Do Voters Punish Political Instability?
Attribution, Electoral Subversion, and Material Disruption in Peru

Carlos Chávez

June 19, 2026
Preliminary Draft

Abstract

Do voters punish political instability at the ballot box? We assemble a district-level electoral panel for Peru (2006–2026) spanning two episodes at attribution's poles: Fuerza Popular's 2021 offensive of over a thousand petitions to annul tallies after Keiko Fujimori narrowly lost the runoff to Pedro Castillo and ran again in 2026, and the paralysis of Lava Jato megaprojects, a shock with no local culprit. Comparing targeted with profile-matched control districts, clustered by province, we find no differential runoff penalty (seven hundredths of a point, robust across thirty-six specifications). We rule out a penalty above five points under conservative trends but not smaller ones. A first-round withdrawal is inseparable from the left's collapse, and the material shock moves nothing. The runoff coordinates voters against the left, so the offensive costs its author little, which helps explain why subversion can pay. Deterrence then falls to courts more than the electorate.

1 Introduction

Political instability is expensive. Coups, contested successions, and the paralysis of public investment lower growth, and they do so through the channels that matter most for the long run, private investment and productivity (Alesina et al., 1996; Aisen and Veiga, 2013). Democracies are meant to contain these costs through a simple mechanism. An actor who disrupts economic life or attacks the integrity of an election should lose votes at the next one, and the prospect of that loss should discipline the conduct in advance. Whether the mechanism works is an empirical question, and the answer is not obvious. The modern study of retrospective voting has come to doubt that electorates sanction even ordinary economic harms accurately, in good part because voters struggle to assign responsibility for outcomes that have many authors and long lags (Achen and Bartels, 2016).

Instability is in one respect harder to price than the economy and in another respect easier. It is harder because its costs are diffuse and arrive years after the act, so a voter who wished to punish would have to trace a slow, broad decline back to a discrete decision. It is easier because the decision itself is often a single, public, attributable event, a coup, a self-coup, an attempt to overturn a count. If the difficulty of pricing a recession is that no one obviously caused it, the sharper question is whether voters punish instability when someone plainly did. Most of what we know about that question comes from survey experiments that show respondents hypothetical candidates who endorse anti-democratic acts and record the hypothetical penalty (Carey et al., 2022; Graham and Svolik, 2020). The penalty is real but small, and it falls as partisanship and polarization rise. What the experimental evidence cannot supply is a high-stakes field setting in which the act was real, the author was named, the voters who bore it can be located, and the author returns to a later ballot to be punished or spared.

Peru between 2016 and 2026 supplies two such episodes, placed at the two extremes of the attribution problem. The first is material and authorless. The Lava Jato corruption investigation halted a set of large infrastructure megaprojects in 2016 and 2017, when the implicated construction firms lost financing and their contracts were suspended, and construction stopped in the districts the corridors crossed. The disruption was real and local, but no local politician caused it. The second is attributable and pointed. After losing the 2021 presidential runoff by a quarter of a percentage point, Fuerza Popular, the party of Keiko Fujimori, filed more than a thousand petitions to annul the tally sheets of individual polling stations, concentrated in the districts that had voted most heavily against it, and pressed claims of fraud for weeks before the result was certified. The petitions were resolved and none was upheld, but the attempt to overturn a national election had a named author, it fell on identifiable districts, and the same candidate ran again in 2026. Together the episodes let us put the accountability question to one electorate twice, once where blame is impossible to assign and once where it is easy.

The attributable episode does not stand alone, and reading it requires the sequence around it. Figure 1 lays that sequence out. Anti-Fujimori coordination had defeated Keiko Fujimori in the 2016 runoff and, narrowly, in 2021, and the annulment offensive followed that second defeat. By 2026 a Fuerza Popular that had lost ground in the first round nonetheless reconsolidated the right in the runoff, against a left alternative that the collapse of the Castillo government had made costly to support. The runoff therefore turns on the risk of the left as much as on the offensive, and that is the frame in which we read a first-round erosion that does not survive to the second round.

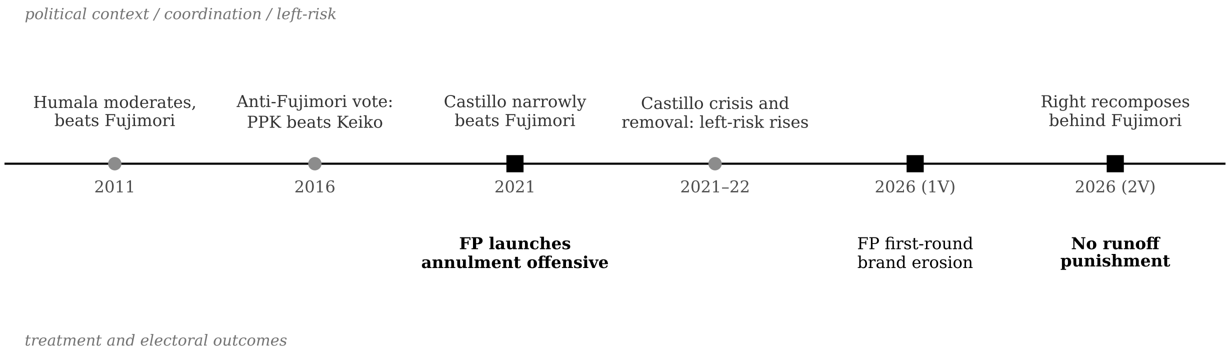


Figure 1: Electoral coordination and the annulment offensive, 2011–2026

Notes. The figure situates Fuerza Popular's 2021 annulment offensive between earlier runoffs in which anti-Fujimori coordination defeated Keiko Fujimori and the 2026 election, where a small first-round erosion of Fuerza Popular did not translate into a runoff penalty. The upper band marks the political context and the lower band the paper's empirical object, the treatment and the outcomes.

The attributable arm carries the paper, and its design is a matched comparison. The offensive was not random. Fuerza Popular petitioned where Castillo had won, so the targeted districts are far more pro-Castillo than the average district, and a comparison against all other districts would confound the offensive with the level of left support and with its own dynamics across two runoffs. We instead model the propensity to be targeted from pre-treatment characteristics and compare treated districts to untargeted districts on the common support of that propensity, the impugnable-but-not-impugned control. The object we estimate is the local, differential response, whether the districts that bore the offensive turned against the actor by more than otherwise similar districts did. A response common to the whole country, in which voters everywhere withdrew support for the subversion, would move the treated and control groups together and is absorbed by the comparison and cannot be identified by it. We are explicit about this throughout. The design recovers the geography of the response, and it reads as a test of the response's existence only under the assumption that voters in a targeted district knew their own sheets had been challenged. The provincial adjudication of the petitions and the local mobilization around them make that assumption plausible, but we cannot measure how far the offensive reached the individual voter, and we report the result both ways.

Targeted districts did not differentially punish the actor in the runoff, our primary outcome. The change in the two-way share between 2021 and 2026 differs from the matched control by seven hundredths of a percentage point, and the same null appears in the protest vote, in two measures of the broader right, and in the decision to vote. It survives thirty-six combinations of the sample cutoff, the matching covariates, and the propensity-score trim, alternative estimators, and a permutation test, and it does not conceal a gradient in the intensity of the offensive. We are deliberate about what it excludes. The outcome carries a pre-treatment trend, so we lead with the conservative end of a sensitivity analysis in the manner of Rambachan and Roth (2023): allowing the post-treatment trend to deviate as much as the 2021 realignment, we rule out a differential punishment larger than about five points. A tighter one-and-a-half point bound holds only under exact parallel trends, which the placebo rejects, so we do not claim to exclude the smaller, experimental-sized effects the literature reports. The first round complicates this reading without overturning it. There the actor's vote falls by about eight tenths of a point more in targeted districts, an erosion absent before the offensive and confined to the post-2021 window, robust to first-difference and ANCOVA bracketing and to matching on structural covariates instead of the lagged vote. But it falls by the same amount as the left does in these same districts. The intensity of the offensive does not predict larger first-round erosion, and the runoff reconstitutes the right onto the actor in targeted and untargeted districts alike, so neither distinguishes a localized penalty from the broader rightward realignment that followed the left's national collapse. We carry it as a mechanism behind the runoff null.

We then ask why no punishment appears, probing five candidate moderators, polarization, development, information, education, and clientelism, as an exploratory exercise corrected for multiple comparisons. We detect no heterogeneity in any of them, and the districts whose 2021 result sat closest to a tie, where partisan-identity theory most expects punishment, are where the null is cleanest. The one moderator significant before correction, household information, runs the wrong way for an account in which voters failed to punish only because the news did not reach them. The material arm, rebuilt on the actual Lava Jato megaprojects, also yields no clean local punishment. Its only positive signals fail pre-treatment placebo checks or coincide with the El Niño Costero disaster of 2017, and the paper does not rest on it.

Because a null is only as informative as the data's power to reveal an effect, we calibrate that power against the incumbency disadvantage that Klašnja and Titiunik (2017) document for weak-party systems: applied to close municipal races the same data recover a six-point fall in re-election, robust to the data-driven bandwidth and the usual validity checks. The power of the annulment comparison itself we read from its equivalence bound in Section 6.

We are deliberate about the limit of the design. The comparison identifies the local, differential response and is silent on a sanction common to the whole country, the channel through which a national penalty would travel. Identifying that channel would require variation we do not have, the post-election trajectories of many candidates who lost close elections, some who conceded and some who subverted. Our claim is the one the data support, about the local, differential response within Peru, and we do not extend it to a national verdict the design cannot deliver.

The paper makes three contributions. It takes the question of whether voters punish anti-democratic conduct out of the survey laboratory and into a high-stakes field setting, where the act was real and the author stood for office again, and finds no differential runoff penalty in the places the subversion targeted. Second, it is careful about what that null can and cannot exclude: the point estimate is essentially zero and we rule out a differential penalty above about five points, but we do not exclude the smaller, experimental-sized effects the literature reports, and a regression discontinuity that recovers a six-point incumbency effect in the same data serves as a check that the data can detect a real effect, and does not license a tighter bound. Third, it links the voter's decision to the economics of political instability, where a penalty the ballot box does not return to its author is a cost the author can impose again, so the act goes underpriced.

The paper proceeds as follows. Sections 2–5 place the question in the literature and lay out the setting, the data, and the matched-comparison strategy. Section 6 reports the annulment offensive, Section 7 the material disruption, and Section 8 the behavioral benchmark. Section 9 examines the mechanisms behind the null, and Section 10 concludes.

2 Related Literature

Our question belongs to the literature on retrospective voting, which holds that voters reward and punish incumbents for the outcomes they preside over and that the anticipation of reward and punishment disciplines those in office (Fiorina, 1981). The mechanism is the foundation of electoral accountability, and it has not aged comfortably. Achen and Bartels (2016) marshal the evidence that voters respond to events their leaders plainly did not cause, from droughts to shark attacks to the local fortunes of a college football team, and conclude that retrospection is too blunt an instrument to discipline much of anything. The reconsideration is more measured. Healy and Malhotra (2013) review the same evidence and find that voters do respond to outcomes, but with biases, a short memory that overweights the election year, a difficulty separating competence from luck, and above all a difficulty assigning responsibility for outcomes that have many hands on them. Attribution is the recurring obstacle, and we remove it by design. One of our episodes has no local author at all, the other a single named one who returned to the ballot, so within a single electorate we observe accountability both where blame is impossible to assign and where it is plain.

The attributable episode is an instance of democratic backsliding, and the literature on backsliding has converged on the point that contemporary erosion rarely arrives as an open coup. It proceeds through legal-seeming instruments, the contested count, the challenged tally, the appeal to a friendly court, which are harder for citizens to read as threats and therefore harder to sanction (Bermeo, 2016; Levitsky and Ziblatt, 2018). When the formal guardrails hold, as the electoral authorities did in Peru in 2021, the voter becomes the residual check, and whether that check binds is exactly the question. The recent answer is discouraging. Graham and Svolik (2020) show that only a small fraction of American voters place democratic principles above party and policy when the two conflict, and that the fraction falls as polarization rises. Svolik (2019) gives the logic, in a polarized electorate voters will trade democratic principle for partisan advantage, so the electoral penalty for subversion thins precisely where subversion is most tempting. Carey et al. (2022) confirm in a conjoint experiment that voters withdraw support from candidates who endorse undemocratic acts, but that the penalty is a few points and is muted by copartisanship. This evidence is overwhelmingly experimental and American. We carry its central question to a high-stakes election in a polarized young democracy, where the subversion was real and the actor faced the voters again, and find no differential runoff penalty, with a point estimate near zero but a bound too wide to exclude an experimental-sized effect itself.

A second literature asks whether voters punish corruption, and it supplies the closest field analogue to our design. Ferraz and Finan (2008) show that Brazilian mayors revealed by random audits to have misused public funds lose votes, but mainly where the finding reached voters through local radio, which places the binding constraint in information. The scandal we study, Lava Jato, is the same regional corruption case at a larger scale, and its consequences for the firms that paid the bribes are documented (Colonnelli and Prem, 2022). We study the voter. The scandal reached the districts we examine as the material disruption left behind when the implicated firms stopped building, a real economic harm with no local author, which is the diffuse counterpart to the annulment offensive and lets us ask whether voters punish a harm to which no politician's name is attached. Artiles et al. (2021) study a related Peruvian institution, the recall referendum, and find that it lowers the quality of the candidates who later run. That is a selection channel operating through who chooses to compete, distinct from the voter's sanction we estimate, and we keep the two apart.

The cost that motivates the exercise comes from the macroeconomics of political instability. Alesina et al. (1996) establish that instability, measured by the propensity for irregular executive turnover, lowers growth across countries, and Aisen and Veiga (2013) trace the effect to private investment and total factor productivity, so the cost is a slow erosion of the economy's capacity. A cost of that shape is the one a retrospective electorate is least equipped to price, diffuse in incidence and delayed in time. The accountability question and the macroeconomic one are two halves of one argument. If the ballot box does not return the cost of instability to the actor who imposes it, instability is underpriced, and an underpriced action recurs. The persistence of instability in the settings where its economic toll is best documented is the puzzle our null helps explain.

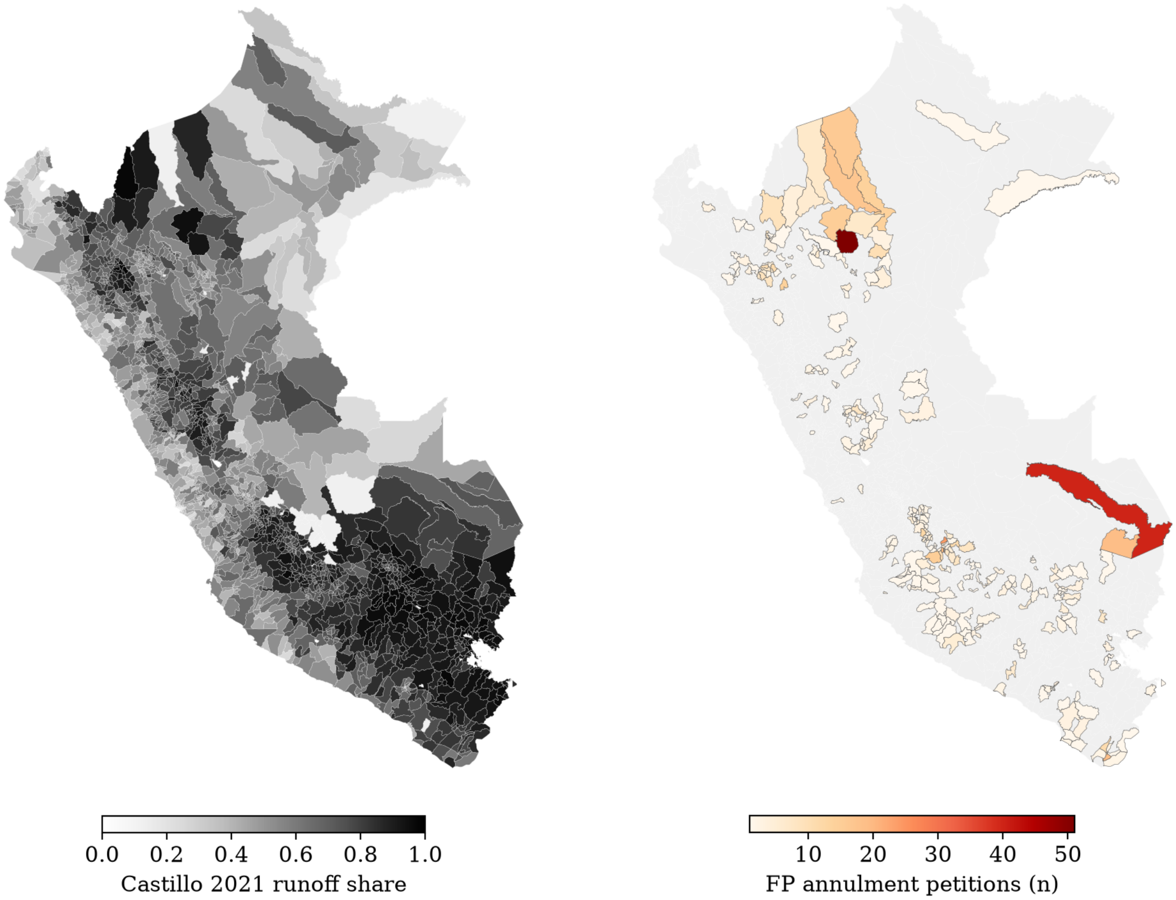
The paper builds, finally, on the study of elections in weak-party democracies, where party labels are unstable and incumbency behaves unlike it does in consolidated systems. Klašnja and Titiunik (2017) document an incumbency disadvantage in such systems, in which a party that barely wins office becomes less likely to hold it next time, for want of the party structures that would otherwise turn incumbency into advantage. We use their result as substance and as a check. Replicated on our data with a regression-discontinuity design at close municipal races, the incumbency disadvantage is a clean six-point effect in the same electoral data, which shows a meaningful electoral effect is recoverable here. It is a different estimand, population, and design from the annulment comparison, so it does not establish that comparison's power. That power we read from the within-design minimum detectable effect and the equivalence test in Section 6, not from the benchmark.

3 Setting and Treatments

The first treatment comes from the Lava Jato corruption scandal. The investigation began in Brazil in 2014 and spread across Latin America, and it revealed that the construction conglomerate Odebrecht and a cartel of allied firms had paid systematic bribes to secure large public-works contracts. In Peru the scheme reached the highest levels of the state and the country's largest infrastructure projects, road corridors and related works awarded over the preceding decade. When the bribery became public in 2016 and 2017, the implicated firms lost financing, partners withdrew, and several contracts were suspended or annulled, and construction on the affected corridors slowed or stopped. The disruption fell on the districts the corridors physically crossed, as halted works, idled local employment, and the withdrawal of the public investment the projects had promised. No local politician caused it. That is what makes these districts a test of whether voters punish an economic harm with no author to hold responsible, and we build the measure of exposure to it in Section 4.

The second treatment comes from the contest that followed. Pedro Castillo, a rural schoolteacher and union leader running for the left, defeated Keiko Fujimori in the 2021 presidential runoff by 0.25 percent of the valid vote, one of the closest results in the country's history.[1] Fuerza Popular did not concede. In the weeks between the vote and the certification of the result, the party filed 1,086 petitions to annul the tally sheets of individual polling stations, alleging fraud, and concentrated them in the districts where Castillo had won most heavily. The electoral authorities adjudicated the petitions and upheld none of them, and the result was certified. The episode had a named author, it fell on identifiable districts, and the same candidate ran again in the 2026 presidential election, which is what makes it the attributable counterpart to the material disruption. Fujimori had conceded a defeat of nearly the same size in 2016, when she lost to Pedro Pablo Kuczynski by 0.24 percent, so the 2021 offensive was a deliberate choice.

[1] This is the official ONPE national margin, which includes the vote cast abroad. The district panel we use for estimation covers domestic tally sheets, where Castillo's 2021 margin is 0.88 percent of the valid vote. The overseas vote, which leans to Fuerza Popular, narrows the national figure below the domestic one. We report national margins in the narrative and estimate on domestic districts throughout, and we treat the 2026 runoff the same way.



(a) Castillo 2021 runoff share

(b) Fuerza Popular annulment petitions

Figure 2: Districts targeted by the annulment offensive

Notes. District maps of Peru. Panel (a) shades each district by Castillo's two-way share in the 2021 runoff, with darker districts more pro-Castillo. Panel (b) shades the districts that Fuerza Popular petitioned to annul by the number of petitions. The two panels share a geography, so the overlap between the dark districts in (a) and the shaded districts in (b) is the selection rule the design exploits. Polygons are from the INEI shapefile and are used for geometry only. The petitions are located by matching each case file's department, province, and district names to the ONPE *ubigeo* crosswalk, after which treatment is joined to the electoral panel by ONPE *ubigeo*, like every other layer (Section 4).

Several features of the Peruvian system matter for the design. Presidential elections are decided in a two-round runoff, which placed Fuerza Popular on the ballot in both the 2021 and 2026 second rounds, against Castillo in the first and Roberto Sánchez in the second, so the actor we track is constant even as the opponent changed. Petitions to annul tally sheets are filed with and adjudicated by the provincial electoral juries, bodies seated in the provincial capitals that hold public hearings, with appeal to the national jury in Lima. The provincial seat of adjudication is the reason we cluster standard errors by province, and the reason the offensive left a footprint in the contested provinces not only in the capital. Local politics runs on weak and short-lived party vehicles, electoral volatility is among the highest in the region, and mayors have been barred from immediate re-election since 2015. The resulting turnover of incumbents is high, and we use it in Section 8 as an external check that an electoral effect of meaningful size is detectable in these data, not as a measure of the annulment comparison's own power.

These features motivate three hypotheses about how a punishment, if one exists, would appear. The accountability logic predicts that the districts the offensive targeted turn against its author at the next election, when she is again on the ballot, so the treated−control difference in her runoff vote should be negative (H1). If the response is graded, the turn should be larger where the offensive was more intense, contesting a larger share of the district's vote (H2). And because the decisive contest is a runoff that forces a binary choice against the left, any penalty may surface less in the runoff than in the first round, where a voter can desert the party brand without conceding the presidency to the opposing camp (H3). The empirical sections take these in turn, and the interpretation in Section 6.4 returns to why the first round and the runoff can diverge.

4 Data

We assemble a district-level panel that combines three sources: electoral results from 2006 to 2026, the judicial record of the 2021 annulment offensive, and a geographic measure of exposure to the disrupted megaprojects. The district is the finest unit at which Peruvian results are public, and the panel covers the country's 1,892 districts. This section describes each source, the variables we construct, and the one feature of the data that materially affects the results.

For the presidential contests we use the official tally-sheet records (*actas electorales*) of the national election office, ONPE, for the 2006, 2011, and 2016 elections in both the first and second rounds, the published district returns for 2021, and the office's results portal for 2026. We collected the 2026 returns directly from the portal as counting neared completion. An initial collection at 98.3 percent of tally sheets was refreshed to the near-final count at 99.4 percent, with all tally sheets processed, the remaining 0.6 percent held for adjudication before the electoral juries, and none pending at the office. The national result stayed within a fraction of a percentage point and had not been officially proclaimed, so we treat every 2026 quantity as near-final and subject to proclamation. The attributable estimate does not turn on this. The treated−control difference in the runoff is −0.0007 and is essentially unchanged between the 98.3 and 99.4 percent counts, with the typical district's runoff share shifting by under two-thousandths, so the result does not depend on the final tally. The uncounted share could in principle correlate with treatment, since late-counted tally sheets tend to be rural, but a district enters the panel at the same rate whether it was targeted or not, with 2.4 percent of treated and 3.2 percent of control districts missing, a difference that is not significant. The 2026 registered roll, fixed before the election, lets us strip out the growth of the electorate and measure the change in turnout directly, the sixth margin in Table 4, and it too does not differ between groups.

Both campaigns contested the 2026 runoff. Juntos por el Perú asked the juries to annul more than two thousand polling stations, most in Lima and a block of overseas stations, alleging fraud in favor of Fuerza Popular, while Fuerza Popular sought annulments of its own in Puno. The juries declared the largest of these petitions inadmissible, as they had in 2021, so few sheets were removed and the near-final count is not materially altered, consistent with the stability of our estimate. This second offensive is national, and the overseas stations lie outside the domestic panel, so we treat it as context. Should the juries annul sheets unevenly across our groups as the remaining cases resolve, we will report the estimate with and without the affected sheets. For the municipal contests we use the returns for the regular elections of 2006 through 2022, drawn from ONPE and from the electoral authority's INFOgob archive. For each district and election we observe the vote for every party or candidate, the registered electorate, turnout, and the counts of blank, null, and contested ballots.

From these records we build the two outcomes the paper uses. The first is the actor's two-way share in the runoff, her votes divided by the sum of the two finalists' votes, which is comparable across the 2021 and 2026 second rounds because she contested both. The second is the invalid-vote share, blank plus null ballots over ballots cast, a measure of protest that does not require voters to coordinate on a named target and that we use for the diffuse arm. We also construct first-round vote shares, a right-bloc share that sums the national right-wing parties, and the actor's share within that bloc, all at the district level.

The treatment comes from the judicial record of the 2021 post-election dispute, retrieved from the historical case system of the national electoral jury, JNE. After the runoff, Fuerza Popular and, on a smaller scale, Perú Libre filed petitions to annul the tallies of individual polling stations. We keep the 1,086 petitions whose subject is the nullity of a polling-station tally, of which 942 were filed by Fuerza Popular and 144 by Perú Libre. Each case file records the district of the contested station, which lets us locate the offensive geographically.[2] Every petition was resolved and none was upheld. We code a district as treated if Fuerza Popular filed at least one petition there, which identifies 285 districts, and we record the number of petitions per district for the dose-response in Section 6.

4.1 Lava Jato material exposure

We measure a district's exposure to the Lava Jato megaprojects from the curated set of core cartel concessions held by the implicated firms: the Interoceánica and IIRSA highways, Red Vial 5 and 6, the Autopista del Sol, the Longitudinal de la Sierra, the Gasoducto Sur Peruano, and Chavimochic III. We georeference the road corridors from their concession traces and clip them to the district map, and locate the two non-road projects, the Gasoducto Sur pipeline and the Chavimochic irrigation works, by their documented footprints, so a district counts as exposed only where a core project physically reached it. We code each project's 2016 construction status as active, partial, or already complete, so that the disrupted districts are those still being built when the firms lost financing. An earlier version of this exposure, built on a broad set of SEACE procurement records, is superseded. The correction and its consequences are documented in the replication package. Figure 3 maps the corridors and the exposed districts by status.

[2] Two Fuerza Popular petitions, both for the same district (Huayllati, in Grau, Apurímac), were initially unmatched because the case file spells the district HUAILLATI while the electoral roll spells it HUAYLLATI. We resolve the variant with a versioned manual name crosswalk, so all 942 petitions locate to a district. The correction reclassifies Huayllati from control to treated, and the runoff estimate is unchanged by it.

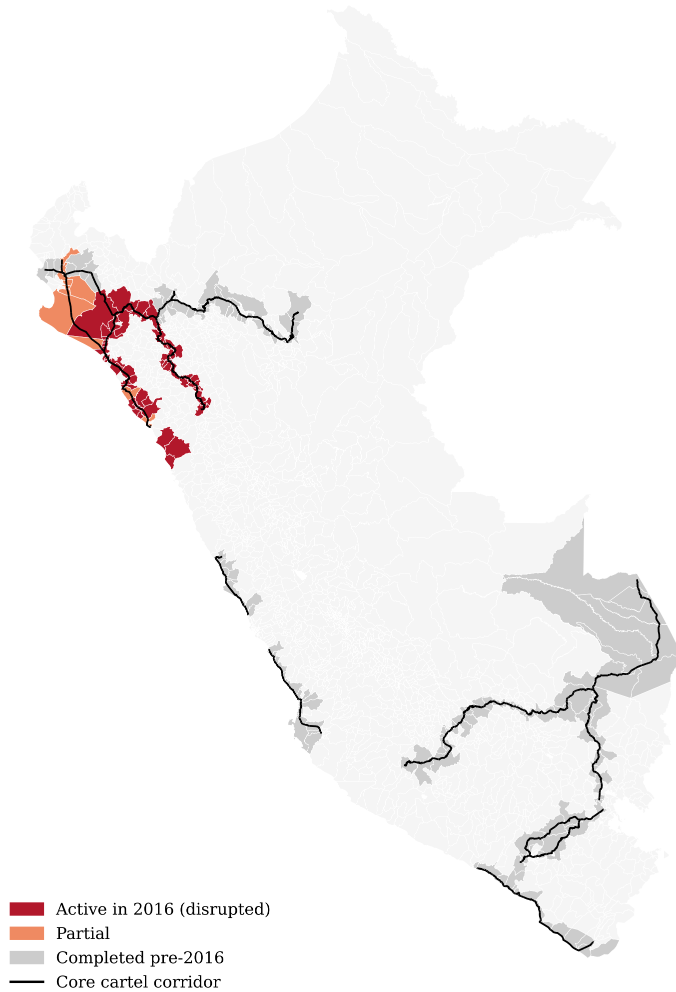


Figure 3: Core Lava Jato corridors and exposed districts

Notes. The georeferenced core cartel road corridors (lines) and the districts they reach, shaded by each project's 2016 construction status. Active-construction districts were still being built when the firms lost financing. Completed corridors finished before the shock. The Gasoducto Sur and Chavimochic non-road projects enter by their documented footprint.

4.2 Household survey (ENAHO)

Our attitudinal evidence comes from the governance module of the national household survey, ENAHO, fielded annually from 2004 to 2025. Each respondent rates confidence in a list of institutions, among them the two electoral authorities the offensive attacked, the National Elections Jury (JNE) and the national elections office (ONPE). We aggregate the JNE and ONPE items to the district and use them in Section 9 to ask whether the offensive moved trust where it did not move the vote.

ONPE and the national statistics institute, INEI, assign different numeric codes to the same district, and their department codes diverge, with ONPE placing Lima at 14 and INEI at 15, for example. The electoral panel, treatment and outcome alike, is linked across years by ONPE district code, so the causal comparison never crosses the two coding systems. We use name matching within province only to bring in the INEI-coded layers, the census and the fiscal records, and we reserve the INEI shapefile for geometry alone. Peru creates and splits districts over time, and the universe grows from roughly 1,830 districts in 2006 to 1,892 in 2026. We link districts across years by their ONPE code, which is fixed for a district once created, and a district created after a given base year simply does not enter a change that spans that year. The attributable arm differences the 2021 and 2026 runoffs, both within the recent and stable part of the universe, so boundary changes touch it little, while the pre-trend chains that reach back to 2011 and 2016 drop the handful of districts that did not yet exist.

Matching the contested districts to impugnable controls on common support and requiring an observed 2026 outcome leaves 249 treated districts and 1,149 controls in the estimation sample. Table 1 traces this funnel from the petitions to the estimated sample. Table 2 then reports means by group on that sample.

Table 1: Sample construction

Step	N
Fuerza Popular annulment petitions (nullity of a polling-station tally)	942
Districts attacked (at least one petition)	285
Treated: on common support with an observed 2026 outcome	249
Matched impugnable controls	1,149
Estimation sample	1,398

Notes. Fuerza Popular filed 942 nullity petitions, which fall in 285 districts. Requiring common support on the targeting propensity and an observed 2026 runoff outcome leaves 249 treated districts, matched to 1,149 impugnable controls, for an estimation sample of 1,398. The full merge audit, including the two recovered Huayllati petitions and the per-step match diagnostics, is in the appendix.

Treated districts and the impugnable control are close on the 2021 Castillo share and on turnout, while the off-support districts, which the design discards, are far lower on the Castillo share. This is the comparison the empirical strategy is built to exploit: a control that resembles the treated districts on the dimensions that drove the offensive.

Table 2: Descriptive statistics, estimation sample

	Treated	Control	SMD
Pre-treatment characteristics			
Castillo 2021 runoff share	0.788	0.730	+0.41
Keiko 2016 runoff share	0.463	0.485	-0.15
Turnout 2021	0.683	0.695	-0.13
Registered voters	11,137	15,851	-0.12
Urban (% census)	31.1	27.2	+0.10
Avg. years of schooling	5.81	5.92	-0.07
FP annulment petitions	3.15	0.00	+0.95
Outcomes (2021→2026 change)			
Δ Keiko runoff share	0.0194	0.0201	-0.02
Δ Keiko first-round share	0.0172	0.0251	-0.22
Δ protest vote	0.0026	0.0022	+0.03
Districts	249	1149	

Notes. Means for treated districts (targeted by the annulment offensive) and the impugnable-but-not-impugned control matched on the estimated propensity to be targeted, with the standardized mean difference (SMD). Targeting loads on the Castillo vote (SMD +0.41). The outcome, the change in the runoff share, is balanced (SMD −0.02). Census composition is from the 2017 census.

5 Empirical Strategy

The estimand for each episode is the difference between affected and unaffected districts in the change in the vote. We follow one rule throughout. We never report a treated-minus-control difference without the control it requires.

Attributable arm. For district d we estimate

$$\Delta y_d = \alpha + \beta T_d + \varepsilon_d, \qquad \Delta y_d \equiv y_{d,2026} - y_{d,2021}, \tag{1}$$

where $T_d = 1$ if district d was targeted by the annulment offensive and y_d is the actor's two-way runoff vote share. The coefficient β is the attributable punishment. We take the change Δy_d not the level, because treated districts differ in level for reasons unrelated to the offensive, and we cluster standard errors by province $j(d)$, the level at which petitions were adjudicated. Eighty-six of the provinces in the sample contain a treated district, which keeps the inference out of the few-clusters regime where cluster-robust standard errors understate uncertainty. The primary outcome and comparison group were specified before the 2026 returns existed, and the time-stamped specification documents are included in the version-controlled replication package.

The control is the set of districts that fit the offensive's selection rule but were not targeted. We estimate the propensity to be targeted from pre-treatment characteristics,

$$\Pr(T_d = 1) = \Lambda(X_d'\gamma)\,, \tag{2}$$

where X_d collects the 2021 Castillo share, turnout, district size, and a rurality proxy, and Λ is the logistic function. We estimate (1) on treated and untreated districts whose propensity lies in the common support. Comparing treated districts to all others would confound the offensive with the level of left support, which has its own dynamics across the two runoffs.

We assess whether the null is precise with a two-one-sided equivalence test (Hartman and Hidalgo, 2018),

$$H_0 : |\beta| \geq \delta \quad \text{against} \quad H_1 : |\beta| < \delta, \qquad \delta = 0.014, \tag{3}$$

where $\delta = 0.014$ is the equivalence bound. It is the minimum detectable effect at 80% power in our sample, the smallest differential punishment the design can detect, a quantity we did not choose in advance from theory. It happens to fall at the low end of the punishments the experimental literature on democratic violations reports, which run from about one to four points (Carey et al., 2022; Graham and Svolik, 2020), and far below the six-point incumbency disadvantage our own benchmark recovers in Section 8. We do not, however, read it as ruling out effects of that experimental size: the equivalence at δ holds only under exact parallel trends, which the runoff placebo rejects, and the trend-robust bound is the wider five points of Appendix D.

Diffuse arm. For the material episode we estimate

$$\Delta y_d = \alpha + \beta\, E_d + \varepsilon_d, \tag{4}$$

where E_d is district d's exposure to disrupted megaprojects and Δy_d is the change in the invalid-vote share. Because exposure is continuous and spatially correlated, we obtain the null distribution of β by permuting treatment across construction corridors, which respects the level at which exposure is assigned.

Benchmark. To test the design against a known behavioral effect, we estimate a regression discontinuity in close municipal races,

$$w_{p,2014} = \alpha + \tau\, \mathbf{1}\{m_{p,2010} \geq 0\} + f(m_{p,2010}) + \varepsilon_p, \tag{5}$$

where $m_{p,2010}$ is party p's margin of victory in 2010, $w_{p,2014}$ indicates that the party wins in 2014, and $f(\cdot)$ is a local-linear polynomial fit on each side of the cutoff with a mean-squared-error-optimal bandwidth and robust bias-corrected inference (Calonico et al., 2014). The coefficient τ is the incumbency effect.

6 The Annulment Offensive

After losing the 2021 runoff by 0.25 percent of the valid vote, Fuerza Popular filed 942 petitions to annul tally sheets, against 144 by Perú Libre, and we assign each petition to a district. The offensive was strategic. Annulling a tally sheet removes its votes from the count, so a party trying to overturn a narrow loss gains the most by annulling sheets in the districts where the winner ran up the largest margins. Fuerza Popular's petitions follow that logic exactly. Figure 4 shows the selection rule. The share of districts the party petitioned rises from 1 percent in the quintile where Castillo was weakest to 23 percent in the quintile where he was strongest. A logit of the petition decision on pre-treatment characteristics confirms it. The 2021 Castillo share and the size of the district, the two quantities that govern how many of the winner's votes a successful annulment would erase, are the strongest predictors of being targeted, while turnout and rurality enter weakly (Appendix E). Perú Libre, the party that had won, petitioned the mirror set of districts, those where it had lost. The offensive thus traced the geography of the vote against its author, and that geography is the variation we exploit. Matching on the determinants of targeting is what makes the comparison informative, because the control districts are those Fuerza Popular would have had a similar reason to contest and did not. No petition was upheld.

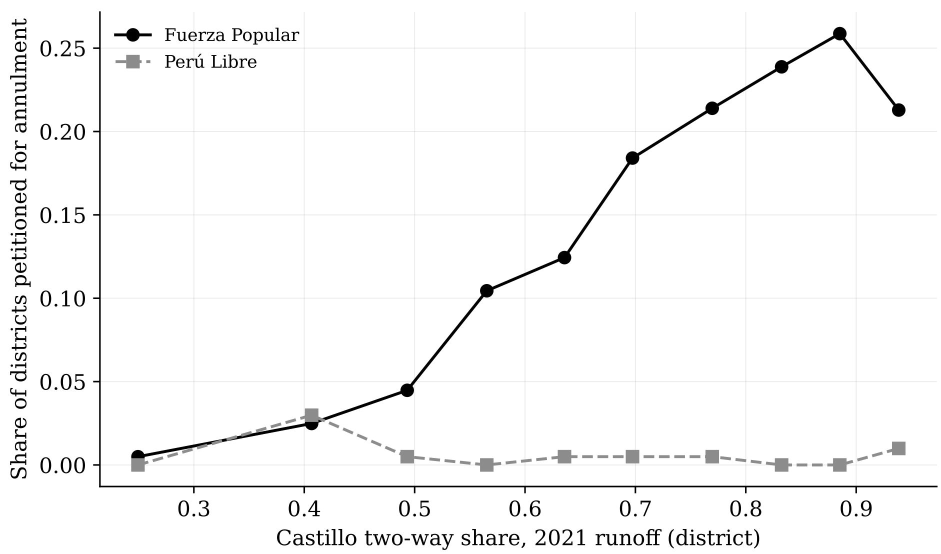
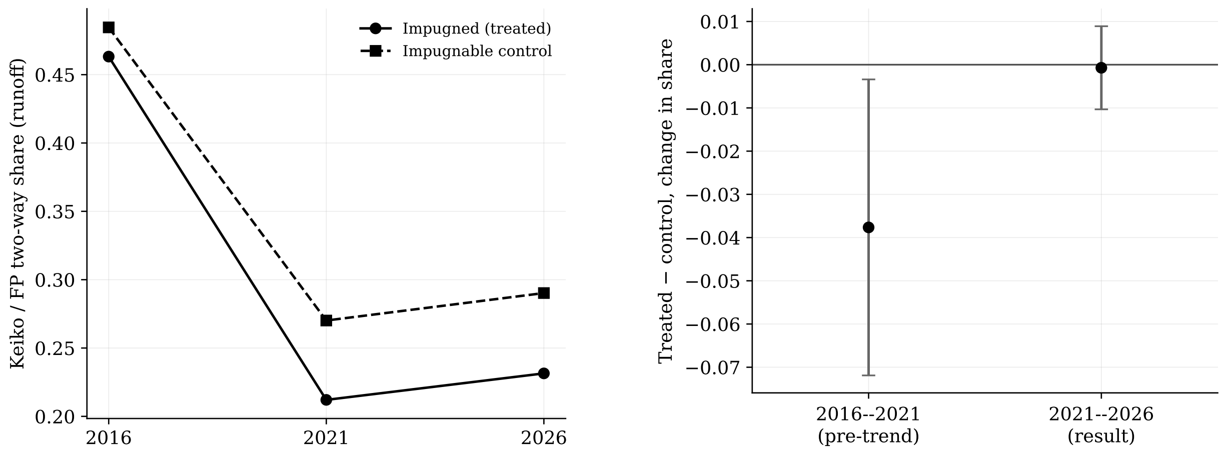


Figure 4: The selection rule

Notes. Binned means of the share of districts petitioned for annulment against Castillo's 2021 runoff share, for Fuerza Popular and for Perú Libre. Fuerza Popular's petitions rise with the Castillo vote. Perú Libre's fall, the mirror pattern. No petition was upheld.

6.1 Main runoff result

Figure 5 reports the change. The actor's runoff vote rose by about two percentage points in both groups between 2021 and 2026, and the difference is -0.0007 with a province-clustered t of -0.14.



(a) Two-way share by group

(b) Treated−control change by period

Figure 5: Runoff vote share by treatment group, 2016–2026

Notes. Panel (a) plots the mean two-way Keiko share among treated districts and the impugnable control at the 2016, 2021, and 2026 runoffs. Panel (b) plots the treated−control difference by period with 95% province-clustered intervals. The 2026 figures reflect the count at 99.4% of tally sheets, pending proclamation.

Figure 6 reports the estimate across five measures of vote choice: the runoff vote, the first-round vote, the protest vote, the right-bloc share, and the actor's share within the right bloc. Every estimate lies within 1.4 percentage points of zero, and five of the six are statistically indistinguishable from it. The exception is the first-round vote, a significant fall of eight tenths of a point that we take up below and that does not survive a pre-trend check. Targeted districts did not turn against the responsible actor relative to the comparison. The decision to vote shows no differential either. Using the 2026 electoral roll to measure turnout, the change in turnout from 2021 differs by −0.6 points between treated and control districts and is not significant, the sixth row of Table 4. At the current count this margin still blends genuine demobilization with any differential under-counting, and it bounds both small.

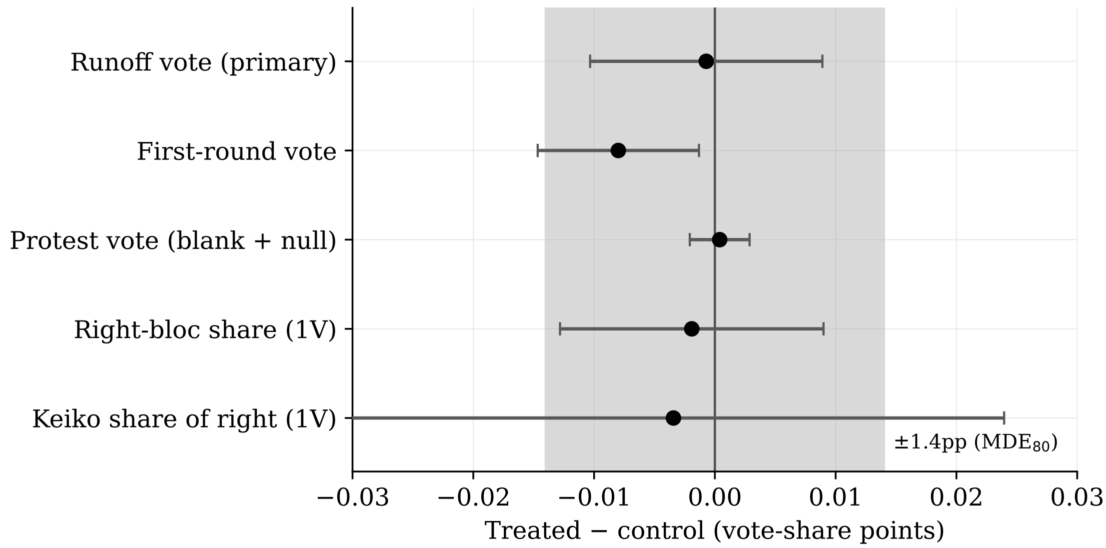


Figure 6: Treated−control differences across five vote measures

Notes. Treated districts are those targeted by the annulment offensive. The control is the set of impugnable-but-not-impugned districts matched on the selection rule. Bars are 95% confidence intervals clustered by province. The shaded band marks ±1.4 percentage points, the minimum detectable effect at 80% power under parallel trends. The bound that survives the pre-trend in the primary outcome is wider, about five points, as discussed in Appendix D.

The null is not the small, noisy estimate of an underpowered test. The point estimate is seven hundredths of a percentage point, on a scale where a backlash of the kind that openly contesting an election might be expected to provoke would have moved the vote by whole points, and the equivalence test rejects effects beyond a point and a half under parallel trends. These are also the districts where a backlash should have been easiest to find. They are the places whose residents' own ballots the actor sought to annul, where the dispute was adjudicated in the provincial electoral jury within sight of the affected voters, and where the actor's vote was already lowest and so had the most room to fall further. The level cuts both ways, and we are careful about it. With the actor's vote already near a fifth in these districts, the pool of remaining supporters left to defect is thin, so a genuine penalty would also register as small here, which is why we lean on the equivalence bound and the dose-response not on the level of the vote alone. A voter inclined to punish an attempt to overturn the election in which her own district was the target had both the occasion and the motive. The targeted districts did not take it, on the runoff vote or on any other margin we observe.

6.2 Robustness and bounds

The result does not depend on how we build the control. Across 36 combinations of the size cutoff, the propensity-score trim, and the logit covariates, traced out as a specification curve (Simonsohn et al., 2020), the estimate stays between −0.004 and +0.003 and never approaches significance. Nor, as far as we can tell, does it come from spillover onto the controls. Because petitions were adjudicated by province, a control that shares a province with a targeted district could absorb a province-level grievance and bias the comparison toward zero. Restricting the controls to districts in provinces where no petition was filed, which drops every such case, leaves the estimate unchanged at +0.001. This reassures without settling it, because the restriction also shifts the geographic composition of the control group, and it speaks only to province-level spillover. A grievance carried by national coverage would fall in the unidentified national response, not here.

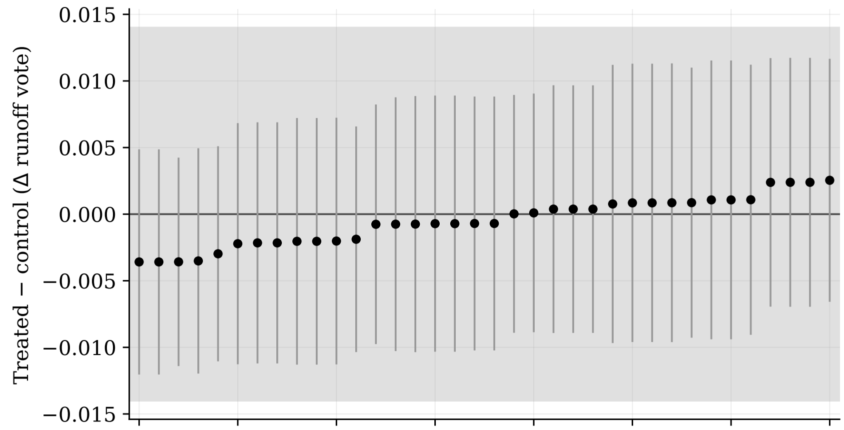
The null is robust to how we estimate it and to how we test it. Reweighting the controls to the treated covariate distribution by inverse-propensity weights, coarsened exact matching, and an augmented doubly-robust estimator all return differences between about −0.010 and zero, every one inside the equivalence band and none significant, so the result does not turn on the propensity-score trimming that the specification curve varies. And because the treatment followed a known selection rule, not chance, we assess it with a propensity-based placebo test, reassigning treatment across the common support in proportion to the estimated propensity and recomputing the difference three thousand times. This is a conditional permutation under the fitted selection model. It is not design-based randomization inference, since the petitions were not assigned by that model. The observed difference falls in the middle of the placebo distribution, with a p-value of 0.99, so the null does not rest on the normal approximation or on the number of effective clusters. Because targeted districts cluster regionally, we also recompute the standard error with a Conley spatial correction over great-circle distances, allowing arbitrary correlation among districts within a chosen radius. Across radii from 50 to 500 kilometers the primary standard error stays between 0.0044 and 0.0052, essentially the province-clustered value of 0.0049, so the null is not an artifact of a smaller effective number of independent treated units. Table 3 collects the standard errors under all of these methods.

Table 3: Inference for the primary estimate

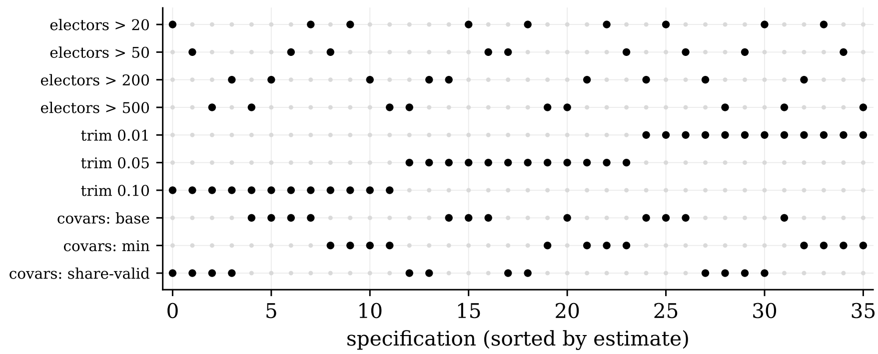
Inference method	Primary outcome ($\beta = -0.0007$)	ANCOVA spline ($\beta = -0.0070$)
OLS / HC robust	0.0029	0.0027
Cluster (province)	0.0049	0.0044
Cluster (department)	0.0046	0.0043
Conley spatial, 50 km	0.0044	0.0040
Conley spatial, 100 km	0.0052	0.0050
Conley spatial, 200 km	0.0044	0.0032
Conley spatial, 300 km	0.0046	0.0028
Conley spatial, 500 km	0.0050	0.0035
Propensity-permutation test	$p = 0.99$ (primary, 3,000 draws)	

Notes. Standard error of the treated−control difference under each inference method, for the primary outcome ($\beta = -0.0007$) and the baseline-adjusted ANCOVA spline ($\beta = -0.0070$). Conley spatial-HAC errors use great-circle distances. The last row reports the propensity-permutation test.

The result also does not hide an effect where the offensive was most intense. The number of petitions a district received varies the dose of the treatment within the treated group, from a single contested sheet to as many as 40, and it tracks how visible the dispute would have been locally, since a district whose count the party fought hardest is a district where the fight was hardest to miss. Regressing the change in the vote on the dose yields no gradient, whether the dose is the raw count of petitions or the share of the district's tally sheets that were contested, which normalizes for size, and if anything the districts the party contested most moved toward it. An account in which the targeted districts failed to punish because the offensive never registered has to explain why the districts where it registered most did not punish either. Two features of the dose qualify it. The common-support trim removes the highest-propensity petitioned districts, including the single most-petitioned one (51 petitions, against a retained maximum of 40), so the dose-response does not reach the very top of the intensity range. Those trimmed districts are not more pro-Castillo than the retained, so the trim does not drop the places where punishment would be largest on that dimension. The first-round vote bears on a separate concern. The two-way runoff share mixes any punishment of the actor with the differential way the left vote transferred between two different opponents, Castillo in 2021 and Sánchez in 2026, and the targeted districts are the most left-leaning and so the most exposed to that transfer. Naming the actor alone and free of the change of opponent, the first-round vote is the one measure negative on its own, at -0.8 percentage points. It does not survive a pre-trend check. We observe the actor's first-round vote in 2011, 2016, and 2021, and the treated-control difference changes sign across these periods, -2.6 points from 2011 to 2016 and $+1.2$ from 2016 to 2021. The pre-treatment path is not linear, so we do not adjust for it and rest the result on the unconditional null in the primary outcome.



(a) Treated−control estimate with 95% confidence intervals



(b) Specification active in each column

Figure 7: Specification curve, 36 specifications

Notes. Each of the 36 specifications varies the electorate-size cutoff, the propensity-score trim, and the logit covariates. Panel (a) plots the treated−control estimate with 95% province-clustered confidence intervals, sorted by estimate, with the shaded band marking ±1.4 percentage points. Panel (b) marks the specification active in each column. The estimate stays between −0.004 and +0.003 throughout.

How large a punishment can we rule out? The answer depends on what we are willing to assume about trends, and we lead with the conservative end. The primary outcome shows a pre-trend, so the credible bound is the one that survives it: allowing the post-treatment violation of parallel trends to be as large as the 2021 realignment itself, we rule out a differential punishment larger than about five percentage points (Appendix D). The sign of the null is robust well beyond that. Under exact parallel trends the bound tightens to the minimum detectable effect of 1.4 points, the optimistic end of the range. The last column of Table 4 reports an equivalence test against this tighter bound. The primary outcome is statistically equivalent to zero within 1.4 points ($p = 0.003$) under parallel trends, and four of the six measures in Table 4 pass the same test. The within-right share and the turnout margin are too imprecise to bound at this threshold, though both point estimates are small and not significant.

Table 4: Treated−control differences across vote measures

Vote measure	(A−B)	Cluster SE	95% CI		Equiv. p
			Lower	Upper	(δ=1.4pp)
Runoff vote (2V), 2021→2026 *[primary]*	-0.0007	0.0049	-0.0103	+0.0089	0.003
First-round vote (1V), 2021→2026	-0.0080	0.0034	-0.0147	-0.0013	0.039
Protest vote (blank+null)	+0.0004	0.0013	-0.0021	+0.0029	0.000
Right-bloc share (1V)	-0.0019	0.0056	-0.0128	+0.0090	0.015
Share within right bloc (1V)	-0.0034	0.0140	-0.0308	+0.0240	0.225
Turnout, 2V 2021→2026	-0.0062	0.0048	-0.0156	+0.0033	0.053

Notes. Each row reports the difference (A−B) between treated districts, those targeted by the annulment offensive, and the impugnable-but-not-impugned control, in the change of the indicated vote measure between 2021 and 2026. Standard errors are clustered by province. The final column reports the p-value of a two-one-sided equivalence test against a bound of ±1.4 percentage points. A small value indicates the estimate is statistically equivalent to zero within that bound. The turnout row uses votes counted over the 2026 registered electoral roll. At the near-final count it does not separate demobilization from counting completeness.

The design turns on how the near-zero holds up under alternative counterfactuals, because the treated and control districts are matched on the same outcome whose change we measure. Table 5 reconstructs the runoff difference under five constructions: the raw matched change, a flexible spline adjustment for baseline support, tight nearest-neighbor matching on the 2021 vote, entropy balancing, and province fixed effects. The estimate stays between zero and about seven tenths of a point and never reaches significance. Baseline conditioning nudges it modestly toward punishment, by half a point at most, a sign of mild regression to the mean. Together these bound any local differential punishment at roughly one to one and a half points, but only under conditional selection on observables. That is a tighter statement than, and does not replace, the conservative trend-sensitivity bound of about five points in Appendix D: baseline conditioning addresses cross-sectional selection but leaves the temporal parallel-trends violation, and the honest bound on a differential punishment is the wider five points.

Table 5: Counterfactual reconstruction of the attributable null

Specification	Estimate	95% CI	Largest punishment ruled out
Raw change, common support	-0.0007	[-0.0104, +0.0090]	1.04 pp
Flexible ANCOVA (baseline spline)	-0.0070	[-0.0157, +0.0016]	1.57 pp
Nearest-neighbor matching	-0.0015	[-0.0085, +0.0055]	0.85 pp
Entropy balancing	-0.0062	[-0.0158, +0.0035]	1.58 pp
Within province (province FE)	+0.0004	[-0.0049, +0.0057]	0.49 pp

Notes. Each row reports the treated−control difference in the change of the actor's two-way runoff share, 2021 to 2026, under a different counterfactual construction, with province-clustered intervals. The last column is the largest differential punishment the 95% interval does not exclude, the negative of the lower bound. The conservative trend-sensitivity bound of about five points (Appendix D) is separate and is not replaced by these.

6.3 Mechanisms: where punishment would surface

The runoff is a blind venue for this question. The same two candidates face off in 2021 and 2026, so the second round reconstitutes the right onto Fujimori as the focal alternative to the left, and any first-round penalty against the actor can be coordinated away before it reaches the decisive vote. A penalty surfaces, if anywhere, in the first round, where voters can defect to a neighbor on the same side before the runoff forces them back. We read the mechanism through the first round and put three questions to the targeted districts. Does the intensity of the offensive predict first-round erosion of the actor, and does that erosion load specifically on Fuerza Popular within the right? Does the block decomposition separate a localized penalty from the broad realignment these heavily ex-Castillo districts underwent after the left collapsed? And is whatever first-round defection occurs reconstituted in the runoff more in targeted districts than in their controls, the pattern a punished-but-absorbed reading requires? All three rest on a measurement we build directly: recovering from the electoral jury's records the individual tally sheets each petition sought to annul, we observe the 768 distinct polling tables Fuerza Popular moved to void across the targeted districts and match every one to its 2021 vote.

Intensity. The binary treatment counts a district as attacked if the party contested at least one of its tally sheets, which says nothing about how much of the result the offensive tried to overturn. We measure that with the threatened vote, the Castillo minus Fujimori margin in the petitioned tables of a district, summed over its uniquely petitioned tables and scaled by the district's valid runoff vote. In the average targeted district the contested tables held about a tenth of the district's net margin for Castillo, and in the most aggressive cases more than four fifths. We relate this intensity to the first-round change in the Fuerza Popular block and, as a mirror, to the change in the left block, since a penalty aimed at the actor should erode the first and spare the second. Figure 8 reports the binned profiles and Table 6 the slopes. Greater intensity does not produce meaningful first-round erosion of the actor. The Fuerza Popular slope is small and not significant under province-clustered inference (-0.20 points per ten points of threatened vote, $p = 0.28$), and the mirror left slope runs the other way ($+0.26$, $p = 0.23$), so the two do not separate ($\gamma = -0.46$, clustered $p = 0.20$). The binned means make the source plain. Across the lower two thirds of intensity the Fuerza Popular change holds near a one to two point gain, and the only downward movement sits in the top intensity tercile, the same sparse high-intensity tail that drives the linear slope. A randomization-inference test that permutes the dose puts the asymmetry at $p = 0.06$, which we report in full, but the effect is economically trivial and confined to a handful of the most aggressively contested districts. The runoff says the same more flatly: even where the offensive threatened most of the margin, Fujimori's second-round share does not move.

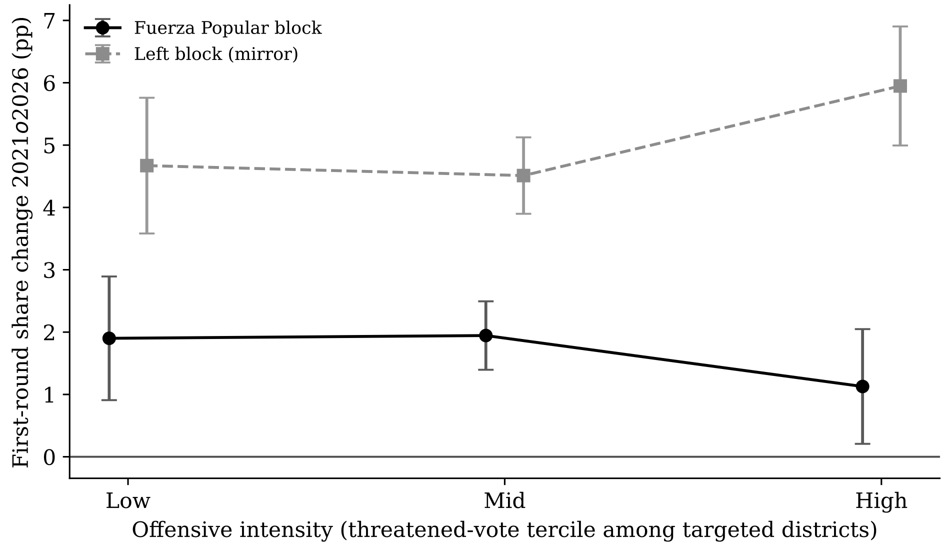


Figure 8: First-round block change and offensive intensity

Notes. Adjusted first-round vote-share change, 2021 to 2026, among targeted districts, by tercile of offensive intensity (the threatened share of the valid vote), for the Fuerza Popular block and the mirror left block. Markers are adjusted means with 95% province-clustered intervals, conditioning on the baseline block share, predetermined covariates, and province fixed effects. A penalty aimed at the actor would steepen the Fuerza Popular profile against intensity and leave the left flat. Both profiles are nearly flat across the lower two terciles, with movement confined to the most aggressively contested districts.

Actor specificity. A penalty aimed at the party that contested the count need not show up in the runoff aggregate, where the same two candidates face off, if it takes the form of first-round defection to a neighbor on the same side. Because party labels and candidates differ across the two elections, we group the first-round vote into comparable political blocks, and compare the targeted districts with their matched control on the change in each block's share. Figure 9 reports the pattern. The Fuerza Popular block falls by about eight tenths of a point more in targeted districts, and a matching rise in the non-Fujimori right absorbs it. The blocks move differently as a group ($p = 0.02$). This is consistent with a first-round withdrawal from the actor, but on its own it is not evidence of actor-specific punishment, and the distinction matters. The left block falls by a statistically indistinguishable amount in the same districts, and a triple difference of the Fuerza Popular block against the left returns essentially zero (-0.1 points, $p = 0.91$). Both blocks drain to the minor right by the same amount. A broad rightward realignment of these heavily ex-Castillo districts would produce exactly that as the national left collapsed after 2021. The offensive could produce parallel falls on the left only through general delegitimization, a disaffection with both poles in the districts it attacked, and not through a penalty aimed at Fuerza Popular. The decomposition on its own therefore cannot separate a localized penalty from realignment, and the intensity test is what adjudicates: under offensive-induced delegitimization the left should erode where the offensive was most intense, and it does not (Table 6). What moves these districts is the collapse of the national left, which bears no relation to how many tables Fuerza Popular contested.

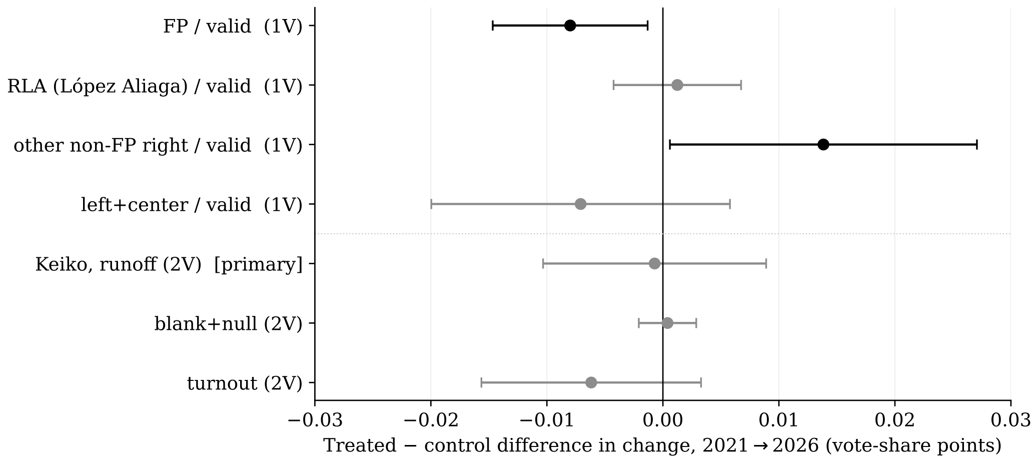


Figure 9: Block-specific electoral change in targeted districts

Notes. Treated minus control differences in the 2021 to 2026 change of each block's vote share, with 95% province-clustered intervals. Because candidates and party labels differ across elections, first-round votes are grouped into comparable political blocks. Filled markers mark intervals that exclude zero. Targeted districts show a modest first-round decline in the Fuerza Popular block and a corresponding rise in the non-Fujimori right. The pattern does not appear in the runoff, where Fujimori is again on the ballot, and the same decline appears on the left, so the decomposition is read as suggestive within-right reallocation, well short of the paper's main causal estimate.

That erosion is diffuse across the right, spreading past the most visible alternative. Separating Renovácíón Popular, the vehicle of Rafael López Aliaga and the strongest right-wing challenger of 2026, from the rest of the non-Fujimori right, its differential change is essentially zero (+0.1 points, not significant), and the entire offsetting gain sits in the remaining minor right parties. The first-round loss also does not concentrate where López Aliaga is locally strong, the targeted-by-his-vote interaction being flat. Targeted districts shed Fuerza Popular toward the broader conservative field, away from the marquee alternative.

A first-round shift from one right party to another need not be punishment. The decline in the Fuerza Popular block could be a limited penalty to the brand or ordinary fragmentation in a more crowded conservative field. Two-round systems invite exactly this divergence, since the first round allows exploratory voting while the runoff forces coordination between the two finalists (Cox, 1997; Bouton, 2013), and Peru's weakly institutionalized, candidate-centered party system makes the right's first-round supply volatile from one election to the next (Levitsky and Cameron, 2003). The pre-treatment window weighs against the pure fragmentation reading. The targeted-control change in Fuerza Popular's first-round share is flat over 2016 to 2021 (+1.2 points, $p = 0.21$), when Peruanos por el Kambio was an equally viable conservative alternative, and turns negative only after the offensive, over 2021 to 2026 (−0.8 points, $p = 0.02$), a difference that is itself significant (−2.0 points, $p = 0.04$). The erosion does not concentrate where the non-Fujimori right is locally larger, the supply interaction being flat and wrong-signed, and it is not a defection to the leading alternative. The decline is not a matching artifact. It holds under first-difference and ANCOVA bracketing, in a district-and-year fixed-effects event study that absorbs each district's level and leaves a clean placebo lead, and when the control is matched on pre-determined structural covariates, not the lagged vote. But robustness is not attribution. The first round is a dirtier comparison than the runoff, the slate of parties and the coordination among them change between elections, and the actor's first-round vote moves non-monotonically across earlier contests (Appendix D). As the block triple difference and the intensity test show, the same withdrawal appears on the left and does not scale with how hard the offensive hit, so a localized penalty cannot be told apart from the broader realignment of these heavily ex-Castillo districts after the left's collapse.

Absorption. The remaining reading is that a real first-round penalty exists but runoff coordination pulls it back, so targeted districts that shed Fuerza Popular in the first round should return to Fujimori in the second more than comparable controls. We test this directly. Within 2026 we measure the first-round to runoff reconstitution, Fujimori's runoff share minus the Fuerza Popular first-round share, and regress it on the first-round erosion, interacting with treatment. The reconstitution is large and rises with erosion for every district (+0.37 points per point of erosion), the mechanical pull of a two-candidate runoff. Absorption requires that targeted districts do this more, and they do not. The differential return is essentially zero and if anything negative ($\theta = -0.09$, clustered $p = 0.15$, randomization-inference $p = 0.37$), and the binned reconstitution profiles of targeted and control districts lie on top of each other (Figure 10). The first-round defection that exists is generic within-right reshuffling that the runoff reconstitutes for targeted and untargeted districts alike, not a penalty of the actor that the second round conceals.

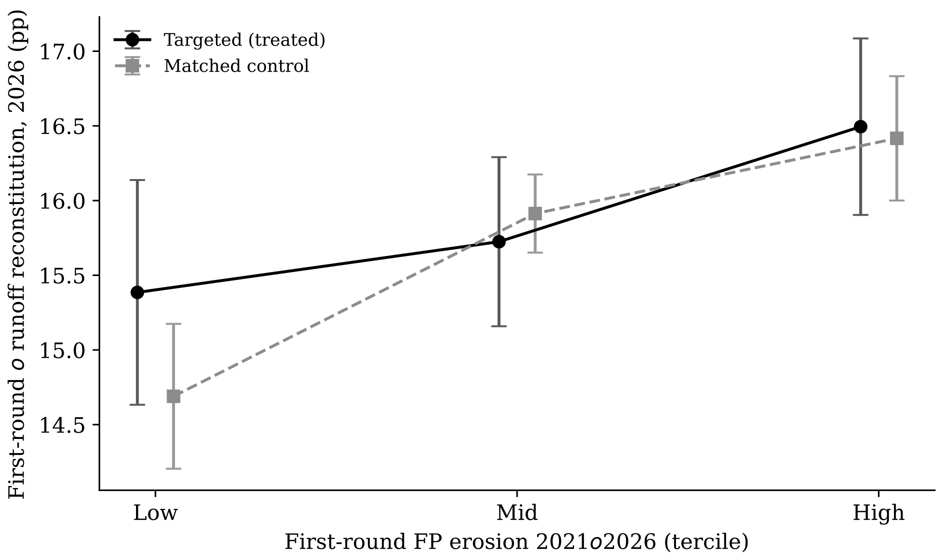


Figure 10: Runoff reconstitution and first-round erosion

Notes. Adjusted first-round to runoff reconstitution in 2026, Fujimori's runoff share minus the Fuerza Popular first-round share, by tercile of first-round Fuerza Popular erosion, for targeted and matched-control districts, with 95% province-clustered intervals. Conditioned on predetermined covariates and province fixed effects. Absorption of a real penalty would lift the targeted profile above the control as erosion rises. The two profiles coincide.

Table 6: First-round intensity and runoff-absorption tests

	Estimate	(SE)	RI p
Panel A. First-round intensity (per 10 pp threatened vote)			
Fuerza Popular block change	-0.20	(0.19)	0.086
Left block change (mirror)	+0.26	(0.21)	–
Asymmetry (FP − Left) = γ	-0.46	(0.36)	0.064
Panel B. Runoff reconstitution (per 1 pp first-round FP erosion)			
First-round→runoff slope (common)	+0.37	(0.07)	–
Differential return in targeted = θ	-0.09	(0.06)	0.370

Notes. Panel A regresses the 2021 to 2026 first-round change of each block on offensive intensity (threatened share of the valid vote), among targeted districts, with province fixed effects and province-clustered standard errors. γ is the differential slope on the Fuerza Popular block relative to the left. Panel B regresses the within-2026 first-round to runoff reconstitution on first-round Fuerza Popular erosion, treated against control. θ is the differential reconstitution in targeted districts, the absorption signature. The RI column reports randomization-inference p-values, permuting the dose in Panel A and the treatment label within province in Panel B, over 2,000 draws.

Together the three tests close the two ways the runoff null could hide a real penalty. The offensive did not induce first-round disaffection that scales with its intensity, symmetrically or on the left, and the first-round defection that does occur is not differentially reconstituted in the runoff. What remains is realignment, exogenous to how many tally sheets Fuerza Popular contested. The runoff null marks the absence of a punishment. Even where the offensive threatened more votes, targeted districts do not turn against the actor, in either round.

6.4 Constrained choice

The absence of a runoff penalty need not mean that voters approved of the annulment offensive. A second-round null is also what we would observe if voters who were unhappy with Fuerza Popular nonetheless returned to it once the alternative was a left-wing candidate they read as risky. The runoff is a choice between two risks, not a referendum on the 2021 offensive, and the sequence in Figure 1 places the 2026 contest at the end of a series in which that tradeoff has repeatedly decided Peruvian runoffs.

Three runoffs make the pattern concrete. In 2011 Ollanta Humala, a nationalist read as radical in 2006, moderated toward the center and won. In 2021 Pedro Castillo was a riskier candidate of the left, and anti-Fujimori coordination still carried him, but by the narrowest of margins. In 2026 Juntos por el Perú fielded Roberto Sánchez, who had served as Castillo's trade minister and ran as an avowed *castillista* (CIDOB, 2026), while the collapse of the Castillo government across 2021 and 2022 had raised the perceived cost of that choice. Humala reduced the perceived risk of the left and Castillo survived it, while Sánchez carried its heaviest version into the 2026 runoff.

This sequence fits the pattern in the data, where the modest first-round movement away from Fuerza Popular reshuffles within the right and does not carry to the runoff. That reconstitution is common to targeted and untargeted districts alike (Section 6.3), so the decisive contest is the one in which any first-round erosion is least likely to surface.

We offer this as an interpretation of the pattern, not as a separately identified effect. The tests in Section 6.3 find no first-round penalty that scales with the offensive and no differential runoff reabsorption in targeted districts, so the evidence favors an electorate that did not punish over one whose punishment was coordinated away. The two are not mutually exclusive in their consequence. Once the election collapses into a binary choice, the coordination that majority-runoff rules force on voters as the field narrows to two (Cox, 1997; Bouton, 2013) reconstitutes the right onto Fujimori whether or not first-round displeasure exists. The runoff null shows only that the decisive vote does not register a verdict on the offensive. In this sense the 2026 outcome reflects constrained choice more than forgiveness.

7 Material Disruption

The material episode asks the accountability question in its hardest form, when the harm is real but no one in particular caused it. We treat it as secondary and conceptual: it marks the opposite pole of the attribution problem from the annulment offensive. Construction on the Lava Jato megaprojects stopped in 2016 and 2017 when the implicated firms lost financing and their contracts were resolved, the largest cases being the cancellation of the Gasoducto Sur Peruano in January 2017 and the paralysis of the Chavimochic III irrigation works in December 2016. We measure a district's exposure as physical crossing by one of these core Lava Jato corridors, georeferenced from the concession traces, and classify each project's 2016 construction status as active, partial, or already complete. The natural outcome is the invalid-vote share, blank and null ballots over ballots cast, which registers protest without requiring voters to name a culprit. The municipal calendar brackets the shock with a regular election in 2014 before it and in 2018 after it.

Rebuilt on the genuine Lava Jato corridors, the material arm yields no clean district-level punishment. Table 7 reports the invalid-vote difference-in-differences by 2016 construction status, each against the pre-shock 2010–2014 placebo. The aggregate contrast between actively-disrupted and not-exposed districts is small, insignificant, and if anything wrong-signed, and the few positive signals do not survive scrutiny. Districts crossed by the paralyzed Chavimochic works show a large rise in invalid voting, but they are the La Libertad coastal valleys that El Niño Costero devastated in March 2017, between the two elections, so the rise cannot be separated from the disaster. The Gasoducto Sur districts in the south, beyond the reach of El Niño, show a rise on the precise built trace that fails a pre-treatment placebo, and a clean placebo only when the exposed set is widened to most of the southern highlands, where it captures a regional trend in null voting, away from the pipeline. No specification isolates a disruption effect that survives both a placebo and a plausible confound, so we read the material arm as a null: an authorless economic shock produces no detectable local electoral penalty, the same answer the attributable arm gives. An earlier version of this exposure, built on a broad set of public-works procurement records instead of the Lava Jato megaprojects, produced a suggestive gradient. That gradient does not survive correct corridor definition, and the full correction is documented in the replication package. This is the other extreme of the attribution problem, and it points the same way as the rest of the paper.

Table 7: Material disruption: invalid-vote difference-in-differences by exposure group

Exposure group	n	Δ14–18 vs ctrl	placebo 10–14	Reading
Active construction (aggregate)	52	-2.2	+5.0	placebo not clean
of which Chavimochic	9	+10.4	+0.9	confounded by El Niño Costero
Partial construction	8	-5.3	+12.8	n small
Completed pre-2016 (placebo)	131	+2.4	-1.5	finished ∼2011, no 2016 shock
Gasoducto Sur (southern route)	135	+1.7	+0.5	regional trend, not pipeline

Notes. Each cell is the change in a district's invalid-vote share (blank plus null, in percentage points) relative to not-exposed districts, across the municipal elections that bracket the shock (2014 to 2018), with the pre-shock 2010 to 2014 window as a placebo. Groups are defined by the 2016 construction status of the core Lava Jato corridor crossing the district. The active aggregate is small and wrong-signed. Chavimochic rises but coincides with the El Niño Costero disaster of 2017, and the southern Gasoducto Sur districts move with a regional trend that a placebo on the built trace does not support. No group isolates a disruption effect that clears both a placebo and a plausible confound.

8 A Behavioral Benchmark

We complement the nulls with an external benchmark. We apply a regression discontinuity to close municipal races, comparing the party that barely wins to the party that barely loses. This is a different identification, estimand, and population from the matched comparison of the annulment arm. It recovers a substantive incumbency effect, which shows the data are not inert, but it does not validate that arm's identifying assumptions or establish its power, which we read from the equivalence bound in Section 6. Figure 11 shows the result. A party that wins the 2010 mayoralty is about six percentage points less likely to win in 2014 than a party that loses narrowly. This is the incumbency disadvantage that Klašnja and Titiunik (2017) document for Brazilian municipalities. The estimate is robust to the data-driven bandwidth and to bias-corrected inference, and it passes the standard density and covariate-balance checks. The published evidence on incumbency in Peru is mixed, so the estimate also contributes a number to that debate, for the period before the 2015 ban on immediate mayoral re-election.

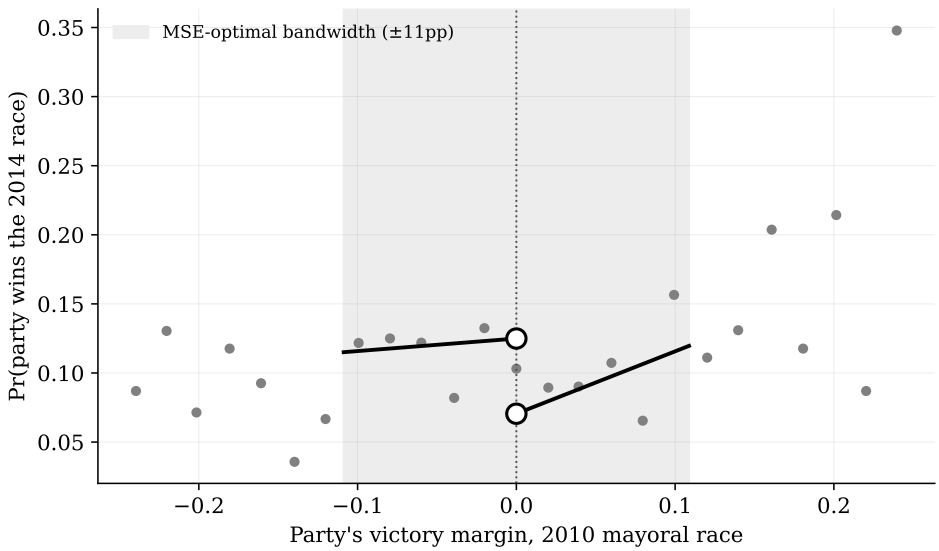


Figure 11: Probability of winning in 2014, by 2010 victory margin

Notes. Binned means of the probability that a party wins the 2014 mayoral race, against its margin of victory in 2010, with local-linear fits estimated within the shaded mean-squared-error-optimal bandwidth. The open circles mark the fitted values on each side of the threshold, and the gap between them is the incumbency disadvantage of Klašnja and Titiunik (2017): a party that barely wins in 2010 is less likely to win in 2014.

The recall referendum provides a second test. Artiles et al. (2021) show that recall lowers the quality of the candidates who run. We find the same force on the extensive margin. The party of a recalled mayor re-enters the next ballot at the same rate as the party of a mayor who survived a recall, and about half of all parties do not re-enter at all. Conditional on re-entering, the party of a recalled mayor loses about five percentage points more than the party of a mayor who survived. That conditional loss is the voter response their design does not measure. The same specifications also recover the realignment of the left vote between the 2021 and 2026 runoffs, with a t above seven. The design detects punishment where it is present and finds no differential punishment on either of our episodes.

9 Mechanisms

If the targeted districts do not turn against the actor, the question is why. We probe five candidate moderators of the treatment effect and treat this as exploratory heterogeneity analysis. For each we report the interaction with the treatment and the treated−control difference within terciles, and we correct for the five comparisons with both the Holm procedure and a Romano-Wolf step-down.

Each moderator stands for an account of the null, and each account predicts where punishment should concentrate if it is the right one. The first is polarization, measured by how close the district's 2021 result sat to an even split. If partisan loyalty is what blocks punishment, the block should loosen where loyalty is weakest, in the competitive districts, and punishment should surface there. The second is development, proxied by urbanization.

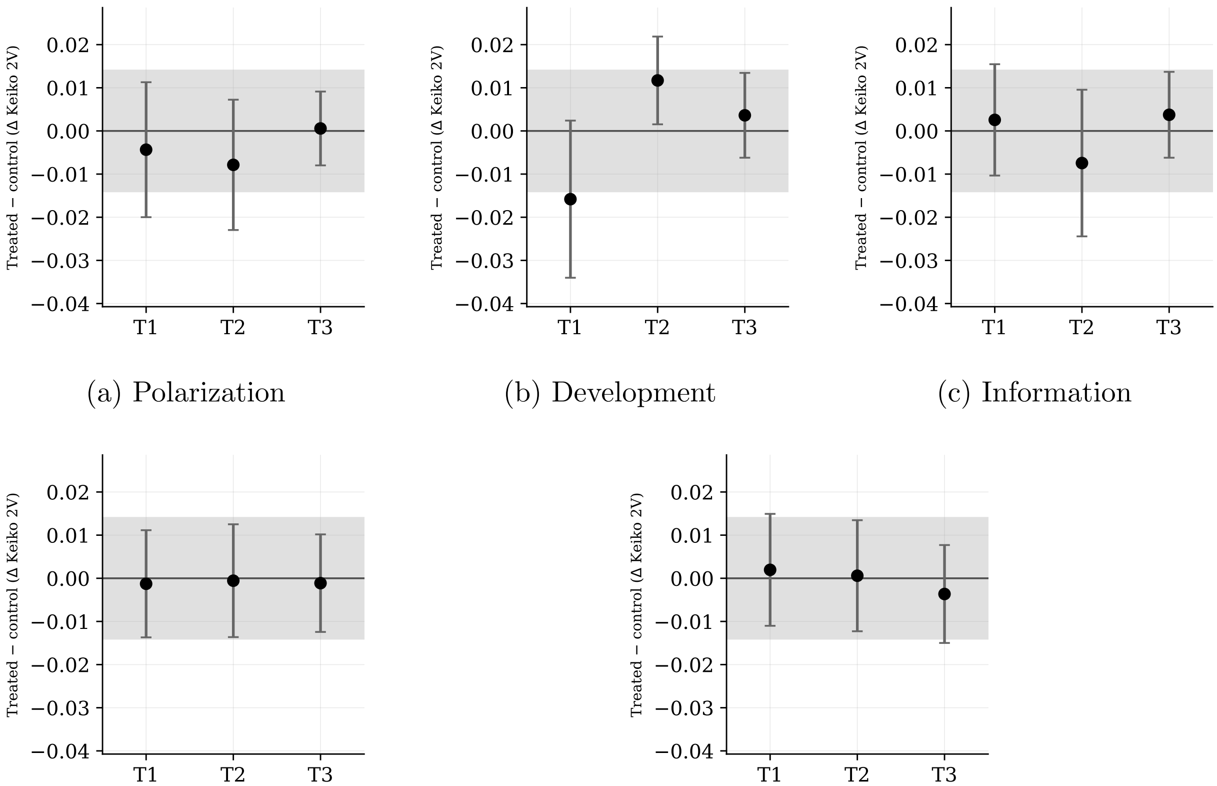
If punishing electoral subversion is a second-order concern that material want crowds out, the sanction should appear in the richer, more urban districts and fade in the poorer ones. The third is information, the share of households with internet access. If voters failed to punish only because the offensive never reached them, the sanction should be present where information is densest. The fourth is education, the district's average schooling, a proxy for the political sophistication that accurate attribution is held to require (Achen and Bartels, 2016), so punishment should rise with it. The fifth is clientelism, the central transfers a district's government receives per capita. If voters in transfer-dependent districts reward the machine that delivers over the principle it violates, the sanction should appear where that dependence is weakest. Polarization comes from the electoral panel, and the remaining four from the 2017 census and the 2020 record of transfers to local governments. We standardize each moderator, interact it with the treatment, and report both the interaction and the treated−control difference within terciles. The census and fiscal layers are linked to the electoral panel by district name, which matches between 85 and 99 percent of districts for the census moderators and 88 percent for transfers. Measurement error in the unmatched remainder would attenuate the interactions toward zero, but the match rates are high enough that it cannot manufacture the flat profile we report.

Table 8: Heterogeneity of the attributive effect by mechanism

Moderator	δ (T×M)	SE	p	Holm p	RW p
Polarization	+0.0072	0.0044	0.102	0.407	0.151
Development (urban)	+0.0051	0.0032	0.112	0.407	0.151
Information (internet)	+0.0081	0.0037	0.029	0.146	0.074
Education (years)	+0.0022	0.0035	0.542	0.542	0.480
Clientelism (transfers pc)	-0.0036	0.0032	0.260	0.519	0.287

Notes. Each row is the coefficient on the interaction of treatment with the standardized moderator in the change of the runoff vote, with province-clustered standard errors. The final column reports the Holm- and Romano-Wolf-adjusted p-values across the five tests. Development is proxied by the urban share and clientelism by the log of 2020 transfers per capita, the latter matched to the electoral panel for 88 percent of districts.

We do not detect heterogeneity in any of the five. No interaction survives either correction, and Figure 12 shows the treated−control difference flat and inside the equivalence band across every tercile. This is a bounded statement, not a claim of exact uniformity. With 249 treated districts split across moderators, each interaction is estimated to within about one point per standard deviation of the moderator, so we can rule out heterogeneity larger than that but not smaller gradients. Each account named a place where the sanction should have surfaced, and in none of those places does it surface. The development account looked for it among the urban and better-off districts, the sophistication account among the more educated, the clientelism account among the least transfer-dependent, and the informational account among the most connected. The treated−control difference is flat and inside the equivalence band across the terciles of all four. This matters because these are precisely the accounts under which the null would be an artifact of who the offensive reached, not a statement about how voters weigh it. Their failure does not prove the remaining account, but it removes the explanations that would let the null be dismissed as a problem of exposure or capacity.



(a) Polarization (b) Development (c) Information

(d) Education (e) Clientelism

Figure 12: Treated−control difference by tercile of each moderator

Notes. Each panel plots the treated−control difference in the change of the runoff vote within terciles of the moderator, with 95% province-clustered intervals. The shaded band marks ±1.4 percentage points. No moderator produces a monotone gradient or a tercile outside the band.

The competitive districts deserve a separate line. If partisan loyalty were what blocked punishment, it would relax where loyalty is weakest, in the districts whose 2021 result sat closest to a tie, and punishment would surface there. It does not. In the tercile nearest a tied result the treated−control difference is −0.004 and statistically indistinguishable from zero. This is the test partisan-identity theory most demands, and it is also the least powered, with only a third of the treated districts: the minimum detectable effect in this tercile is 2.2 points, not the 1.4 of the full sample, so we can rule out punishment larger than 2.2 points where the theory most expects it, not smaller gradients.

The information channel is the only moderator significant before correction, and its sign is positive: where household internet access is denser, the actor's vote if anything held up better. We read this as the expected direction under motivated reasoning. Where partisans engage more with political information in a polarized environment, counter-attitudinal signals are discounted or rejected, not acted on, and integrity information can entrench partisan support instead of eroding it (Kunda, 1990; Little, 2019). The implication for the null is the one that matters: the account in which targeted districts failed to punish only because the offensive never reached them requires the opposite sign. Greater information should have produced *more* punishment under an information-deficit story. It produced none, which removes that explanation instead of supporting it. The same conclusion follows from the attributable arm, where punishment did not grow with the intensity of the offensive (Section 6). The reach of information and the intensity of the treatment are independent margins, and neither produces punishment.

A natural worry is that targeted districts registered the offensive in attitudes even if not in votes. They did not. The governance module of the national household survey asks confidence in the two electoral authorities the offensive attacked, the National Elections Jury and the national elections office, and we compare trust in targeted and matched control districts before the offensive (2018–2019) and after it (2022–2025), with district and year fixed effects and province-clustered errors. The treated−control difference in the change is -0.002 on the institutions' four-point scale ($p = 0.94$) and is null for each body separately. Trust fell by about the same amount in both groups, tracking a national decline instead of a targeted one. The offensive did not move trust in the electoral institutions where it landed, just as it did not move the vote. This is consistent with, not contrary to, the finding that contested counts erode confidence in the result chiefly among the *loser's own* supporters (Hernández-Huerta and Cantú, 2022): the districts the offensive targeted are the ones that voted against its author, the population for whom its claim of fraud was never credible. Disputing a count is in this sense a strategic act addressed to a candidate's own base and to the legitimacy of the winner, not a bid for the voters where the contested sheets happen to lie (Hernández-Huerta, 2020). A null in those districts, in trust as in vote, is what that reading predicts. Two limits keep the claim narrow. The survey result is attitudinal and cannot speak to vote choice. And because the control districts are themselves anti-Fujimori places matched to the treated, the comparison bounds the *differential* trust response across comparable populations, not whether the offensive moved trust nationally. That national question is the same boundary the rest of the design respects. That the offensive moved neither vote nor trust where it landed, relative to comparable places, is the contribution here.

With the accounts that would make the null a feature of circumstance set aside, information, want, sophistication, clientelism, and partisan loyalty in its sharpest form, what remains is that the null reflects how these voters weigh post-electoral subversion against the other stakes of the election. Svolik (2019) formalizes the tradeoff: even voters who value fair competition decline to sanction a co-ideological transgressor when the sanction means electing a platform they oppose, and the price of punishing rises with polarization. In a polarized contest a voter who agrees with the actor on those stakes does not treat an attempt to overturn the count as disqualifying, and the experimental literature, in which the penalty for anti-democratic conduct is small and falls with copartisanship (Graham and Svolik, 2020; Carey et al., 2022), points the same way. We cannot rule out that the offensive simply registered too faintly to move any voter, and we do not claim to separate faint salience from settled preference. What the mechanisms establish is the narrower and more useful point that the null is not the residue of poor or disconnected districts failing to react. It is the micro-foundation for the reading we propose in the conclusion, in which an electorate that does not visibly price subversion would leave its deterrence to institutions other than the vote.

10 Conclusion

Voters in our setting do not differentially punish political instability. The districts a no-author economic disruption hit show no greater rise in protest voting than comparable unexposed districts, and the districts an identifiable actor targeted in attempting to overturn a national election did not turn against that actor relative to comparable untargeted ones. A regression discontinuity at close municipal races recovers the incumbency disadvantage of weak-party systems, which shows the electoral data can recover effects of economically meaningful magnitude under a strong local design. A sanction common to the whole country we cannot see, but the local, differential sanction is absent, and it is absent most cleanly in the very districts that bore the offensive most directly and those whose 2021 result sat closest to a tie.

Why the sanction does not appear we cannot settle, but the evidence narrows the field. The two readings that would make the null an artifact of circumstance, that voters never learned the offensive had reached them or that poverty crowds out the luxury of punishing subversion, both fail their tests. The information channel runs positive, the actor's vote if anything higher where information is denser, the direction motivated reasoning predicts, and the null is uniform across development, education, and the competitiveness of the district. What remains is a reading in terms of preferences. In a polarized electorate, voters who side with an actor on the stakes of the election do not abandon the actor for contesting its result, and the experimental literature that finds a small penalty for anti-democratic conduct, attenuated by copartisanship, points the same way (Graham and Svolik, 2020; Svolik, 2019). The flat response to the intensity of the offensive is consistent with that reading and with limited salience, and we do not claim to separate the two. A structural reading goes further. The runoff, our clean margin, is by the same coordination logic the margin least able to express a penalty, since the second round forces the right to consolidate against the left, while the first round, where displeasure could surface, cannot be separated from the realignment of these ex-Castillo districts. Two-round competition reconstitutes the right against the left whether or not a subversion provokes first-round displeasure, so any such displeasure does not reach a decisive vote, a reason such conduct may go locally unpunished in majority-runoff systems quite apart from whether voters privately condemn it.

The finding connects the voter's decision to the cost of instability. Instability lowers investment and productivity, a cost spread thin across the country and realized over years (Alesina et al., 1996; Aisen and Veiga, 2013). The return to the actor who produces it, from contesting a narrow defeat or from the rents behind a rival's paralyzed projects, is concentrated and immediate. When the ballot box does not return the diffuse and delayed cost to the actor who took the concentrated and immediate benefit, the price the actor faces is too low, and an action that is underpriced recurs. This is the sense in which the null helps explain a pattern the macroeconomic literature documents without accounting for, the persistence of instability in the places where its economic toll is highest. The deterrent that accountability is meant to supply is, at least in the local response we can observe, not supplied by the electorate.

What binds, if not the voter? Here we can only propose an answer, because our evidence speaks to the local electoral response and the rest is imported. In 2021 the formal guardrails held. The electoral juries adjudicated the annulment petitions and upheld none, and the result was certified despite the campaign against it, and the pattern repeated in 2026. The backsliding literature casts the courts and electoral authorities as the guardrails and the voter as the line behind them (Levitsky and Ziblatt, 2018). Our local null, together with the small and copartisan-attenuated penalties the experimental literature reports (Graham and Svolik, 2020; Carey et al., 2022), is consistent with that line being slack in a polarized young democracy, which would leave the burden of deterring electoral subversion on the institutions in front of it. We offer this as the reading our evidence supports, not as a verdict our design delivers. What the design delivers is narrower and firmer. Where the offensive landed, the voters it targeted did not turn against the actor, and a national reckoning, if one occurred, is beyond what we can see.

The result should travel best to settings that resemble ours, young democracies with weak parties, high electoral volatility, and polarization, where the experimental penalty for anti-democratic conduct is already smallest. Three limitations bound it. The design speaks to the local, differential response and cannot see a sanction common to the whole country, which would require time-series or cross-national variation we do not have. The local reading rests on the assumption that targeted voters knew their sheets had been challenged, which the provincial adjudication makes plausible but which we do not measure. And the runoff outcome carries a pre-treatment trend, so the precise equivalence bound holds under parallel trends while only the sign of the null survives substantial violations of them. What would most sharpen the picture is direct evidence on whether the offensive was perceived locally, from regional media or from survey measures of fraud belief tied to geography, which we leave to later work.

Accountability presumes that voters price the conduct of those who seek to govern them. For the most consequential conduct of all, the attempt to overturn an election, we find that in this setting they do not.

References

Achen, C. H. and Bartels, L. M. (2016). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton University Press.

Aisen, A. and Veiga, F. J. (2013). How does political instability affect economic growth? *European Journal of Political Economy*, 29:151–167.

Alesina, A., Özler, S., Roubini, N., and Swagel, P. (1996). Political instability and economic growth. *Journal of Economic Growth*, 1(2):189–211.

Artiles, M., Kleine-Rueschkamp, L., and León-Ciliotta, G. (2021). Accountability, political capture, and selection into politics: Evidence from peruvian municipalities. *The Review of Economics and Statistics*, 103(2):397–411.

Bermeo, N. (2016). On democratic backsliding. *Journal of Democracy*, 27(1):5–19.

Bouton, L. (2013). A theory of strategic voting in runoff elections. *American Economic Review*, 103(4):1248–1288.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Carey, J., Clayton, K., Helmke, G., Nyhan, B., Sanders, M., and Stokes, S. (2022). Who will defend democracy? evaluating tradeoffs in candidate support among partisan donors and voters. *Journal of Elections, Public Opinion and Parties*, 32(1).

CIDOB (2026). Roberto sánchez palomino: biografía. Barcelona Centre for International Affairs (CIDOB), Biografías Líderes Políticos. Accessed 18 June 2026.

Colonnelli, E. and Prem, M. (2022). Corruption and firms. *The Review of Economic Studies*, 89(2):695–732.

Cox, G. W. (1997). *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge University Press, Cambridge.

Ferraz, C. and Finan, F. (2008). Exposing corrupt politicians: The effects of brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics*, 123(2):703–745.

Fiorina, M. P. (1981). *Retrospective Voting in American National Elections*. Yale University Press, New Haven.

Graham, M. H. and Svolik, M. W. (2020). Democracy in america? partisanship, polarization, and the robustness of support for democracy in the united states. *American Political Science Review*, 114(2):392–409.

Hartman, E. and Hidalgo, F. D. (2018). An equivalence approach to balance and placebo tests. *American Journal of Political Science*, 62(4):1000–1013.

Healy, A. and Malhotra, N. (2013). Retrospective voting reconsidered. *Annual Review of Political Science*, 16:285–306.

Hernández-Huerta, V. (2020). Disputed elections in presidential democracies: Contexts of electoral "blackmail". *The Journal of Politics*, 82(1):89–103.

Hernández-Huerta, V. and Cantú, F. (2022). Public distrust in disputed elections: Evidence from latin america. *British Journal of Political Science*, 52(4):1923–1930.

Klašnja, M. and Titiunik, R. (2017). The incumbency curse: Weak parties, term limits, and unfulfilled accountability. *American Political Science Review*, 111(1):129–148.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3):480–498.

Levitsky, S. and Cameron, M. A. (2003). Democracy without parties? political parties and regime change in fujimori's peru. *Latin American Politics and Society*, 45(3):1–33.

Levitsky, S. and Ziblatt, D. (2018). *How Democracies Die*. Crown, New York.

Little, A. T. (2019). The distortion of related beliefs. *American Journal of Political Science*, 63(3):675–689.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5):2555–2591.

Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11):1208–1214.

Svolik, M. W. (2019). Polarization versus democracy. *Journal of Democracy*, 30(3):20–32.

Appendices

These appendices are supplementary to the main text. They document the data and its construction (Appendix A), the specification curve (Appendix B), the validity of the behavioral benchmark (Appendix C), the pre-treatment trends and trend-sensitivity analysis (Appendix D), and the selection model (Appendix E).

Appendix A Data and Construction

This appendix documents the electoral panel, the treatment, and the merges behind the estimation sample.

Electoral returns. We use the official tally-sheet records (*actas electorales*) of the national election office, ONPE, for the presidential contests of 2006, 2011, and 2016 in both rounds and for the municipal elections of 2006 through 2022. For the 2021 runoff we use the official final district returns published by INFOgob, the information service of the national jury, JNE. The ONPE acta scrape for 2021 stalled near 95 percent and left about forty heavily pro-Castillo districts in the VRAEM at zero, which biases the baseline downward for the left. The INFOgob file recovers them, and Castillo's domestic margin on the complete file is 0.88 percent of the valid vote against 0.25 percent nationally, the difference being the overseas vote, which leans to Fuerza Popular. For 2026 we collected the runoff returns from the ONPE results portal at the near-final count of 99.4 percent of tally sheets, with all sheets processed and the remainder before the electoral juries. The result was not proclaimed at that vintage and we assert no winner. The attributable estimate is unchanged between the 98.3 and 99.4 percent counts.

Treatment. The petitions to annul tally sheets come from the expediente records of the electoral juries. Fuerza Popular filed 942 and Perú Libre 144, 1,086 in all. We geolocate each petition to its district of origin from the polling-station identifier on the expediente. A versioned manual name crosswalk resolves the one spelling variant the automatic match misses, the district the jury typed as Huaillati against the roll's Huayllati, so all 942 Fuerza Popular petitions locate to a district. The petitions fall in 285 districts, the treatment universe.

Estimation sample. We model the propensity to be petitioned from pre-treatment characteristics and keep the treated districts and the impugnable-but-not-impugned controls that share common support. This leaves 249 treated districts matched to 1,149 controls, an estimation sample of 1,398. The trim drops 36 of the 285 petitioned districts, the highest-propensity ones, which are not more pro-Castillo than the retained, and it removes the single most-petitioned district, so the dose-response does not reach the very top of the intensity range.

Ubigeo and merges. ONPE and INEI number districts differently, so we never merge electoral returns to the INEI shapefile or to the household survey by code. We bridge them by normalized district name, with the crosswalk above for the exceptions. Each merge is reconstructed from its upstream source and the unmatched records on either side are classified, not dropped silently.

Material arm and trust. The Lava Jato exposure uses the curated set of core cartel concessions georeferenced to the district map, as described in Section 7. The trust measure is the governance module of the national household survey, ENAHO, items on confidence in the National Elections Jury and the national elections office, fielded annually from 2004 to 2025.

Appendix B The Specification Curve

Figure 7 traces the attributive estimate across the reasonable specifications as a curve. Table 9 lists the underlying 36. Each row varies the electorate-size cutoff that defines the sample, the propensity-score trim that sets the common support, and the covariates in the selection logit. The treated−control estimate stays within a narrow band around zero throughout, and none of the 36 reaches significance.

Table 9: The 36 specifications of the attributive estimate

electors >	trim	covariates	n_T	(A−B)	t
20	0.01	base	271	+0.0008	+0.16
20	0.01	minimal	271	+0.0024	+0.50
20	0.01	share-valid	271	+0.0011	+0.20
20	0.05	base	249	-0.0007	-0.15
20	0.05	minimal	249	+0.0004	+0.08
20	0.05	share-valid	249	-0.0007	-0.14
20	0.10	base	221	-0.0022	-0.47
20	0.10	minimal	221	-0.0020	-0.43
20	0.10	share-valid	221	-0.0036	-0.83
50	0.01	base	271	+0.0008	+0.16
50	0.01	minimal	271	+0.0024	+0.50
50	0.01	share-valid	271	+0.0011	+0.20
50	0.05	base	249	-0.0007	-0.15
50	0.05	minimal	249	+0.0004	+0.08
50	0.05	share-valid	249	-0.0007	-0.14
50	0.10	base	221	-0.0022	-0.47
50	0.10	minimal	221	-0.0020	-0.43
50	0.10	share-valid	221	-0.0036	-0.83
200	0.01	base	271	+0.0008	+0.14
200	0.01	minimal	271	+0.0024	+0.50
200	0.01	share-valid	271	+0.0009	+0.16
200	0.05	base	249	-0.0007	-0.15
200	0.05	minimal	249	+0.0004	+0.08
200	0.05	share-valid	249	-0.0008	-0.15
200	0.10	base	221	-0.0022	-0.48
200	0.10	minimal	221	-0.0020	-0.43
200	0.10	share-valid	221	-0.0035	-0.81
500	0.01	base	268	+0.0011	+0.21
500	0.01	minimal	268	+0.0025	+0.55
500	0.01	share-valid	268	+0.0009	+0.17
500	0.05	base	246	+0.0001	+0.02
500	0.05	minimal	246	+0.0000	+0.00
500	0.05	share-valid	246	-0.0008	-0.17
500	0.10	base	218	-0.0030	-0.72
500	0.10	minimal	218	-0.0019	-0.44
500	0.10	share-valid	218	-0.0036	-0.90

Notes. Each row is one specification of equation (1). The first three columns give the electorate-size cutoff, the propensity-score trim, and the covariate set in the selection logit. The last two give the number of treated districts and the treated−control estimate with its cluster t-statistic.

Appendix C Validity of the Behavioral Benchmark

The benchmark in Section 8 recovers the incumbency disadvantage at close municipal races, a clean estimate for pre-2015 Peru and an external check that an electoral effect is detectable in these data, not a validation of the annulment design or a statement about its power. Table 10 reports the regression-discontinuity estimates with a mean-squared-error-optimal bandwidth and robust bias-corrected inference, together with the validity checks. The density test finds no sorting of races across the cutoff, and the pre-determined covariates are balanced across it, so the close races behave as good as randomly assigned.

Table 10: Regression-discontinuity estimates and validity placebos

Outcome	coef.	bw (pp)	N_h	robust p	robust 95% CI
Wins 2014 (Pr)	-0.0632	10.9	2376	0.025	[-0.118, -0.008]
Vote share 2014	+0.0001	11.7	2443	0.916	[-0.021, +0.024]
Validity placebos					
Density (McCrary, p)	0.948				
Balance, 2006 party share	+0.0091	robust $p = 0.338$			
Balance, party ran in 2006	-0.0075	robust $p = 0.872$			

Notes. The running variable is the party's 2010 margin of victory. The top panel estimates equation (5) for two outcomes. The lower panel reports the McCrary density test and the balance of two pre-determined covariates, the party's 2006 vote share and whether it ran in 2006.

Appendix D Pre-Treatment Trends

Two pre-treatment checks bear on the result. The first concerns the first-round arm, which is negative on its own. Table 11 shows the treated−control difference in the actor's first-round vote across 2011, 2016, and 2021. The changes are large and switch sign, so the first-round path is not a stable trend that an adjustment could remove, and we do not lean on that arm.

The second concerns the primary runoff outcome, and it is the more demanding check. The last row of Table 11 is a placebo on the runoff change between 2016 and 2021, before the offensive. It is negative and significant, at −3.8 points. Targeted districts were already moving away from the actor in the runoff before the treatment, because the 2021 realignment pulled the most pro-Castillo districts left, and those are the districts the offensive later targeted. This is a level shift instead of a continuing trend. Between 2021 and 2026 the two groups move together, each gaining about two points for the actor (Figure 5), so the pre-treatment gap neither continues nor reverses differentially, which is what a mean-reversion account masking real punishment would require. We therefore rest the result on the unconditional null in the primary outcome, not on a trend adjustment, and we read the equivalence bound as a statement about the differential change between the two runoffs.

We make this sensitivity explicit in the style of Rambachan and Roth (2023). The runoff outcome has three district-level periods, 2016, 2021, and 2026, and the offensive falls after the 2021 runoff, so an event study with district and year fixed effects has one pre-period coefficient and one post-period effect. In the unadjusted event study the pre-period coefficient is +3.8 points ($t = 1.8$) and the post-period effect is zero. The single pre-period sets a demanding benchmark, because the only interval available, 2016 to 2021, contains the largest electoral realignment in recent Peruvian history, the change of opponent from Kuczynski to Castillo, which pulled the most pro-Castillo districts left, and those are the districts the offensive later targeted. Assuming that the 2021-to-2026 violation, with the same candidate and no comparable realignment in these districts, could be as large as that realignment ($\bar{M} = 1$) is conservative in the extreme.

This motivates a flexible adjustment instead of a replacement. We add interactions of the matching covariates with each period, which lets the trend differ by the 2021 Castillo share, turnout, district size, and rurality, and so removes the realignment directly. The adjusted pre-period coefficient falls to +1.7 points and is no longer significant ($t = 1.1$), and the adjusted post-period effect is -0.6 points and insignificant. Across both specifications and every $\bar{M}$, the robust interval is centered near zero and never excludes it, so the sign of the result, no punishment, is robust to substantial departures from parallel trends. The precision is weaker than the unadjusted equivalence bound suggests. At $\bar{M} = 1$ the adjusted robust interval runs to about -3.4 points and the unadjusted to about -5 points, so what we state with confidence is that any differential punishment is smaller than about five points, not that it is below the 1.4-point equivalence bound, which holds only under exact parallel trends. With a single usable pre-period the relative-magnitude restriction is an assumption anchored to a benchmark violation, the 2021 realignment, not a quantity learned from a sequence of pre-trends, so we report the whole ladder from assumption to bound, not a single $\bar{M}$. Under exact parallel trends the bound on differential punishment is 1.4 points, allowing flexible trends in the matching covariates it is 1.7 points, and allowing a violation as large as the 2021 realignment itself it is about five points (Figure 13). The unadjusted and adjusted effects, near zero and -0.6 points, are estimates under different trend adjustments, not competing claims, and both are small and insignificant. We report the covariate adjustment as a modification made after the sensitivity analysis, not before it.

Table 11: Pre-treatment changes and the runoff placebo

Period change	(A−B)	cluster SE	t	p
1V change 2011→2016 (pre)	-0.0256	0.0107	-2.40	0.016
1V change 2016→2021 (pre)	+0.0125	0.0101	+1.23	0.217
1V change 2021→2026 (outcome)	-0.0079	0.0034	-2.29	0.022
2V change 2016→2021 (placebo P3)	-0.0376	0.0175	-2.15	0.031

Notes. Each row regresses a period change in the vote on the treatment indicator, on treated and impugnable-control districts, with province-clustered standard errors. The first two rows are pre-treatment first-round changes, the third is the first-round outcome, and the last is a pre-treatment placebo on the runoff change.

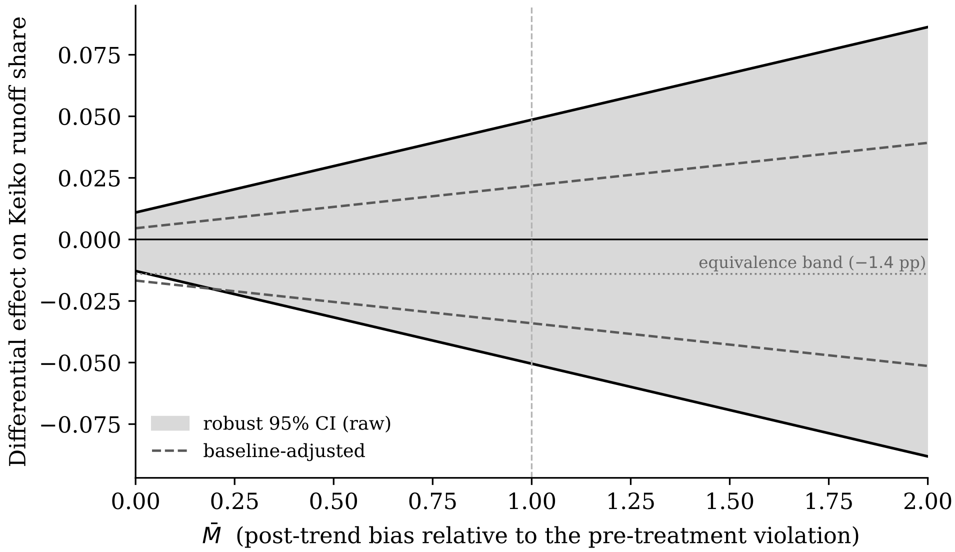


Figure 13: Trend-sensitivity of the runoff effect (relative magnitudes)

Notes. Robust 95% confidence interval for the 2021–2026 treated−control runoff difference as a function of $\bar{M}$, the size of the post-treatment parallel-trends violation relative to the single observed pre-treatment violation (2016 to 2021), in the manner of Rambachan and Roth (2023). The shaded band is the unadjusted interval and the dashed lines the interval after interacting the matching covariates with period. The interval is centered near zero throughout and never excludes it. At $\bar{M} = 1$ it reaches about five points unadjusted and three and a half adjusted.

Appendix E The Selection Model

The estimation compares treated districts to controls on the common support of the propensity to be targeted, modeled by the logit in equation (2). Table 12 reports it on the full set of districts. The 2021 Castillo share and district size are the strongest predictors, which matches the gradient in Figure 4. Turnout enters negatively and rurality weakly, both smaller in magnitude. The model is what defines the control group used throughout the attributive arm.

Table 12: Logit of the targeting decision

Pre-treatment covariate	coef.	SE	p
Castillo 2V-2021 share	+0.980	0.120	0.000
Turnout 2021	-0.371	0.108	0.001
District size (log electors)	+0.825	0.118	0.000
Rurality	+0.275	0.101	0.006
$N = 1874$, pseudo-$R^2 = 0.122$			

Notes. The outcome is an indicator for at least one Fuerza Popular petition in the district. The covariates are standardized, so the coefficients are comparable across rows. Standard errors are province-clustered.

Two diagnostics describe the matched comparison. Figure 14 plots the distribution of the estimated propensity by group. Treated districts and the impugnable control share a common support, while the off-support districts the design discards sit at lower propensities. Figure 15 reports standardized mean differences before and after matching. Matching narrows the gap on every covariate, most sharply on district size and turnout. It does not erase the gap on the 2021 Castillo share, which falls from about eight tenths of a standard deviation to four, because the control is matched on the propensity, not on the Castillo share itself. This residual imbalance on the baseline outcome is precisely what the flexible-baseline and balancing specifications in Table 5 are built to address.

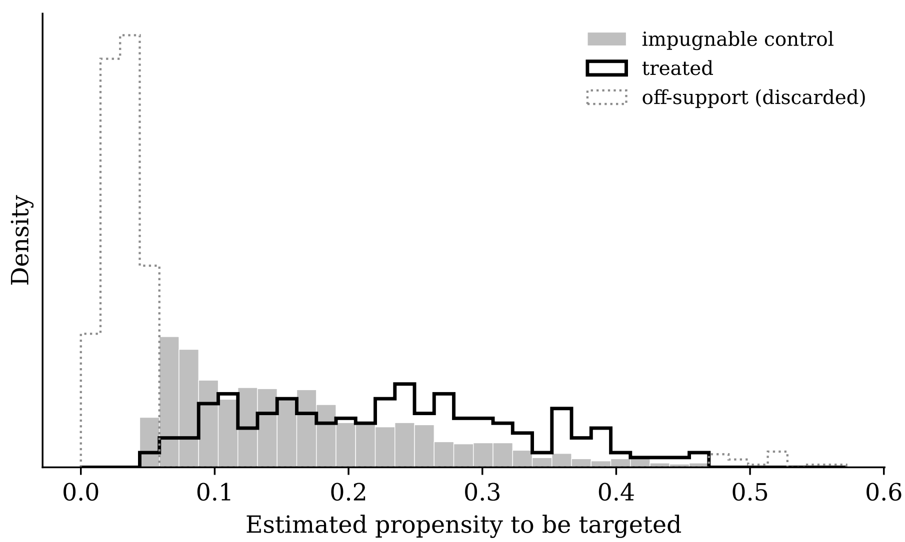


Figure 14: Propensity-score overlap by group

Notes. Distribution of the estimated propensity to be targeted (equation (2)) for treated districts, the impugnable control, and the off-support districts the design discards. Treated and control overlap on the common support.

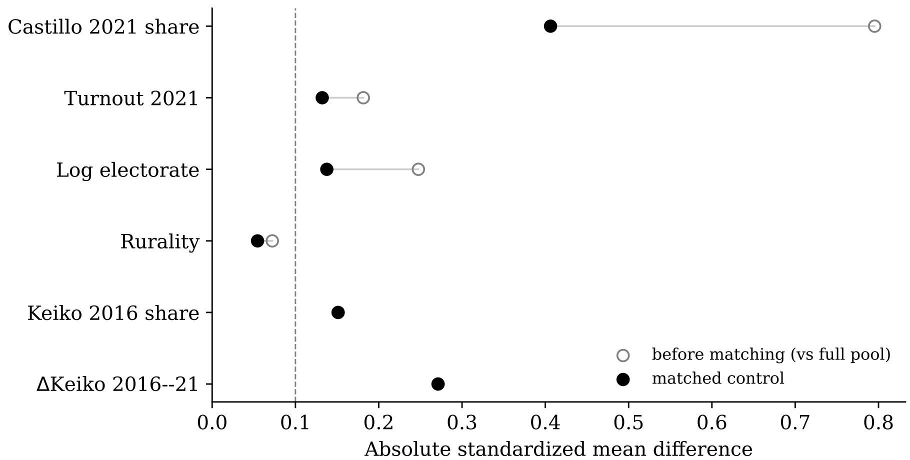


Figure 15: Covariate balance before and after matching

Notes. Absolute standardized mean differences between treated and control districts, before matching (treated versus the full pool of non-targeted districts) and after matching (treated versus the impugnable control). The dashed line marks the conventional 0.1 balance threshold.