Auditing the Challenger: Legal Challenge Records as Adversarially Selected Forensic Samples*

Abstract

Peru's 2021 presidential runoff was decided by 44,263 votes, after which the losing campaign challenged 768 polling stations as fraudulent. The courts dismissed almost all on procedure, never examining the ballots. This paper links all 1,086 mesa-level nullity cases to the official results and audits the challenged mesas as an adversarially selected diagnostic. The challenge record is strategic targeting: the challenged mesas capture more net Castillo votes than any of 10,000 draws matched on geography or first-round composition, and Perú Libre's own petitions mirror the selection from the other side. Beyond the extremity they were selected for, the challenged mesas carry no statistical signature of manipulation. Injection simulations bound anything hidden, within the audited mesas and for the simulated technologies, an order of magnitude below the margin, and digit tests detect none of the simulated manipulations. The evidence is consistent with a clean count in the challenged mesas.

Keywords: election forensics, electoral integrity, fraud allegations, vote-count manipulation, last-digit tests, adversarial selection, Peru

JEL classification: C18, D72, K42, O54

*Replication materials are available in a public repository. The link is withheld here to preserve anonymity for peer review. *Artificial-intelligence disclosure.* The author used Claude Code (Anthropic, Claude Opus 4.8) in June and July 2026 for data collection and cleaning and for proofreading.

1 Introduction

After losing Peru's 2021 presidential runoff by 44,263 votes, Fuerza Popular filed 942 nullity petitions against specific polling stations, alleging manipulation mesa by mesa. The electoral courts never examined whether those allegations were true. Under the JNE's rules, a nullity claim grounded in events at the polling station had to be raised there by the party's own personero and recorded in the acta,[1] so the post-election filings died on procedure: 49 percent were dismissed as time-barred, most of the rest as inadmissible, and only 8 percent ever received a ruling on the merits, every one of them denied. International observers assessed the process. They did not examine the targeted mesas. The result is a gap at the center of the most consequential fraud dispute in recent Latin American history: the 768 polling stations that the losing campaign itself identified as the site of the alleged fraud have never been systematically examined. This paper examines them.

From the JNE's public case files I assemble the complete first-instance record of mesa-level nullity for the runoff, 1,086 expedientes, 942 filed by Fuerza Popular and 144 by Perú Libre, link every petition to the specific mesas it names, and audit the 768 mesas Fuerza Popular challenged with the tools of election forensics. The design treats the challenge record as an adversarially selected diagnostic. Case selection is the standing weakness of forensic screening, an auditor must decide where among 86,488 tally sheets to look, and here the search was performed by the actor with the largest stake in finding fraud and the best information about where to allege it. The stakes were real: the challenged mesas held 148,166 votes and a net Castillo margin of 86,688, so the petitions, had they all been granted, would have reversed the election. The audit asks one question. Are the mesas the campaign itself identified enriched in the statistical signatures of manipulation, relative to the closest mesas nobody challenged?

Three findings come out. First, the challenge record is strategic targeting of the opponent's strongholds, and measurably so. The challenged mesas carry a larger net Castillo margin

[1] Resolución N.º 0086-2018-JNE distinguishes nullity grounded in events at the mesa, which must be raised at the mesa itself, from nullity grounded in facts external to it, which follows a separate route (Jurado Nacional de Elecciones, 2018). Section 2 details both.

than any of 10,000 random draws matched to their department or first-round composition, and within their own districts they run 0.9 percentage points higher in Castillo share and 2.3 points higher in unanimity than matched neighbors. Those coefficients measure which mesas the campaign chose. They do not test for manipulation. The placebo arm makes the point: the 138 mesas challenged only by Perú Libre sit on the other side, 1.9 points below their neighbors in Castillo share. Each losing side aimed at the mesas most extreme for its opponent.

Second, beyond the extremity they were selected for, the challenged mesas carry no statistical signature of manipulation. Turnout sits slightly below matched controls where stuffing would push it above, the Klimek corner of joint extreme turnout and winner share is no fuller than in the controls, and injection simulations convert the nulls into a bound stated in votes. Ballot stuffing above roughly 3,000 votes, seven percent of the margin, is caught in every replication. Vote switching is caught at 746 votes, the smallest scale simulated. The worst case reads the entire matched share coefficient as manipulation, and even then it comes to 1,361 net votes, 2,110 at the upper end of the confidence interval. The bound is deliberately scoped, to the challenged mesas and to the manipulation technologies actually simulated, and Section 6 states both limits in full. Within that scope the evidence is consistent with a clean count in the challenged mesas, and anything hidden in that count is an order of magnitude too small to have changed the result.

Third, digit tests, the most portable tool in the forensic kit, have almost no power here. Across 13,000 injection replications spanning both manipulation technologies up to outcome-flipping scale, last-digit tests never detect at more than the nominal size of the test. At Peruvian mesa sizes, roughly two hundred ballots cast, a null digit result cannot distinguish an election-flipping manipulation from a clean count. The failure is a general limit on last-digit tests at small polling stations, not a fact about Peru, and it is the paper's most portable result.

The paper also makes two methodological points. The first concerns measurement. The

natural reading of the public data takes ONPE's COMPUTADA RESUELTA status as the footprint of the campaigns' challenges. It is not. Of the actas carrying the status, 98.6 percent were challenged by nobody, only 10 of the 768 challenged mesas carry it, and the two populations sit on opposite sides of the country and of the electoral cleavage. Administrative status variables are workflow artifacts of electoral bureaucracies, and challenge-based designs need the challenger's own case files, linked unit by unit, before any comparison is run. The second concerns what adversarial selection does to inference. A challenge record solves forensic case selection, but the same adversarial quality disqualifies every outcome the filer could observe, vote shares, unanimity, swings, as evidence of manipulation. A challenge audit must split its outcomes into those that document the selection and those that can still discriminate, and it must demonstrate the power of the discriminating tests by injection rather than assume it. The paper claims no general framework beyond that discipline. It is an applied forensic case study. But the discipline travels to any election where a losing side names the units it believes were stolen.

The study joins a forensic literature built largely on screening without case selection: fingerprint methods (Klimek et al., 2012; Myagkov et al., 2009), digit tests (Beber and Scacco, 2012; Deckert et al., 2011), integer round-offs (Kobak et al., 2016), batteries applied across Latin American elections (Chumacero, 2025), including quick statistical readings of this very runoff (Isea and Isea, 2021). The nearest neighbors are audits of specific fraud allegations. Cantú (2014) audits Mexican local elections by comparing contiguous polling stations that differ only in their assigned surnames, and Cantú (2019) reconstructs the 1988 Mexican presidential count from photographs of the tally sheets and finds alteration in about a third of them. Eggers et al. (2021) test the statistical claims made about the 2020 United States election and find that none survives scrutiny, and Bolivia in 2019 shows the stakes of getting such an audit wrong: the late-count shift read as fraud (Escobari and Hoover, 2024) dissolves under scrutiny, the shift explained by the composition of late-reporting mesas (Idrobo et al., 2022) and the discontinuity design failing on its own terms (Idrobo et al., 2026), after the

accusation had already helped topple a government (Gallego et al., 2023). The record audited here has a feature none of those settings offered. The losing campaign itself, through 942 legal filings, specified exactly where the fraud was supposed to be, and the courts' procedural dismissal left that specification untested. Statistical audit is no substitute for legal review or for observation (Hyde, 2007), but where the legal process never reached the merits, it is the only systematic examination available.

Section 2 lays out the institutions, Section 3 the challenge record, and Section 4 why the administrative status variable does not identify the challenges. Section 5 presents the data and the empirical strategy. Section 6 reports the forensic battery and the bound in votes, and states what the evidence rules out and what it cannot. Section 7 shows the powerlessness of digit tests, and Section 8 concludes.

2 The election and its institutions

Peru elects its president by two-round majority. The April 2021 first round fragmented across eighteen candidates: Pedro Castillo of Perú Libre led with 18.9 percent of the valid vote, Keiko Fujimori of Fuerza Popular advanced with 13.4, and the third-placed candidate took 11.8 (Oficina Nacional de Procesos Electorales, 2021a). The June 6 runoff collapsed that fragmentation onto the country's oldest political cleavage. Castillo carried the rural highlands and much of Amazonia, Fujimori the coast and metropolitan Lima (Figure 1). Turnout rose from 70.0 to 74.6 percent of the 25.3 million registered voters, 18.9 million ballots were cast at 86,488 mesas, and Castillo won by 44,263 votes (Oficina Nacional de Procesos Electorales, 2021b). Fuerza Popular did not concede. It alleged mesa-level fraud, and it turned to the legal machinery that the rest of this section describes, under the eyes of observation missions from the OAS and the European Union (Organization of American States, 2021; European Union Election Observation Mission, 2021).

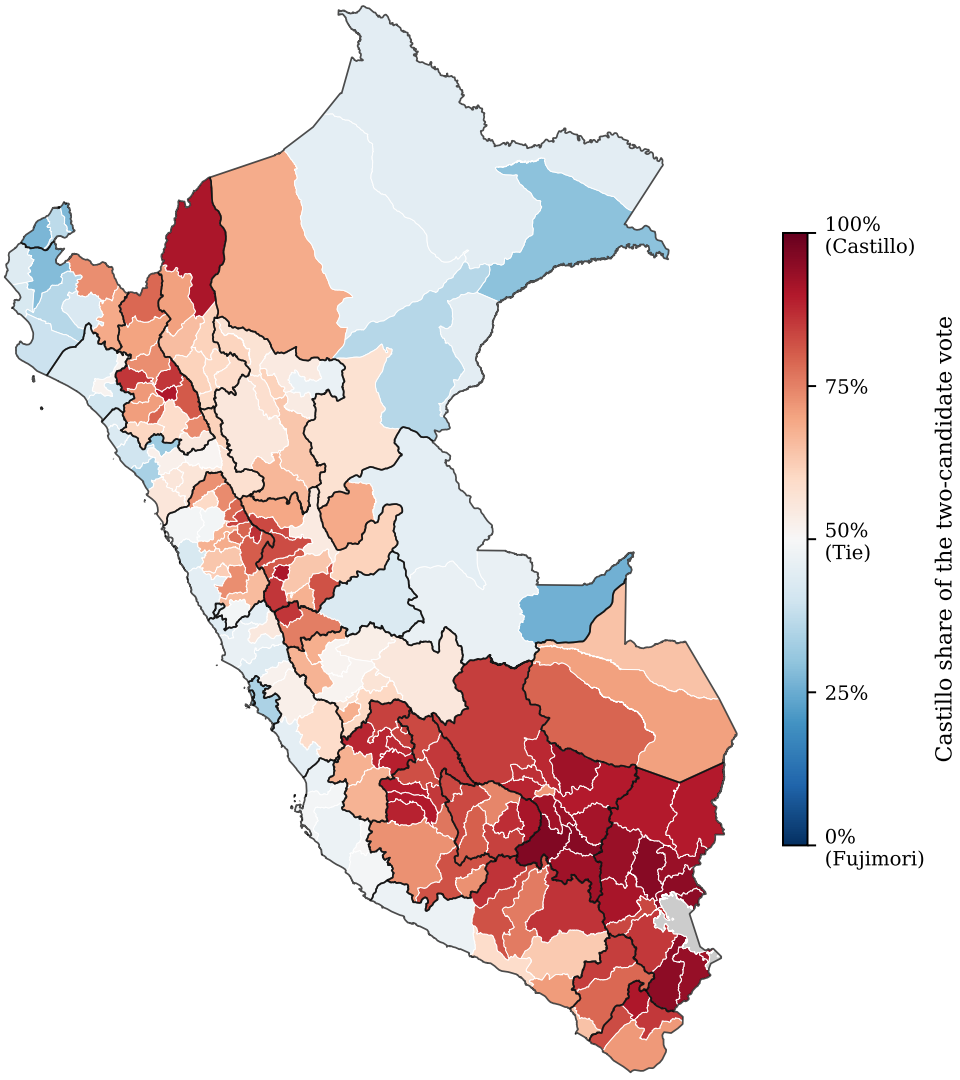

Figure 1: *Castillo share of the two-candidate vote by province, 2021 runoff*

Castillo's share of the two-candidate vote from the ONPE acta-level results, aggregated to province. Department boundaries in black. Lago Titicaca carries no voters and is uncolored. Datem del Marañón and Putumayo, created after the GADM 4.1 polygons, display within their parent provinces. Source: Oficina Nacional de Procesos Electorales (2021b).

Two bodies run Peruvian elections. The Oficina Nacional de Procesos Electorales (ONPE) administers the vote and the count: its decentralized offices (ODPE) receive the tally sheets, process them, and publish results. The Jurado Nacional de Elecciones (JNE) adjudicates:

temporary Jurados Electorales Especiales (JEE) resolve first-instance electoral disputes across the country, with appeal to the JNE's plenary in Lima. The unit both bodies work in is the mesa de sufragio. Three citizens drawn by lot administer each mesa, at most about three hundred registered voters and 292 on average in 2021. They receive the ballots, count them at closing, and record the tally in a signed acta that travels to the ODPE for processing.

The parties sit inside that machinery through personeros, accredited representatives entitled to be present at the mesa through voting and counting, to sign the acta, to have their observations recorded in it, and to litigate for the party before the JEE and the JNE. The personero is the legal channel connecting events at a polling station to anything a campaign can later claim about them.

Nullity of a mesa's vote is governed by Article 363 of the Ley Orgánica de Elecciones, a closed list of grounds that includes irregular installation of the mesa, suplantación of voters or mesa members, and violence or intimidation over them during the voting (Jurado Nacional de Elecciones, 2018). Resolución 0086-2018-JNE fixes the procedure and splits it into two routes. Nullity grounded in events at the mesa, where every ground Fuerza Popular would later plead sits, must be raised at the mesa itself, by the party's own personero, and recorded in the acta at the time: a mesa-events claim surfacing for the first time days after the count is procedurally dead on arrival.[2] Nullity grounded in facts external to the mesa follows a separate route with its own filing window (Jurado Nacional de Elecciones, 2018). The rule has a purpose the design of this paper inherits. It forces fraud claims to be raised at the time and documented, because post-hoc allegations against unfavorable tallies are cheap, and it is the rule against which 49 percent of Fuerza Popular's petitions died as time-barred in Section 3.

Separate from nullity altogether, the administration runs its own quality-control track. Under Resolución 0331-2015-JNE the ODPE flags actas observadas during processing, tally

[2]The tribunals' own reasoning confirms the classification. Ninety percent of the 461 time-barred dismissals cite Resolución 0086-2018-JNE as the governing rule, and petitions filed on the evening of June 9, three days after the vote, were already out of time.

sheets with missing signatures, illegible entries, or material errors, and forwards them to the JEE, which resolves the observation so the acta can enter the count (Jurado Nacional de Elecciones, 2015). An acta that passes through this track appears in the published file with the status COMPUTADA RESUELTA, computed after resolution. The track is administrative. It runs whether or not any party contests anything, and Section 4 shows why it must not be mistaken for the challenge record. A third mechanism, narrower than either, lets personeros impugn individual ballots at the mesa, and the impugned ballots ride the acta to the JEE for resolution (Jurado Nacional de Elecciones, 2021). The published record closes this track at zero: any ballot impugned during processing had been resolved before the count was final, and the results file records zero impugned votes nationwide (Oficina Nacional de Procesos Electorales, 2021b).

The 2021 dispute therefore left three institutional trails that must not be confused. The campaigns filed mesa-nullity expedientes, the challenge record of Section 3. The administration generated actas observadas, the status variable of Section 4. And ballot-level impugnations left a trail the published record closes at zero. The paper's treatment variable comes from the first. The earlier version's came, mistakenly, from the second.

3 The challenge record

The JNE's public case registry lists 1,518 expedientes filed by the two runoff campaigns in connection with the 2021 second round. Setting aside 370 appeals, 60 complaints against denied appeals, and 2 nullity requests aimed at whole districts or provinces, the first-instance record of mesa-level nullity comprises 1,086 expedientes: 942 filed by Fuerza Popular and 144 by Perú Libre. Filing came in a burst in the week after the June 6 vote. Of the 897 Fuerza Popular expedientes with a recorded filing date, 145 were filed on June 9, 593 on June 10, 25 on June 11, and 134 on June 12. The petitions name 768 distinct mesas: 939 of the 942 name at least one mesa, three name none, and 114 mesas appear in more than one

expediente, up to six filings against a single mesa. Perú Libre's 144 expedientes name 141 distinct mesas, three of which Fuerza Popular also challenged. Figure 2 maps the challenged mesas by province. The filings concentrate in the rural highland south, Ayacucho above all, with a second mass in the Amazonian north and east and near-absence on the coast and in metropolitan Lima.[3]

[3]Eleven challenged mesas sat at overseas voting sites, ten in the Americas and one in Egypt, and are not shown. Castillo led in all eleven, against an overseas vote that went 66 percent to Fujimori: even abroad, the campaign challenged the rare mesas it lost, the same outcome targeting that Section 5 documents at scale.

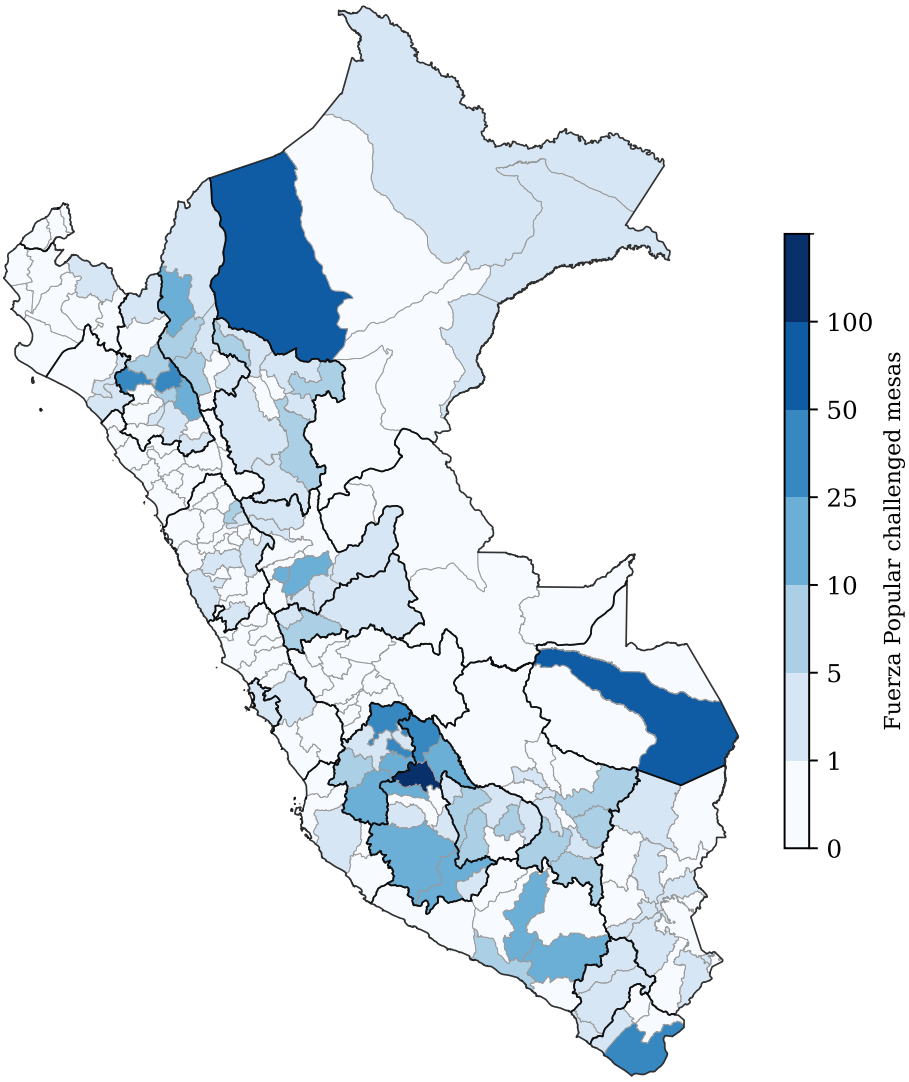

Figure 2: *Fuerza Popular's mesa-nullity petitions by province*

The 757 domestic mesas among the 768 named in the petitions, counted by province (11 challenged mesas at overseas voting sites are not shown). Counts are binned. Datem del Marañón, created in 2005 and absent from the GADM 4.1 polygons, displays within Alto Amazonas, its parent province. Source: JNE mesa-nullity expedientes linked to ONPE mesa geography.

The aggregate stakes were not symbolic. The 768 challenged mesas hold 215,401 registered voters and cast 148,166 votes, with a net Castillo margin of 86,688 votes against an official national margin of 44,263. Had every petition been granted, the annulled mesas would have

erased nearly twice the margin: the challenge record, taken at face value, was a petition to reverse the result of the election.

The courts disposed of that record almost entirely on procedure. Of Fuerza Popular's 942 petitions, 461 (49 percent) were dismissed as time-barred, 274 were declared inadmissible on other liminar grounds, 107 were absorbed into a parallel expediente on the same mesa (91 of the 107 name a traceable parent, and every traceable parent was itself declared inadmissible), 17 were archived after the party consented to the resolution, and 6 closed on other procedural terms. Only 77 petitions, 8 percent, received a ruling on the merits, and all 77 were denied as unfounded. Perú Libre fared no better: 142 of its 144 petitions were inadmissible and the other 2 denied on the merits. Across all 1,086 expedientes, not one petition was granted. Fuerza Popular's merits success rate is exactly zero, and roughly 92 percent of its allegations were never examined on their substance by anyone.

What the petitions actually alleged is recoverable for only part of the record. Dismissals rarely restate the petition's substantive claim, so the Article 363 ground can be recovered for 268 of the 942 Fuerza Popular expedientes (28.5 percent), and all grounds findings are accordingly labeled partial. Within that recoverable subset the record is lopsided: 199 petitions (74 percent) plead suplantación, the forged-signatures template of Article 363 literal b, 64 (24 percent) plead violence or intimidation at the mesa, 3 plead an outcome-based argument (zero or implausible Fuerza Popular votes), and 2 plead irregular installation. The repetition of wording across expedientes is measurable, with a caveat the measurement carries: the petitions themselves are not in the public record, so the statistic runs on the resolutions' restatements of them. Among the 483 expedientes whose resolutions restate any of the petition's own language, a wider set than the 268 with a classifiable ground, 89 percent share near-verbatim text with another petition filed in the same JEE, at a Jaccard similarity of at least 0.60 on five-word shingles, while only 5 percent match a petition in a different JEE. The duplication is therefore organized by tribunal, consistent with batch-drafted filings processed docket by docket, but because every restatement passes through the tribunal's

own pen, the record cannot apportion the boilerplate between the filer's templates and the tribunals' summarizing conventions. What it can establish is narrower and sufficient here: the filings arrived in bulk, on four days, in near-identical form within each tribunal. I had pre-registered the opposite expectation, that the outcome-based argument would be the modal ground, matching the campaign's public statements about statistically impossible results. That prediction is disconfirmed by the legal record: the statistical-impossibility argument lived in press conferences, and almost never in the filings themselves. The selection of mesas on extreme outcomes, documented in Sections 5 and 6, is therefore established by the estimates and the yield benchmark, and not by the text of the petitions.

Two limitations of the record matter for what follows. The published tallies are final figures. Because not one petition was granted, the challenge process itself altered no tally, and the only challenged mesas whose figures passed through adjudication are the ten that also entered the administrative resolution track, which the estimation drops as a pre-specified robustness check, and the annulled mesas, one challenged by Fuerza Popular and five by Perú Libre, which are excluded from the vote-based tests in their respective arms. And ONPE published no mesa-level processing timestamps, so nothing in this paper can test claims about the order in which actas entered the count, the count-sequence family of suspicions that circulated in 2021. The audit below is cross-sectional by necessity.

4 Measurement: administrative flags are not challenge records

Every mesa in the ONPE results file carries a final status. Most actas enter the count directly as CONTABILIZADA. A smaller set, COMPUTADA RESUELTA ("computed after resolution"), enters only after some authority resolves a problem with the tally sheet. The natural reading takes that status as the empirical footprint of the campaigns' legal challenges. Table 1 crosses the status variable against the challenge record linked from the JNE case files.

The two populations barely touch.

Table 1: *Acta status and the challenge record*

Final status	FP only	PL only	Both	Unchallenged	Total
CONTABILIZADA	754	124	3	83,982	84,863
COMPUTADA RESUELTA	10	9	0	1,367	1,386
ANULADA	1	5	0	207	213
EN PROCESO	0	0	0	20	20
SIN INSTALAR	0	0	0	6	6
Total	765	138	3	85,582	86,488

All 86,488 presidential actas in the ONPE final file, crossed against the mesa-nullity expedientes linked from JNE case files. FP only and PL only count mesas challenged by one party, Both counts the three mesas challenged by each. Fuerza Popular challenged 768 distinct mesas (765 alone plus the 3 joint), Perú Libre 141. Of the 1,386 COMPUTADA RESUELTA actas, 1,367 (98.6 percent) were challenged by nobody.

The numbers leave no room for that reading. Of the 1,386 actas with the status, 98.6 percent appear in no party's petition. Of the 768 mesas Fuerza Popular challenged, 10 carry the status, 757 entered the count as ordinary CONTABILIZADA actas, and one was annulled.[4] The status variable identifies a different institution altogether: the actas-observadas track of Resolución 0331-2015-JNE, under which the ODPE flags tally sheets during processing for missing signatures, illegible entries, or material errors, and a JEE resolves the observation so the acta can enter the count (Jurado Nacional de Elecciones, 2015). It is a workflow artifact of the electoral administration, generated whether or not any party contests anything.

Geography says the same thing. The unchallenged COMPUTADA RESUELTA actas concentrate in Lima (550 of the 1,367) and abroad (all 309 overseas actas, none of them challenged), in mesas averaging 40 percent Castillo. The challenge record concentrates in

[4]The 19 challenged mesas inside COMPUTADA RESUELTA (10 Fuerza Popular, 9 Perú Libre) are consistent with actas that drew both an ODPE observation and a party challenge. The pre-specified robustness in Section 6 excludes the 10 and nothing moves.

Ayacucho above all (205 mesas), then Huancavelica (91) and Cajamarca (83), highland mesas averaging 81 percent Castillo. The status group and the challenge record sit on opposite sides of the country and on opposite sides of the electoral cleavage.

The consequence for any analysis built on the flag is total. Contested-versus-uncontested comparisons measure the geography of ordinary acta processing, the apparent concentration of challenges inverts, and the count of flagged units, 1,077 domestic COMPUTADA RESUELTA actas (1,386 in all, less 309 overseas), bears no relation to the 768 mesas the campaign actually challenged. The crosswalk is the whole point: an administrative status and a legal challenge record are disjoint populations, and only the linked case files identify the second.[5]

The lesson is worth stating as method, because nothing about it is specific to Peru. Administrative status variables are generated by electoral bureaucracies for their own workflow, their names describe processing steps, and no name, however suggestive, substitutes for linking the legal record itself. Challenge-based forensic designs need the challenger's case files, matched unit by unit, before any comparison is run. The remainder of the paper is built on that linkage.

5 Data and empirical strategy

Two records meet in this study, and each arrives by its own process. The returns are ONPE's acta-level results for both rounds of the 2021 general election, published as open data at datosabiertos.gob.pe. Each of the 86,488 presidential tally sheets is one row. It carries each candidate's vote count, the registered electorate, and the final processing status. The extracts are checksummed in the replication repository.

The challenge record is built from the JNE's public case registry at plataformahistorico.jne.gob.pe. Every expediente the two campaigns filed in connection with the runoff was retrieved, and the first-instance mesa-nullity cases were isolated by the registry's own

[5]The 213 ANULADA actas include one Fuerza-Popular-challenged mesa, in Combapata, Cusco, annulled through the administrative track and not by any granted petition.

materia field, the 1,086 expedientes whose subject is a mesa de sufragio, separated from the appeals and the district-level requests that the crosswalk of Section 3 accounts for. Each expediente names the polling stations it contests, and those names were linked to ONPE mesa identifiers. The linkage is complete: all 768 Fuerza Popular and 141 Perú Libre mesas match an acta in the 86,488-row second-round file (Table 1). The estimation sample keeps the two vote-bearing statuses, CONTABILIZADA and COMPUTADA RESUELTA, with positive cast votes. Six challenged mesas ended ANULADA, one challenged by Fuerza Popular and five by Perú Libre, and enter the file with tallies zeroed, so they drop from the vote-based tests and appear in the aggregate accounting through their 1,782 registered voters. Because not one petition was granted, the published tallies are untouched by the challenge process itself, and the ten challenged mesas that passed through the administrative resolution track are removed in a pre-specified robustness check without moving anything. Summed over every acta, the extract gives a national margin of 44,240 votes against the proclaimed 44,263, a 23-vote gap, 0.05 percent of the margin, whose acta-level source remains unidentified and is noted here rather than attributed to rounding.

The runoff outcomes need no merge. The swing and turnout-change outcomes and the matching covariates require the first-round file, which merges for 11,568 of the 11,931 observations in the Fuerza Popular estimation sample and for the placebo sample in full. Challenged mesas without a first-round merge stay in the fixed-effects sample and drop from the matched sample, leaving 767 vote-bearing challenged mesas, 662 of them matched to 2,560 controls across 282 districts, and 133 placebo mesas, 124 of them matched to 620 controls across nine districts.[6] As a final check on the merge itself, redefining every first-round covariate from CONTABILIZADA actas only, a stricter first-round panel, moves the matched share coefficient from 0.92 to 0.85 points ($p = 0.002$) and changes no conclusion. Table 2

[6]The challenged count differs across displays by the filter each applies. Of the 768 named mesas, 767 are vote-bearing, 760 carry a first-round merge, and 662 match to a within-district control, leaving 106 without a close one. The forensic-space figures apply display filters of their own. The first-round scatter (Figure B.2) shows the 754 domestic mesas with a merge, and the Klimek plots (Figure 3) the 737 of those with turnout above 5 percent and interior candidate shares.

traces the full descent from the published file to the estimation samples.

Table 2: *Sample construction*

	Actas
All presidential actas, second-round file	86,488
CONTABILIZADA	84,863
COMPUTADA RESUELTA	1,386
ANULADA	213
EN PROCESO	20
SIN INSTALAR	6
Vote-bearing frame	86,244
Fuerza Popular arm	
Challenged mesas, vote-bearing	767
with a first-round merge	760
matched (treated)	662
matched controls	2,560
Districts in the fixed-effects sample	282
Perú Libre placebo arm	
Challenged by Perú Libre alone, vote-bearing	133
matched (treated)	124
matched controls	620
Districts in the fixed-effects sample	9

From the full ONPE second-round file to the estimation samples. The vote-bearing frame keeps the two statuses that carry counts, CONTABILIZADA and COMPUTADA RESUELTA with positive cast votes, dropping the 239 annulled, in-process, or uninstalled actas and five more with no cast votes. Six challenged mesas ended ANULADA, one Fuerza Popular and five Perú Libre, and leave the vote-based samples here. Matched columns use within-district nearest-neighbor matching on first-round Perú Libre share and turnout. Source: replication output `sample_flow.txt`.

With the data in place, the design that operates on them follows. The challenge record solves the case-selection problem that defeats most election forensics, and it solves it adversarially. An auditor screening 86,488 actas must decide where to look, and any decision made after seeing the results contaminates whatever test is run on the chosen cases. Here the losing campaign made the decision. Fuerza Popular filed after the results were public, against 768 specific mesas, with the full returns in hand and the strongest possible incentive to aim wherever manipulation was likeliest to be found or plausibly alleged (Section 3). The design treats those mesas as an adversarially selected diagnostic. The estimand is enrichment: if the count was manipulated toward Castillo at any scale worth litigating, the mesas that the aggrieved campaign itself identified should be enriched in the statistical signatures of manipulation relative to the closest mesas nobody challenged.

The unchallenged neighbors are not a no-fraud counterfactual, and nothing in the design certifies any mesa clean. The comparison answers a narrower question: whether the mesas the campaign flagged look different, in the directions manipulation would push them, from mesas in the same district with the same first-round profile. A null answer does not prove the election clean. It establishes that the one systematic body of specific fraud allegations, examined against its own target set, carries no statistical support there, and the injection simulations of Section 6 convert that null into a bound in votes.

Selection is also the design's central discipline, because it splits the outcomes into two classes. The campaign chose mesas after the results were public, so any outcome it could observe and maximize is disqualified as a test. Castillo's vote share is one such outcome. The unanimity indicator and the between-round swing are others. All of them measure which mesas the campaign chose, and the paper reads them only as selection. Unanimity carries one complication the paper states rather than hides. It also reads as a fabrication signature in the digit-forensics tradition, and unlike share and swing it has no placebo mirror, since no Perú Libre-challenged mesa records zero Castillo votes. Neither point reaches the bound. Fabrication on the scale that drives a mesa to unanimity pushes the winner's share

to its ceiling and into the Klimek corner, which the injection simulations of Section 6 catch. The unconditioned contrast shows the size of the trap. Challenged mesas average a Castillo share of 0.766 against 0.472 in the rest of the country, a gap that measures the campaign's choices and enters the paper only as the naive benchmark the design exists to avoid. The discriminating outcomes are the ones a filer gains nothing by targeting and manipulation cannot avoid moving. Ballot stuffing mechanically raises turnout and its change between rounds, wholesale acta fabrication lands in the Klimek corner of joint extreme turnout and winner share, and invented tallies distort the digit composition of the counts. The forensic question is whether the challenged mesas carry those signatures once the extremity they were selected for is conditioned away.

The primary specification conditions on geography and on first-round profile. Index mesas by m and their districts by $d(m)$, and let $C_m = 1$ when Fuerza Popular challenged mesa m. For each outcome y_m,

$$y_m = \beta\, C_m + \alpha_{d(m)} + \varepsilon_m, \tag{1}$$

with district fixed effects $\alpha_{d(m)}$ and standard errors clustered by district, so β is the average within-district contrast between the challenged mesas and the rest.

The headline sample sharpens that contrast by matching. Each challenged mesa is paired with up to five unchallenged mesas in the same district, nearest neighbors on the first-round covariates $x_m = (s_m, t_m)$, Perú Libre's first-round share and first-round turnout, without replacement and within a caliper of 0.10 on s_m, treated mesas processed in fixed order so the matched set is deterministic. Estimating (1) on the matched sample gives the enrichment estimand

$$\beta = \mathbb{E}[y_m \mid C_m = 1] - \mathbb{E}[y_m \mid C_m = 0,\ m \text{ matched to a challenged mesa}]\,, \tag{2}$$

the difference in each signature between the challenged mesas and their closest unchallenged neighbors, written for readability as a plain contrast of means. On the matched sample I still

fit (1) with its district fixed effects and district-clustered standard errors, so the reported β is the district-demeaned version of this contrast, a variance-weighted average of the within-district differences that coincides with the simple difference only under a balanced design. The first-round covariates predate the runoff, so no runoff-day manipulation can move them. For a discriminating outcome the null of a clean count in the challenged mesas is $\beta = 0$, and for a selection outcome β measures the targeting and is not a test of manipulation. The same machinery runs twice: once with Fuerza Popular's 768 mesas as treatment and once with the mesas Perú Libre challenged, whose petitions allege manipulation in the opposite direction and which serve as the design's placebo arm. The placebo estimates were produced before any treatment-arm output, per the frozen specification, and rest on nine district clusters, so their inference uses the bootstrap reported with Table 4.

Whether the record meets these requirements is an empirical question, taken up with the results in Section 6.

6 Results

The design needs a record that is adversarial and that the courts never filtered, and the evidence that this one qualifies comes first. Table 3 gathers it. Panels A and B recall from Section 3 that not one of the 1,086 petitions was granted and that, where a substantive ground is recoverable, it is the templated forged-signatures plea rather than the outcome-based argument the campaign pressed in public (the grounds validation is Appendix A). Because the petitions' text cannot establish what the campaign targeted, the design must establish it statistically.

Panel C is that statistical statement. The challenged mesas contain a net Castillo margin of 86,688 votes, of which 86,630 sit in the 760 challenged mesas inside the benchmark universe of vote-bearing mesas with a first-round merge. The draws sample from that universe, so 86,630 is the figure the comparison targets. Drawing 10,000 random sets of mesas with the

same department composition yields a mean net margin of 68,520 votes with a standard deviation of 1,279, and the largest of the 10,000 draws reaches 72,980, far short of the target. Matching the draws instead on deciles of first-round Perú Libre share, so that the benchmark concedes the targeting of Castillo-leaning terrain and asks only about selection within it, raises the mean to 72,717 and the largest draw to 77,007, still short. The actual selection sits above every draw under both benchmarks. Fuerza Popular did more than file where it lost. It filed on the mesas that maximized the net votes an annulment would recover, which is the behavior of a strategic actor pursuing reversal and the reason every outcome the filer could observe is barred as evidence of manipulation.

Table 3: *Selection and stakes of the challenge record*

Panel A: disposition of the 1,086 mesa-nullity expedientes

	Fuerza Popular	Perú Libre
Dismissed as time-barred	461	—
Inadmissible on other liminar grounds	274	142
Absorbed into a parallel expediente	107	—
Denied on the merits	77	2
Archived on the party's consent	17	—
Other procedural closure	6	—
Granted	0	0
Total	942	144

Panel B: Article 363 grounds, recoverable subset

	Petitions	Percent
Forged signatures (suplantación)	199	74.3
Violence or intimidation at the mesa	64	23.9
Outcome-based (zero or implausible FP votes)	3	1.1
Irregular installation	2	0.7

Panel C: net-vote yield of the selection vs 10,000 random draws

	Dept-matched	Share-decile
Mean net Castillo margin of draws	68,520	72,717
Standard deviation of draws	1,279	1,275
Largest draw	72,980	77,007
Actual net Castillo margin, same universe	86,630	86,630
Draws at or above actual	0	0

Panel A counts first-instance mesa-nullity expedientes by final procedural outcome from the JNE resolution record, with the outcome classifier hand-validated on 50 sealed expedientes at 92 percent agreement and no merits ruling misclassified. Panel B covers the 268 Fuerza Popular expedientes whose resolutions restate the substantive ground, 28.5 percent of the 942, labeled partial under the pre-registered coverage rule. Panel C compares the vote-weighted net Castillo margin in the challenged mesas with random draws of equal mesa counts from the same departments (first column) or the same

Table 4 reports the forensic battery, estimated under the specification and samples of Section 5, with the within-district matched estimates as the headline and the district fixed-effects estimates alongside.

The first panel records the selection itself. Challenged mesas run 0.9 percentage points higher in Castillo share than their matched neighbors (SE 0.3), are 2.3 points more likely to record zero Fujimori votes (SE 0.8), and swung 0.8 points further toward Castillo between rounds (SE 0.3). These coefficients measure which mesas Fuerza Popular chose to challenge. They carry no weight as evidence of manipulation. The placebo arm confirms the reading. Mesas challenged only by Perú Libre sit 1.9 points *lower* in Castillo share and 1.9 points lower in swing than their matched neighbors, the mirror image of the Fuerza Popular arm. Each losing side targeted the mesas most extreme for its opponent. One caution on placebo inference: the Perú Libre arm spans only nine district clusters, too few for reliable cluster-asymptotic standard errors, and under a wild cluster bootstrap with the null imposed the matched placebo p-values rise from 0.04 to 0.13. The placebo's work in the design is done by its point estimates, the mirror image, and no claim in the paper rests on placebo significance.

The second panel holds the outcomes that can discriminate manipulation from selection. Turnout in challenged mesas is 0.5 points below matched controls ($p = 0.20$). The change in turnout between rounds is 0.2 points lower ($p = 0.10$ matched, $p = 0.04$ in the unmatched within-district specification), and it points toward lower turnout, where ballot stuffing would raise it. Challenged mesas are no more likely than their matched controls to land in the Klimek corner of turnout above 0.90 combined with winner share above 0.90 (coefficient 0.0003, $p = 0.87$).

The third panel reports the digit tests, which come in two forms that ask different questions. Within the challenged mesas, last digits of Castillo and Fujimori counts are consistent with uniformity ($p = 0.17$ and $p = 0.09$), the test of departure from a flat digit distribution. The two-sample comparison, which instead asks whether the challenged and control digit distributions differ from each other, rejects homogeneity for Castillo counts at

$p = 0.04$, while Fujimori counts give $p = 0.20$ and the two placebo-arm comparisons give $p = 0.24$ and $p = 0.09$. Two near-uniform distributions can still differ at this sample size, so the pairing is not a contradiction. I report the rejection as it stands. Section 7 shows that digit tests detect almost none of the manipulation technologies simulated below even at outcome-flipping scale, so this comparison carries little diagnostic weight in either direction.

Figure 3 places the two challenge arms in the Klimek space of turnout against Castillo share, which summarizes the design in one view. Fuerza Popular's challenged mesas cluster high above the 50 percent line, deep in Castillo territory, and Perú Libre's cluster below it. Both sit inside the density contours of the unchallenged mass, away from the upper-right corner that wholesale fabrication produces. The mirror is the adversarial selection made visible, the empty corner the null. Figure B.1 in Appendix B repeats the plot in Fujimori share.

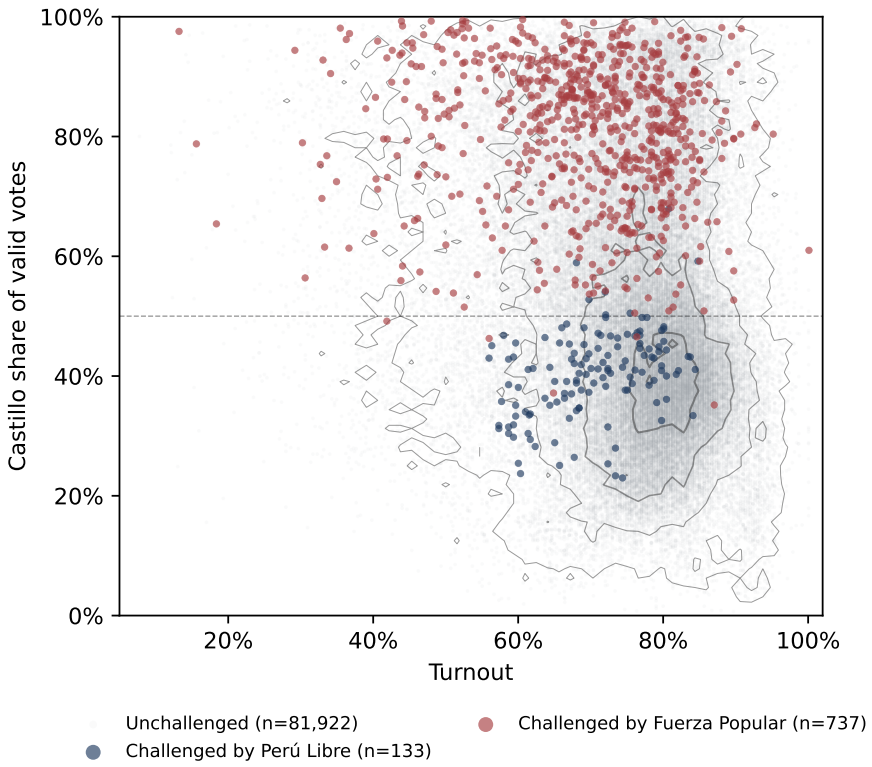

Figure 3: *Challenge arms in the Klimek space: Castillo share*

Domestic vote-bearing mesas with turnout above 5 percent and interior candidate shares, which is why the challenged count here is below the 767 vote-bearing total and the 754 with a first-round merge in Figure B.2. The corner indicator of Table 4 is a deliberately crude summary of this fingerprint, the joint upper-right mass the full Klimek method models in two dimensions, and the scatter carries the content the indicator compresses. Contours trace the density of the unchallenged mass. The dashed line marks 50 percent.

The multiplicity works in the null's favor. Each arm runs seven battery tests beyond the selection panel, the three discriminating outcomes and the four digit tests. Under a clean count at the 0.05 level, the expected number of false rejections is 0.35 by linearity, which needs no assumption about how the tests relate. Even treating the seven as independent, which if anything overstates the false-positive risk because they are positively dependent, the chance of at least one rejection is about 0.30 and of two or more about 0.04. The headline matched battery returns exactly one nominal rejection, the Castillo digit-homogeneity test at $p = 0.043$, which Section 7 shows carries no power and no direction. The only other

significance anywhere in the battery is the between-round turnout change, and it appears in the district fixed-effects specification, not the matched headline, pointing toward lower turnout where ballot stuffing would raise it. One powerless rejection against a 0.30 baseline, and a separate-specification coefficient pointing against manipulation, is what a clean count predicts.

The pre-specified robustness checks leave the pattern in place. Coding treatment by the number of expedientes naming the mesa gives 0.8 points of Castillo share per expediente ($p = 0.0004$) with a corner coefficient indistinguishable from zero ($p = 0.61$). Excluding the three mesas challenged by both parties, or the ten challenged mesas that also passed through the administrative resolution track, moves the within-district share coefficient from 1.27 points to 1.25 and 1.31 ($p < 0.001$ throughout). Matching itself does not select the finding. The 106 unmatched challenged mesas, which lack a close within-district control and are plausibly the more extreme, stay in the district fixed-effects sample, and that specification reaches the same conclusions as the matched one. The matched estimate is also stable to the matching order, which the frozen specification fixes for exact replication. Randomizing the order in which treated mesas are processed, over 200 permutations, moves the share coefficient within a range of 0.82 to 1.14 points, all positive, around a mean of 1.00, with the fixed mesa-order baseline of 0.92 at the conservative end.

Table 4: *Forensic battery in the challenged mesas*

	Fuerza Popular arm		Perú Libre placebo	
	District FE	Matched	District FE	Matched
Selection outcomes				
Castillo share (2V)	0.0127	0.0092	−0.0175	−0.0191
	(0.0031)	(0.0026)	(0.0076)	(0.0080)
Unanimity (zero Fujimori votes)	0.0209	0.0226	—	—
	(0.0073)	(0.0077)		
Swing toward Castillo (2V−1V)	0.0039	0.0080	−0.0179	−0.0190
	(0.0032)	(0.0025)	(0.0080)	(0.0079)
Discriminating outcomes				
Turnout (2V)	0.0004	−0.0047	0.0029	0.0034
	(0.0040)	(0.0036)	(0.0042)	(0.0050)
Turnout change (2V−1V)	−0.0029	−0.0023	0.0001	0.0005
	(0.0014)	(0.0014)	(0.0029)	(0.0026)
Klimek corner (>0.90, >0.90)	−0.0005	0.0003	—	—
	(0.0015)	(0.0017)		
Digit tests (p-values, matched samples)				
Castillo last digit: uniformity	0.166		0.070	
Castillo last digit: homogeneity vs controls	0.043		0.243	
Fujimori last digit: uniformity	0.093		0.206	
Fujimori last digit: homogeneity vs controls	0.205		0.095	
Treated mesas (vote-based sample)	767	662	133	124
Matched control mesas		2,560		620
Districts	282		9	
Observations	11,931	3,222	3,351	744

Mesa-level regressions of each outcome on the challenge indicator with district fixed effects and standard errors clustered by district (in parentheses). Matched columns use within-district nearest-neighbor matching (up to five controls) on first-round Perú Libre share and turnout, caliper 0.10 on share. The

Table 5 turns the null results into a bound stated in votes. The simulations seed synthetic manipulation favoring Castillo into the actual challenged mesas, 500 replications per cell, with per-mesa magnitudes drawn Poisson so that digit structure is realistically perturbed. Ballot stuffing adds votes to Castillo and to the cast total, capped by each mesa's unused registration. Vote switching moves votes from Fujimori to Castillo, capped by each mesa's Fujimori count. For the turnout and corner statistics and the digit tests, detection means rejection at $p < 0.05$. For the share and swing coefficients, whose observed values already carry the selection documented above, detection means the injected coefficient escapes the confidence interval actually estimated, so the criterion asks whether the manipulation could hide inside the estimates in Table 4.

The bound comes out an order of magnitude below the outcome. Stuffing is invisible to the battery at 2,301 total votes and caught in every replication at 3,071, so the detection threshold sits inside that bracket: stuffing below roughly 3,000 votes can escape, and anything from 3,071 votes, seven percent of the official margin, is caught with probability one. The steepness of that jump has a mechanical cause worth stating. Challenged mesas run 0.5 points below their controls in turnout, so injected stuffing must first fill that observed deficit before the turnout test can reject, which makes the turnout-based detection conservative and the bracket, if anything, generous to hidden manipulation. Switching is caught at the smallest magnitude simulated. At 746 switched votes the swing coefficient escapes its estimated interval in 95 percent of replications and the share coefficient in 90 percent, and 1,489 votes move both in all of them. A switched vote moves the margin by two, so even the top of the switching grid, 25,368 votes, exceeds the scale needed to overturn the official margin of 44,263. Every manipulation large enough to matter is caught long before it does.

These bounds are conditional on one reading: they treat the observed share and swing coefficients as selection, and ask only whether manipulation could hide on top of them. The worst case drops that reading and attributes the entire matched share coefficient to manipulation. The coefficient is Castillo's share of the votes cast, so applying 0.92 points

to the 148,166 votes cast in the challenged mesas gives 1,361 net Castillo votes, 2,721 in margin-equivalent terms if realized by switching, and 2,110 and 4,220 at the upper end of the confidence interval.[7] The placebo mirror is the reason selection is the right reading of that coefficient. But even the reading that concedes all of it to manipulation stays an order of magnitude below the 44,263-vote margin. The two readings do not conflict. The injection bound catches manipulation added on top of the observed coefficients, while the worst case reads the observed coefficient itself as manipulation, and both land far below it.

Two scope caveats travel with the bound wherever it is read. It applies only inside the 768 challenged mesas and says nothing about the mesas nobody challenged. And it covers only the two simulated technologies, ballot stuffing and vote switching favoring Castillo. Pre-electoral conduct lies outside a vote-count audit by construction, as does forgery that altered no count. Within that scope, the evidence is consistent with a clean count in the challenged mesas, and anything hidden in that count is an order of magnitude too small to change the result. The null bounds what could hide in the challenged mesas. It does not clear the election. Unchallenged mesas were never audited, and a clean reading inside the challenged set says nothing about the whole.

The campaign's own pleadings connect the alleged fraud back to this bound. The modal recoverable ground is suplantación under Article 363 literal b, a template whose language covers the suplantación of voters and of mesa members without distinguishing, so the pleading admits two readings, and each maps onto the test family with power against it. Read as impostor voting, ballots cast under forged voter signatures, the fraud moves the counts: ballots genuine electors did not cast either inflate the cast total, the signature ballot stuffing leaves, or displace votes that would have been cast otherwise, the signature of switching, and Table 5 excludes both at decisive scale. Read as forged mesa-member signatures, an acta completed by hands other than the drawn citizens', the allegation is wholesale invention of tally sheets, the one class of manipulation Section 7 shows digit tests do reach, and

[7] Vote figures are computed from the unrounded coefficient and standard error, so they sit slightly below the product of the rounded 0.0092 and 148,166.

within the challenged mesas the last digits of both candidates' counts are consistent with uniformity. In either reading, the test with power against the pled fraud finds nothing at scale. A count-based audit reaches sheet invention only when the forged numbers leave a digit or arithmetic trace, and reading the tally-sheet images directly, as Cantú (2019) does for Mexico's 1988 count, tests that class without the assumption, a route Peru's public acta images leave open. Forged signatures that altered no count remain possible, but then the alleged conduct, whatever its legal gravity, did not move votes, and the dispute this paper addresses is a dispute about votes. As pled, the fraud implies count alteration, count alteration at decisive scale is excluded, and what survives is the version of the allegation that could not have changed the result.

Table 5: *Injection simulations: detection probability and the equivalence bound*

Injected votes	Share	Swing	Turnout	Corner	Digits	Battery
Ballot stuffing (adds Castillo votes and raises the cast total)						
766	0.00	0.00	0.00	0.00	0.02	0.02
2,301	0.00	0.00	0.00	0.00	0.01	0.01
3,071	0.74	0.92	1.00	0.00	0.02	1.00
3,834	1.00	1.00	1.00	0.00	0.03	1.00
7,654	1.00	1.00	1.00	0.00	0.04	1.00
15,262	1.00	1.00	1.00	0.97	0.03	1.00
42,181	1.00	1.00	1.00	1.00	0.01	1.00
Vote switching (moves votes from Fujimori to Castillo)						
746	0.90	0.95	0.00	0.00	0.01	0.95
1,489	1.00	1.00	0.00	0.00	0.02	1.00
2,949	1.00	1.00	0.00	0.00	0.04	1.00
7,105	1.00	1.00	0.00	0.00	0.04	1.00
25,368	1.00	1.00	0.00	0.00	0.00	1.00
Equivalence bound (votes)	stuffing: 2,301–3,071		switching: below 746			

Each cell is the share of 500 replications in which the statistic detects manipulation seeded into the challenged mesas of the matched sample. Injected votes are realized means scaled to the 768-mesa petition set. Detection means $p < 0.05$ for every statistic except share and swing, where it means the injected coefficient reaches the observed estimate plus 1.96 standard errors. Battery detects when any statistic does. The stuffing detection threshold lies between 2,301 (detection 0.01) and 3,071 (detection 1.00), the bracketing grid points. The worst case attributes the observed matched share coefficient entirely to manipulation: 1,361 net votes, 2,721 margin-equivalent by switching. The official margin is 44,263 votes. The full grid and the seeding mechanics are in Appendix C.

7 The powerlessness of digit tests

Digit tests hold a central place in the election-forensics toolkit. Their premise is behavioral. Officials who invent tallies write numbers, and humans write nonuniform digits, so a last-digit distribution that departs from uniformity flags fabrication (Beber and Scacco, 2012). The tests travel easily because they require nothing beyond the published counts, which has made them a standard first screen in contested elections, alongside warnings that their theoretical baselines are fragile (Deckert et al., 2011) and that coarser fingerprints like integer round-offs may do better (Kobak et al., 2016). What the literature rarely supplies is their power against a concrete manipulation of known size in real data.

The injection design measures that power directly. Every simulated manipulation in Table 5 recomputes the two-sample digit-homogeneity test on the perturbed counts, so the digits column reports how often the test catches manipulation that is actually there, seeded into the actual challenged mesas. The answer is almost never. Across both technologies and every magnitude on the grid, 13,000 replications in all, detection never exceeds 5 percent, the nominal size of the test. At outcome-flipping scale it is worse: 42,181 added votes are caught in under 1 percent of replications, and 25,368 switched votes in none. A digit test applied to these mesas cannot distinguish an election-flipping manipulation from a clean count.

The mechanism is arithmetic. The challenged mesas cast roughly two hundred ballots each, 148,166 votes across 768 mesas, so candidate counts are genuine two- and three-digit numbers whose last digits are close to uniform whether or not anyone tampered with them.[8] Stuffing or switching a Poisson-distributed number of ballots adds noise to a true count, and noise pushes last digits toward uniformity, never away from it. The simulations show this directly: at the largest magnitudes detection falls below the test's false-positive rate, because

[8]The tests keep candidate counts of at least ten, because below that the last digit is tied to the magnitude rather than free under the uniform null. The cutoff is not doing the work. The marginal Castillo homogeneity rejection is unchanged as the threshold falls to one and to zero ($p = 0.043$, 0.042, 0.042), while including zero counts manufactures a spurious Fujimori rejection, its uniformity p falling from 0.093 to 0.033 as the many mesas with no Fujimori votes pile onto the last-digit-zero cell, the mechanical artifact the cutoff removes rather than a signature of manipulation. Full grid in the replication output.

the injection smooths whatever digit irregularity the observed counts carry. The marginal homogeneity rejection for Castillo counts in Table 4 ($p = 0.04$) should be read in that light. Every simulated manipulation would have erased that pattern and none would have created it, so it carries no diagnostic weight against them in either direction.

The scope of the conclusion matters. Digit structure comes from inventing numbers without arithmetic behind them, the setting the tests were built for (Beber and Scacco, 2012), and wholesale fabrication of tally sheets remains inside their reach. What lies outside it, at Peruvian mesa sizes, is any manipulation that moves real ballots and keeps the arithmetic consistent, the class of vote-count manipulation at the center of the 2021 public dispute. For that class, a null digit result is not evidence of a clean count, here or in any election with polling stations this small. The bound in Section 6 therefore rests on the statistics with demonstrated power, turnout, the Klimek corner, and the share and swing coefficients, and the digit tests are reported for completeness only. Two scope limits are worth stating plainly. The argument is about last-digit tests, the family run here. Whether the same contraction covers second-digit Benford or integer-percentage fingerprints (Kobak et al., 2016) is not established here, since this paper does not run them, though the integer-percentage fingerprint is itself coarse at these mesa sizes, where a mesa casts roughly two hundred votes and a single vote already moves the winner's share by about half a percentage point. And the powerlessness holds at these mesa sizes. Last-digit tests may still inform in settings with far larger vote counts.

8 Conclusion

After losing the 2021 runoff, Fuerza Popular filed nullity petitions against 768 mesas that together held nearly twice the national margin, and the courts disposed of the record on procedure, denying the few petitions they reached and never testing the substance of the rest. This paper examined the allegations on their merits. The challenged mesas turn out to have

been selected for their extremity more effectively than any random benchmark allows. Beyond that selection they carry no statistical signature of manipulation. Turnout, the fabrication corner, and the digit counts all look like the matched controls. Anything still hidden in their count, within the audited mesas and for the simulated technologies, is bounded an order of magnitude below the margin. Where the tribunals did reach the merits, on the 77 petitions that drew a ruling, they denied every one as unfounded, an independent legal echo of the forensic null.

Two lessons travel beyond Peru. The first is about measurement. Administrative status variables are workflow artifacts of electoral bureaucracies, named for processing steps that carry no information about which mesas a party challenged. No forensic design built on a challenge record can skip linking the challenger's own case files, unit by unit. The status variable here is the cautionary case: 98.6 percent of the actas it flags were challenged by nobody. The second is about selection. A challenge record is adversarially selected on outcomes, so every outcome the filer could observe is disqualified as manipulation evidence the moment the record exists, and a challenge-based audit must split its outcomes into those that document the selection and those that can discriminate, with power demonstrated by injection rather than assumed.

The digit-test result travels furthest of all. At polling stations of a few hundred ballots, last-digit tests have no power against vote-count manipulation, so a null digit result there is not evidence of a clean count, in this election or any other run on units this small. The limit is in the tool, not in Peru, and it applies wherever the standard digit screen is reached for on small polling stations.

One data release would make the next audit stronger than this one. ONPE publishes acta-level results but no processing timestamps, which leaves the count-sequence family of suspicions permanently untestable, and the timestamp designs used elsewhere would not settle them: the late-count shift read as fraud in Bolivia (Escobari and Hoover, 2024) dissolves once the composition of late-reporting mesas is handled (Idrobo et al., 2022) and fails on

its own terms under closer scrutiny (Idrobo et al., 2026). Publishing mesa-level processing timestamps would cost little and would close an entire family of allegations to speculation, in either direction.

The dispute over these mesas outlived the proclamation and settled into the country's politics as an open question. It should not be one. The record the losing campaign itself assembled, the mesas it chose, examined on the merits for the first time, is consistent with a clean count in the challenged mesas.

References

Beber, B. and Scacco, A. (2012). What the numbers say: A digit-based test for election fraud. *Political Analysis*, 20(2):211–234.

Cantú, F. (2014). Identifying irregularities in Mexican local elections. *American Journal of Political Science*, 58(4):936–951.

Cantú, F. (2019). The fingerprints of fraud: Evidence from Mexico's 1988 presidential election. *American Political Science Review*, 113(3):710–726.

Chumacero, R. A. (2025). Election forensics: Statistical tests for detecting electoral manipulation in Latin America. *SAGE Open*, 15(1).

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Deckert, J., Myagkov, M., and Ordeshook, P. C. (2011). Benford's Law and the Detection of Election Fraud. *Political Analysis*, 19(3):245–268.

Eggers, A. C., Garro, H., and Grimmer, J. (2021). No evidence for systematic voter fraud: A guide to statistical claims about the 2020 election. *Proceedings of the National Academy of Sciences*, 118(45):e2103619118.

Escobari, D. and Hoover, G. A. (2024). Late-arriving votes and electoral fraud: A natural experiment and regression discontinuity evidence from Bolivia. *World Development*, 173:106418.

European Union Election Observation Mission (2021). Final report: Peru presidential and congressional elections 2021. Technical report, European External Action Service, Brussels.

Gallego, J., Long, J. D., and Weingast, B. R. (2023). Fraud, uncertainty, and the legitimacy of electoral outcomes: Evidence from Bolivia. *Journal of Conflict Resolution*, 67(6):1129–1157.

Hyde, S. D. (2007). The observer effect in international politics: Evidence from a natural experiment. *World Politics*, 60(1):37–63.

Idrobo, N., Kronick, D., and Rodríguez, F. (2022). Do shifts in late-counted votes signal fraud? evidence from bolivia. *Journal of Politics*, 84(2):1128–1132.

Idrobo, N., Kronick, D., and Rodríguez, F. (2026). On unfounded claims of electoral fraud. *World Development*, 198:107155.

Isea, J. and Isea, R. (2021). Was there voter fraud in the 2021 Peru presidential elections? *Journal of Model Based Research*, 1(4):1–5.

Jurado Nacional de Elecciones (2015). Resolución n.º 0331-2015-jne. reglamento del procedimiento aplicable a las actas observadas en elecciones generales y de representantes peruanos ante el parlamento andino. Technical report, Jurado Nacional de Elecciones, Lima.

Jurado Nacional de Elecciones (2018). Resolución n.º 0086-2018-jne. reglas sobre solicitudes de nulidad de votación de mesa de sufragio. Technical report, Jurado Nacional de Elecciones, Lima.

Jurado Nacional de Elecciones (2021). Resoluciones sobre actas observadas e impugnadas,

elecciones generales 2021 segunda vuelta presidencial. Technical report, Jurado Nacional de Elecciones del Perú, Lima, Perú.

Klimek, P., Yegorov, Y., Hanel, R., and Thurner, S. (2012). Statistical detection of systematic election irregularities. *Proceedings of the National Academy of Sciences*, 109(41):16469–16473.

Kobak, D., Shpilkin, S., and Pshenichnikov, M. S. (2016). Integer percentages as electoral falsification fingerprints. *Annals of Applied Statistics*, 10(1):54–73.

Myagkov, M., Ordeshook, P. C., and Shakin, D. (2009). *The Forensics of Election Fraud: Russia and Ukraine*. Cambridge University Press, Cambridge.

Oficina Nacional de Procesos Electorales (2021a). Resultados por mesa de las Elecciones Presidenciales 2021 Primera Vuelta. Datos abiertos, https://datosabiertos.gob.pe. File: Resultados_1ra_vuelta_Version_PCM.csv.

Oficina Nacional de Procesos Electorales (2021b). Resultados por mesa de las Elecciones Presidenciales 2021 Segunda Vuelta. Datos abiertos, https://datosabiertos.gob.pe. File: Resultados_2da_vuelta_Version_PCM.csv.

Organization of American States (2021). Preliminary report of the OAS electoral observation mission to the general elections of Peru, second round. Technical report, OAS Electoral Observation Mission, Washington, D.C.

A Grounds coding and validation

The grounds coding was pre-registered before any resolution text was read: categories fixed to the Article 363 literals, a 50-expediente validation sample sealed in advance, a 90 percent agreement threshold for the outcome classifier, and a 60 percent coverage threshold below which grounds findings carry a permanent partial label.

The first classifier was rejected before validation on a mechanical ground: it matched keywords anywhere in the resolution, and the resolutions quote the Article 363 statute text, which itself names grounds like violence and forgery. Its output measured the statute text instead of the petition's own claim. The validated classifier codes only from restatement windows, the passages where a resolution restates what the petition alleged, with the statute quotation stripped.

Against the sealed sample the outcome classifier agrees in 46 of 50 cases. The four disagreements are two absorbed duplicates whose direct-versus-inherited outcome was conflated, one JNE remand whose final outcome is ambiguous in the record and is conservatively counted as a miss, and one mooted case coded as consent archival. None crosses the merits line: no granted petition was misclassified, and the zero-granted result is unaffected. A separate cross-check on the 107 absorbed expedientes finds 91 with a machine-readable parent, all 91 parents resolving to inadmissibility.

Substantive grounds are recoverable for 268 of the 942 Fuerza Popular expedientes, 28.5 percent, far below the coverage threshold, so all grounds findings are partial. Within the recoverable subset, 199 petitions plead suplantación, 64 violence or intimidation, 3 an outcome-based argument, and 2 irregular installation. The pre-registered prediction that outcome-based grounds would be modal is disconfirmed.

The templated character of the filings is quantified over the 483 expedientes whose resolutions restate any of the petition's language: 89.4 percent share near-verbatim text with another petition in the same JEE (Jaccard similarity of at least 0.60 on five-word shingles), while 5.0 percent match a petition in a different JEE. Because the statistic runs on

the resolutions' restatements, the petitions themselves not being in the public record, the within-JEE duplication cannot be apportioned between the filer's templates and the tribunals' summarizing conventions, and the paper rests the bulk-filing claim on the filing-date record instead.

B The challenge record in the forensic space

Figure B.1 repeats the Klimek plot of the main text (Figure 3) in Fujimori share. The two challenge arms exchange places: Fuerza Popular's mesas now sit low and Perú Libre's high, the same mirror seen from the other candidate's side.

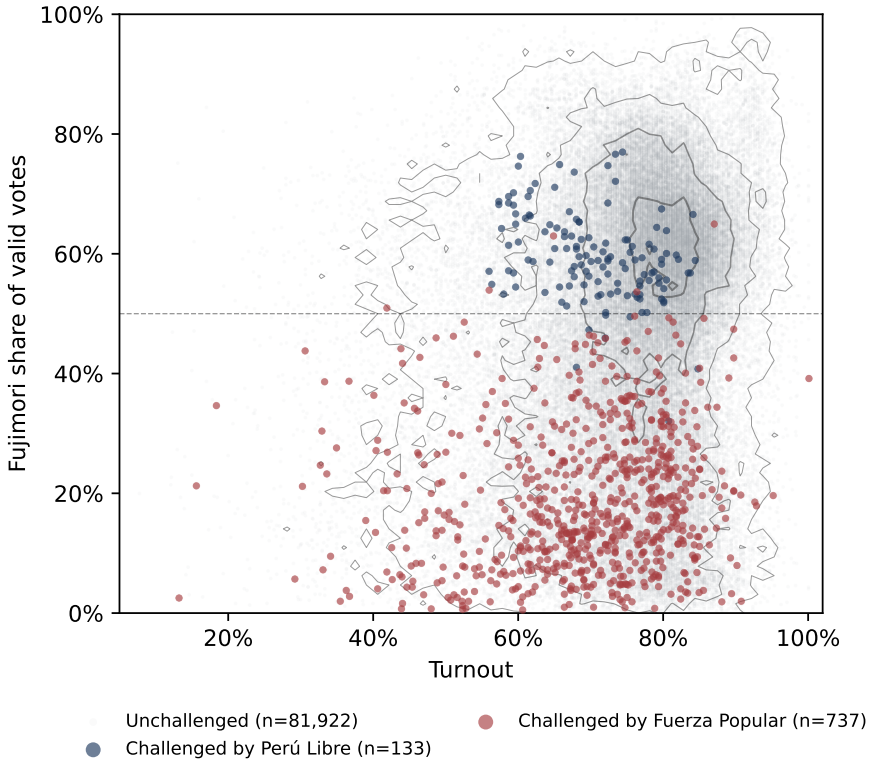

Figure B.1: *Challenge arms in the Klimek space: Fujimori share*

Same construction as Figure 3 with Fujimori's share on the vertical axis, so the two challenge arms exchange places.

Figure B.2 shows the same selection in first-round space: Fuerza Popular's challenges sit above the national first-to-second round relationship, Perú Libre's below it, each side deep in

its opponent's terrain. Figures B.3 and B.4 compare the swing distributions of unchallenged and challenged mesas, whose difference is the matched swing coefficient of Table 4.

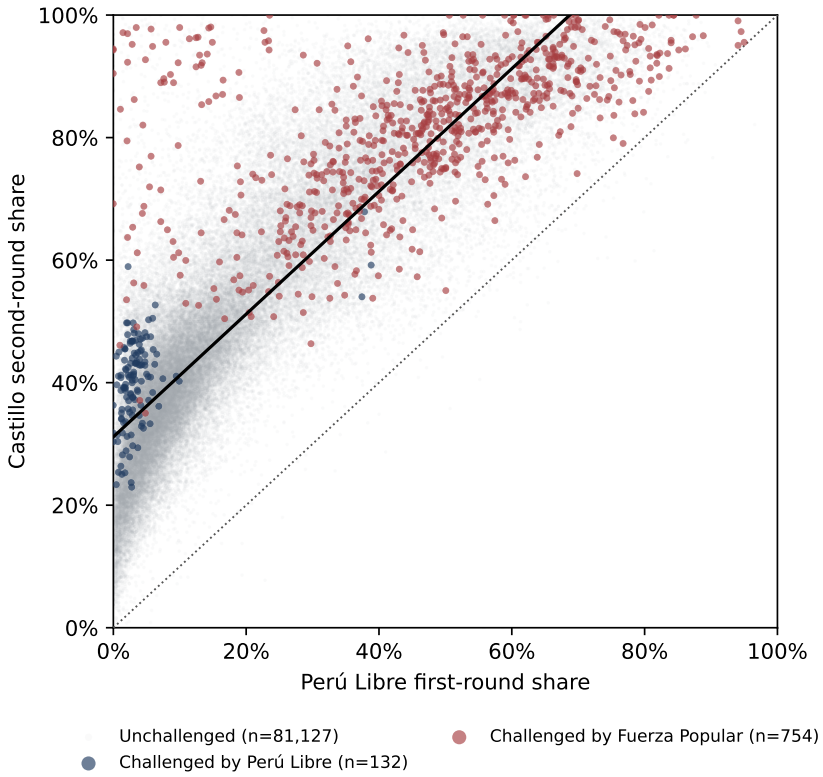

Figure B.2: *First-round share against second-round share by challenge arm*
Domestic vote-bearing mesas with a first-round merge, so the challenged count (754) exceeds the interior-share subset of Figure 3. Challenges sit above the fitted line, the positive residual that Table 4's swing coefficient measures. Because first-round level and between-round swing are correlated, this is the same targeting the yield benchmark reads on level, and no separate swing criterion is at work. The solid line is the OLS fit over all mesas, the dotted line the diagonal.

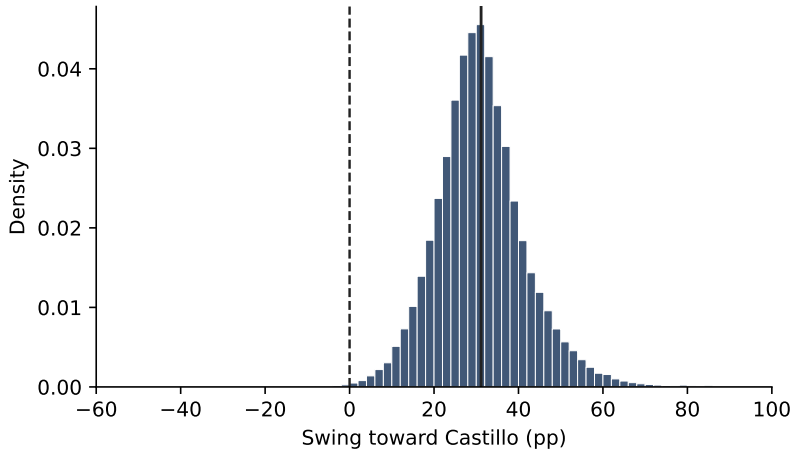

Figure B.3: *Swing toward Castillo between rounds: unchallenged mesas*

Mesa-level change from Perú Libre's first-round share to Castillo's two-candidate runoff share. The solid line marks the mean, the dashed line zero.

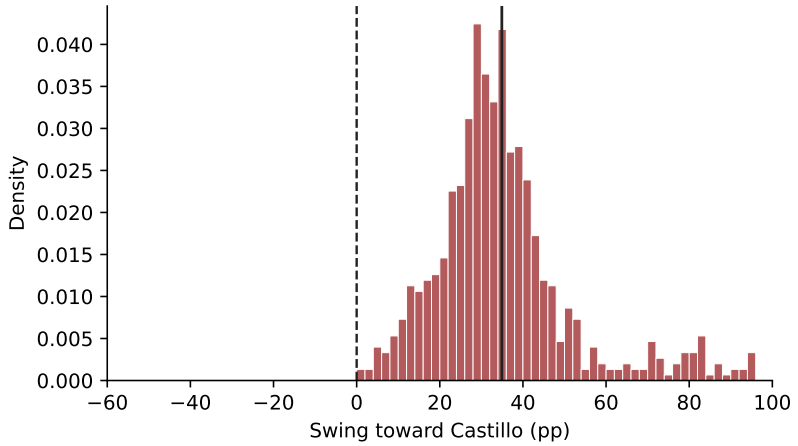

Figure B.4: *Swing toward Castillo between rounds: challenged mesas*

Same construction as Figure B.3 for the Fuerza Popular-challenged mesas. The difference between the two distributions, conditioned within districts, is the swing row of Table 4.

Figure B.5 closes the appendix by locating the extreme swings. Of the 167 mesas whose swing toward Castillo reached 80 percentage points or more, 27 were challenged, 16 percent against a base challenge rate of 0.9 percent in the swing sample, so extreme-swing mesas are targeted at roughly eighteen times the average rate. Yet they are not the campaign's main quarry, and the reason sharpens the selection story rather than complicating it. A swing of 80 points requires, by arithmetic, a first-round Perú Libre share of at most 20 percent, so

these are mesas where the party started from almost nothing and that carry few net votes to recover. The bulk of the strategic value sits in high-*level* mesas, the large Castillo margins the yield benchmark of Table 3 rewards. The low-yield tail of dramatic swings contributes little. The campaign optimized the level of the opponent's margin, and the swing tail is over-challenged only because level and swing are correlated.

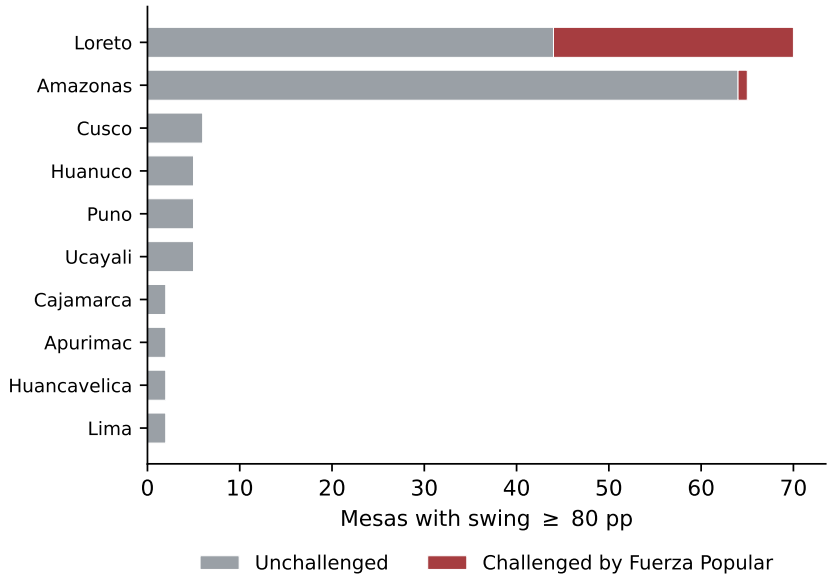

Figure B.5: *Mesas with swings of 80 points or more, by department and challenge status*
The ten departments with the most extreme-swing mesas. Challenged segments count Fuerza Popular petitions.

C Injection detail and the robustness of inference

Each simulation seeds synthetic manipulation favoring Castillo into the actual challenged mesas of the matched sample and re-runs every test, 500 replications per cell. Per-mesa magnitudes are drawn Poisson with mean m, so digit structure is realistically perturbed rather than shifted by a constant. Stuffing adds the draw to Castillo's count and to the cast total, capped by the mesa's unused registration. Switching moves the draw from Fujimori to Castillo, capped by the mesa's Fujimori count. The caps make realized totals fall short of the nominal $m \times 768$, badly so for switching at high magnitudes where Fujimori's count binds, so

every total reported in the paper is the realized mean injection scaled to the full petition set. The seeding and the detection tests run on the 662 matched treated mesas, and the totals are scaled by 768/662 to state the bound over the whole petition set. That scaling assumes the 106 unmatched challenged mesas, which lacked a close within-district control, are comparable in size to the matched ones. At $m = 4$ stuffing, for instance, the realized injection is about 2,647 votes on the 662 mesas and 3,071 scaled to 768. Reporting the matched-sample total instead would lower every stuffing figure by the same factor and leave the order-of-magnitude conclusion unchanged.

Detection is defined per outcome. For turnout, the corner indicator, and digit homogeneity, whose baselines do not reject, detection is rejection at $p < 0.05$. For the Castillo share and swing coefficients the $p < 0.05$ criterion is degenerate, because the observed coefficients already reject at zero injection, selection having been on the outcome, and a criterion that detects manipulation in unmanipulated data measures nothing. Detection there means the injected coefficient reaches the observed estimate plus 1.96 standard errors, thresholds of 0.0142 on share and 0.0130 on swing: the question is whether the manipulation could hide inside the confidence intervals actually estimated. The battery detects when any test does. Table C.1 reports the full grid.

Table C.1: *Full injection grid: detection probability by test*

m	Realized votes	Share	Swing	Turnout	Corner	Digits	Battery
Ballot stuffing							
1	766	0.00	0.00	0.00	0.00	0.02	0.02
2	1,531	0.00	0.00	0.00	0.00	0.01	0.01
3	2,301	0.00	0.00	0.00	0.00	0.01	0.01
4	3,071	0.74	0.92	1.00	0.00	0.02	1.00
5	3,834	1.00	1.00	1.00	0.00	0.03	1.00
6	4,598	1.00	1.00	1.00	0.00	0.02	1.00
8	6,126	1.00	1.00	1.00	0.00	0.03	1.00
10	7,654	1.00	1.00	1.00	0.00	0.04	1.00
15	11,459	1.00	1.00	1.00	0.16	0.03	1.00
20	15,262	1.00	1.00	1.00	0.97	0.03	1.00
25	19,067	1.00	1.00	1.00	1.00	0.03	1.00
40	30,184	1.00	1.00	1.00	1.00	0.02	1.00
58	42,181	1.00	1.00	1.00	1.00	0.01	1.00
Vote switching							
1	746	0.90	0.95	0.00	0.00	0.01	0.95
2	1,489	1.00	1.00	0.00	0.00	0.02	1.00
3	2,221	1.00	1.00	0.00	0.00	0.03	1.00
4	2,949	1.00	1.00	0.00	0.00	0.04	1.00
5	3,662	1.00	1.00	0.00	0.00	0.04	1.00
6	4,374	1.00	1.00	0.00	0.00	0.05	1.00
8	5,765	1.00	1.00	0.00	0.00	0.04	1.00
10	7,105	1.00	1.00	0.00	0.00	0.04	1.00
15	10,282	1.00	1.00	0.00	0.00	0.05	1.00
20	13,136	1.00	1.00	0.00	0.00	0.03	1.00
25	15,671	1.00	1.00	0.00	0.00	0.03	1.00
40	21,340	1.00	1.00	0.00	0.00	0.01	1.00
58	25,368	1.00	1.00	0.00	0.00	0.00	1.00

Each cell is the share of 500 replications in which the test detects. m is the Poisson mean of the per-mesa injection, realized votes the mean injected total scaled to the 768-mesa petition set. Share and swing detection means the injected coefficient escapes the observed confidence interval, other tests reject at $p < 0.05$. The turnout column tests the runoff turnout level. The switching grid ends at 25,368 realized votes because Fujimori's counts in the challenged mesas bind the caps.

Spatial dependence. The primary standard errors cluster by district, allowing arbitrary

correlation among the mesas within each of the 282 districts, and the pre-specified province-level clustering and the placebo wild cluster bootstrap probe that correlation at a coarser and a finer resolution. A spatial heteroskedasticity- and autocorrelation-consistent correction (Conley, 1999) would additionally absorb correlation across district boundaries. Two facts make it immaterial here. First, the data preclude its sharp form: the ONPE extract locates each acta only to its district, with no mesa coordinates, so any spatial kernel would have to place every mesa at its province centroid, a correction coarser than the district clustering already imposed. Second, a spatial correction can only widen standard errors, and the paper's conclusions rest on three quantities a wider interval cannot overturn. The nulls on the discriminating outcomes only strengthen as their intervals widen. The bound in votes comes from the injection simulations, which use no analytic standard error at all. And the selection rests on the yield benchmark, a non-parametric comparison against 10,000 draws that has no standard error to correct. The analytic significance of the share and swing coefficients is not what carries it. A spatial correction could move the reported significance of those two coefficients, but they are stated as selection evidence and carry no conclusion of their own, so no result in the paper depends on it.