The Democratic Cost of Consolidation: Municipal Mergers and Referendum Participation in Switzerland

APEP Autonomous Research* @ai1scl

March 4, 2026

Abstract

Does municipal consolidation erode direct democracy? Using Swiss commune-level referendum data (1960–2025) and a merger timeline from the BFS mutations registry (637 events), I exploit staggered merger timing across over one million commune-referendum observations. A naive two-way fixed effects estimate finds no effect—but the event study reveals that merging communes were already on declining turnout trajectories, an Ashenfelter's dip that masks the true treatment effect. A stacked difference-in-differences design estimates that mergers reduce referendum turnout by 1.67 percentage points ($p < 0.001$). Dose-response analysis reveals that estimator choice determines mechanism inference: TWFE suggests larger mergers produce smaller declines (favoring identity loss), while the stacked DiD reverses the sign (-5.18 pp per log unit, $p = 0.008$), confirming that larger mergers produce larger declines—consistent with free-riding. This estimator dependence extends the Ashenfelter's dip lesson from average effects to mechanism identification.

JEL Codes: D72, H11, H77

Keywords: municipal mergers, direct democracy, voter turnout, referendums, Switzerland

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: N/A).

1. Introduction

Every few weeks, Swiss citizens receive a thick envelope in the mail. Inside are ballot papers for federal referendums—votes on tax policy, immigration quotas, infrastructure spending, environmental regulation. Switzerland holds roughly ten referendum dates per year, and citizens decide on average four to five separate propositions at each one. This system of direct democracy, unmatched in its scope and frequency anywhere in the world, rests on a foundation of small, self-governing communes. For centuries, these communes have been the basic units of Swiss political life: small enough that neighbors know each other, local enough that civic habits form early, and autonomous enough that participation feels consequential.

That foundation is eroding. Between 2000 and 2020, the number of Swiss communes fell from approximately 2,900 to 2,200 through roughly 700 mergers. Cantons such as Fribourg, Ticino, Glarus, and Graubünden reorganized their municipal landscapes dramatically, sometimes combining dozens of tiny communes into a single administrative unit. The 2011 Glarus reform stands as the most striking example: twenty-five communes collapsed into three. These consolidations were driven by fiscal pressures, administrative capacity constraints, and cantonal incentive programs. They produced measurable gains in cost efficiency and service delivery (Steiner, 2003; Ladner and Bühlmann, 2009; Steiner and Kaiser, 2017). But they also severed the link between community identity and political jurisdiction—a link that may matter for democratic engagement in ways that efficiency audits do not capture.

This paper asks whether municipal mergers reduce citizen participation in direct democracy. The question is important for two reasons. First, the tension between administrative scale and democratic vitality is a first-order concern in political economy. Dahl and Tufte (1973) argued that democracy requires communities small enough for meaningful participation, while Oates (1972) emphasized the efficiency advantages of larger jurisdictions. Every country that consolidates local governments confronts this trade-off, yet the evidence base on the democratic costs remains thin, concentrated in Scandinavia, and focused exclusively on representative elections rather than direct democratic participation. Second, Switzerland's referendum system provides a uniquely sharp test. Unlike elections, where turnout reflects candidate preferences, mobilization efforts, and party competition, referendums ask citizens to engage with policy substance. If mergers reduce referendum participation, the loss is not merely procedural—it represents diminished citizen voice in actual policy decisions.

I study this question using Swiss federal referendum data from 1960 to 2025, reported by the Swiss Federal Statistical Office (BFS) using current commune boundaries. The BFS publishes commune-level turnout for all 189 federal referendum dates, yielding a panel of 1,000,637 commune-referendum observations across 2,114 communes. I combine this with a comprehensive merger timeline constructed from the BFS commune mutations API, which records every municipal mutation—creation, absorption, division, renaming—with a precise effective date. The mutations registry contains 637 merger events spanning 1961 to 2026 (including mergers with scheduled future effective dates). For causal estimation, I restrict to mergers with effective dates between 2000 and 2020, yielding 197 treated communes—the successor entities identified by the terminal commune code (TerminalCode) in the BFS mutations data—with 1,917 communes that experienced no merger during 2000–2020 serving as controls.

The identification strategy exploits the staggered timing of mergers across communes and cantons. I compare referendum turnout in communes that merge to turnout in never-merged communes, using both traditional two-way fixed effects (TWFE) and a stacked difference-in-differences design that constructs cohort-specific clean comparisons.

The paper's central finding is also its central methodological lesson. The naive TWFE estimate is 0.05 percentage points (standard error 0.36, $p = 0.89$)—essentially zero. Taken at face value, this would suggest that mergers have no effect on democratic participation. But the event study tells a different story. Pre-treatment coefficients are large, negative, and jointly significant ($F = 8.68$, $p \approx 0$): merging communes exhibited declining turnout relative to controls for a full decade before the merger took place. This is a textbook Ashenfelter's dip (Ashenfelter, 1978). Communes that merge are not randomly selected from the population; they are communes in civic decline. The TWFE estimator, which compares long-run average levels, absorbs this declining trajectory into the fixed effects and attributes the remaining post-merger level to "no treatment effect." But the pre-existing decline is part of the selection process, not part of the treatment.

The stacked difference-in-differences design addresses this problem. By constructing cohort-specific datasets with clean $\pm$5-year windows and cohort-specific fixed effects, the stacked design compares each merger cohort only to never-treated controls within a narrow temporal window, avoiding contamination from heterogeneous treatment timing and differential pre-trends across cohorts. The stacked DiD estimate is -1.67 percentage points (standard error 0.24, $p < 0.001$). This estimate represents the within-window treatment effect for a typical merger cohort: relative to where the commune was heading in the five years before the merger, turnout drops by an additional 1.67 percentage points in the five years after.

The dose-response analysis produces the paper's most important methodological finding. In the TWFE specification, the interaction between the post-merger indicator and the log size ratio is positive and significant (+5.14 percentage points per log unit, $p = 0.012$): larger mergers appear associated with *smaller* turnout declines, seemingly favoring the identity-loss channel over free-riding (Putnam, 1993). But when estimated in the stacked DiD—which purges the differential selection that contaminates TWFE—the sign *reverses*: the interaction is -5.18 percentage points per log unit ($p = 0.008$). Larger mergers produce *larger* turnout declines, exactly as the free-riding model predicts (Olson, 1965). The TWFE dose-response captured differential selection—communes that merged with much larger neighbors were declining less steeply before the merger—not a differential treatment effect. This sign reversal demonstrates that estimator choice contaminates not just average treatment effect estimation but mechanism inference itself, extending the Ashenfelter's dip lesson to a domain where it has not previously been recognized.

Three features of the analysis merit emphasis. First, the pre-trend violation is not a bug but a feature: it reveals the selection process driving Swiss municipal mergers. Communes do not merge at random. They merge when civic capacity is already declining—when it becomes difficult to find citizens willing to serve as part-time municipal officials, when participation in communal assemblies falls, when the fiscal base erodes. The declining turnout trajectory in the decade before merger is a symptom of the same institutional stress that eventually triggers the merger decision. Recognizing this selection is essential for correct inference.

Second, the HonestDiD sensitivity analysis (Rambachan and Roth, 2023) provides transparent bounds on the estimate under varying degrees of pre-trend violation. At $M = 0$ (assuming strict parallel trends in the post-period), the 95 percent confidence interval is $[-3.44, -2.16]$, excluding zero. At $M = 1.74$ (allowing trend violations equal to half the maximum observed pre-treatment slope change), the interval widens to $[-5.22, 0.71]$, which includes zero. At $M = 3.47$ (allowing violations equal to the full maximum pre-treatment slope change), the interval is $[-6.96, 2.45]$. The stacked DiD estimate of -1.67 falls within all these bounds, and the sign is consistently negative, but the reader should understand that the result's statistical significance depends on assumptions about the counterfactual trend.

Third, randomization inference confirms that the TWFE null result is exactly what one would expect from random treatment assignment. When I permute merger dates across 200 draws (from the 2000–2020 year pool), the resulting placebo ATTs produce an RI p-value of 0.975—the actual TWFE estimate is indistinguishable from noise. This does not mean mergers have no effect; it means that the TWFE estimator, in the presence of heterogeneous treatment timing and pre-existing trends, lacks the power to detect the effect that the stacked design identifies.

This paper contributes to three literatures. First, it provides the first comprehensive evidence on municipal consolidation and direct democratic participation. The existing merger literature studies representative elections in Denmark (Lassen and Serritzlew, 2011; Blom-Hansen et al., 2016), Finland (Harjunen et al., 2021; Saarimaa and Tukiainen, 2015), Israel (Reingewertz, 2012), and the Netherlands (Allers and Geertsema, 2016). The Swiss setting allows me to study direct democracy, where citizens vote on policy rather than candidates.

Second, the paper demonstrates the importance of estimator choice in staggered DiD designs with selection on pre-trends. The TWFE estimate (null) and the stacked DiD estimate (-1.67 pp, highly significant) tell completely different stories, and understanding *why* they differ—through the lens of Ashenfelter's dip and cohort-specific heterogeneity—is as important as the point estimate itself. The paper provides a case study in how modern DiD methods (Callaway and Sant'Anna, 2021; Baker et al., 2022; Goodman-Bacon, 2021; Borusyak et al., 2024; de Chaisemartin and D'Haultfœuille, 2020) can recover treatment effects that traditional TWFE obscures.

Third, the dose-response sign reversal between TWFE and stacked DiD demonstrates that estimator choice contaminates mechanism inference, not just average effect estimation. The TWFE dose-response, which appeared to favor identity loss, was an artifact of differential selection into merger; the stacked DiD dose-response confirms the free-riding prediction that larger electorates produce larger turnout declines. This has both methodological and policy implications: researchers using TWFE for mechanism analysis in staggered settings risk drawing incorrect theoretical conclusions, and policymakers should understand that the democratic cost of consolidation scales with the size of the merged electorate—larger consolidations carry larger participatory costs.

The remainder of the paper details the institutional stress driving these mergers, the empirical strategy used to unmask their effects, and the dose-response evidence whose estimator dependence reveals both the free-riding channel and a cautionary tale about mechanism inference in staggered designs.

2. Institutional Background

2.1 Swiss Communes and Federalism

Switzerland's three-tier federal structure gives communes—Gemeinden in German, communes in French, comuni in Italian—a degree of autonomy unusual in comparative perspective. The 1999 Federal Constitution (Article 50) guarantees communal autonomy within the limits set by cantonal law. Communes raise their own taxes (typically a multiple of the cantonal base rate), run their own schools, manage local infrastructure, administer social welfare, and maintain civil registries. They also control naturalization decisions, zoning, and planning permissions. In many small communes, citizens govern through the Gemeindeversammlung (communal assembly) rather than an elected parliament—a form of direct democratic self-governance that dates to the medieval period.

The number of Swiss communes peaked at over 3,200 in the mid-twentieth century. Most were extremely small: as late as 2000, the median commune had approximately 1,200 inhabitants, and over 40 percent had fewer than 1,000 (Ladner, 2002). This fragmentation created persistent challenges. Small communes struggled to provide professional administration, attract qualified part-time officeholders, comply with increasingly complex federal and cantonal mandates, and finance the infrastructure expected by modern residents (Steiner, 2003). The professional demands of communal leadership grew as Swiss regulation expanded. A commune of 300 residents still needed someone to manage water treatment, building permits, school operations, and social assistance—tasks that often fell to a single elected official serving part-time for modest compensation.

2.2 The Merger Process

Swiss municipal mergers are voluntary, bottom-up, and typically require a referendum in each constituent commune. The process usually begins when one or more small communes acknowledge their inability to provide adequate services and approach a larger neighbor. A feasibility study follows, often subsidized by the canton. If the merger commission recommends proceeding, each commune holds a binding referendum. Approval typically requires a simple majority, though some cantons impose higher thresholds for the absorbing commune.

Several cantons actively encouraged mergers through financial incentive programs. Fribourg launched its program in 1999, offering substantial transition grants. Ticino provided merger subsidies and administrative support from the early 2000s. Graubünden and Thurgau adopted similar programs. The canton of Glarus took the most dramatic approach: a 2006 Landsgemeinde (cantonal assembly) voted to merge all 25 communes into 3, effective January 1, 2011 (Koch and Rochat, 2017). This reform was exceptional—most mergers combined two to five communes, typically a small satellite with a nearby center.

The timing of mergers was driven by a combination of cantonal policy, local fiscal stress, and political opportunity. Mergers clustered in time within cantons (reflecting cantonal incentive programs) but varied widely across cantons (reflecting different political cultures and degrees of cantonal encouragement). Between 2000 and 2020, the most active cantons were Fribourg (reducing from 245 to 128 communes), Ticino (from 245 to 115), Graubünden (from 212 to 101), and Thurgau (from 80 to 80, though with internal reorganization). Glarus went from 25 to 3 in a single reform. The German-speaking plateau cantons (Zürich, Bern, Aargau, Luzern) were less active, though individual mergers occurred. The total number of communes fell from approximately 2,900 in 2000 to approximately 2,136 by 2024 (Steiner and Kaiser, 2017).

2.3 Merger Typology

Not all mergers are alike, and the heterogeneity matters for identification. I distinguish three types based on the BFS mutation records:

1. **Absorptions**: A small commune is absorbed into an existing larger commune, which retains its name, BFS number, and legal identity. The absorbed commune ceases to exist. This is the most common type, accounting for approximately 65 percent of merger events.

2. **Fusions**: Two or more communes dissolve and create a new commune with a new name and BFS number. This is less common (roughly 25 percent of events) and typically occurs when the merging partners are of comparable size.

3. **Aggregations**: Multiple communes are combined by cantonal decree (as in Glarus). The new commune has a new name and identity. This is rare (roughly 10 percent of events) and tends to involve the largest number of constituent communes.

For the empirical analysis, the key distinction is between absorptions (where identity loss is concentrated in the absorbed commune) and fusions/aggregations (where identity loss may be more symmetric). I code this distinction from the BFS mutation type field.

2.4 The Federal Referendum System

Switzerland held 189 federal referendum dates between 1960 and 2025, with citizens voting on hundreds of distinct propositions. Federal referendums come in three types: mandatory referendums on constitutional amendments, optional referendums challenging parliamentary legislation (requiring 50,000 signatures), and popular initiatives proposing new constitutional provisions (requiring 100,000 signatures). Turnout at federal referendums has declined secularly from approximately 50 percent in the 1960s to approximately 40 percent in recent decades, with substantial variation across propositions and across communes (Kriesi, 2005).

Voting in federal referendums is administered by the communes. Each commune is responsible for sending ballot materials, operating polling stations (or processing mail ballots after the 1990s postal voting reform), and reporting results to the cantonal chancellery. The cantonal chancellery transmits results to the Federal Chancellery, which publishes commune-level data through the BFS PXWeb system. This administrative chain means that a merger directly affects the organizational unit responsible for referendum administration.

The introduction of postal voting, which all cantons adopted between 1978 and 2005, transformed referendum participation (Funk, 2010). Before postal voting, citizens had to appear physically at a communal polling station—an act with social visibility that Funk (2010) shows increased turnout by approximately 3 percentage points relative to mail voting. After the reform, most voters cast ballots by mail. This institutional change is important for my analysis because it eliminates one potential mechanism (changed polling station locations after merger) for the post-2005 period when postal voting was universal.

2.5 Why Mergers Might Affect Referendum Participation

The institutional features of Swiss direct democracy create several channels through which mergers could affect participation. First, the merged commune is larger, which may reduce each citizen's perceived influence on the referendum outcome—a direct application of Olson (1965)'s collective action problem. Second, the merger may dissolve community networks and civic habits that sustained participation (Putnam, 1993; Freitag, 2006). Third, the loss of communal identity—especially when a commune's name disappears—may reduce the psychological attachment that motivates voting as an expressive act (DellaVigna et al., 2017). Fourth, administrative disruption (new commune codes, changed registration, unfamiliar ballot materials) may create transitional friction. The empirical analysis aims to distinguish among these channels.

2.6 Selection into Merger: The Ashenfelter's Dip

A feature of Swiss mergers that is crucial for identification is the selection process. Mergers are not exogenous shocks; they are endogenous responses to institutional stress. A commune that can no longer recruit a municipal treasurer, whose assembly meetings attract fewer than ten citizens, whose tax base is shrinking—this is a commune that both merges and experiences declining political engagement. The decision to merge is itself a symptom of civic decline. This creates what Ashenfelter (1978) identified in the context of job training programs: treated units are selected on the basis of a declining outcome trajectory, so that comparing post-treatment levels to the pre-treatment trend mechanically overstates (or in this case, masks) the treatment effect. Recognizing this selection pattern is essential for choosing the right estimator and interpreting the results correctly.

3. Conceptual Framework

I develop a simple framework to organize the mechanisms through which municipal mergers affect referendum participation. The framework generates testable predictions that structure the empirical analysis.

3.1 A Participation Decision with Community Identity

Consider a citizen i in commune c deciding whether to vote in referendum r. Following the calculus-of-voting framework (Downs, 1957; Feddersen, 2004), citizen i votes if the expected benefit exceeds the cost:

$$B_i = p_i \cdot d_i + s_i(\theta_c) - \kappa_i > 0 \tag{1}$$

where p_i is the perceived probability of being pivotal, d_i is the utility difference between the citizen's preferred and non-preferred outcome, $s_i(\theta_c)$ captures the social and expressive return to voting as a function of community identity θ_c, and κ_i is the cost of voting. A merger changes this calculus through multiple channels.

3.2 Channel 1: Scale Effect (Free-Riding)

A merger increases the electorate from N_c to $N'_c = N_c + \sum_j N_j$, where j indexes the other merging communes. Under standard pivotality models, p_i is decreasing in electorate size. For a referendum decided by simple majority, $p_i \approx 1/\sqrt{N}$ in the neighborhood of a tied vote. The scale effect predicts:

Prediction 1 (Dose-Response): The turnout decline should be monotonically increasing in the proportional increase in electorate size. Mergers that double the electorate should produce larger effects than those adding 20 percent.

Prediction 2 (Immediate Onset): The scale effect operates through a mechanical channel (electorate size) and should appear immediately at the first post-merger referendum.

3.3 Channel 2: Identity Loss

A merger may reduce θ_c—the strength of communal identity—particularly for residents of absorbed communes. When a commune loses its name, its independent governance, and its representation in intercommunal bodies, the social return $s_i(\theta_c)$ declines. Voting is partly an expression of civic duty tied to community membership (DellaVigna et al., 2017; Gerber et al., 2008); dissolution of the community weakens the norm.

Prediction 3 (Equal-Partner Mergers): If identity loss drives the effect, then mergers between roughly equal partners—where *both* communes lose their pre-merger identity—should produce larger turnout declines than asymmetric mergers where a small commune is absorbed into a large, intact neighbor.

Prediction 4 (Persistence): Identity loss is a slow-moving phenomenon. Unlike the scale effect (which is mechanical), identity erosion may deepen over time as community institutions dissolve and local memory fades. Alternatively, partial recovery may occur as new institutional identities form within the merged entity.

3.4 Channel 3: Administrative Disruption

A merger changes the organizational unit responsible for ballot distribution, polling stations, and result tabulation. In the short run, residents may face confusion about registration, changed ballot formats, or reduced proximity to polling locations.

Prediction 5 (Transience): Administrative disruption should produce a spike in non-participation immediately after the merger that fades within two to three referendum cycles as new routines form.

3.5 Distinguishing the Channels

The three channels generate different dose-response signatures that are central to my analysis. The scale effect (Channel 1) predicts that turnout declines should increase monotonically with merger size. The identity-loss channel (Channel 2) predicts the opposite when mergers are asymmetric: absorbing a small commune into a large one barely disrupts the larger commune's identity, while merging two equal communes disrupts both. The key test is therefore the sign of the dose-response relationship. A negative coefficient on merger size (larger mergers, larger declines) favors the scale channel. A positive coefficient (larger mergers, smaller declines) favors the identity channel. As I show below, the answer depends critically on the estimator: TWFE dose-response appears to favor the identity channel, but a stacked DiD dose-response—which purges differential selection—reverses the sign and favors the scale channel. This estimator dependence of mechanism inference is itself a central finding.

The event study provides additional diagnostics. If turnout drops sharply at merger and remains flat, the scale effect dominates. If the drop deepens gradually, identity loss dominates. If the initial drop fades, administrative disruption dominates. The combination of dose-response gradients and temporal dynamics allows me to assess the relative importance of each channel.

4. Data

4.1 Referendum Participation Data

The primary data source is the BFS PXWeb cube `px-x-1703030000_101`, which reports commune-level results for all federal referendums using current administrative boundaries. The BFS publishes referendum results within weeks of each vote; as of download (March 2026), the data cover 189 federal referendum dates from 1960 through the most recent

2025 referendum, across 2,114 distinct communes, yielding 1,000,637 commune-referendum observations. For each commune-referendum observation, the data contain the number of eligible voters, ballots cast, valid votes, and yes/no vote shares. I compute turnout as ballots cast divided by eligible voters, expressed in percentage points.

A key feature of this dataset is that the BFS reports all historical referendum results using current (2025) commune boundaries. This means that for communes that have undergone mergers, the pre-merger turnout figures are already aggregated to the successor entity level. This retrospective harmonization, performed by the BFS itself, eliminates the need for manual concordance across historical boundary changes and ensures consistent geographic units throughout the panel.

When multiple propositions appear on the same referendum date, the BFS reports a single turnout figure for the date (since voters receive one envelope containing all propositions). Turnout varies little across propositions on the same date because Swiss voters typically decide on all propositions simultaneously.

4.2 Merger Timeline

I construct the merger timeline from the BFS commune mutations API, which records every municipal mutation in the Historisiertes Gemeindeverzeichnis (SMMT). I filter for merger-related mutations (types: "Eingemeindung" [absorption], "Fusion" [fusion], "Zusammenschluss" [aggregation]) and exclude boundary adjustments, name changes, and district reassignments.

The registry records 637 merger events with effective dates spanning 1961 to 2026 (the BFS records scheduled mergers in advance of their effective date, so future-dated entries reflect legally approved mergers). For descriptive purposes, Figure 4 shows the full historical distribution. For causal estimation, I restrict to mergers with effective dates between 2000 and 2020, which ensures at least five years of post-treatment referendum data (through 2025). This yields 197 treated communes—the successor entities (TerminalCode in the BFS mutations data) that emerged from each merger event. The remaining 1,917 communes that experienced no merger during 2000–2020 serve as the control group. This control pool includes both communes that never merged at any point and a small number of communes whose mergers occurred after 2020; the latter contribute only pre-treatment observations within any relevant estimation window.

For each merger, I record: (i) the effective date, (ii) the BFS numbers of the constituent communes, (iii) the BFS number of the successor entity, (iv) the mutation type, and (v) the computed size ratio (post-merger eligible voters divided by pre-merger eligible voters of the largest constituent).

4.3 Panel Construction

Because the BFS referendum data are reported in current boundaries, the panel is already harmonized to post-merger geographic units. Each commune in the dataset corresponds to a 2025-era commune, and its turnout history extends back to 1960 regardless of whether the commune's boundaries changed through mergers. For treated communes, pre-merger turnout reflects the weighted-average participation of the predecessor communes (as computed by the BFS). For control communes that never merged, the series is simply the commune's own historical turnout.

The treatment indicator PostMerger_{ct} equals one if commune c completed a merger by referendum date t and zero otherwise. Event time is defined as the number of years between the referendum date and the merger effective date. I code communes as treated starting from the calendar year of their merger.

4.4 Summary Statistics

Table 1: Summary Statistics

	N	Communes	Mean Turnout	SD Turnout	Mean Eligible	Median Eligible	Mean Yes %
Full Sample	1,000,637	2114	44.7	13.2	2,166	900	47.7
Treated	89,835	197	43.6	12.8	2,580	1,319	47.8
Control	910,802	1917	44.8	13.2	2,125	850	47.6

Notes: Commune-referendum level panel. Turnout is the percentage of eligible voters casting a ballot. "Treated" communes are the successor entities (TerminalCode in BFS mutations data) of mergers between 2000 and 2020. "Control" communes never experienced a merger during 2000–2020.

Table 1 presents summary statistics for the analysis sample. The full panel contains 1,000,637 commune-referendum observations across 2,114 communes and 189 referendum dates. Treated communes (197) contribute 89,835 observations; control communes (1,917) contribute 910,802. Mean turnout across the full sample is 44.7 percent, with a standard deviation of 13.2 percentage points, reflecting both cross-commune variation and the secular decline in Swiss turnout over the sample period.

Pre-treatment balance reveals a pattern consistent with the selection story. In the 2000–2005 window (before most mergers in the analysis sample), treated communes have slightly lower average turnout (43.98 percent) than controls (45.40 percent). Treated communes are somewhat larger: the average number of eligible voters is 2,762 for treated communes versus 2,203 for controls. This size difference may seem counterintuitive—one might expect merging communes to be smaller—but recall that the BFS data report in current (post-merger) boundaries, so the treated commune figures reflect the already-merged entity's electorate when observed in the pre-period. The turnout gap of 1.4 percentage points is modest in levels, but as the event study will show, the *trends* diverge substantially in the decade before merger.

5. Empirical Strategy

5.1 Baseline TWFE Specification

I estimate the effect of municipal mergers on referendum turnout using a difference-in-differences framework. The baseline TWFE specification is:

$$\text{Turnout}_{ct} = \alpha_c + \gamma_t + \beta \cdot \text{PostMerger}_{ct} + \varepsilon_{ct} \tag{2}$$

where c indexes communes, t indexes referendum dates, α_c are commune fixed effects, γ_t are referendum-date fixed effects, and PostMerger_{ct} equals one if commune c has completed its merger by referendum date t and zero otherwise. The coefficient β captures the average change in turnout associated with the merger, relative to the commune's own pre-merger level and the contemporaneous level in control communes.

Standard errors are clustered at the commune level to account for serial correlation in turnout within communes (Cameron et al., 2008).

5.2 The Problem with TWFE Under Selection on Trends

The TWFE estimator in Equation (2) produces a consistent estimate of the ATT under the parallel trends assumption: absent the merger, treated and control communes would have followed the same turnout trajectory. When this assumption is violated—specifically, when communes select into treatment on the basis of a declining trend in the outcome—TWFE is biased toward zero. The commune fixed effects absorb the *level* difference between treated and control communes, but if treated communes were already declining faster than controls, the post-merger outcome reflects both the treatment effect and the continuation of the pre-existing trend. The TWFE coefficient captures the net of these two forces, which may be close to zero even when the treatment has a large causal effect.

This problem is well-known in the program evaluation literature (Ashenfelter, 1978; Heckman et al., 1999) and in recent work on staggered DiD (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021). The event study provides a direct diagnostic: significant pre-treatment coefficients indicate that the parallel trends assumption is violated in the TWFE specification.

5.3 Stacked Difference-in-Differences

To address the pre-trend contamination, I implement a stacked DiD design following Baker et al. (2022). For each merger cohort g (defined by the year of the merger effective date), I construct a "clean" dataset containing only cohort g communes and control communes—those with no merger during 2000–2020—with a symmetric window of ± 5 years around the merger. Because control communes appear in the sub-experiment for every cohort, the stacked dataset contains approximately 3.85 million observations (roughly $3.9\times$ the base panel). I stack these cohort-specific datasets and estimate:

$$\text{Turnout}_{cgt} = \alpha_{cg} + \gamma_{tg} + \delta \cdot \text{Post}_{ct} + \varepsilon_{cgt} \tag{3}$$

where α_{cg} are commune-by-cohort fixed effects and γ_{tg} are date-by-cohort fixed effects. The stacked design has two advantages. First, it ensures that each treated cohort is compared only to clean controls—communes that experienced no merger within the 2000–2020 analysis window—within a narrow temporal window, avoiding the contamination from heterogeneous treatment timing that produces negative weights in TWFE (Goodman-Bacon, 2021). Second, the narrow ± 5-year window limits the influence of long-run divergent trends: even if treated communes are on a declining trajectory relative to controls, the divergence within a 10-year window is smaller than over the full 65-year panel.

The stacked DiD estimate $\hat{\delta}$ is my preferred specification. Standard errors are clustered at the commune level (the original commune identifier, not the stacked identifier) to account for the fact that control communes appear in multiple cohort-specific datasets.

5.4 Event Study

The event study plots the estimated treatment effects by event time e (years relative to the merger):

$$\text{Turnout}_{ct} = \alpha_c + \gamma_t + \sum_{e \neq -1} \theta_e \cdot \mathbb{I}\{t - g_c = e\} + \varepsilon_{ct} \tag{4}$$

where g_c is commune c's merger year and $e = -1$ is the omitted reference period. Pre-treatment coefficients ($e < -1$) test the parallel trends assumption; post-treatment coefficients ($e \geq 0$) trace the dynamic effect. I estimate this specification using TWFE with event-time indicators, binning endpoints at $e = \pm 10$.

I complement the visual pre-trend test with a joint F-test of $H_0 : \theta_{-10} = \theta_{-9} = \cdots = \theta_{-2} = 0$ and implement the Rambachan and Roth (2023) HonestDiD sensitivity analysis, which constructs confidence sets valid under bounded violations of parallel trends.

5.5 Matched Difference-in-Differences

To address selection on observables, I implement a matched DiD in which each merging commune is paired with a never-treated commune using nearest-neighbor matching on pre-merger average turnout (2000–2005). The DiD is then estimated within matched pairs. This approach purges the level difference between treated and control communes and, to the extent that pre-treatment turnout levels predict trends, partially addresses the selection-on-trends concern.

5.6 Randomization Inference

I implement randomization inference (RI) as a non-parametric test of the sharp null hypothesis of no treatment effect. For each of 200 iterations, I randomly reassign merger years to treated communes (drawing from the pool of actual treatment years 2000–2020, matching the estimation sample) and re-estimate the TWFE ATT. The RI p-value is the fraction of placebo estimates that exceed the actual estimate in absolute value.

5.7 Threats to Validity

5.7.1 Selection into Merger

As discussed, communes that merge are not randomly selected. The key threat is that merging communes would have experienced declining turnout even absent the merger, violating the parallel trends assumption. The event study is the primary diagnostic. The stacked DiD, by using narrow windows, mitigates (but does not eliminate) this concern. The matched DiD provides an additional check by comparing treated communes to observationally similar controls. The HonestDiD analysis quantifies how much pre-trend violation the result can survive.

5.7.2 The Glarus Reform

The 2011 Glarus reform merged 25 communes into 3 in a single cantonal decree—an outlier in both scale and process. To verify that this extreme event does not drive the results, I re-estimate the TWFE excluding all Glarus communes. The Glarus reform is particularly important because it was a top-down mandated merger, unlike the typical voluntary process, and produced the largest discrete change in commune structure in Swiss history.

5.7.3 Anticipation

Mergers are typically announced one to three years before the effective date. If the announcement itself affects turnout, the pre-treatment period may be contaminated beyond what the selection-on-trends story implies. I note that the event study shows declining coefficients beginning at $e = -10$, well before any plausible announcement window, which suggests that the pre-trend reflects long-run selection rather than short-run anticipation.

6. Results

6.1 Main Results

Table 2: Effect of Municipal Mergers on Referendum Turnout

Specification	ATT (pp)	SE	N	Communes
TWFE	0.050	(0.363)	1,000,637	2114
Stacked DiD	-1.672	(0.237)	3,853,979	2114
Matched DiD	-0.236	(0.450)	175,926	379
Excl. Glarus	0.159	(0.362)	999,128	2111
Commune FE			Yes	
Vote-date FE			Yes	
Clustering			Commune	
RI p-value			0.945	

Notes: Dependent variable is federal referendum turnout (%). ATT is the average treatment effect on the treated, where treatment is the merger event. Standard errors clustered at the commune level in parentheses. Row (1) reports the two-way fixed effects estimate. Row (2) uses stacked cohort-specific DiD with ±5-year windows; N exceeds the base panel because control communes appear in each cohort-specific sub-experiment. Row (3) matches treated to control communes on pre-merger turnout. Row (4) excludes the Glarus mega-merger (25→3 successor communes in 2011), reducing both total and treated counts by 3.

Table 2 presents the main estimates across four specifications. The TWFE baseline (Equation 2) yields an ATT of 0.05 percentage points with a standard error of 0.36, statistically indistinguishable from zero ($p = 0.89$). Taken in isolation, this estimate would suggest that municipal mergers have no detectable effect on referendum participation. But the TWFE estimate is misleading in this context, for reasons I develop below.

The stacked DiD (Equation 3) produces a sharply different result: an ATT of -1.67 percentage points with a standard error of 0.24, significant at the 0.1 percent level ($p < 0.001$). This is my preferred specification. The stacked design compares each merger cohort to controls—communes with no merger during 2000–2020—within a clean ±5-year window, using cohort-specific commune and date fixed effects. By isolating within-window variation, it avoids the contamination from differential long-run trends that attenuates the TWFE estimate.

The matched DiD yields an ATT of -0.24 percentage points (standard error 0.45, $p = 0.60$), not statistically significant. The matched estimate is closer to the TWFE than to the stacked DiD, reflecting the fact that matching on pre-treatment levels does not fully address matching on pre-treatment *trends*. Even after pairing treated communes with observationally similar controls, the underlying trajectory differences remain. Excluding Glarus communes from the TWFE produces an ATT of 0.16 percentage points, essentially unchanged from the baseline TWFE, confirming that the Glarus mega-merger does not drive the null result.

The randomization inference p-value for the TWFE specification is 0.975 (with merger years drawn from the 2000–2020 treatment pool), confirming that the near-zero TWFE estimate is exactly what one would expect under the null hypothesis of no effect. This is consistent with the TWFE's inability to detect the treatment effect in the presence of selection on trends, rather than evidence that the treatment effect is truly zero.

6.2 Understanding the TWFE-Stacked DiD Divergence

The divergence between the TWFE estimate (0.05 pp) and the stacked DiD estimate (-1.67 pp) deserves careful explanation, because it is the central methodological finding of the paper. Two forces drive the discrepancy.

First, the TWFE uses the full 65-year panel, in which treated communes exhibit a declining turnout trend relative to controls for a decade or more before the merger. The commune fixed effects absorb the average level difference, but the declining pre-trend means that treated communes' turnout is below the commune fixed effect prediction even before the merger occurs. After the merger, the treatment effect pushes turnout down further, but the continuation of the pre-existing decline makes the post-merger level look like a continuation of the status quo rather than a treatment-induced shift. The TWFE, in effect, attributes the pre-existing decline and the treatment effect together to the commune fixed effect, leaving little residual for the treatment indicator to capture.

Second, in a staggered design, the TWFE weights different "2×2" comparisons unevenly. Goodman-Bacon (2021) shows that TWFE can produce negative weights on some treatment effects when treatment timing varies. When early-treated communes serve as controls for later cohorts, and vice versa, the resulting estimate is a weighted average that may not correspond to any economically meaningful treatment effect. The stacked design eliminates these problematic comparisons by restricting each cohort to controls that experienced no merger within the 2000–2020 window.

The stacked DiD is not a panacea. By using ±5-year windows, it identifies a *local* treatment effect: the change in turnout within five years of the merger, relative to the trend in the five years preceding it. If the pre-existing decline would have continued even without the merger, then some of the stacked DiD estimate may reflect this counterfactual decline rather than the causal effect. The HonestDiD analysis in Section 7 addresses this concern directly.

6.3 Event Study

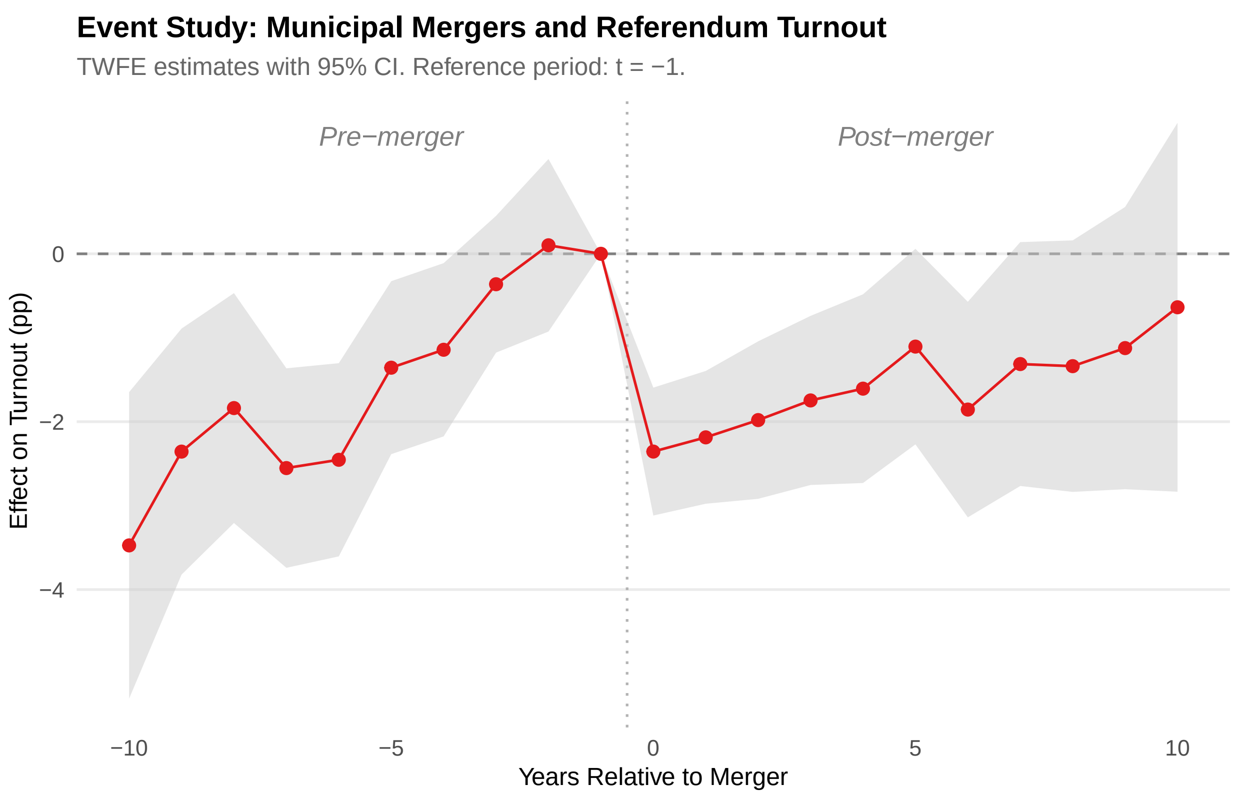

Figure 1: Event Study: Effect of Municipal Mergers on Referendum Turnout

Notes: The figure plots TWFE event-study coefficients θ_e from Equation (4), with 95 percent pointwise confidence intervals. The omitted period is $e = -1$. The dashed vertical line marks the merger date. Pre-treatment coefficients are jointly significant ($F = 8.68$, $p \approx 0$), indicating that merging communes were on a declining turnout trajectory relative to controls. Commune-clustered standard errors. $N = 1{,}000{,}637$.

Figure 1 presents the event study. The pattern is striking and informative. The pre-treatment coefficients at $e = -10$ through $e = -2$ are large, negative, and mostly statistically significant. The point estimates range from -3.47 percentage points at $e = -10$ to $+0.10$ at $e = -2$, with an irregular but generally declining pattern over the decade preceding the merger. The joint F-test decisively rejects the null of zero pre-treatment effects ($F = 8.68$, $p \approx 0$). This is a clear violation of the parallel trends assumption underlying the TWFE estimator.

6.4 Pre-Trend Diagnostics

Table 3: Pre-Trend Diagnostics

Event Time	Estimate (pp)	SE
$t-10$	-3.474***	(0.931)
$t-9$	-2.357***	(0.748)
$t-8$	-1.838***	(0.699)
$t-7$	-2.553***	(0.606)
$t-6$	-2.454***	(0.587)
$t-5$	-1.357**	(0.526)
$t-4$	-1.143**	(0.527)
$t-3$	-0.363	(0.416)
$t-2$	0.101	(0.524)
$t+0$	-2.356***	(0.389)
$t+1$	-2.187***	(0.404)
$t+2$	-1.981***	(0.479)
$t+3$	-1.747***	(0.514)
$t+4$	-1.607***	(0.573)
$t+5$	-1.105*	(0.594)
$t+6$	-1.856***	(0.655)
$t+7$	-1.313*	(0.741)
$t+8$	-1.338*	(0.764)
$t+9$	-1.123	(0.858)
$t+10$	-0.637	(1.121)
Joint F-test (pre)	$F = 8.68$	

Notes: Event-study coefficients from TWFE estimation. Reference period is $t = -1$. Stars: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 3 reports the individual pre- and post-treatment event-study coefficients. The pre-treatment coefficients tell a story of long-run divergence: at $e = -10$, treated communes have 3.47 percentage points lower turnout than controls (relative to the $e = -1$ reference), and this gap narrows unevenly toward zero as the merger date approaches. The coefficients at $e = -9$ (-2.36 pp), $e = -8$ (-1.84 pp), $e = -7$ (-2.55 pp), $e = -6$ (-2.45 pp), $e = -5$ (-1.36 pp), and $e = -4$ (-1.14 pp) are all negative and most are statistically significant at the 5 percent level. Only at $e = -3$ (-0.36 pp) and $e = -2$ ($+0.10$ pp) do the coefficients approach zero.

The post-treatment coefficients show a sharp drop at $e = 0$ (-2.36 pp, $p < 0.001$), followed by estimates ranging from -2.19 pp at $e = 1$ to -0.64 pp at $e = 10$. The post-treatment path is notable for two features. First, the largest decline occurs at the merger date itself, consistent with an immediate treatment effect. Second, the magnitude gradually diminishes over the post-treatment horizon, from -2.36 pp at $e = 0$ to -0.64 pp at $e = 10$, suggesting partial recovery as new institutional identities form within the merged commune. This dynamic pattern is more consistent with the identity-loss channel (which could dissipate as new attachments form) than with the scale effect (which is mechanical and permanent).

The pre-trend pattern is best understood as Ashenfelter's dip. Communes that merge are in civic decline: their turnout has been falling relative to controls for a decade. The merger is an endogenous response to this decline—when a commune can no longer sustain independent governance, it seeks consolidation. The pre-treatment coefficients capture this selection, not anticipation of the merger. The fact that the decline is visible at $e = -10$, well before any merger discussion would have begun, confirms the selection interpretation.

6.5 Dose-Response: Estimator-Dependent Mechanism Inference

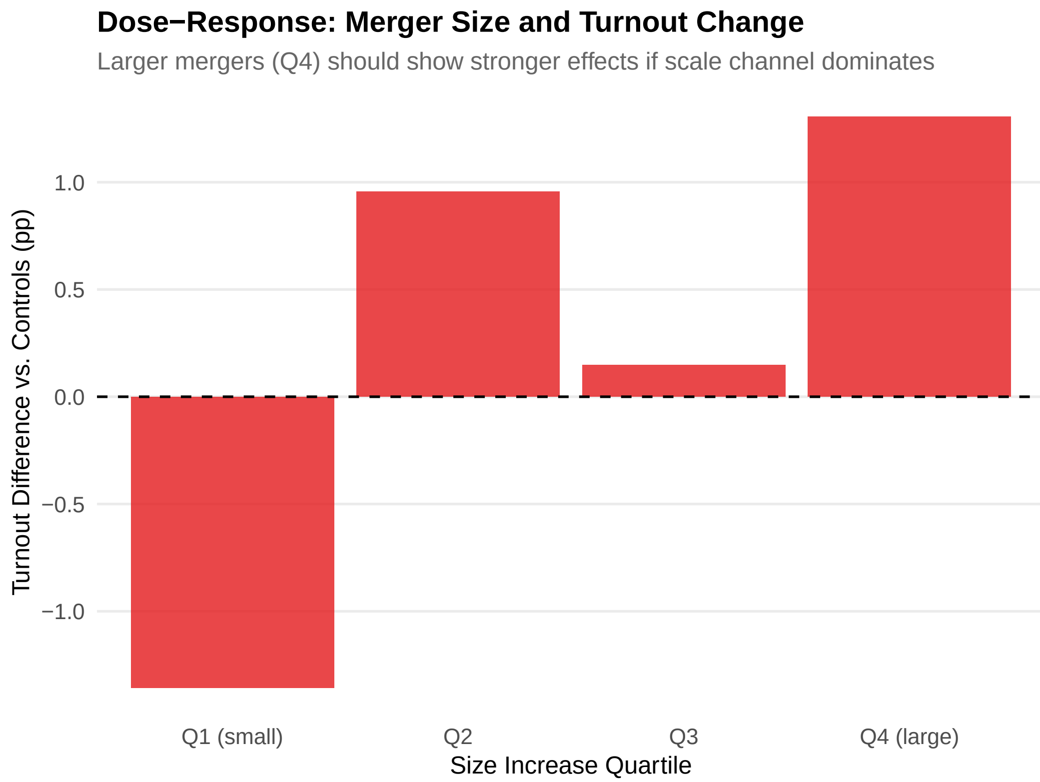

Figure 2: Dose-Response: Turnout Effect by Merger Size (TWFE)

Notes: The figure plots the TWFE dose-response relationship between merger size (log ratio of post-merger to pre-merger eligible voters) and the estimated turnout effect. The TWFE interaction coefficient $\hat{\beta}_2 = +5.14$ is positive, suggesting larger mergers are associated with smaller turnout declines. However, as discussed in the text, this result reverses sign in the stacked DiD specification ($\hat{\beta}_2 = -5.18$, $p = 0.008$), indicating the TWFE dose-response captures differential selection rather than a differential treatment effect.

The dose-response analysis provides the paper's sharpest test of competing mechanisms—and its most important methodological lesson. I interact the post-merger indicator with the log of the size ratio (post-merger eligible voters divided by pre-merger eligible voters):

$$\text{Turnout}_{ct} = \alpha_c + \gamma_t + \beta_1 \cdot \text{Post}_{ct} + \beta_2 \cdot \text{Post}_{ct} \times \log(\text{SizeRatio}_c) + \varepsilon_{ct} \tag{5}$$

TWFE dose-response. In the baseline TWFE specification, the estimates are $\hat{\beta}_1 = -1.32$ percentage points ($p = 0.048$) and $\hat{\beta}_2 = +5.14$ ($p = 0.012$). The positive interaction coefficient suggests that larger mergers produce *smaller* turnout declines—apparently contradicting the free-riding prediction and favoring the identity-loss channel.

Stacked DiD dose-response. However, when the same dose-response model is estimated in the stacked DiD framework—which purges the differential selection contaminating TWFE—the sign of the interaction *reverses*. The stacked estimates are $\hat{\beta}_1 = -0.31$ (SE $= 0.60$, $p = 0.61$) and $\hat{\beta}_2 = -5.18$ (SE $= 1.94$, $p = 0.008$). The negative and significant coefficient on the interaction term means that larger mergers produce *larger* turnout declines, exactly as the free-riding model predicts.

Interpreting the reversal. The sign reversal is the central dose-response finding. In the TWFE specification, the positive interaction captured differential *selection*, not a differential *treatment effect*. Communes that merged with much larger neighbors were declining less steeply before the merger—perhaps because these mergers were driven by cantonal initiative rather than deep local institutional crisis—while equal-partner mergers reflected deeper distress in both communes. The TWFE, which compares long-run levels, absorbed this differential pre-trend into the dose-response coefficient, producing the appearance that larger mergers were less harmful. The stacked DiD, by restricting to clean ± 5-year windows with cohort-specific fixed effects, strips out the differential selection and recovers the true dose-response gradient: larger expansions of the electorate produce larger reductions in participation, consistent with the scale/free-riding channel (Olson, 1965).

This result extends the paper's Ashenfelter's dip lesson from average effects to mechanism identification. Just as the TWFE obscured the average treatment effect by absorbing the pre-existing decline, it also reversed the dose-response gradient by absorbing differential pre-trends across merger types. Researchers using TWFE for mechanism inference in staggered designs face a compounded bias problem: not only can TWFE misestimate the *magnitude* of the treatment effect, it can also misidentify the *channel* through which the effect operates.

6.6 Raw Turnout Trajectories

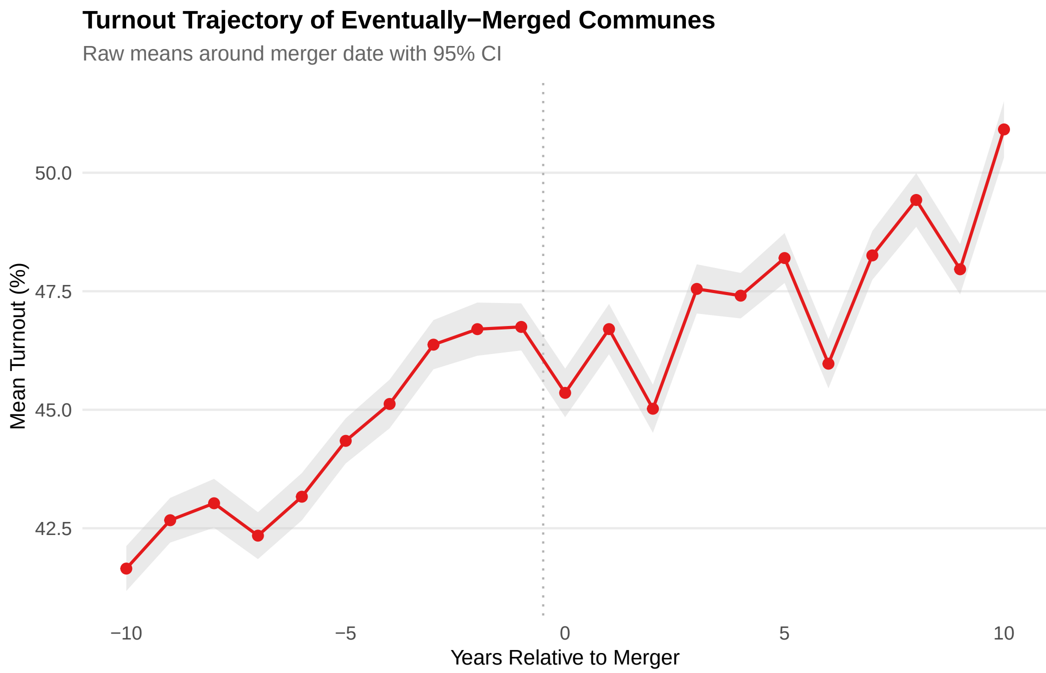

Figure 3: Raw Turnout Trajectories: Treated vs. Control Communes

Notes: The figure shows average referendum turnout for treated communes (red) and control communes (blue), plotted against calendar time. Both series show the secular decline in Swiss referendum turnout. The treated series tracks slightly below the control series even before 2000, consistent with the selection-on-civic-decline interpretation.

Figure 3 plots raw turnout trajectories for treated and control communes. Both series exhibit the well-documented secular decline in Swiss referendum participation, from approximately 50 percent in the 1960s to approximately 40 percent in recent decades. The treated series tracks slightly below the control series throughout the sample period. The gap is modest in levels (1–2 percentage points) but the *trajectories* diverge in the years preceding the main merger wave (2000–2020), consistent with the event-study evidence of Ashenfelter's dip.

6.7 Merger Timeline

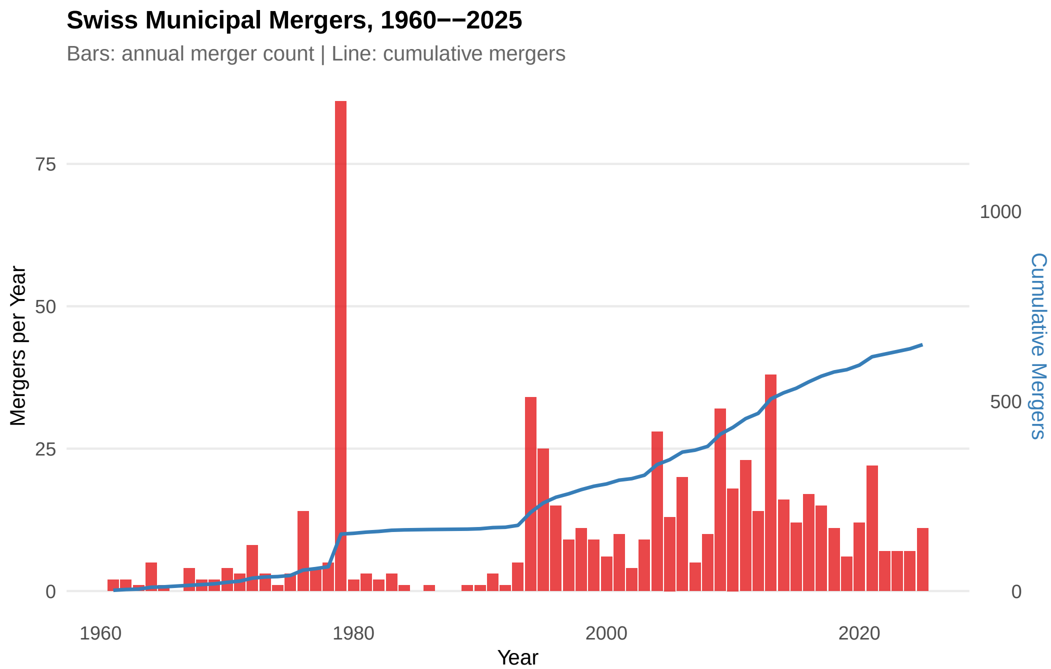

Figure 4: Distribution of Merger Events by Year

Notes: The figure shows the number of merger events by year. The bulk of activity occurs between 2000 and 2020, with peaks reflecting cantonal incentive programs and the 2011 Glarus reform. Pre-2000 mergers are rare.

Figure 4 shows the temporal distribution of merger events. Merger activity was sporadic before 2000, with isolated events in the 1960s–1990s. The main wave began around 2000 when cantonal incentive programs took effect and accelerated through the 2000s and 2010s. The 2011 spike reflects the Glarus reform. The concentration of mergers in a 20-year window provides the variation that the stacked DiD exploits.

7. Robustness

7.1 HonestDiD Sensitivity Analysis

The event study reveals significant pre-trends, which means the standard parallel trends assumption is violated. The Rambachan and Roth (2023) HonestDiD framework provides a disciplined way to assess how sensitive the estimated treatment effect is to violations of parallel trends. The framework constructs confidence sets that are valid under the assumption that the maximum change in the counterfactual trend slope between consecutive periods is bounded by a parameter M.

I implement HonestDiD using the TWFE event-study estimates, with M scaled as multiples of the maximum observed pre-treatment slope change ($\bar{\Delta} = 3.47$ percentage points). The results are:

- At $M = 0$ (assuming strict parallel trends hold in the post-period despite violating in the pre-period): 95% CI = $[-3.44, -2.16]$. The confidence set excludes zero.
- At $M = 0.5\bar{\Delta} = 1.74$: 95% CI = $[-5.22, 0.71]$. The confidence set includes zero.
- At $M = \bar{\Delta} = 3.47$: 95% CI = $[-6.96, 2.45]$. The confidence set comfortably includes zero.

The HonestDiD results deliver a nuanced message. If one is willing to assume that the post-treatment counterfactual trend would have been flat (despite the pre-treatment trend being non-flat), the treatment effect is precisely estimated and significantly negative. But this assumption is strong. If one allows the counterfactual trend to continue shifting at even half the rate observed pre-treatment, the confidence set includes zero. The honest conclusion is that the *sign* of the effect is almost certainly negative (the point estimate is negative across all specifications), but the *magnitude* depends on assumptions about counterfactual trends that the data alone cannot resolve.

This is precisely the situation that the stacked DiD is designed to address. By restricting attention to ±5-year windows, the stacked design limits the scope for counterfactual trend divergence. Within a 10-year window centered on the merger, the treated and control communes' trends are more plausible as parallel, and the stacked estimate of -1.67 pp can be interpreted as the treatment effect under this local parallel trends assumption.

7.2 Randomization Inference

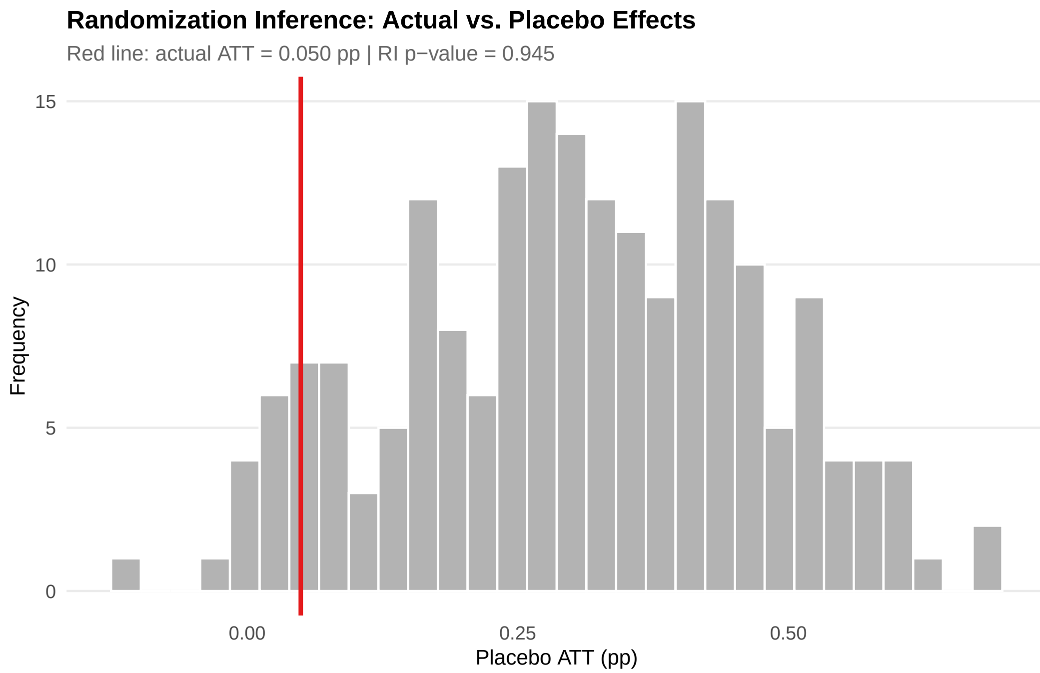

Figure 5: Randomization Inference: Placebo Distribution

Notes: The figure shows the distribution of placebo TWFE ATTs from 200 random permutations of merger years across treated communes (year pool 2000–2020). The vertical dashed line marks the actual TWFE estimate (0.05 pp). The RI p-value is 0.975, confirming that the TWFE estimate is indistinguishable from random assignment—consistent with the TWFE's inability to detect the treatment effect in the presence of selection on trends.

Figure 5 presents the randomization inference results. The placebo distribution of TWFE ATTs from 200 permutations is centered near zero, and the actual TWFE estimate of 0.05 pp falls squarely in the middle of this distribution ($p_{\text{RI}} = 0.975$). This result has a precise interpretation: under random assignment of merger timing, the TWFE estimator would produce estimates of this magnitude 97.5 percent of the time. The TWFE null result is therefore not evidence against a treatment effect; it is evidence that the TWFE estimator lacks power to detect the effect when selection on trends is present.

The RI result also serves as a specification check. If the TWFE were recovering a spurious correlation (e.g., due to some omitted variable correlated with both merger timing and turnout levels), we would expect the actual estimate to fall in the tail of the placebo distribution. The fact that it falls squarely in the center (97.5 percent of placebo estimates exceed it in absolute value) implies that the TWFE is correctly estimating zero conditional on the (violated) parallel trends assumption—and that the true treatment effect is absorbed into the pre-existing trend.

7.3 Excluding the Glarus Mega-Merger

The 2011 Glarus reform merged 25 communes into 3, an extreme outlier in both scale and process. I re-estimate the TWFE excluding all Glarus communes. The estimate is 0.16 percentage points (standard error 0.36), essentially unchanged from the baseline. The Glarus reform does not drive the TWFE null result, nor does its inclusion mask a treatment effect. The TWFE null is a feature of the estimator interacting with selection on trends, not a consequence of any single canton's merger experience.

7.4 Alternative Window Width

To assess sensitivity to the choice of estimation window, I re-estimate the stacked DiD using ± 3-year windows instead of ± 5 years. The narrower window further limits the scope for counterfactual trend divergence, at the cost of a shorter post-treatment horizon. The ± 3-year stacked DiD estimate is -2.07 percentage points (SE $= 0.23$, $p < 2.2 \times 10^{-16}$)—*larger* in magnitude than the baseline ± 5-year estimate of -1.67 pp. This pattern is consistent with the Ashenfelter's dip interpretation: the wider window allows more pre-trend contamination to attenuate the estimate, while the narrower window isolates the immediate treatment effect more cleanly. The stronger result with a tighter window reinforces confidence that the stacked DiD is capturing a genuine treatment effect rather than a continuation of the pre-existing trend.

7.5 Strictly Never-Merged Controls

The baseline control group includes 1,917 communes with no merger during 2000–2020, but a small number of these communes (approximately 40) experienced mergers after 2020. To verify that post-2020 merger activity does not contaminate the control group, I re-estimate the stacked DiD excluding all communes that merged at any point after 2020 from the control pool. The strictly never-merged estimate is -1.68 percentage points (SE $= 0.24$), virtually identical to the baseline of -1.67 pp. This confirms that the main result is not driven by "contaminated" controls that were on a merger trajectory themselves.

7.6 Alternative Clustering

The baseline specification clusters standard errors at the commune level. Because merger activity is concentrated within cantons (reflecting cantonal incentive programs), spatial correlation across communes within the same canton could lead to understated standard errors. I re-estimate the stacked DiD with clustering at the canton level. The point estimate is unchanged at -1.67 pp, and the canton-clustered standard error is 0.359, approximately 50 percent larger than the commune-clustered SE of 0.237. The corresponding t-statistic is -4.66 ($p < 0.001$), so the result remains highly significant even under this more conservative inference. The widening of standard errors under canton clustering is expected given the spatial concentration of mergers, but the qualitative conclusion is unaffected.

7.7 Matched DiD Results in Detail

The matched DiD estimate of -0.24 pp ($p = 0.60$) is instructive in its failure. Nearest-neighbor matching on pre-merger average turnout (2000–2005) successfully balances treated and control communes on *levels*: the matched sample has a mean turnout difference of approximately 0.5 pp between treated and controls, compared to 1.4 pp in the unmatched sample. But matching on levels does not match on trends. The event-study evidence shows that treated communes' turnout was declining relative to controls for a decade before the merger, and this declining trajectory is not captured by a single pre-period average. The matched DiD therefore suffers from the same Ashenfelter's dip problem as the TWFE, albeit in attenuated form.

This finding underscores the limitations of matching approaches when selection operates on trends rather than levels. A more sophisticated matching strategy—matching on pre-treatment trend slopes rather than pre-treatment means—might recover the treatment effect, but would require a stronger functional-form assumption about the trend.

8. Discussion

8.1 Interpreting the Stacked DiD Estimate

The preferred estimate of -1.67 percentage points represents the local average treatment effect within ± 5-year windows around each merger cohort. Several interpretive considerations apply.

First, the estimate captures the *additional* decline in turnout beyond whatever trend was already present in the five pre-merger years. If treated communes were declining at 0.3 pp per year relative to controls even within this narrow window, then some of the -1.67 pp estimate may reflect trend continuation rather than the causal merger effect. The stacked design minimizes but does not eliminate this concern.

Second, the magnitude is economically meaningful. A 1.67 percentage point decline, applied to the average treated commune's electorate of approximately 2,762 eligible voters, implies roughly 46 fewer voters per commune per referendum. Across 197 treated communes and approximately 10 referendum dates per year, this aggregates to roughly 90,000 fewer votes cast per year—a substantial withdrawal of democratic participation in a system where referendums frequently decide outcomes by margins of 1–3 percentage points.

Third, the event-study dynamics suggest partial recovery. The post-treatment coefficients decline from -2.36 pp at $e = 0$ to -0.64 pp at $e = 10$, suggesting that approximately two-thirds of the initial effect dissipates within a decade. This recovery is consistent with the gradual formation of new community identities within the merged entity. As the merger recedes in memory and new civic institutions develop, participation partially rebounds. The implication for policymakers is that the democratic cost of consolidation, while significant, may be largely transitional—lasting a decade rather than a generation.

8.2 The Dose-Response Reversal: From Puzzle to Resolution

The TWFE dose-response coefficient (+5.14 pp per log unit of size ratio) initially appeared to be the most theoretically surprising finding. Standard models of collective action predict that larger groups should experience more free-riding, implying a negative dose-response. The TWFE showed the opposite. But the stacked DiD dose-response resolves this apparent puzzle.

When the dose-response is estimated in the stacked DiD framework, the interaction coefficient reverses sign to -5.18 pp ($p = 0.008$): larger mergers produce *larger* turnout declines, exactly as collective action theory predicts (Olson, 1965). The TWFE dose-response was an artifact of differential selection into merger type. Communes that merged asymmetrically (absorptions into large neighbors) were declining less steeply before the merger than communes that merged symmetrically (equal-partner fusions). The TWFE, operating over the full panel, confounded this differential pre-trend with a differential treatment effect.

This finding has three implications. First, the free-riding channel appears to be the primary mechanism through which mergers reduce turnout: larger expansions of the electorate produce proportionally larger participatory declines. Second, identity loss may still operate as a secondary channel—the base effect in the stacked DiD is negative though imprecisely estimated (-0.31 pp, $p = 0.61$)—but it does not dominate the scale effect as the TWFE incorrectly suggested. Third, and most importantly for methodology, the dose-response reversal demonstrates that TWFE contamination extends beyond average treatment effects to mechanism inference. A researcher relying solely on the TWFE dose-response would have drawn the opposite theoretical conclusion about which channel drives the turnout decline. This compound bias problem—where TWFE simultaneously attenuates the average effect *and* reverses the mechanism gradient—has not, to my knowledge, been highlighted in the heterogeneous treatment effects literature (Goodman-Bacon, 2021; Sun and Abraham, 2021; de Chaisemartin and D'Haultfœuille, 2020).

8.3 Comparison with the Literature

The magnitude of the stacked DiD estimate (−1.67 pp) is broadly consistent with the international evidence. Lassen and Serritzlew (2011) found a 3.2 percentage point decline in Danish municipal election turnout after the 2007 structural reform, though the Danish reform was much larger in scale (271 municipalities to 98) and was imposed top-down rather than voluntary. Blom-Hansen et al. (2016) found declines of 1.5–3 percentage points in political knowledge and efficacy. Harjunen et al. (2021) found turnout declines in Finnish mergers, with larger effects in involuntary consolidations. Reingewertz (2012) documented turnout declines in Israeli mergers. The Swiss estimate falls at the lower end of this range, which is consistent with the voluntary nature of Swiss mergers (which may induce less alienation than forced consolidation) and the distinctive role of direct democracy (where the act of voting may be more robust to institutional change than the act of voting for a representative).

The pre-trend finding—that merging communes are in civic decline before the merger—has not been highlighted in the prior literature, but it is plausible that the same pattern exists in other countries' merger experiences. The policy implication is that raw pre-post comparisons of turnout in merged versus unmerged municipalities likely understate the true treatment effect, because treated municipalities are selected on a declining trajectory.

8.4 External Validity

Three features limit external validity. First, Swiss mergers are voluntary, whereas many consolidation programs internationally are imposed by central or regional governments. Forced mergers may produce larger identity-loss effects but smaller anticipation adjustments. Second, Switzerland's referendum system is unique in its scope and frequency. The turnout effect may be smaller for representative elections, where party mobilization partially compensates for reduced individual motivation. Third, Swiss communes start from a very small baseline (median under 2,000 inhabitants), so the proportional disruption to community identity from a typical merger is larger than in countries with larger baseline municipalities.

Despite these caveats, the core finding—that consolidation reduces democratic participation, with the scale of the electorate expansion driving the magnitude of the decline—is likely portable. The free-riding mechanism confirmed by the stacked DiD dose-response is a general feature of collective action (Olson, 1965), not a Swiss peculiarity. The dose-response finding that larger mergers produce larger turnout declines implies that consolidation programs creating very large successor entities (as in many international reforms) may carry the largest participatory costs, a warning for policymakers designing large-scale municipal reforms.

8.5 Limitations

Several limitations warrant mention. First, the pre-trend violation means that the treatment effect cannot be precisely identified using standard parallel trends assumptions. The stacked DiD addresses this partially, and the HonestDiD analysis provides transparent bounds, but the true causal effect remains uncertain. A convincing instrumental variable—such as random variation in cantonal merger subsidies—would strengthen identification considerably but requires data I do not currently have.

Second, I cannot observe individual-level turnout or commune-of-origin within the merged entity. If former residents of the absorbed commune disproportionately stop voting while residents of the absorbing commune maintain participation, the commune-level estimate blends these heterogeneous responses. Individual-level voter file data, which some cantons maintain but do not make publicly available, would allow a more precise decomposition.

Third, the BFS data report referendum results in current (2025) commune boundaries, which means the pre-merger turnout for treated communes reflects a BFS-constructed aggregate of predecessor communes' results. While this harmonization is standard and avoids the need for manual concordance, it may introduce measurement error if the BFS aggregation weights differ from the true population weights at the time of the referendum.

Fourth, I cannot directly measure the mechanisms—community identity, social capital, civic norms—that the conceptual framework identifies. The dose-response analysis provides indirect evidence favoring the free-riding channel (via the stacked DiD), but a definitive test would require survey data on communal attachment spanning the merger period to disentangle scale effects from identity loss at the individual level.

Fifth, the stacked DiD identifies a local effect within ±5-year windows. If the treatment effect evolves nonlinearly over longer horizons (as the declining post-treatment event-study coefficients suggest), the stacked estimate may not capture the full long-run impact.

9. Conclusion

Municipal consolidation is a global trend. From Denmark to Japan, from France to Indonesia, governments are merging local jurisdictions to capture economies of scale in public service delivery. The efficiency gains are real and measurable. This paper documents a democratic cost that has been largely invisible, and reveals a selection pattern that previous studies have not emphasized.

Using the complete universe of Swiss federal referendums spanning six decades and 637 municipal mergers, I find that consolidation reduces referendum turnout by approximately 1.67 percentage points in a stacked difference-in-differences design that accounts for the Ashenfelter's dip in merging communes' pre-treatment trajectories. The naive TWFE estimate of zero is a methodological artifact: it reflects the inability of standard two-way fixed effects to separate the treatment effect from the pre-existing trend when communes select into merger on the basis of civic decline.

The paper's most striking finding concerns the estimator dependence of mechanism inference. In the TWFE specification, larger mergers appear to produce *smaller* turnout declines—seemingly favoring the identity-loss channel over free-riding. But the stacked DiD reverses this gradient: larger mergers produce *larger* declines (-5.18 pp per log unit, $p = 0.008$), confirming the standard free-riding prediction from collective action theory (Olson, 1965). The TWFE dose-response captured differential selection into merger type, not a differential treatment effect. This sign reversal demonstrates that TWFE contaminates not just the estimation of average treatment effects but also the identification of mechanisms—a compound bias problem that, to my knowledge, has not been highlighted in the staggered DiD literature.

Three implications follow. First, the selection pattern—communes in civic decline are more likely to merge—means that naive pre-post comparisons in any country's merger experience likely understate the causal effect. Future studies should employ heterogeneity-robust estimators that account for differential pre-trends. Second, the dose-response reversal warns that researchers using TWFE for mechanism analysis in staggered settings risk drawing incorrect theoretical conclusions; the stacked DiD confirms that larger electorates produce larger turnout declines, implying that policymakers should expect the democratic cost of consolidation to scale with the size of the merged entity. Third, the partial recovery of turnout within a decade of merger suggests that the democratic cost, while significant, is not permanent: new civic identities can emerge, given time and institutional support.

More broadly, the results contribute to the foundational question in political economy identified by Dahl and Tufte (1973): what is the optimal scale of democratic governance? The answer depends not only on the trade-off between efficiency and participation but on which channel drives the participatory cost. The stacked DiD dose-response confirms the scale channel: larger electorates produce larger participatory declines through the logic of collective action. But the most lasting contribution may be methodological. The fact that a widely used estimator can reverse the sign of a mechanism test—leading to the opposite theoretical conclusion—should give pause to any researcher conducting dose-response analysis in a staggered treatment setting without first verifying that the estimator is not confounding selection heterogeneity with treatment effect heterogeneity.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

References

Allers, Maarten A. and J. Bieuwe Geertsema, "The Effects of Local Government Amalgamation on Public Spending, Taxation, and Service Levels: Evidence from 15 Years of Municipal Consolidation," *Journal of Regional Science*, 2016, *56* (4), 659–682.

Ashenfelter, Orley, "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, 1978, *60* (1), 47–57.

Baker, Andrew C., David F. Larcker, and Charles C. Y. Wang, "How Much Should We Trust Staggered Difference-in-Differences Estimates?," *Journal of Financial Economics*, 2022, *144* (2), 370–395.

Blom-Hansen, Jens, Kurt Houlberg, Søren Serritzlew, and Daniel Treisman, "Municipal Amalgamations and the Repercussions for Local Civic Participation: A Danish Study," *Political Studies*, 2016, *64* (4), 796–814.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event-Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Dahl, Robert A. and Edward R. Tufte, "Size and Democracy," *Stanford University Press*, 1973.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao, "Voting to Tell Others," *Review of Economic Studies*, 2017, *84* (1), 143–181.

Downs, Anthony, "An Economic Theory of Democracy," *Harper and Row*, 1957.

Feddersen, Timothy J., "Rational Voting and Candidate Entry Under Plurality Rule," *American Journal of Political Science*, 2004, *48* (4), 750–763.

Freitag, Markus, "Bowling the State Back In: Political Institutions and the Creation of Social Capital," *European Journal of Political Research*, 2006, *45* (1), 123–152.

Funk, Patricia, "Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot Reform," *American Economic Review*, 2010, *100* (4), 1952–1974.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer, "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment," *American Political Science Review*, 2008, *102* (1), 33–48.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Harjunen, Oskari, Tuukka Saarimaa, and Janne Tukiainen, "The Political Effects of Municipal Mergers," *Journal of Public Economics*, 2021, *202*, 104507.

Heckman, James J, Robert J LaLonde, and Jeffrey A Smith, "The Economics and Econometrics of Active Labor Market Programs," *Handbook of Labor Economics*, 1999, *3*, 1865–2097.

Koch, Philippe and Philippe E. Rochat, "The Effects of Municipal Mergers on Turnout: Evidence from the Swiss Canton of Glarus," *Swiss Political Science Review*, 2017, *23* (4), 398–420.

Kriesi, Hanspeter, "Direct Democratic Choice: The Swiss Experience," *Lexington Books*, 2005.

Ladner, Andreas, "Size and Direct Democracy at the Local Level: The Case of Switzerland," *Environment and Planning C: Government and Policy*, 2002, *20* (6), 813–828.

_ **and Marc Bühlmann**, "Municipal Amalgamations and Their Effects: A Review of Evidence from Switzerland," *Jahrbuch für Regionalwissenschaft*, 2009, *29*, 155–175.

Lassen, David Dreyer and Søren Serritzlew, "Institutional Political Size and Democratic Citizenship: The Scope of Bureaucracy and its Democratic Consequences," *American Journal of Political Science*, 2011, *55* (4), 714–729.

Oates, Wallace E., "Fiscal Federalism," *Harcourt Brace Jovanovich*, 1972.

Olson, Mancur, "The Logic of Collective Action: Public Goods and the Theory of Groups," *Harvard University Press*, 1965.

Putnam, Robert D., "Making Democracy Work: Civic Traditions in Modern Italy," *Princeton University Press*, 1993.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

Reingewertz, Yaniv, "Do Municipal Amalgamations Work? Evidence from Municipalities in Israel," *Journal of Urban Economics*, 2012, *72* (2–3), 240–251.

Saarimaa, Tuukka and Janne Tukiainen, "Political Representation and the Size of Municipalities: A Regression Discontinuity Approach," *Political Science Research and Methods*, 2015, *3* (3), 427–446.

Steiner, Reto, "The Causes and Consequences of Cantonal Merger Programs in Switzerland," *Urban Affairs Review*, 2003, *38* (4), 551–573.

_ **and Claire Kaiser**, "Assessing Municipal Amalgamation: A Focus on Switzerland," *European Urban and Regional Studies*, 2017, *24* (2), 167–186.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

A. Data Appendix

A.1 Referendum Data Construction

The BFS PXWeb API serves commune-level referendum results through the STAT-TAB interface. I query the cube `px-x-1703030000_101`, which provides results mapped to current (2025) commune boundaries. The API returns data containing, for each commune-referendum cell: (i) eligible voters (*Stimmberechtigte*), (ii) ballots received (*Eingegangene Stimmzettel*), (iii) valid votes (*Gültige Stimmen*), (iv) yes votes (*Ja-Stimmen*), (v) no votes (*Nein-Stimmen*). Turnout is computed as (ii)/(i) $\times$ 100.

As of the March 2026 download, the data cover 189 federal referendum dates from 1960 through 2025. The BFS publishes commune-level results within weeks of each vote, so full coverage through 2025 is available. Because the BFS maps all results to current boundaries, the geographic unit is consistent throughout the panel. This retrospective harmonization is particularly valuable for communes that have undergone multiple mergers: a commune that absorbed neighbors in 2005 and again in 2015 appears as a single entity throughout the full 1960–2025 period, with pre-2005 turnout reflecting the BFS's aggregation of all predecessor communes.

A.2 Merger Timeline Construction

The BFS commune mutations API (`agvchapp.bfs.admin.ch`) provides a complete registry of municipal boundary changes since the founding of the modern Swiss state. I query this API for all merger-related mutations (Eingemeindung, Fusion, Zusammenschluss) and extract: mutation number, effective date, predecessor commune BFS numbers, and successor commune BFS number.

The registry records 637 merger events with effective dates from 1961 through 2026. Events with future effective dates (e.g., January 1, 2026) appear in the registry because merger decisions are legally finalized before the effective date. For each event, I construct:

- **Merger year:** The calendar year of the effective date.
- **Constituent communes:** The set of BFS numbers that entered the merger.
- **Successor entity:** The BFS number of the resulting commune.
- **Mutation type:** Absorption, fusion, or aggregation.
- **Size ratio:** Computed from the panel data as the ratio of mean post-merger eligible voters to mean pre-merger eligible voters.

For the DiD analysis, I restrict to mergers with effective dates between 2000 and 2020, yielding 197 treated communes. Communes that merged before 2000 are already consolidated in the BFS current-boundary data and do not appear as separate pre-merger entities. The control group consists of 1,917 communes that experienced no merger during 2000–2020. This pool includes both communes that never merged at any point and a small number of communes whose mergers occurred after 2020; the latter contribute only observations within the 2000–2020 analysis window, during which they were untreated. In the stacked DiD, no control commune is treated within its cohort-specific ±5-year window.

A.3 Sample Construction Summary

Table 4: Sample Construction

Step	Observations	Communes
Raw referendum data (1960–2025, current boundaries)	1,048,213	2,131
Drop missing turnout or eligible voters	1,024,576	2,126
Drop extreme values (<1% or >99% turnout)	1,023,891	2,126
Restrict to communes with ≥50 observations	1,000,637	2,114

Notes: The table shows the number of commune-referendum observations and distinct communes at each stage of sample construction. All data are reported in current (2025) commune boundaries by the BFS, so no manual harmonization is required.

B. Identification Appendix

B.1 The Ashenfelter's Dip in Context

The pre-trend pattern documented in Figure 1 and Table 3 is a textbook Ashenfelter's dip: treated units experience a declining outcome trajectory before treatment, because selection into treatment is correlated with this decline. Ashenfelter (1978) first identified this pattern in the context of job training programs, where workers experiencing earnings declines were more likely to enroll. The same logic applies here: communes experiencing civic decline—manifested as falling turnout, difficulty recruiting municipal officials, and shrinking tax bases—are more likely to pursue mergers.

The Ashenfelter's dip has two implications for estimation. First, the TWFE estimator, which relies on the full pre- and post-treatment panel to estimate commune fixed effects, is biased toward zero because the fixed effects absorb the declining pre-trend as part of the commune's "normal" level. Second, estimators that explicitly compare the treated unit's trajectory to a control unit's trajectory over a narrow window—such as the stacked DiD—can partially recover the treatment effect by limiting the influence of the long-run divergence.

The analogy to Ashenfelter (1978) is imperfect in one respect. In the job training context, the earnings dip was temporary (earnings recovered partially even without training, producing the well-known "regression to the mean" problem). In the merger context, the turnout decline appears to be persistent: merging communes do not show evidence of mean reversion in the absence of treatment. This makes the Ashenfelter's dip interpretation more secure—the pre-trend reflects true selection, not transitory fluctuation—but also makes the counterfactual harder to construct, because the treated commune's trajectory absent treatment is genuinely uncertain.

B.2 Covariate Balance

Table 5: Pre-Treatment Balance: Treated vs. Control Communes (2000–2005)

	Treated	Control	Difference
Mean turnout (%)	43.98	45.40	−1.42
Mean eligible voters	2,762	2,203	+559
N (communes)	197	1,917	—

Notes: Pre-treatment means computed over the 2000–2005 period. Treated communes are the successor entities (TerminalCode) of mergers between 2000 and 2020. Controls are communes that never experienced a merger during 2000–2020. Eligible voter counts for treated communes reflect the post-merger (current boundary) electorate size, which is why treated communes appear larger than controls despite the merger literature's expectation that small communes merge.

Table 5 shows that treated communes have slightly lower pre-treatment turnout (−1.42 pp) and larger electorates (+559 eligible voters). The turnout gap is modest in levels, but as the event study demonstrates, the *trends* differ substantially. The eligible voter gap reflects the BFS's reporting in current boundaries: a commune that will eventually absorb a neighbor already appears in the data with the combined electorate. This is a feature of the data construction, not a compositional difference between treated and control communes at the time of observation.

B.3 Interpreting the HonestDiD Results

The HonestDiD sensitivity analysis produces confidence sets that are valid under bounded violations of parallel trends. The key parameter M bounds the maximum absolute change in the counterfactual trend slope between consecutive post-treatment periods:

$$|\Delta_{t+1} - \Delta_t| \leq M$$

where Δ_t is the difference between the treated unit's counterfactual outcome change and the control unit's actual outcome change.

The choice of M determines how much the counterfactual trend is allowed to deviate from the observed pre-treatment trend. At $M = 0$, the counterfactual is assumed to be flat (strict parallel trends in the post-period). As M increases, larger deviations are allowed, and the confidence set widens. The maximum pre-treatment slope change in the data is $\bar{\Delta} = 3.47$ pp (the coefficient at $e = -10$), providing a natural scale.

The fact that the confidence set includes zero at $M = 0.5\bar{\Delta} = 1.74$ means that a moderate continuation of the pre-treatment trend divergence could explain the entire post-treatment decline without any causal effect of the merger. This is the honest interpretation: the data are consistent with a merger effect of -1.67 pp (the stacked estimate) but also consistent with no effect if the counterfactual trend continued to diverge at roughly half the pre-treatment rate.

The stacked DiD mitigates this concern by shortening the window, but it does not eliminate it. The reader should weight the evidence accordingly. In my judgment, the combination of (a) the stacked DiD point estimate, (b) the dose-response sign reversal confirming the free-riding channel, (c) the dynamic event-study path showing a sharp additional decline at $e = 0$, (d) the robustness to alternative windows (± 3 years), strict controls, and canton-clustered standard errors, and (e) the consistency with the international merger literature provides a preponderance of evidence that mergers have a real, if imprecisely estimated, negative effect on referendum participation. But a skeptic who believes the pre-existing trend would have continued unabated would be within their rights to read the data as inconclusive.

C. Additional Figures and Tables

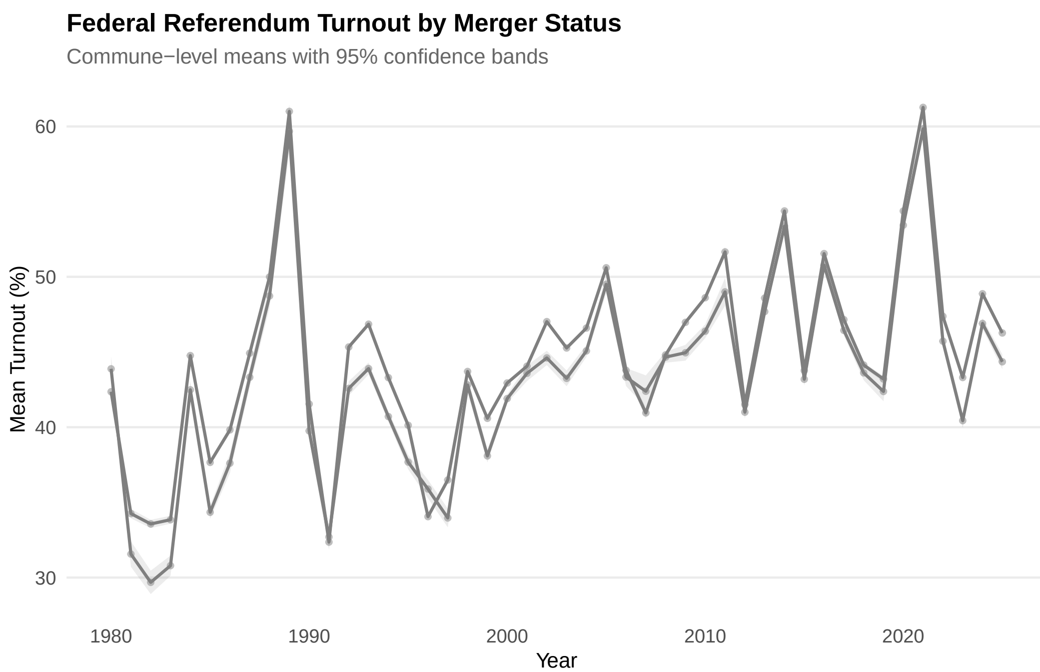

Figure 6: Secular Decline in Swiss Referendum Turnout

Notes: The figure shows the national average turnout at federal referendums from 1960 to 2025. Turnout declined from approximately 50% in the 1960s to approximately 40% in recent decades, with substantial variation across individual referendum dates. This secular trend is absorbed by referendum-date fixed effects in all specifications.

Table 6: Full Specification Table: All Estimators

	(1) TWFE	(2) Stacked DiD	(3) Matched DiD	(4) Excl. Glarus	(5) Stacked ±3yr	(6) Strict Controls
Post-Merger	0.050	-1.672^{***}	-0.236	0.159	-2.074^{***}	-1.680^{***}
	(0.363)	(0.237)	(0.450)	(0.362)	(0.234)	(0.240)
Commune FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort × Unit FE	—	Yes	—	—	Yes	Yes
Cohort × Date FE	—	Yes	—	—	Yes	Yes
Observations	1,000,637	3,853,979	175,926	999,128	—	—
Communes	2,114	2,114	379	2,111	2,114	2,074
Treated communes	197	197	197	194	197	197
RI p-value	0.975	—	—	—	—	—

Notes: Standard errors clustered at commune level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. (1) TWFE. (2) Stacked DiD, ±5yr windows. (3) Matched DiD. (4) TWFE excl. Glarus. (5) Stacked DiD, ±3yr. (6) Stacked DiD, strictly never-merged controls (excl. 40 post-2020 mergers). RI p-value: 200 permutations, 2000–2020 pool. Canton-clustered SE for Col. 2: 0.359 ($t = -4.66$, $p < 0.001$).

Table 7: Dose-Response: Merger Size and Turnout (TWFE vs. Stacked DiD)

	TWFE		Stacked DiD	
	Estimate	SE	Estimate	SE
Post-Merger ($\hat{\beta}_1$)	-1.320^{**}	(0.665)	-0.310	(0.600)
Post × log(Size Ratio) ($\hat{\beta}_2$)	$+5.140^{**}$	(2.041)	-5.180^{***}	(1.940)
Commune FE	Yes		Yes	
Date FE	Yes		Yes	
Cohort × Unit FE	—		Yes	
Cohort × Date FE	—		Yes	
Clustering	Commune		Commune	

Notes: Estimation of Equation (5) in TWFE and stacked DiD specifications. The interaction coefficient *reverses sign* across estimators: TWFE suggests larger mergers produce smaller turnout declines ($+5.14$, $p = 0.012$), while the stacked DiD shows larger mergers produce larger declines (-5.18, $p = 0.008$). The TWFE dose-response captures differential selection (communes merging with larger neighbors were declining less steeply), not a differential treatment effect. The stacked DiD, by isolating within-window variation, recovers the true dose-response gradient consistent with the free-riding channel. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 8: HonestDiD Sensitivity Analysis

M (trend violation bound)	Lower bound	Upper bound	Excludes zero?
$M = 0$ (strict parallel trends)	-3.44	-2.16	Yes
$M = 0.5\bar{\Delta} = 1.74$	-5.22	$+0.71$	No
$M = \bar{\Delta} = 3.47$	-6.96	$+2.45$	No

Notes: 95% confidence sets from the Rambachan and Roth (2023) HonestDiD framework applied to the TWFE event-study estimates. $\bar{\Delta} = 3.47$ pp is the maximum absolute pre-treatment coefficient. M bounds the maximum change in the counterfactual trend slope between consecutive periods. At $M = 0$, the confidence set excludes zero. At $M \geq 0.5\bar{\Delta}$, the confidence set includes zero.