, matched compliers within the treated stratum show no aggregate ATT and no consistent single-function ATT, and the published service-quality effects shrink to a fraction of their size and turn null under RD at the threshold. To sustain a causal interpretation, one would need to explain why the result vanishes with flexible size controls, why the estimate attenuates to zero in narrow bands around the threshold, and why actual compliers behave like similar non-compliers.

A France placebo: data, design, and institutional context

A.1 Why French commune-level thresholds at 5,000 are clean placebos

The French placebo's validity rests on the absence of any French reform that creates a treatment-control contrast at the 5,000-inhabitant commune-level cutoff during the 2002–2022 window. Three pieces of legislation could in principle pose such a contrast, and all three operate at the *établissement public de coopération intercommunale* (EPCI, intercommunal grouping) level rather than the commune level.

The *loi Chevènement* (*Loi n° 99-586* du 12 juillet 1999) restructured intercommunalité, introducing the *communauté d'agglomération* (minimum 50,000 inhabitants in the grouping, anchored on at least one commune over 15,000) and the *communauté urbaine* (minimum 500,000 in the grouping). Its *communauté de communes* category—the form covering most rural areas—imposed no minimum population. None of these thresholds attaches a differential obligation to communes crossing 5,000 inhabitants individually.

The *loi de réforme des collectivités territoriales* (*Loi n° 2010-1563* du 16 décembre 2010, “loi RCT”) made EPCI membership universal: every commune must belong to an EPCI of at least 5,000 inhabitants. The 5,000 floor here applies to the EPCI's *total* population summed across its constituent communes, not to any single commune. A commune of 1,000 inhabitants belonging to an EPCI totalling 8,000 faces the same legal regime as a commune of 6,000 inhabitants in the same EPCI; the running variable for the placebo (own-commune population) does not align with this French rule.

The *loi NOTRe* (*Loi n° 2015-991* du 7 août 2015) raised the EPCI floor from 5,000 to 15,000 inhabitants, with derogations to a 5,000 floor for mountain communes, island communes, and EPCI in zones of low population density. These derogations preserve the 5,000 figure but again at the EPCI level, never at the commune level. The number of EPCI fell from 2,062 to 1,266 between 2016 and 2017 in consequence—but again, this changes which EPCI a commune belongs to, not what the commune itself is obliged to provide.

None of the three laws creates a sharp obligation that activates when a commune individually crosses 5,000 inhabitants. The French DID running variable (own-commune *populations légales* 2006) is therefore orthogonal to the legal regime under any of these laws, and the placebo can be interpreted as a clean test of the design’s behaviour in the absence of any reform.

A.2 Data and design

I use first-round presidential election results for 2002, 2007, 2012, 2017, and 2022, obtained from the data.gouv.fr aggregated elections dataset (Ministère de l’Intérieur). Polling-station results are aggregated to the commune level. The outcome is far-right vote share: votes for Le Pen (FN/RN) and Mégret (MNR, 2002) and Zemmour (Reconquête, 2022) divided by valid votes. Commune codes are harmonized to the 2026 Code Officiel Géographique (current edition) using the INSEE commune movements table. The panel covers 34,710 metropolitan communes observed in all five elections.

Population is fixed at the 2006 census (*populations légales*), before the post period. The design mirrors the Italian specification: treatment is $\mathbf{1}(\text{pop}_{2006} < \text{cutoff})$, with $\text{pre} = \{2002, 2007\}$ and $\text{post} = \{2012, 2017, 2022\}$. As established in Section A.1 above, no French public-service reform creates a commune-level discontinuity at any population threshold during this period.

Every anomaly documented for Italy replicates in France. Tables 4 and 5 of the main paper show the narrow-band attenuation and placebo-threshold results. Table A.1 below shows that the pre-period placebo (2002 vs. 2007) is significant, and a late-break placebo ($\text{pre} = 2002\text{--}2012$, $\text{post} = 2017\text{--}2022$) produces an even larger estimate. Figure A.1 shows the DID estimate is proportional to the log-population gap across thresholds ($R^2 = 0.96$).

Table A.1: Temporal placebo estimates at 5,000 (France)

Specification	Estimate
Pre-period only (2002 vs. 2007)	0.025 (0.001)
Full sample (pre = 2002–2007, post = 2012–2022)	0.039 (0.001)
Late break (pre = 2002–2012, post = 2017–2022)	0.037 (0.001)

Notes: TWFE with commune and year FE, SEs clustered by commune. Treatment is $1(\text{pop} < 5,000)$. The first row uses only pre-period elections. The last row shifts the break point forward.

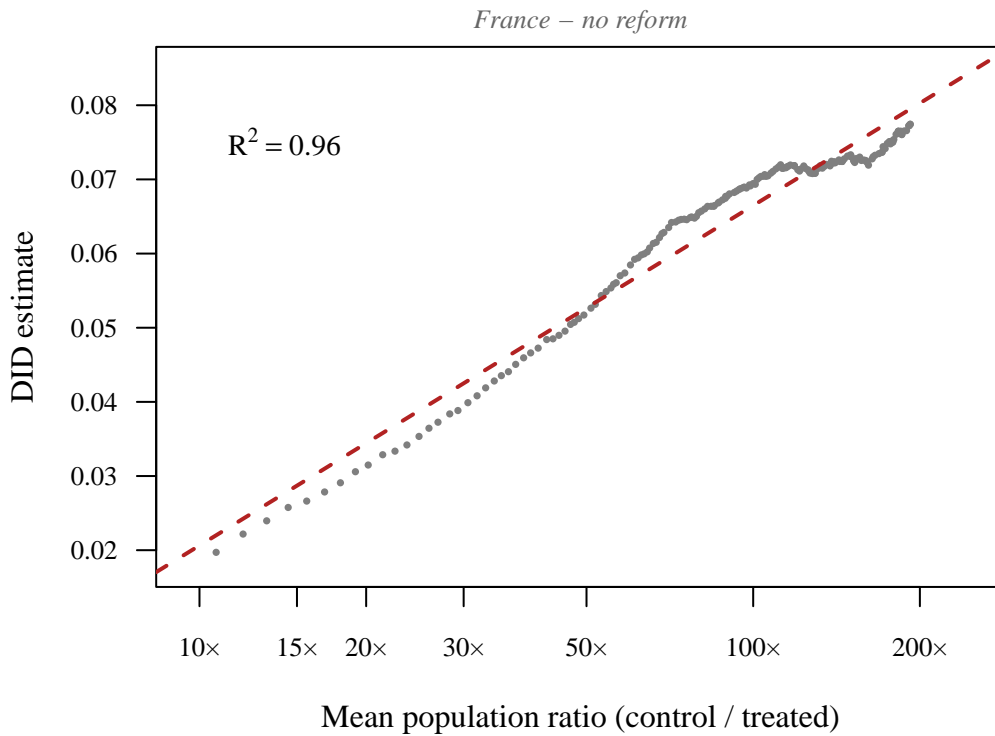


Figure A.1: DID estimate versus mean log-population gap (control – treated) at 197 placebo thresholds in France (2002–2022). Each point is a separate TWFE regression. The dashed line is the OLS fit ($R^2 = 0.96$).

B Formal decomposition of the DID coefficient

Consider the following data-generating process for far-right vote share:

$$Y_{it} = \alpha_i + \gamma_t + \delta_t \cdot f(\text{pop}_i) + \varepsilon_{it}, \quad (\text{A.1})$$

where Y_{it} is far-right vote share in municipality i in election t , α_i is a municipality fixed effect, γ_t is a year fixed effect, pop_i is municipality population (fixed at the pre-reform census), $f(\cdot)$ is a monotone function of population, and δ_t is the slope of the size gradient in election t . The decomposition below does not depend on the choice of f ; I use $f = \log$ throughout as a convenient parametric choice, and Table 6 of the main paper shows that a nonparametric alternative (population percentile rank \times year) yields the same conclusions.

The key feature is that the size gradient changes over time: Figure 1 of the main paper shows the slope steepening from near-zero in 2001 to sharply negative by 2018.

The paper estimates a standard DID specification that omits the size-gradient term:

$$Y_{it} = \alpha_i + \gamma_t + \beta \cdot T_i \times \text{Post}_t + \varepsilon_{it}, \quad (\text{A.2})$$

where $T_i = \mathbf{1}(\text{pop}_i < c)$ equals one for municipalities below the population cutoff c and Post_t equals one for elections after the reform. The coefficient β is the estimated treatment effect.

In a two-period setup with one pre-reform and one post-reform election, the DID estimand is:

$$\beta^{\text{DID}} = (\bar{Y}_{T=1,\text{post}} - \bar{Y}_{T=1,\text{pre}}) - (\bar{Y}_{T=0,\text{post}} - \bar{Y}_{T=0,\text{pre}}). \quad (\text{A.3})$$

Under (A.1), the municipality fixed effects α_i cancel in the first differences (pre vs. post within each group), and the year fixed effects γ_t cancel in the second difference (treated vs.

control):

$$\beta^{\text{DID}} = \underbrace{(\delta_{\text{post}} - \delta_{\text{pre}})}_{\text{gradient change}} \cdot \underbrace{(\overline{f(\text{pop})}_{T=1} - \overline{f(\text{pop})}_{T=0})}_{\text{size gap}}. \quad (\text{A.4})$$

The first term is how much the size gradient steepened between the pre- and post-reform periods. The second is the difference in mean $f(\text{pop})$ between treated municipalities (below the cutoff, smaller) and control municipalities (above it, larger). For any monotonically increasing f , treated municipalities have lower $f(\text{pop})$ than controls, so the size gap is negative. Whenever the gradient steepens ($\delta_{\text{post}} < \delta_{\text{pre}}$), both factors are negative, and their product yields a positive β^{DID} —even without any reform effect. The two-period case illustrates the mechanism; in the multi-period TWFE, the estimand is a variance-weighted average of period-specific comparisons (Goodman-Bacon 2021), each contaminated by the same gradient term.

This decomposition generates several testable predictions, none of which depend on the choice of f :

1. *Placebo thresholds should produce nonzero estimates*, because different cutoffs create different size gaps. Higher thresholds create larger gaps and should produce larger estimates.
2. *Estimates should attenuate in narrow bands*, because municipalities just above and just below any cutoff are similar in size, shrinking the size gap toward zero.
3. *Adding $f(\text{pop}) \times \text{year}$ controls should eliminate the spurious component*, because those controls directly absorb $\delta_t \cdot f(\text{pop}_i)$.
4. *The DID estimate should be proportional to the size gap across thresholds*. A meta-regression of the DID estimate on the log-population gap confirms this, with $R^2 = 0.92$ in Italy and $R^2 = 0.96$ in France.

The meta-regression supporting prediction 4 fits 197 placebo cutoffs from 1,000 to 50,000 in steps of 250:

$$\widehat{\text{DID}} = \underset{(0.0004)}{0.0046} + \underset{(0.0001)}{0.0057} \times \text{log-pop gap}, \quad R^2 = 0.92 \quad (\text{A.5})$$

where the figures in parentheses are heteroskedasticity-consistent standard errors. At the actual reform threshold (5,000), the mean log-pop gap is 2.10 (control communes are $8.2\times$ larger on average), so the fitted line predicts a spurious DID of $0.0046 + 0.0057 \times 2.10 = 0.0166$ against an observed Cremaschi-style estimate of 0.0154—a 7% discrepancy. The decomposition therefore not only predicts the qualitative pattern of the placebo sweep but quantitatively reproduces the headline number from the size-gradient mechanism alone. The corresponding French scatter (Figure A.1) yields $R^2 = 0.96$.

C Formal RD tests of compliance at the threshold

Section 4.2 of the main paper presents Figure 7 (fig:compliance_gradient), showing that compliance with the joint-service-delivery mandate is a continuous function of population, with no visible jump at the 5,000-inhabitant threshold (3,000 for mountain communes). Table A.2 reports formal local-linear regression-discontinuity estimates at the cutoff for each of the nine compliance measures shown in that figure: Ministry-registry post-2010 *unione* membership and the eight Article 14 fundamental-function associated-delivery indicators (OpenCivitas 2015). Each is estimated via `rdrobust` with degree-1 polynomial, triangular kernel, and MSE-optimal bandwidth.

Eight of the nine point estimates fail to reject the null of no discontinuity at the conventional 5% level. The one significant estimate at $p < 0.05$ is local police (-0.110 , robust $p = 0.026$); this estimate does not survive a Bonferroni correction for nine tests ($|t|$ threshold ≈ 2.77 , while $|t|$ here is ≈ 2.39). The formal tests confirm the visual reading of Figure 7: compliance behaviour does not exhibit a discontinuity at the legal cutoff.

Table A.2: Local-linear RD estimates of the discontinuity in compliance at the 5,000-inhabitant cutoff

Compliance measure	Estimate	95% CI	p (robust)	MSE-bw (log-pop)
Post-2010 unione (Ministry)	-0.037	[-0.097, +0.023]	0.236	0.58
Local police	-0.110	[-0.200, -0.019]	0.026	0.54
Garbage	-0.034	[-0.098, +0.029]	0.361	0.81
Social services	+0.047	[-0.050, +0.144]	0.325	0.58
Civil registry	+0.008	[-0.029, +0.046]	0.602	0.61
Education (any of 6)	+0.015	[-0.051, +0.082]	0.505	0.71
Territorial planning	-0.066	[-0.143, +0.011]	0.164	0.62
Road maintenance	-0.004	[-0.035, +0.026]	0.996	0.62
Tax collection	-0.016	[-0.067, +0.035]	0.553	0.60

Notes: Each row is a separate `rdrobust` regression of a 0/1 compliance indicator on the running variable $\log(\text{pop}/\text{threshold})$, with degree-1 local polynomial, triangular kernel, and MSE-optimal bandwidth. The Ministry registry indicator (post-2010 *unione* membership) and the eight OpenCivitas associated-delivery indicators are reported. The conventional point estimate is shown with its 95% confidence interval; the p -value is the robust bias-corrected p from [Calonico et al. \(2014\)](#).

C.1 Extension to reweighted estimators

The same logic applies to any DID estimator that uses weighted group means. Let w_i^T and w_i^C denote nonnegative weights on treated and control units, each summing to one. Substituting (A.1) into the weighted DID and canceling fixed effects, the spurious component under the null of no treatment effect (superscript 0) is:

$$\beta^0(W) = (\delta_{\text{post}} - \delta_{\text{pre}}) \cdot \left[\sum_{i:T_i=1} w_i^T f(\text{pop}_i) - \sum_{i:T_i=0} w_i^C f(\text{pop}_i) \right]. \quad (\text{A.6})$$

This is the weighted analogue of (A.4). Reweighting changes only the second term—the weighted size gap—but does not remove the omitted-variable problem.

Matching (MTWFE). Matching is a special case of (A.6) in which the control weights w_i^C are chosen to balance pre-treatment covariates. Matching can shrink the weighted size gap, but it cannot eliminate it (absent model-based extrapolation) when treatment is defined by a population cutoff: if $T_i = \mathbf{1}(\text{pop}_i < c)$, all treated units lie below c and all controls above

it, so the weighted treated mean of $f(\text{pop}_i)$ remains strictly below the weighted control mean for any weights.

Empirically, Mahalanobis matching narrows the mean population gap from 16,250 to 4,552 (Table 6 of the main paper). The matched estimate is smaller than TWFE (0.004 vs. 0.015) but remains significant; adding $\log(\text{pop}) \times \text{year}$ controls reverses it to -0.007 ($p < 0.01$).

Synthetic DID (SDID). SDID chooses unit weights ω_i and time weights λ_t to reproduce the treated group’s pre-treatment outcome path. Under (A.1), the spurious component is:

$$\beta_{\text{SDID}}^0 = \left(\bar{\delta}_{\text{post}} - \sum_{t \in \text{pre}} \lambda_t \delta_t \right) \cdot \left(\overline{f(\text{pop})}_{T=1} - \sum_{i: T_i=0} \omega_i f(\text{pop}_i) \right), \quad (\text{A.7})$$

where $\bar{\delta}_{\text{post}}$ is the average post-treatment gradient. A constant gradient ($\delta_t = \delta$) zeros the first term—but a constant gradient generates no DID bias in the first place, since $\delta \cdot f(\text{pop}_i)$ is absorbed by municipality fixed effects. SDID eliminates the bias exactly when the bias is already zero.

When the gradient steepens, pre-period balance does not guarantee post-period balance. The weights that matched $\delta_{\text{pre}} \cdot f(\text{pop}_i)$ will not match $\delta_{\text{post}} \cdot f(\text{pop}_i)$ unless they also balance $f(\text{pop}_i)$ itself—which is impossible when treatment is assigned by a population cutoff. Empirically, SDID’s unit weights ω close almost none of the population gap (at the 5,000 cutoff, the weighted-control log-pop gap is -2.18 versus TWFE’s -2.19 ; within 1% across all twenty placebo cutoffs). However, SDID’s *time* weights λ_t partially neutralise the first factor in (A.7): when the pre-period gradient evolves smoothly, $\sum_t \lambda_t \delta_t$ approximates $\bar{\delta}_{\text{post}}$ and the time-weighted bracket shrinks. The growing population gap is then multiplied by a small number at every cutoff, leaving a residual estimate that stays near $+0.014$ across the placebo grid rather than scaling with the cutoff the way TWFE’s does. The flat SDID placebo therefore reflects partial protection from the gradient mechanism, not closure of the population gap that drives the bias.

C.2 What causes the gradient?

One might object that the size gradient is not a primitive confound but is *caused* by observable covariates—selective outmigration of educated residents, demographic aging, declining labor markets, or pre-existing attitudes differentially activated by salient events. If the gradient is fully mediated by matchable covariates, proper matching should eliminate it.

The formal response is that $f(\text{pop}_i)$ in (A.1) need not represent a causal effect of population. It can denote any size-linked score whose electoral relevance changes over time. The bias formula (A.6) holds regardless of what causes the gradient—its microfoundation is irrelevant to the identification failure. But matching on *baseline* covariate values does not absorb *changes* in those covariates. If brain drain from small municipalities is accelerating, then 2008 education levels do not predict 2018 education levels within population strata. And if the gradient partly reflects attitudes—such as distrust of institutions—whose electoral salience changes over time, matching on their baseline levels faces the same problem. In principle, sufficiently rich time-varying covariate data could absorb the gradient for some of these channels, but the identification failure persists as long as any component remains uncontrolled.

Table A.3 tests this. Adding year-specific slopes for the six covariates used in Mahalanobis matching (population, foreign-born share, over-65 share, mean income, university education share, and maximum altitude) reduces the estimate from 0.015 to 0.006—the covariates capture roughly 60% of the gradient but leave a highly significant residual ($p < 0.001$). Adding $\log(\text{pop}) \times \text{year}$ alone eliminates the estimate entirely (-0.000 , $p = 0.78$). The size gradient captures confounding that the matched covariates do not.

The gradient itself cannot be a consequence of the reform: Figure 1 of the main paper shows the slope steepening from 2001 onward, well before the 2010 reform, and the same pattern appears in France where no comparable reform exists (Figure 2 of the main paper).

Table A.3: Covariate controls versus population controls

Specification	Estimate
TWFE (baseline)	0.015 (0.001)
TWFE + covariates \times year	0.006 (0.001)
TWFE + $\log(\text{pop}) \times$ year	0.000 (0.002)
TWFE + covariates \times year + $\log(\text{pop}) \times$ year	0.003 (0.001)

Notes: All specifications include municipality and year FE with SEs clustered by municipality. “Covariates \times year” adds year-specific slopes for foreign-born share, over-65 share, mean income, university share, female share, and maximum altitude—the same variables used for Mahalanobis matching.

References

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.
- Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455.
- Conferenza Stato-Città ed Autonomie Locali (2025). L’obbligo di esercizio associato delle funzioni fondamentali. Approfondimento tematico, Presidenza del Consiglio dei Ministri. Last updated 29 January 2025. <https://www.conferenzastatocitta.gov.it/home/approfondimenti/tematiche-di-interesse/associazionismo-comunale/l-obbligo-di-esercizio-associato-delle-funzioni-fondamentali/>.

- Corte Costituzionale (2019). Sentenza n. 33 del 2019. Corte Costituzionale della Repubblica Italiana. Partial unconstitutionality of Article 14 of decree-law 78/2010. <https://www.cortecostituzionale.it/>.
- Crevaschi, S., Rettl, P., Cappelluti, M., and De Vries, C. E. (2024). Geographies of discontent: Public service deprivation and the rise of the far right in Italy. *American Journal of Political Science*, 69(4):1581–1599.
- Di Ielsi, G., Fiorillo, F., and Porcelli, F. (2022). Le unioni di comuni in Italia: modelli di gestione associata a confronto. *Moneta e Credito*, 75(297).
- Ferraresi, M., Migali, G., and Rizzo, L. (2018). Does intermunicipal cooperation promote efficiency gains? Evidence from Italian municipal unions. *Journal of Regional Science*, 58(5):1017–1044.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.
- Gori, G. F., Lattarulo, P., Porcelli, F., Rizzo, L., and Secomandi, R. (2025). Does intermunicipal cooperation affect spending vs output? Evidence from Italy. *Regional Studies*, 59(1).
- Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236.
- Luca, D. and Modrego, F. (2021). Stronger together? Assessing the causal effect of intermunicipal cooperation on the efficiency of small Italian municipalities. *Journal of Regional Science*, 61(1):261–293.
- Ministero dell'Interno (2020). Anagrafe delle Unioni di Comuni. Dipartimento per gli Affari Interni e Territoriali. Composition registry of currently-active inter-municipal

consortia (*unioni di comuni*). Includes *codice unione*, member municipalities, and date of constitution. 2020 vintage accessed March 2026 at https://dait.interno.gov.it/territorio-e-autonomie-locali/sut/open_data.php.

Repubblica Italiana (2010). Decreto-legge 31 maggio 2010, n. 78. Misure urgenti in materia di stabilizzazione finanziaria e di competitività economica. *Gazzetta Ufficiale* n. 125 (Suppl. Ordinario n. 114), converted into Law 122/2010. <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legge:2010-05-31;78>.

Repubblica Italiana (2024). Decreto-legge 27 dicembre 2024, n. 202. Disposizioni urgenti in materia di termini normativi. *Gazzetta Ufficiale* n. 302. Article 21 abolishes the mandatory associated exercise of fundamental functions for small municipalities. <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legge:2024-12-27;202>.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244.

Spano, A. and Tagliagambe, V. (2018). Inter-municipal cooperation and cost savings: Myth or reality? Conference paper, IRSPM XXII Annual Conference, Edinburgh, 11–13 April 2018. https://cpt.regione.sardegna.it/wp-content/uploads/2018/04/35_335_20190131143402.pdf.