Do shifts in late-counted votes signal fraud? Evidence from Bolivia[†‡]

Nicolás Idrobo[§] Dorothy Kronick[¶] Francisco Rodríguez[‖]

January 28, 2022

Abstract

Shifts in late-counted votes often spark unfounded claims of electoral fraud. These claims exploit *the early-count mirage*: the expedient illusion that, absent fraud, an early advantage will persist. We characterize the early-count mirage and evaluate associated fraud claims in four disputed elections, focusing on the case of Bolivia in 2019. When late-counted votes delivered a narrow victory for the incumbent, fraud accusations followed—with dramatic political consequences. But we find that the vote-share trend can be explained without invoking fraud, and that the allegedly suspicious shift in late-counted votes was actually an artifact of methodological and coding errors on the part of electoral observers. We document similar patterns in the other three cases. The details are context-specific, but the core insights are general: time trends from legitimate vote-counting processes are far more varied—and errors in influential analysis far more frequent—than election skeptics allege.

[†]The empirical analysis has been successfully replicated by the JOP replication analyst. Replication files are available in the JOP Data Archive on Dataverse (http://thedata.harvard.edu/dvn/dv/jop).

[‡]We are grateful to Mike Alvarez, Santiago Anria, Marc Meredith, and Rocío Titiunik for detailed guidance. For research assistance, we thank Mateo Arbeláez, Estefanía Bolívar Méndez, Giancarlo Bravo, Adolfo De Lima, and Juan Vera. For comments, we thank María Eugenia Boza, Ernesto Calvo, Francisco Cantú, Matías D. Cattaneo, Tulia Falleti, Camilo García-Jimeno, Francisco Garfias, Etan Green, Justin Grimmer, Guy Grossman, David Hausman, Dan Hopkins, Richard Kronick, Yphtach Lelkes, Matt Levendusky, Michele Margolis, Edward Miguel, Jonathan Mummolo, David Rosnick, Josh Simon, Uri Simonsohn, Tara Slough, and workshop participants at the University of Pennsylvania, Berkeley, Tulane, Stanford, Data Colada, Polmeth, and UCSD. We thank Anatoly Kurmaneav and the New York *Times* for sharing data. Rodríguez acknowledges the generous support for this research of the Center for Inter-American Policy and Research of Tulane University and the Kellogg Institute for International Studies of the University of Notre Dame.

[§]Department of Political Science, University of Pennsylvania

[¶]Corresponding author (kronick@upenn.edu). Dept. of Political Science, University of Pennsylvania

[‖]Council on Foreign Relations

The order in which votes are counted is anything but random. Voters who have sat through election night in Brazil or Colombia or the United States, for example, know that results from the first few precincts rarely resemble the final outcome.[1]

Yet politicians often point to this fact as evidence of fraud. In doing so, they encourage and exploit what we call *the early-count mirage*: the expedient illusion that, absent fraud, an early advantage will persist through the end of the count.

After the 2020 presidential election in the United States, for example, Donald Trump denounced the "brazen and outrageous election theft" that allegedly occurred after 10:00 p.m. on election night, when late-counted votes delivered key states for his opponent (NPR, 2021). In Mexico in 2006, likewise, presidential candidate Andrés Manuel López Obrador charged that his margin shrank "in an inexplicable way" on the morning after the poll, leading to a narrow loss (Antenangeli and Cantú, 2019, p. 142). And in Ecuador's 2017 presidential election, Guillermo Lasso's running mate claimed that "with 82.2% of the vote counted, it was impossible to revert the trend [in our favor]"—as effectively happened in the final tally (Páez Benalcázar, 2017).

This dynamic is dangerous. In the United States, the early-count mirage contributed to the storming of the Capitol and to enduring doubts about the integrity of the 2020 presidential election (Eggers, Garro and Grimmer, 2021). In Mexico, it fueled months of protests (El País, 2006). In Ecuador, it so thoroughly undermined perceptions of electoral integrity that, in the first round of the subsequent election, the third-place candidate denounced a "satanic pact" behind fraud (Nodal, 2021). The early-count mirage can deepen the well-documented tendency to disbelieve and delegitimize electoral losses (Anderson et al., 2007; Cantú and García-Ponce, 2015; Bush and Prather, 2017; Calvo et al., 2020), which in turn can spark protests and violence (Tucker, 2007; Little, Tucker and LaGatta, 2015; Luo and Rozenas, 2018). Ultimately, the early-count mirage can erode a fundamental tenet of democracy: the acceptance of legitimate electoral results (Przeworski, 1991).

We characterize the early-count mirage and reevaluate associated fraud claims in

[1]One example from each case illustrates the point. Brazil: In the 2018 presidential election, Fernando Haddad earned just 27% of the first 93% of votes counted but more than 43% of the last 7% (Eleitoral, 2018). Colombia: In the 2018 presidential election, Iván Duque earned a 13-point lead in the first 93% of the vote but just a 5-point lead in the last 7% (Registraduría de Colombia, 2018). United States: In the 2018 congressional election, Young Kim (candidate for California's 39th district) held a 3-point lead with 65% of the vote counted—but ultimately lost by 3 points (Li, Hyun and Alvarez, 2021).

four disputed elections, focusing on one especially consequential case. While some of the details are context-specific, the fraud claims suffer from similar flaws and are therefore susceptible to common correctives. In all four cases, fraud claims tied to the early-count mirage fail in the face of one general insight: time trends from legitimate vote-counting processes are far more varied—and errors in influential analysis far more frequent—than election skeptics allege.

We bring this insight to the Bolivian presidential election of 2019, where late-counted votes secured a razor-thin first-round reelection victory for left–populist incumbent Evo Morales. The runner-up alleged fraud (Mesa, 2019). The Organization of American States (OAS), which observed the election, expressed "deep concern and surprise at the drastic and hard-to-explain change in the trend of the preliminary results revealed after the closing of the polls" (October 21, 2019*c*). Subsequent OAS analysis cited "a massive and unexplainable surge in the final 5% of the vote count"—in other words, a discontinuous jump in the incumbent's vote share after 95% of votes had been counted (OAS, 2019*a*, p. 94).

The political consequences were far-reaching. In large part because of the fraud allegations, and in an environment marked by polarization and incumbent power grabs, the Bolivian military asked Morales to resign; he fled to Mexico.

Our analysis reveals two problems with these consequential claims of fraud. The first is a problem of theory: Bolivian electoral administration is such that we would *expect* a leftward shift in vote share, and, moreover, we would *expect* sharp changes in vote share at specific moments in the count—though not, as it happens, at 95% of the count, the point at which the OAS claimed to find evidence of a jump (Section 2.1). The second is a problem of method: the "massive and unexplainable surge in the final 5% of the count" (OAS, 2019*a*, p. 94) was, we show, actually an artifact of methodological and coding errors (Section 2.2). We also discuss issues with other quantitative claims of fraud in this election, later published by academic researchers.

These problems are not unique to Bolivia. We consider the early-count mirage in three additional cases: Mexico in 2006, the United States in 2020, and Honduras in 2017. In all three cases, fraud claims based on late-counted votes undermined the legitimacy of a presidential election; in all three cases, the fraud claims were unfounded.[2] The precise causes of shifts in late-counted votes are context-specific.

[2]For the United States and Mexico, the fraud claims have been discussed elsewhere, and our

But flaws in the fraud claims follow a common pattern: either they fail to acknowledge an innocuous explanation for the shift in late-counted votes, or they fail to analyze the electoral data correctly. Recognizing the diversity of time trends from legitimate vote-counting processes addresses the first failure; prompt review and replication addresses the second.

We advance a growing body of work on fraud-claim forensics. Political scientists have developed sophisticated tools for fraud detection (e.g. Alvarez, Hall and Hyde, 2009; Myagkov, Ordeshook and Shakin, 2009; Hicken and Mebane, 2017). A smaller literature assesses politicians' (often unsophisticated) claims of fraud. Recent efforts debunk the myth of widespread double-voting in the United States (Goel et al., 2020), Donald Trump's fraud allegations (Cottrell, Herron and Westwood, 2018; Eggers, Garro and Grimmer, 2021), and claims of fraud in the Mexican presidential election of 2006 (Antenangeli and Cantú, 2019). Similarly, work on the "blue shift" in the United States provides an innocuous explanation for the fact that late-counted votes often favor Democrats (Foley, 2013; Cottrell, Herron and Westwood, 2018; Foley and Stewart, 2020; Li, Hyun and Alvarez, 2021; Curiel, Stewart III and Williams, 2021). We build on this work by conceptualizing a widespread source of fraud claims—the early-count mirage—and by illustrating how to assess such claims.

We also contribute to an empirical debate over the Bolivian presidential election of 2019.[3] Our analysis does not establish the absence of fraud in this election; rather, we establish that we do not *require* fraud in order to explain the quantitative patterns that helped indict Evo Morales—and changed the course of Bolivian history.

The paper proceeds as follows. Section 1 characterizes the early-count mirage and identifies common flaws in associated fraud claims. Section 2 studies the early-count mirage in Bolivia, documenting the theoretical and empirical flaws with fraud claims in the 2019 presidential election. Section 3 considers three additional cases (Mexico, the United States, and Honduras), emphasizing not only that the early-count mirage is widespread and dangerous, but also that legitimate vote-counting processes produce

contribution is to present them in a unified framework; for Honduras (and for Bolivia), to the best of our knowledge, we are the first to highlight certain errors underlying the accusations.

[3] OAS (2019*a*); Escobari and Hoover (2019); Johnston and Rosnick (2020); Williams and Curiel (2020); Nooruddin (2020*c*); Minoldo and Quiroga (2020); Newman (2020); Rosnick (2020*a*); Nooruddin (2020*a*,*b*); Rosnick (2020*b*). In "Evidence Against Fraudulent Votes Being Decisive in the Bolivia 2019 Election," Mebane (2019) uses his own toolkit to estimate the extent of fraud (rather than responding to or building on claims from the Organization of American States).

a wide variety of time trends even in the absence of malfeasance—and that errors in quantitative claims of fraud are widespread. Section 4 concludes with a discussion of policy recommendations and directions for future work.

1 The Early-Count Mirage

We define the early-count mirage as *the expedient illusion that, absent fraud, an early advantage will persist through the end of the count.*

In some cases, this early advantage takes the form of a simple lead in votes counted early. Trump, for example, claimed fraud merely by noting that, in several key states, his initial lead over Biden disappeared by the time all votes were counted (NPR, 2021). In other cases, the early advantage is more subtle. In Ecuador in 2017, for instance, the losing candidate complained that an early *trend* in his favor should have continued (Páez Benalcázar, 2017). Whether based on early-vote levels, trends, or more sophisticated extrapolation, the early-count mirage appears when the first electoral returns create the deceptive image of one competitor's eventual victory.

One way to understand the early-count mirage is to place it in the framework of hypothesis testing (Eggers, Garro and Grimmer, 2021). Fraud claims that exploit the early-count mirage implicitly formulate a null hypothesis—there was no fraud—and reject it in favor of the alternative that there *was* fraud, using a test statistic tied to time trends in vote share. Returning to Trump's statements, we might think of his implicit test statistic as *the difference between a candidate's final margin and his early margin*, and the rejection region as all negative values of the test statistic: a shrinking margin over the course of the count constitutes grounds for rejecting the null hypothesis of no fraud. In Ecuador in 2017, the implicit test statistic was closer to *the difference between a candidate's actual final margin and the margin obtained by extrapolating early trends*, again with a rejection region < 0.

This framework suggests a two-step approach to assessing fraud claims tied to the early-count mirage.

The first step is to evaluate whether the (implicit) test statistic provides a valid test of the null hypothesis of no fraud. Researchers have done this work for the "blue shift" in the United States (Foley, 2013; Cottrell, Herron and Westwood, 2018; Foley and Stewart, 2020; Li, Hyun and Alvarez, 2021; Curiel, Stewart III and Williams, 2021), showing that late-counted votes often lean Democrat because of ballot type: mail-in ballots are disproportionately Democrat, and mail-in ballots are often counted at the end. This finding invalidates using the blue shift itself as a test statistic.

Vagaries of vote-counting processes in other contexts can similarly provide innocuous explanations for shifts in late-counted votes. In the 2021 presidential runoff in Peru, for example, the vote-share trend was nonmonotonic, moving leftward for most of the count as the share of rural votes increased, and then ticking rightward in the final 3% of the count (Figure H.4)—a reversal that might appear anomalous were it not for the fact that right-leaning votes from abroad are counted last (because embassies send them by diplomatic pouch; Infobae, 2021). Elsewhere, vote-share trends are driven by factors such as physical distance from a precinct to electoral authorities (Antenangeli and Cantú, 2019); poll-worker characteristics (Spencer and Markovits, 2010; Challú, Seira and Simpser, 2020); and/or communication technology, which can determine the speed of electronic transmission of results (as in Venezuela; Martinez, 2021).

These causes are context-specific, but the associated fraud claims suffer from a common flaw: they fail to recognize that an allegedly suspicious shift in late-counted votes might arise even in the absence of fraud.

The second step is to evaluate whether the testing procedure was executed correctly. Even when fraud claims employ a valid testing procedure—that is, even when they highlight a result that would be sufficiently unlikely under the null hypothesis of no fraud—the analysis may suffer from methodological or coding errors (as Eggers, Garro and Grimmer, 2021, document for several U.S. fraud claims). In one sense, this is unremarkable: fraud claims often rest on time-sensitive analysis of complex data sets, and mistakes plague even peer-reviewed research undertaken in the absence of deadlines or public scrutiny (e.g. Ziemann, Eren and El-Osta, 2016). Yet the pervasiveness of basic errors in politically influential analysis is surprising, as the cases we study emphasize. Again, the errors are context-specific, but the fraud claims suffer from a common flaw: they fail to analyze the electoral data correctly.

In short, fraud claims tied to the early-count mirage are vulnerable to one of two general correctives: documenting the (sometimes strange) time trends that emerge from legitimate vote-counting processes, and reviewing or replicating the quantitative analysis.

In what follows, we bring these correctives to the controversial case of Bolivia in 2019. We find, first, that details of Bolivia's vote-counting process invalidate the testing procedures used to claim evidence of electoral fraud—and, moreover, that the test statistics were calculated incorrectly.

2 The Early-Count Mirage in Bolivia

Tensions ran high in advance of Bolivia's 2019 presidential election. The incumbent, Evo Morales, first elected in 2005 as part of Latin America's pink tide (Falleti and Parrado, 2018), was controversially seeking a fourth term in office.[4] He led in the polls, but it was not clear whether he would avoid a runoff by earning a ten-point margin over the runner-up. A runoff would likely have been close (ANF, 2019).

At 7:50 p.m. on election night, Bolivia's electoral authority reported that, with approximately 83% of voting booths counted, Morales (the incumbent) held a 7.87-point lead over the runner-up—not enough to avoid a runoff (Bolivia tv). Publication of updated results was then unexpectedly suspended for nearly twenty-four hours.[5] By that time, the electoral authority announced, Morales had gained a 10.15% lead (Los Tiempos, 2019*a*)—just enough to avoid a runoff.[6]

This sequence of events conjured the early-count mirage. Opposition leaders cried fraud (AFP, 2019), electoral observers sounded alarm bells (OAS, October 21, 2019*c*), and Bolivia "exploded in protest" (Kurmaneav and Castillo, 2019).

The Organization of American States (OAS) audited the election and, three weeks after the poll, released a preliminary report that raised a number of concerns, one of which was a suspicious shift in late-counted votes (OAS, November 10, 2019*b*, p. 9). That afternoon, Morales agreed to new elections (Collyns, 2019). But just hours later, under intense public pressure, Bolivia's military and police chiefs asked him to

[4]Bolivia's 2009 constitution imposed a two-term limit—itself more favorable to the incumbent than the previous rule of no reelection (Corrales, 2016, p. 8)—but in 2013 courts allowed Morales to run for a third term, on the grounds that his first term did not count because it began prior to the new constitution. In 2016, Morales held a referendum on his proposal to eliminate term limits all together—and voters defeated it. He was able to run in 2019 only because Bolivia's highest court ruled that term limits violated the American Convention on Human Rights (Anria and Cyr, 2019).

[5]Antenangeli and Cantú (2019) emphasize that such interruptions often cause controversy—even when they are the result of technical difficulty. In Bolivia, critics charged that the government used "the shutdown" to tamper with results; the government claimed that they never intended to tally 100% of the vote in the preliminary results system (Los Tiempos, 2019*b*), while others cited an "enormity of technical fuck-ups" and "lack of expertise" (*impericia*) (Cambara Ferrufino, 2019).

[6]The final results, announced on October 25, gave Morales 47.08% to Mesa's 36.51%, a margin of 10.57 points.

resign (Kurmanaev, Machicao and Londoño, 2019). He flew to political asylum in Mexico, claiming that he had been ousted in a coup d'etat.

To assess these influential claims of fraud, we begin by describing the vote-share time trends that we would *expect* to observe in Bolivia in the absence of malfeasance.

2.1 The time trends we would expect in the absence of fraud

Three details of Bolivian electoral administration are key to understanding the vote-share time trends that we would expect to observe in the absence of fraud.

First, the poll workers who count ballots are randomly selected from among voters registered *at each voting booth*. They are not self-selected volunteers, or government employees, or even randomly selected from among voters registered in the precinct as a whole (there are six voting booths per precinct,[7] on average). Rather, these poll workers—called jurors (*jurados*)—are legally required to serve. This means that jurors' socio-economic characteristics are highly correlated with those of the voters whose ballots they tabulate: where voters are highly educated, for example, jurors are likely highly educated, too. The jurors are responsible for checking voters' names against the registration list, distributing and receiving ballots, and, most importantly, counting the paper ballots (Appendix Figure H.1) and writing the totals on a paper tally sheet (*acta*, Appendix Figure H.2).

The second key detail concerns the order in which these tally sheets are submitted to Bolivia's preliminary results system.[8] An on-site operator at each precinct takes a photo of each tally sheet—again, one per voting booth—and transmits the image to the electoral authority via a cell-phone app. The operator also types the vote totals into the cell-phone app. Conceptually, we can think of this *transmission time* as each voting booth's *reporting time*.[9]

[7] We use "voting booth" for *mesa* and "precinct" for *recincto*.

[8] The preliminary system is called the TREP: Transmisión de Resultados Electorales Preliminares. Bolivia also has a definitive results system, Cómputo, which is much slower and counts tally sheets in an entirely different order. Discussion of late-counted votes centered on time trends from the preliminary results system, so that is our focus here; however, we briefly describe the definitive system below.

[9] Even though the time stamps include minutes and seconds, only 8% of tally sheets have unique transmission time stamps. This makes sense given the overall speed of the preliminary system: more than 97% of the 34,555 tally sheets arrived between 4:30 p.m. and 9:00 p.m., an average of more than two per second (see Fig. A.2). We sort tally sheets in a random order within each time stamp.

But the vote totals typed into the cell-phone app are not immediately added to the preliminary count. Instead, after *transmission*, *verifiers* at a central location view each tally-sheet image and re-type the vote totals; if the figures match those entered by the on-site operator at each precinct, the tally sheet is *verified* and added to the preliminary count.

This is the third key detail of Bolivia's vote-counting process: tally sheets are not verified in the order in which they are received (i.e., transmitted through the cell-phone app). Rather, verification operators view tally sheet images *in a random order* from the pool of tally sheet images transmitted thus far (NEOTEC, October 28, 2019, p. 5). This means that, during periods when verification largely kept up with transmission, the verification order would largely preserve the transmission order (because the pool of unverified images would remain small). But at any moment when verification lagged behind transmission, the verification order would not reflect the transmission order (see Appendix Figure A.2a). This is especially relevant after 8:06:59 p.m. on election night, when verification paused overnight (even as transmission continued). When verification resumed the following day, verifiers received tally-sheet images in a random order from a large pool.

In the absence of malfeasance, we would expect this system to produce three patterns that, in fact, we observe in the data:

(a) *A pro-incumbent shift in vote share as* transmission *progressed (i.e., as voting booths transmitted results through the app).* The reason is straightforward. Voters and voting-booth jurors with higher levels of education should complete the voting and counting process marginally faster, and, in Bolivia, education is negatively correlated with support for incumbent Evo Morales (Madrid, 2012, p. 36). (Recall that jurors are randomly selected from among voters registered at each booth.) Critically, small *absolute* differences in reporting time correspond to very large *relative* differences in reporting time because the system is so fast: nearly all voting booths (97.35%) transmit results between 4:30 p.m. and 9:00 p.m. (see Figure A.2). In the densest part of this reporting window, a mere ten-minute delay in transmission time would move a voting booth from the 36th to the 46th percentile of transmission time. This is all to say that the education–reporting time gradient need not be very steep, in absolute terms, in order to produce a marked pro-incumbent shift in vote share as reporting progressed.

And indeed, we observe a strong pro-incumbent shift in vote share as reporting progressed. The blue line in Figure 1a plots Morales's average margin (not cumulative margin) over the runner-up as a function of voting booth reporting time (i.e., transmission time); Morales's average margin rose through most of the active reporting window, from near zero to approximately 40 percentage points by the end of the evening.[10] Consistent with the notion that this trend arises because less-educated voting-booth jurors take longer to report—and because less-educated voters support Morales at higher rates—we observe that the trend flattens considerably when we simply control for a small set of socio-economic characteristics, crudely measured at the municipality level. Specifically, the grey line in Figure 1a plots our estimate of $f\left(\text{Time}_{bpm}\right)$ from:

$$\text{Margin}_{bpm} = f\left(\text{Time}_{bpm}\right) + \boldsymbol{\beta}\boldsymbol{X}_{bpm} + \epsilon_{bpm} \tag{1}$$

where $\boldsymbol{X}$ are characteristics of booth b in precinct p in municipality m: measures of education at the municipality level, a vector of proxies for rurality, the number of registered voters, and an indicator for all voting booths located in the Bolivian lowlands (which lean against the incumbent: Anria, 2018, p. 64–67).[11] We estimate $f\left(\text{Time}\right)$ using the semi-parametric estimator proposed by Robinson (1988). Additional controls further flatten the vote margin–time gradient (Appendix B). In short, Bolivia's "blue shift" is predictable rather than anomalous.

(b) *Discontinuous changes in vote share at certain* verification *times* (though not at the time highlighted by the OAS). Because verification operators view tally-sheet images in a random order from the pool of tally sheets transmitted thus far—and because one candidate's vote share increases over time—we would *expect* a discontinuous change in vote share whenever verification lags behind transmission.[12]

[10]The blue line in Figure 1 is fit only to observations in the estimation sample for which we have socio-economic covariates, and thus excludes precincts outside Bolivia (such as embassies). The time trend including all observations is quite similar; see Appendix Figure B.1e.

[11]Specifically, the measures of education at the municipality level are: (i) proportion of adults who are literate, (ii) proportion of adults who completed primary school, (iii) proportion of adults who completed secondary school. The measures of rurality are: (a) the (log) number of registered voters per square kilometer around precinct p, (b) the (log) proportion of the population employed in agriculture in municipality m, (b) the (log) population of municipality m, and (c) an indicator taking a value of 1 if municipality m is the capital of its department.

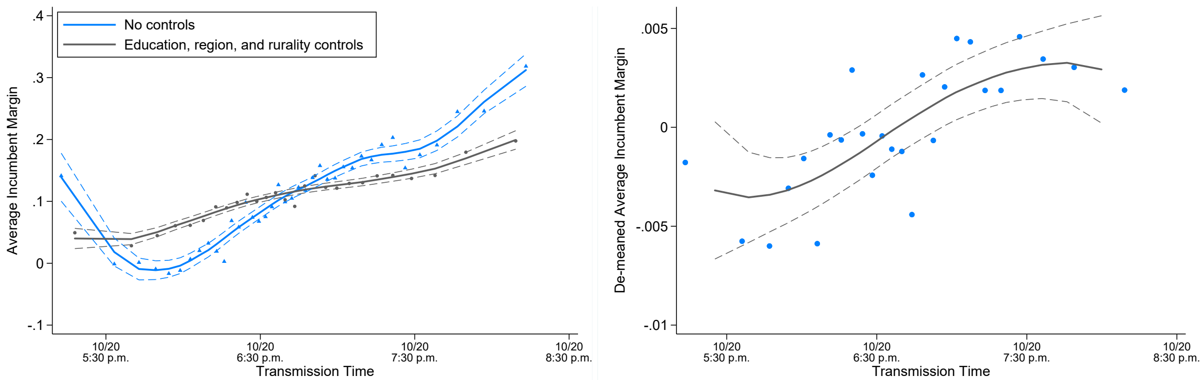
[12]To see this, consider a simplified example. Imagine that there are 100 tally sheets, and that the incumbent's vote share increases in transmission order, at a constant rate: he earns, say, 10% of votes on the tally sheet transmitted first, 10.5% of votes in the tally sheet transmitted second, 11% on the

Figure 1: Predictable Pro-Incumbent Trends in Vote Share

In (a), the blue line marks estimates of the incumbent's average margin over time; the gray line marks our estimate of this relationship after controlling for local educational attainment and other characteristics (Eq. 1). Fig. (b) reveals that there is a slight pro-Morales trend even *within* precincts.

(a) Education, Rurality Largely Explain the Pro-Incumbent Shift

(b) Pro-Incumbent Shift Even Within Precincts



Points mark the average MAS (incumbent) margin over Civic Community (runner-up) in optimal (data-driven) bins of the transmission time (Cattaneo et al., 2019), using data from Bolivia's $N = 34,555$ tally sheets. For the 377 tally sheets without transmission times, we assign the median transmission time in the respective municipality; other reasonable choices do not change the result (Appendix Figure B.1). The solid blue line in (a) and gray line in (b) mark local linear fits following Calonico, Cattaneo and Farrell (2018). The solid grey line in (a) marks estimates of $f(\text{Time})$ from Eq. 1, using the estimator proposed in Robinson (1988). Dashed lines mark 95% confidence intervals. Figure (a) trims the top and bottom 2% of observations; for a version without trimming, see Appendix Figure B.1a.

And indeed, the overnight pause in the process of *verifying* tally sheets naturally created a discontinuous jump in vote share (Appendix A), because tally sheets were verified in a random order after the interruption (though again, this was not the cutoff studied in OAS, 2019*a*).

(c) *A slight pro-incumbent shift even across voting booths within the same precinct.* The distribution of voter education is unlikely to be exactly identical across voting booths within a precinct. This would be true even if voters were assigned to booths randomly within a precinct; the differences across booths are likely greater because Bolivian voters are assigned to booths by alphabetical order of surname—not randomly—and because surnames are tied to ethnicity in Bolivia (we elaborate this point below and in Appendix D). Ethnicity, in turn, is corre-

tally sheet transmitted third, and so on until earning 60% of votes on the tally sheet transmitted last. Now imagine that tally sheets 1–50 are verified in the order in which they were received, but that tally sheets 51–100 are verified in a random order (because verification paused between tally sheet 50 and tally sheet 51). If we plot the incumbent's vote share against the *verification* order, we should expect a jump between the 50th-verified sheet (vote share = 25%) and the 51st-verified sheet, which is drawn at random from the remaining tally sheets (expected vote share = 42.5%).

lated both with education and with support for the incumbent (Madrid, 2012, pp. 36, 69; Klein, 2011, p. 282). As noted above, we would *expect* less-educated voters (and jurors) to report later and to support the incumbent at higher rates; education varies not only across precincts but across booths within a precinct. It would therefore be unsurprising to find a positive within-precinct correlation between the incumbent's margin and reporting time (i.e. transmission time).

And indeed, we observe a slight pro-incumbent shift in vote share even *within* precincts. Figure 1b plots de-meaned incumbent margin—that is, the margin in voting booth b minus the average margin in precinct p, or $(\text{Margin}_{bp} - \overline{\text{Margin}}_p)$—against transmission time. Again, this pattern is consistent with the notion that booths with less-educated voters favored the incumbent and also tend to report later, because the jurors are less educated, too (it is also consistent with other explanations; see below and Appendix D for additional discussion). In short, the within-precinct trend is not necessarily anomalous.

2.2 The time trends cited as indicative of fraud

With these expectations in mind, we follow the two steps outlined in Section 1 above to assess major claims of fraud in Bolivia's 2019 presidential election. For each claim, we first evaluate whether alleged anomalies in late-counted votes were actually anomalous; second, we revisit the execution of the empirical analysis. To paraphrase Eggers, Garro and Grimmer (2021, p. 2), we find that purportedly anomalous facts about Bolivia's late-counted votes were either not anomalous or not facts.

An "inexplicable surge" in the last 5% of the preliminary count. Electoral observers from the Organization of American States (OAS) pointed to late-counted votes as indicative of fraud. On the evening after the election, when the Bolivian authorities first announced that incumbent Evo Morales had cleared the ten-point margin required to avoid a runoff, the OAS mission issued a statement expressing "deep concern and surprise at the drastic and hard-to-explain change in the trend of the preliminary results revealed after the closing of the polls" (OAS, October 21, 2019*c*). The preliminary report of the OAS audit, in turn, sounded alarm bells about a "highly unlikely trend in the last 5% of the vote count" (OAS, November 10, 2019*b*, p. 9), a concern echoed at length in the final audit report (OAS, 2019*a*).

Specifically, the final report of the OAS audit team claimed that Morales's vote share jumped discontinuously after 95% of the vote had been verified (OAS, 2019*a*, p. 88).[13] The report did not articulate a theory of fraud that would produce the alleged jump at 95% of the vote verified—indeed, the report labels 95% "an arbitrary point"[14] (p. 10)—but the implicit notion was one of centralized tampering: realizing that Morales was not on track to win by more than 10 points, presumably, his agents crudely added votes in all booths that had yet to be verified in the preliminary system. Hence his victory was "only made possible by a massive and unexplainable surge in the final 5% of the vote count. Without that surge . . . he would not have crossed the 10% margin that is the threshold for outright victory" (OAS, 2019*a*, p. 94).

In other words, the OAS implicitly tested the null hypothesis of no fraud—or, at least, of no reason for concern about fraud—against the alternative of cause-for-concern, using the test statistic *change in vote share at 8:03:59 p.m., when 95% of the vote was verified in the preliminary system*. A large pro-incumbent change "strained credulity" (p. 94) sufficiently to reject the null. We argue both that the proposed hypothesis test is invalid and, moreover, that the test statistic was miscalculated.

The test is problematic because of the nature of the *verification* time series. As noted in the previous section, tally sheets were verified in a random order conditional on transmission, so any pause or delay in the verification process would naturally generate a jump in vote share (as in fact occurred at 8:06:59 p.m.; see Appendix A). To justify the hypothesis test proposed in the audit report, therefore, the OAS would not only have to articulate a theory of fraud consistent with a jump at 95% but would also need to refute the alternative explanation that the jump arose as the benign and predictable consequence of the design of Bolivia's vote-counting system. Regardless, as we establish in the remainder of this section, the test statistic was miscalculated: there is no jump in vote share at the cutoff studied in OAS (2019*a*, p. 88).

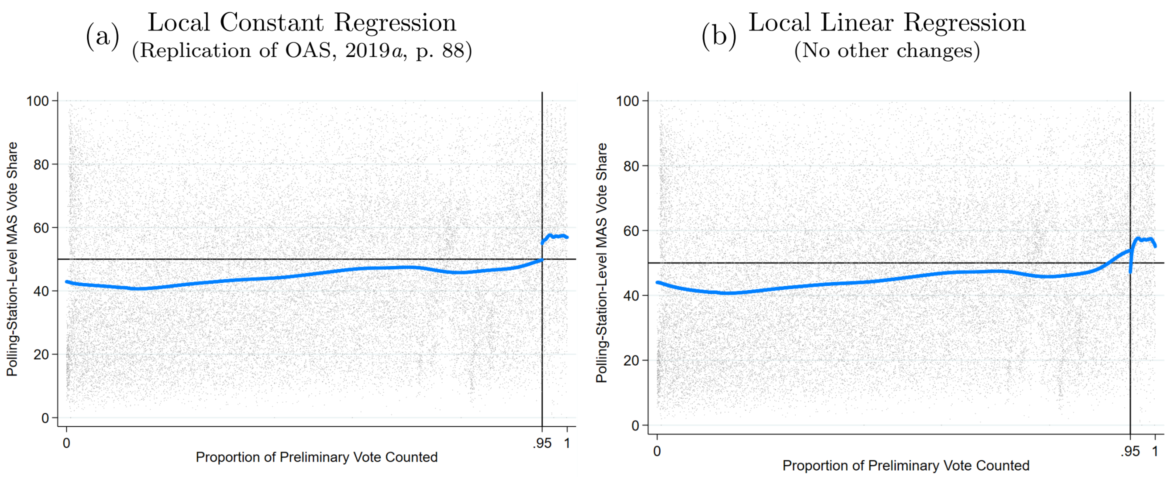
In support of the claim of "a massive and unexplainable surge in the final 5% of the count," the OAS presented Figure 2a (p. 88). But the apparent discontinuous jump in this figure—at 95% of votes verified in the preliminary system—is the artifact

[13] To be precise, the report claimed to find a discontinuous jump after 95% *of the eventually verified vote* was verified. 4.4% of votes were never verified in the preliminary results system; these are excluded from the figure in the OAS report (Nooruddin, 2020*a*) and from our replication exercise.

[14] The 95% cutoff studied in the OAS final report does not correspond either to 7:40:57 p.m., when the government stopped publishing updated results on the electoral authority website, or to 8:06:59 p.m., when the electoral authority stopped verifying tally sheets until the following day. See Appendix A for additional discussion.

Figure 2: A Methods Error and the Alleged Jump at 95%

Figure (a) reproduces OAS (2019*a*) (p. 88). Figure (b) shows that the apparent jump disappears when we simply use local linear rather than local constant regression.



The gray dots mark the underlying raw data, with the preliminary results system *verification time* on the x axis and the final (definitive results system) incumbent vote shares on the y axis, following Nooruddin (2020*b*). The lines mark lowess estimates with handpicked bandwidths, as implemented by Nooruddin (2020*b*). Again following Nooruddin (2020*b*), both figures use data from the $N = 33,038$ tally sheets verified in the preliminary results system, excluding the 4.4% of tally sheets that were never verified; see Appendices A and C for additional discussion.

of using an estimator inappropriate for regression discontinuity analysis. The OAS created the smoothed line in Figure 2a by estimating one local constant regression at each data point and connecting the predicted values.[15] One problem with this approach is that local constant regression often misrepresents the data at boundary points (that is, at the edges). This boundary bias problem is well documented: "a polynomial of order zero—a constant fit—has undesirable theoretical properties at boundary points, which is precisely where regression discontinuity estimation must occur" (Cattaneo, Idrobo and Titiunik, 2020, p. 38).[16] In Figure 2b, we instead use a local polynomial of degree one (i.e., local linear regression); this change alone is sufficient to eliminate the appearance of a jump.

The use of local constant rather than local linear regression is not the only problem with Figure 2a. For one thing, this figure excludes the 4.4% of observations that never

[15]In particular, the OAS used Stata's `lowess` function, with the `mean` option, which implements local constant regression rather than local linear regression ("running-mean smoothing" rather than "running-line least-squares smoothing," which is the default).

[16]See also Yu and Jones (1997), who conclude, "Detrimental boundary influence indeed exists when using local constant fitting in some cases, and it is this aspect which clinches the argument in favour of local linear smoothing" (p. 165); as well as Fan and Gijbels (1996), Sections 2.2.3, 3.2.5, and 3.4.2, and Imbens and Kalyanaraman (2011), p. 935.

Table 1: No Evidence of Discontinuous Changes at Two Cutoffs

We cannot reject the null hypothesis of no treatment effect at two cutoffs: 95% of the vote verified in the preliminary system, the cutoff studied in OAS (2019*a*); and 7:40:57 p.m. on election night, when the government stopped posting updated results ("the shutdown").

				Robust		Observations	
Cutoff	Date & Time	RD Estimate	BW	p-val	95% C.I.	Left	Right
0.950	10/20/2019 20:03:59	0.027	0.040	0.634	[-0.044, 0.073]	1,291	1,345
0.889	10/20/2019 19:40:57	-0.018	0.039	0.721	[-0.065, 0.045]	1,248	1,245

The running variable is *percentiles of verification time* in the $N = 33,038$ tally sheets verified in Bolivia's preliminary results system; the outcome is *incumbent (MAS) vote share*, as recorded in the definitive results system (following Nooruddin, 2020*b*). We use the non-parametric regression discontinuity estimator proposed by Calonico, Cattaneo and Titiunik (2014).

made it in to the preliminary count—contrary to the OAS's claim that "all analysis conducted below includes these additional [observations]" (OAS, 2019*a*, p. 86). When we append these observations to the end of the preliminary results data, as the OAS claimed to do, there is no discontinuity at 95% (see Appendix C). For another, the local regressions underlying Figure 2a use handpicked, arbitrary bandwidths. Moreover, the OAS presented no formal test of the null hypothesis of continuity at 95% of the preliminary count. Our simple modification in Figure 2b does not solve these problems; it merely illustrates the severe boundary bias problem created by the use of local constant regression in Figure 2a.

To estimate the size of the jump at 95% of the preliminary count, we use the data-driven regression discontinuity estimator proposed by Calonico, Cattaneo and Titiunik (2014). This approach estimates the treatment effect by running two local linear regressions precisely at the cutoff (one to the left, one to the right). We test for discontinuities at two points: (1) 95% of the preliminary count, i.e., the point studied by the OAS in Figure 2a; and (2) 7:40:57 p.m. on election night, when the electoral authority stopped publishing updated results.[17] We cannot reject the null of continuity at either of these two points (Table 1). In Appendix F, we show that this finding is robust to (a) the (random) sort order within identical time stamps, (b) polynomial degree, and (c) bandwidth.

In sum, we find no evidence of the alleged discontinuous jump in Morales's vote share at 95% of the vote counted—the "surge" to which the OAS attributed his first-round

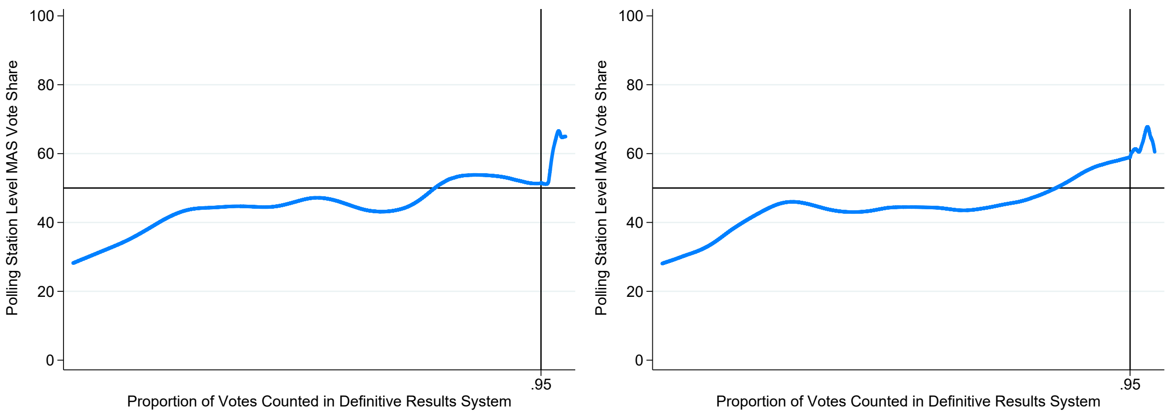
[17] The last website update occurred at 7:40:57 p.m., in advance of the 7:50 p.m. press conference mentioned above.

Figure 3: A Coding Error and the Alleged "Striking Trend" (Rosnick, 2020*b*)

Figure (a) reproduces OAS (2019*a*) (p. 92), for which time stamps were mistakenly sorted alphabetically (7:01 p.m. follows 7:01 a.m.). Figure (b) shows that the apparent "striking upward trend" disappears when time stamps are sorted chronologically, as noted in the press (Rosnick, 2020*b*).

(a) Time Stamps Sorted Alphabetically (Replication of OAS, 2019*a*, p. 92)

(b) Time Stamps Sorted Chronologically (No other changes)



The lines mark lowess estimates with handpicked bandwidths, as implemented by Nooruddin (2020b). Both lines are fit to data from the $N = 34,555$ tally sheets in Bolivia's definitive results system.

victory (OAS, 2019*a*, p. 94).

A "striking upward trend" in the last 5% of the definitive count. As additional evidence of the anomalous character of late-counted votes, the OAS also presented graphs using a separate set of time stamps (in other words, using a different x-axis): time stamps from Bolivia's *definitive* results system, or *Cómputo*. The definitive system is slower and, at least in theory, more accurate than the preliminary system; it relies on physical delivery of tally sheets to each of nine electoral authority offices, rather than tally-sheet images transmitted through the cell-phone app. The last 5% of observations in the preliminary-system verification time series are thus different from the last 5% of observations in the definitive-system time series. The OAS nevertheless presented graphs using the definitive-system time stamps as a kind of robustness check, stating that "we should analyze if the same patterns emerge if we use only the [definitive system] time stamps" (p. 91).

The OAS concluded that similar patterns *do* emerge in the analysis using the definitive-system time stamps (p. 91). In support of this statement, the OAS presented Figure 3a, in which there is a "striking upward trend" (p. 92) in Morales's vote share after 95% of votes are counted in the definitive results system. But this pattern is the artifact of a coding error. The OAS sorted the definitive-system time stamps alphabetically, such that 7:01 p.m. comes right after 7:01 a.m., rather than chronologically (as Rosnick, 2020*b*, noted, after OAS replication materials were posted in response to an earlier draft of this paper). Correcting this error eliminates the appearance of an anomalous late-breaking surge in the incumbent's vote share (Figure 3b).

Within-precinct variation as evidence of "a statistically significant case of electoral fraud." Researchers outside the OAS also pointed to late-counted votes as indicative of fraud in the Bolivian presidential election. Escobari and Hoover (2019) highlight within-precinct variation. Specifically, they note that MAS performed better in voting booths reporting after the government stopped publishing results (*post-shutdown*) than in voting booths *from the same precinct* that reported earlier (*pre-shutdown*). Escobari and Hoover view the within-precinct variation as evidence of "a statistically significant case of electoral fraud" (p. 1).

In our view, these inferences are unjustified. The analysis in Escobari and Hoover (2019) compares two periods (*pre* and *post*) without accounting for a secular trend. As noted above, Bolivian electoral administration would lead us to *expect* a within-precinct correlation between vote share and reporting time, and, indeed, we observe this correlation in the data (see Figure 1b above). We show in Appendix D that accounting for this secular trend eliminates the appearance of an anomalous within-precinct pre-post difference in vote shares. In Appendix D we also elaborate on the explanation proposed above—that education varies sufficiently across voting booths within a precinct to produce a slight within-precinct correlation between reporting time and incumbent vote share—providing additional evidence based on discussion in Escobari and Hoover (2020).

In short, the pre–post difference estimated by Escobari and Hoover (2019) is not a test statistic that can be used to evaluate the null hypothesis of no fraud. Neither the secular trend nor our proposed explanation establishes the absence of tampering with late-reporting booths in this election; rather, they imply that we do not *need* electoral manipulation in order to explain the within-precinct differences that these authors cited as evidence of foul play.

2.3 Consequences of Bolivia's Early-Count Mirage

Statements from the OAS played an important role in the evolution of Bolivia's political crisis (Crisis Group, 2020, p. 3–4). Key political actors cited the OAS in calls for new elections and for Morales's resignation. The party of runner-up Carlos Mesa, for example, summarized the OAS reports as "evidencing the violation of basic principles essential for the transparency of this electoral process and a sudden and inexplicable change of the irreversible trend towards a second round" (Comunidad Ciudadana, November 8, 2019). The opposition Committee for Santa Cruz even drafted a resignation letter for Morales and asked him to sign it; first on the Committee's list of reasons was the fact that "as the OAS delegate said, [the preliminary results transmission system] resumed with an inexplicable change in the vote trend" (CSC, November 4, 2019). Moreover, following the publication of the preliminary report of the electoral audit, OAS Secretary General called it "irrefutable," saying that "yes, there was a coup d'etat in Bolivia . . . when electoral fraud was committed," and that those who committed fraud "had blood on their hands" (EFE, 2019). U.S. Secretary of State Mike Pompeo expressed his full support for the report's findings (Pompeo, 2019).

This is all to say that fraud allegations stemming from the early-count mirage contributed to Morales's exit and to the consolidation of an interim government under a little-known conservative politician. Within one week, security forces killed twenty Morales supporters in two incidents (Human Rights Watch, 2020; Anderson, 2020); the interim government also charged dozens of former Morales government officials and allies for corruption, sedition, and "making illegal appointments" (Chauvin and Faiola, 2020; Human Rights Watch, 2020), leading to an expression of concern from U.N. human rights chief Michelle Bachelet (Europa Press, 2020). New elections were called for early 2020 but postponed twice, citing COVID. When they were finally held in October, 2020, Morales's former economy minister Luis Arce won in the first round with 55.1% of the vote, nearly doubling the vote share of the runner-up.

3 The Early-Count Mirage Beyond Bolivia: Evidence from Three Cases

The early-count mirage is not unique to Bolivia. In Mexico, the United States, and Honduras, presidential candidates, electoral observers, and/or the international press made unfounded claims of fraud based on shifts in late-counted votes. The details vary across cases, but the fraud claims suffer from common flaws. Either they fail to recognize how legitimate vote-counting processes could produce an allegedly suspicious time trend, or they fail to correctly execute the empirical analysis.

Mexico, 2006. The Mexican presidential election of 2006 was decided by less than six-tenths of one percentage point. The runner-up, Andrés Manuel López Obrador (AMLO)—who went on to win the 2018 presidential election—refused to acknowledge the result, accusing the government of fraud and even setting up a makeshift government of his own in Mexico City's central plaza. AMLO's fraud accusations stemmed from two features of the trend in vote share (Antenangeli and Cantú, 2019). First, he noted that his rival pulled ahead very early on—after just 26 voting booths had reported—and never again fell behind. AMLO and his allies viewed this as suspicious: in such a close election, they reasoned, wouldn't you expect more reversals over the course of the count? Second, AMLO questioned a nonmonotonicity in the early part of the count, asking why his rival took an initial lead and then fell behind, only to regain the advantage "in an inexplicable way" (Antenangeli and Cantú, 2019, p. 142). Specifically, AMLO and other observers accused the government of sorting the voting-booth-level tallies so as to give his rival an early boost, rather than counting them in the order in which they were received.

Antenangeli and Cantú (2019) study both accusations, finding that neither the number of reversals nor the initial nonmonotonicity were unusual under the null hypothesis of no fraud. First, using simulations, they show that, contrary to AMLO's intuition, we would *not* expect many reversals over the course of the count—even in such a close race. Second, they model the reporting time of each voting booth as a function of covariates (such as distance-to-district-office, to which tally sheets are physically delivered) and examine observations with apparently anomalous reporting times. We might think of these voting booths as the population of potentially re-ordered observations—that is, those that might provide evidence in favor of AMLO's accusation. In fact, Antenangeli and Cantú find that removing or re-ordering these voting booths is of little consequence for the vote-share trends. Which is all to say that, if the government attempted to sort voting-booth-level tallies in order to give one candidate an early lead, it failed. Taken together, these findings underscore the point that legitimate vote-counting processes produce a wide variety of time trends.

The United States, 2020. The experience of the United States in the 2020 presidential election provides yet more evidence of this variation. In North Carolina, for example, Joe Biden held a lead 90 minutes after polls closed on election day—only to fall behind by dawn the next morning, and ultimately to lose the state (Bronner, Wiederkehr and Rakich, 2020). This pattern was the predictable outcome of the state's decision to count mail-in ballots in advance (mail-in ballots typically lean Democrat; Foley, 2013; Foley and Stewart, 2020). In the critical state of Pennsylvania, in contrast, Donald Trump pulled ahead early on election night and held his lead not only the next morning but for more than two days; it took 61 hours after the poll-closing time for Biden's vote count to surpass Trump's (Bronner, Wiederkehr and Rakich, 2020). Again, this late-breaking Biden win was the predictable outcome of the fact that Pennsylvania does not begin processing mail-in ballots until election day. That fact did not stop Trump from referring to late-counted votes as "explosions of bullshit" with which the Democrats stole the election (NPR, 2021). Nor did it stop a lawsuit against Pennsylvania from claiming that, in the absence of fraud, there was only a "one in a quadrillion" chance of the state's vote-share trend occurring (Eggers, Garro and Grimmer, 2021, p. 12). As Eggers, Garro and Grimmer (2021) explain, this statement was based on the erroneous premise that Biden would capture the same *number* (not share) of early- and late-counted votes—despite the fact that less than 10% of votes are defined as "late." In other words, in order for the data to pass the lawsuit's flawed test for electoral integrity, Biden's share of the vote would have had to increase by a factor of at least nine between early- and late-counted votes (from 11% to 99%, for example). Here again, election skeptics rejected the null hypothesis of no fraud using test statistics entirely consistent with that null.

Honduras, 2017. The Honduran presidential election of 2017 provides additional examples of fraud claims likely based on coding errors *and* of fraud claims based on conceptual errors—both stemming from the early-count mirage.

The context bears several similarities to the Bolivian case. First, tensions ran high prior to election day; the right-wing incumbent had overseen a constitutional amendment to allow reelection. Second, there were reasons to doubt the incumbent's commitment to electoral integrity (among other issues, incumbent-party poll workers were apparently encouraged to facilitate double-voting for the incumbent and to nullify unfavorable ballots, among other tactics; The Economist, 2017*a*). Third, early votes favored the challenger—the day after the election, with 57% of the vote counted, he held a five-point lead—and then, after a series of unexplained delays in the count and deadly protests, electoral authorities announced that the incumbent had won by 1.3 percentage points (The Economist, 2017*b*).

The Organization of American States observed the poll, documented several irregularities, and, as in Bolivia, called for new elections (OAS, 2017). This conclusion was based in part on quantitative analysis of the time-trend in vote share. The trend was non-monotonic; the opposition candidate's cumulative margin increased through the first part of the count and then declined, ultimately falling below zero. In and of itself, the report viewed this trend as potentially innocuous—"possibly the result of [opposition]-favoring areas reporting results earlier and being counted sooner"—rather than suspicious. But the report also claimed to show that "in every department [i.e. state], the same pattern [of collapse in the opposition's margin] is evident," a fact that "raised real doubts in [their] mind" (Nooruddin, 2017, p. 4).

The apparent geographic homogeneity of the trend in vote margin made international news. The New York *Times*, for example, noted that a sharp swing away from the opposition candidate "occurred across all regions" (Malkin, 2017).

But the graphical evidence for this point, reproduced in Appendix G, appears to be the result of an error. Graphs presented as time trends specific to each of Honduras's eighteen departments (plus one graph for votes at embassies abroad) actually seem to report the *national* time trend in vote share, projected on to the department-specific time stamps. The shape and scale of the vote-margin trend are not *similar* across departments in these graphs; they are *identical*—a pattern inconsistent with the reality of department-level vote margins. Rather, these graphs appear to re-print the national vote-margin trend for each department, varying only the x-axis (the time stamps).

Confusion about the Honduran electoral returns was not limited to this apparent error or to the OAS. Prior to the OAS report, the Economist magazine analyzed municipality-level electoral returns and also found "reasons to worry" (The Economist, 2017*b*). In particular, the Economist noted that there was a late-breaking swing toward the incumbent in the majority of municipalities. Of course, as the Economist acknowledged, this could be the result of within-municipality variation in covariates such as urbanization: even within a small municipality (i.e. district), towns might report before villages. To refute this, the Economist might have shown that the most homogenous municipalities—those that are almost-all rural or almost all-urban—have swings just as large as those of more mixed municipalities (those that are, e.g., 50% urban). Instead, the Economist showed that the swing toward the incumbent is uncorrelated with the overall rurality of the municipality, a fact that has no bearing on the article's stated hypothesis.

None of this is to say that there was no fraud in the 2017 Honduran presidential election, or even to say that there is no quantitative evidence of fraud. Rather, we highlight two separate quantitative analyses that, in our view, erroneously presented late-counted votes as indicative of possible malfeasance. The Economist neglected to establish that the observed pattern was inconsistent with the null hypothesis of no fraud, while the OAS appears to have made a simple coding error. We can therefore refute these claims from Honduras with the same insight that we apply to other cases: time trends from legitimate vote-counting processes are far more varied—and errors in influential analysis far more frequent—than election skeptics allege.

4 Conclusion

Governments rarely announce election results all at once. Instead, they release partial tallies as they trickle in, telling the public how things stand with 30% of precincts reporting, 70%, 90%, and so on. These updates respond to popular demand for information. But they can also entail an important and seldom-studied cost: what we call *the early-count mirage*, or the expedient illusion that, absent fraud, an early advantage will persist.

The early-count mirage generates a tradeoff between transparency and certainty. Incremental reporting provides transparency, but waiting to announce the final result provides certainty. Dispelling the early-count mirage lowers the costs of transparency and thereby softens the government's tradeoff between these two objectives.

We establish that fraud claims tied to the early-count mirage suffer from common flaws. Either they neglect the fact that legitimate vote-counting processes can produce apparently anomalous shifts in vote share over the course of the count, or their analysis suffers from methodological and/or coding errors. Refuting these claims therefore requires describing the vote-share trends that we would expect to observe in the absence of fraud, as well as careful replication of quantitative results.

We study the Bolivian presidential election of 2019, in which fraud claims tied to the early-count mirage played an important role in reversing the outcome of the election (Crisis Group, 2020)—with large social consequences. Our analysis reveals that these influential fraud claims rested on conceptual, methodological, and coding errors. And even where fraud claims tied to the early-count mirage do not overturn an election—as in Honduras, Ecuador, Peru, Kenya, local and national elections in Mexico, and local and national elections in the United States, for example—these claims can spark conflict and erode perceptions of democracy.

The Bolivian case underscores key findings from the literature on international election observation (e.g. Donno, 2010, 2013; Hyde, 2007, 2011; Beaulieu and Hyde, 2009; Hyde and Marinov, 2014; Simpser and Donno, 2012; Bush and Prather, 2018; Kavakli and Kuhn, 2020). As in Bush and Prather (2017), third-party monitors powerfully shaped local perceptions of electoral credibility—especially those of political losers. Moreover, the controversial role of the OAS in Bolivia likely affected attitudes toward the OAS across the Americas, as well as decisions about whether and when to engage with the OAS on unrelated issues (consistent with Corstange and Marinov, 2012; Bush and Prather, 2020). On the other hand, the literature finds that intergovernmental organizations in general—and the OAS in particular—are *less* likely to question electoral integrity than nongovernmental organizations (Kelley, 2012, 2009, p. 779). In Bolivia, we find, the early-count mirage was so powerful as to overcome that hesitancy. Beyond electoral observation, our findings speak to work that connects the agendas of international organizations with those of domestic political actors, especially in the shadow of elections (e.g. Schneider and Slantchev, 2018).

Our analysis also has policy implications for electoral observers employing statistical analysis. First, an ex-ante decision rule about *when* to use quantitative data would improve transparency and consistency. Second, using established tools for election forensics (e.g. Hicken and Mebane, 2017; Alvarez, Hall and Hyde, 2009; Myagkov, Ordeshook and Shakin, 2009), with established indicators of fraud, would reduce the prevalence of invalid tests. Third, instituting an internal and/or external review process for quantitative analysis would lower the probability of errors. Of course, review processes take time, and electoral observers often work on tight deadlines. One way to navigate this tension between time and quality would be to post replication data and code together with the electoral observers' report. Contested elections generate sufficient interest from quantitative researchers that conceptual or methodological errors would likely be caught quickly and communicated to politicians, journalists, the public, and the electoral observers themselves. Timely correction could avert some of the political consequences of unfounded claims of fraud.

Researchers can protect the legitimacy of the electoral process not only by detecting fraud, but also by debunking false claims of fraud. We advance this agenda by conceptualizing the early-count mirage and articulating insights that can dispel it.

References

AFP. 2019. "Morales opponent rejects poll results in Bolivia." Agence France Presse.

Alvarez, R.M., T.E. Hall and S.D. Hyde. 2009. *Election Fraud: Detecting and Deterring Electoral Manipulation*. Brookings Series on Election Administration and Reform Brookings Institution Press.

Anderson, C.J., S. Bowler, A. Blais, T. Donovan and O. Listhaug. 2007. *Losers' Consent: Elections and Democratic Legitimacy*. COMPARATIVE POLITICS SERIES CEP Non-Continuation Series Oxford University Press.
URL: *https://books.google.com/books?id=4RJVAAAAYAAJ*

Anderson, Jon Lee. 2020. "The Fall of Evo Morales." The New Yorker.

ANF. 2019. "Elecciones 2019: 14 mil encuestados perfilan apretada victoria de Mesa sobre Evo en balotaje." Agencia de Noticias Fides.

Anria, Santiago. 2018. *When Movements Become Parties: The Bolivian MAS in Comparative Perspective*. Cambridge Studies in Comparative Politics Cambridge University Press.

Anria, Santiago and Jennifer Cyr. 2019. "Is Bolivia's democracy in danger? Here's what's behind the disputed presidential election." Washington Post.

Antenangeli, Leonardo and Francisco Cantú. 2019. "Right on Time: An Electoral Audit for the Publication of Vote Results." *Statistics, Politics and Policy* 10(2):137–186.

Beaulieu, Emily and Susan D. Hyde. 2009. "In the Shadow of Democracy Promotion: Strategic Manipulation, International Observers, and Election Boycotts." *Comparative Political Studies* 42(3):392–415.

Bronner, Laura, Anna Wiederkehr and Nathaniel Rakich. 2020. "What Blue And Red 'Shifts' Looked Like In Every State." FiveThirtyEight.

Bush, Sarah Sunn and Lauren Prather. 2017. "The Promise and Limits of Election Observers in Building Election Credibility." *The Journal of Politics* 79(3):921–935.

Bush, Sarah Sunn and Lauren Prather. 2018. “Who’s There? Election Observer Identity and the Local Credibility of Elections.” *International Organization* 72:659–92.

Bush, Sarah Sunn and Lauren Prather. 2020. “Foreign meddling and mass attitudes toward international economic engagement.” *International Organization* 74(3):584–609.

Calonico, Sebastian, Matias D. Cattaneo and Max H. Farrell. 2018. “On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference.” *Journal of the American Statistical Association* 113(522):767–779.

Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2014. “Robust nonparametric confidence intervals for regression-discontinuity designs.” *Econometrica* 82(6):2295–2326.

Calonico, Sebastian, Matias D Cattaneo and Rocio Titiunik. 2015. “Optimal data-driven regression discontinuity plots.” *Journal of the American Statistical Association* 110(512):1753–1769.

Calvo, Ernesto, Silvio Waisbord, Tiago Ventura and Natalia Aruguete. 2020. “Winning! Electoral Adjudication and Dialogue in Social Media.” Working Paper.

Cambara Ferrufino, Pablo Cesar. 2019. “Costas atribuye a una ‘impericia’ la paralización de la transmisión de votos del TREP.” El Deber.

Cantú, Francisco and Omar García-Ponce. 2015. “Partisan losers’ effects: Perceptions of electoral integrity in Mexico.” *Electoral Studies* 39:1–14.

Cattaneo, Matias D., Nicolás Idrobo and Rocío Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell and Yingjie Feng. 2019. “On Binscatter.” *arXiv* .

Challú, Cristian, Enrique Seira and Alberto Simpser. 2020. “The Quality of Vote Tallies: Causes and Consequences.” *American Political Science Review* 114(4):1071–1085.

Chauvin, Lucien and Anthony Faiola. 2020. "As the U.S.-backed government in Bolivia unleashes a wave of political persecution, the Trump administration remains silent." The Washington Post.

CNE Honduras. 2017. "Resultados del Proceso Electoral General 2017." Consejo Nacional Electoral de Honduras.

Collyns, Dan. 2019. "Evo Morales agrees to new elections after irregularities found." The Guardian.

Comunidad Ciudadana. November 8, 2019. "Carta abierta al Vicepresidente Álvaro García Linera." Twitter.

Corrales, Javier. 2016. "Can Anyone Stop the President? Power Asymmetries and Term Limits in Latin America, 1984–2016." *Latin American Politics and Society* .

Corstange, Daniel and Nikolay Marinov. 2012. "Taking sides in other people's elections: The polarizing effect of foreign intervention." *American Journal of Political Science* 56(3):655–670.

Cottrell, David, Michael Herron and Sean Westwood. 2018. "An exploration of Donald Trump's allegations of massive voter fraud in the 2016 General Election." *Electoral Studies* 51:123–142.

Crisis Group, International. 2020. "Bolivia Faces New Polls in Shadow of Fraud Row." Crisis Group Latin America Briefing N°43.

CSC. November 4, 2019. "Carta al Presidente Evo Morales." Comité pro Santa Cruz.

Curiel, John, Charles Stewart III and Jack Williams. 2021. "One Shift, Two Shifts, Red Shift, Blue Shift: Reported Election Returns in the 2020 Election." SSRN.

Donno, Daniela. 2010. "Who Is Punished? Regional Intergovernmental Organizations and the Enforcement of Democratic Norms." *International Organization* 64:593–625.

Donno, Daniela. 2013. *Defending Democratic Norms: International Actors and the Politics of Misconduct*. Oxford University Press.

EFE. 2019. "Evo fue el que cometió "golpe de Estado", dice Luis Almagro." El Milenio.

Eggers, Andrew C., Haritz Garro and Justin Grimmer. 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).
URL: *https://www.pnas.org/content/118/45/e2103619118*

El País. 2006. "Más de un millón de personas apoyan a López Obrador en la Plaza del Zócalo en México." El País.

Eleitoral, Tribunal Superior. 2018. "Repositório de dados eleitorais." https://www.tse.jus.br/.

Escobari, Diego and Gary Hoover. 2019. "Evo Morales and Electoral Fraud in Bolivia: A Natural Experiment Estimate." Working Paper.

Escobari, Diego and Gary Hoover. 2020. "Evo Morales and Electoral Fraud in Bolivia: A Natural Experiment and Discontinuity Evidence." Working paper.

Europa Press. 2020. "Bachelet expone la preocupación de la ONU por el "procesamiento" de exfuncionarios del Gobierno de Morales." Notimérica.

Exeni Rodríguez, José Luis. 2020. "Interview with José Luis Exeni Rodríguez." *Former Vice President of the Tribunal Supremo Electoral* .

Falleti, Tulia and Emilio Parrado. 2018. *Latin America Since the Left Turn*. Univeristy of Pennsylvania Press.

Fan, J. and I. Gijbels. 1996. *Local Polynomial Modelling and its Applications*. CRC Press.

Foley, Edward B. 2013. "A big blue shift: Measuring an asymmetrically increasing margin of litigation." *Journal of Law and Politics* .

Foley, Edward B. and Charles Stewart. 2020. "Explaining the Blue Shift in Election Canvassing." Working Paper.

Forebears.io. 2020. "Most Common Last Names in Bolivia." Forebears 2012–2020.

Fujiwara, Thomas. 2015. "Voting technology, political responsiveness, and infant health: evidence from Brazil." *Econometrica* 83(2):423–464.

Garrido, Sebastián. 2021. "Porcentaje acumulado de actas capturadas y contabilizadas en los PREPs de la elección de diputaciones federales de 2006 a 2021." Twitter.

Goel, Sharad, Marc Meredith, Michael Morse, David Rothschild and Houshmand Shirani-Mehr. 2020. "One Person, One Vote: Estimating the Prevalence of Double Voting in U.S. Presidential Elections." *American Political Science Review* pp. 1–14.

Hicken, Allen and Walter Mebane. 2017. "A Guide to Elections Forensics: Research and Innovation Grants Working Papers Series." USAID.

Human Rights Watch. 2020. "La justicia como arma: Persecución política en Bolivia." HRW.org.

Hyde, Susan D. 2007. "The Observer Effect in International Politics: Evidence from a Natural Experiment." *World Politics* 60(1):37–63.

Hyde, Susan D. 2011. *The Pseudo-Democrat's Dilemma: Why Election Observation Became an International Norm*. Cornell University Press.

Hyde, Susan D. and Nikolay Marinov. 2014. "Information and Self-Enforcing Democracy: The Role of International Election Observation." *International Organization* 68(2):329–359.

Imbens, Guido and Karthik Kalyanaraman. 2011. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79(3):933–959.

Infobae. 2021. "Llegan a Perú las actas electorales de los votos en el exterior, claves para definir el ajustado ballotage entre Keiko Fujimori y Pedro Castillo." Infobae.

Johnston, Jake and David Rosnick. 2020. "Observing the Observers: The OAS in the 2019 Bolivian Elections." Center for Economic and Policy Research.

Kavakli, Kerim Can and Patrick M. Kuhn. 2020. "Dangerous Contenders: Election Monitors, Islamic Opposition Parties, and Terrorism." *International Organization* 74(1):145–164.

Kelley, Judith. 2009. "D-Minus Elections: The Politics and Norms of International Election Observation." *International Organization* 63:765–87.

Kelley, Judith. 2012. *Monitoring Democracy: When International Election Observation Works, and Why it Often Fails.* Princeton University Press.

Klein, Herbert S. 2011. *A Concise History of Bolivia.* Cambridge Concise Histories 2 ed. Cambridge University Press.

Kurmanaev, Anatoly, Mónica Machicao and Ernesto Londoño. 2019. "Military Calls on President to Step Down After Election Dispute in Bolivia." The New York Times.

Kurmaneav, Anatoly and Cesar Del Castillo. 2019. "How and Unknown Female Senator Came to Replace the Bolivian President Evo Morales." The New York Times.

Li, Yimeng, Michelle Hyun and R. Michael Alvarez. 2021. "Why Do Election Results Change After Election Day? The "Blue Shift" in California Elections." *Political Research Quarterly* Forthcoming.

Little, Andrew T, Joshua A Tucker and Tom LaGatta. 2015. "Elections, protest, and alternation of power." *The Journal of Politics* 77(4):1142–1156.

Los Tiempos. 2019*a*. "El sistema de Transmisión de Resultados Electorales Preliminares (TREP) se actualizó al 94,99una diferencia de 10 puntos entre los candidatos del MAS y CC." Twitter.

Los Tiempos. 2019*b*. "TSE dice que se paralizó el TREP para "evitar la confusión" con resultados de departamentos." Los Tiempos.

Luo, Zhaotian and Arturas Rozenas. 2018. "The Election Monitor's Curse." *American Journal of Political Science* 62(1):148–160.

Madrid, Raúl. 2012. *The Rise of Ethnic Politics in Latin America.* Cambridge University Press.

Malkin, Elizabeth. 2017. "Honduran President Declared Winner, but O.A.S. Calls for New Election." The New York Times.

Martinez, Eugenio. 2021. "Personal communication with Eugenio Martinez.".

Mebane, Walter. 2019. "Evidence Against Fraudulent Votes Being Decisive in the Bolivia 2019 Election." Working Paper.

Mesa, Carlos D. 2019. "Post-election thread." Twitter.

Minoldo, Sol and Rodrigo Quiroga. 2020. "¿Hubo fraude en las elecciones de Bolivia?" Cenital.

Myagkov, M., P.C. Ordeshook and D. Shakin. 2009. *The Forensics of Election Fraud: Russia and Ukraine*. Cambridge University Press.
URL: *https://books.google.com/books?id=Ffkjl4Toa68C*

NEOTEC. October 28, 2019. "Informe Final Elecciones Generales 2019." NEOTEC.

Newman, John. 2020. "The OAS Conclusions about the election integrity of the Bolivian election are correct." Working Paper.

Nodal. 2021. "Yaku Pérez insiste en que hubo fraude y afirma que no apoyará a Guillermo Lasso." Noticias de América Latina y el Caribe.

Nooruddin, Irfan. 2017. "Analysis for the Organization of American States." Organization for American States (OAS).

Nooruddin, Irfan. 2020*a*. "Comment on the 2019 Bolivia Presidential Election and OAS Statistical Analysis.".

Nooruddin, Irfan. 2020*b*. "Replication Data for: OAS Audit of Bolivia General Election 2019." Harvard Dataverse.

Nooruddin, Irfan. 2020*c*. "Yes, Bolivia's 2019 election was problematic. Here's why." Washington Post Monkey Cage.

NPR. 2021. "Read Trump's Jan. 6 Speech, A Key Part Of Impeachment Trial." National Public Radio.

OAS. 2016*a*. "Informe al Consejo Permanente, Misión de Observación Electoral, Referendo Constitucional." *Organization of American States* .

OAS. 2016*b*. "Informe Preliminar: Misión OEA felicita la alta participación en el referendo constitucional de Bolivia." *Organization of American States* .

OAS. 2017. "Statement by the OAS General Secretariat on the Elections in Honduras." Organization for American States (OAS).

OAS. 2019*a*. "Electoral Integrity Analysis: General Elections in the Plurinational State of Bolivia, October 20, 2019, FINAL REPORT." Organization for American States (OAS), Secretariat for Strenthening Democracy (SSD), Department of Electoral Cooperation and Observation (DECO).

OAS. November 10, 2019*b*. "Electoral integrity analysis: General Elections in the Plurinational State of Bolivia, PRELIMINARY FINDINGS." *Organization of American States* .

OAS. October 21, 2019*c*. "Statement of the OAS Electoral Observation Mission in Bolivia." *Organization of American States* .

Páez Benalcázar, Andrés. 2017. "Los ecuatorianos tienen derecho al recuento." The New York Times.

Pompeo, Mike. 2019. "Tweet in support of the OAS report." Twitter.

Przeworski, A. 1991. *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America.* Studies in Rationality and Social Change Cambridge University Press.

Registraduría de Colombia. 2018. "Elecciones 2018." https://www.registraduria.gov.co/.

Robinson, P. M. 1988. "Root-N-Consistent Semiparametric Regression." *Econometrica* 56(4):931–954.

Rodriguez-Larralde, Alvaro, José Dipierri, Emma Alfaro Gomez, Chiara Scapoli, Elisabetta Mamolini, Germano Salvatorelli, Sonia De Lorenzi, Alberto Carrieri and Italo Barrai. 2011. "Surnames in Bolivia: A study of the population of Bolivia through isonymy." *American Journal of Physical Anthropology* 144(2):177–184.

Rosnick, David. 2020*a*. "The Ends Don't Justify the Means: Yet Another Analysis Fails to Support the OAS' Delegitimization of Bolivian Elections." Center for Economic and Policy Research.

Rosnick, David. 2020*b*. "The OAS Accusation of Electoral Fraud Against Evo Morales Is Bullshit — And Now We Have the Data to Prove It." *Jacobin* .

Schneider, Christina J and Branislav L Slantchev. 2018. "The domestic politics of international cooperation: Germany and the European debt crisis." *International Organization* 72(1):1–31.

Simpser, Alberto and Daniela Donno. 2012. "Can International Election Monitoring Harm Governance?" *The Journal of Politics* 74(2):501–513.

Spencer, Douglas M and Zachary S Markovits. 2010. "Long lines at polling stations? Observations from an election day field study." *Election Law Journal* 9(1):3–17.

The Economist. 2017*a*. "Is Honduras's ruling party planning to rig an election?" The Economist.

The Economist. 2017*b*. "What the data say about the integrity of Honduras's election." The Economist.

Tucker, Joshua A. 2007. "Enough! Electoral Fraud, Collective Action Problems, and Post-Communist Colored Revolutions." *Perspectives on Politics* 5(3):535–551.

UNICEF. 2014. "El derecho a la educación y a la protección en Bolivia." UNICEF.

Williams, Jack and John Curiel. 2020. "Analysis of the Quick Count in the 2019 Bolivia Election." Working Paper.

Yu, Keming and M.C. Jones. 1997. "A comparison of local constant and local linear regression quantile estimators." *Computational Statistics & Data Analysis* 25(2):159 – 166.

Ziemann, Mark, Yotam Eren and Assam El-Osta. 2016. "Gene name errors are widespread in the scientific literature." *Genome biology* 17(1):1–3.

Appendix

A Why 95%?

Section 2.2 of this paper revisits the OAS's claim that the incumbent's vote share jumped discontinuously at 95% of the count,[18] and that this alleged jump is suggestive of some form of electoral manipulation (OAS, 2019*a*, p. 88).

This appendix considers why the OAS selected 95% as the focal cutoff in their final audit report, despite referring to 95% as "an arbitrary point" (OAS, 2019*a*, p. 10).

As noted in Section 2.1 of our main text, verification of tally sheets in the preliminary results system halted at 8:06:59 p.m. on election night. This can be seen in Figure A.1a, which plots the density both of *transmission time* and of *verification time* for tally sheets in the preliminary results system.[19] Transmission occurred almost entirely between 4:30 p.m. and 9:00 p.m. on election day and continued through the night;[20] in contrast, verification paused abruptly at 8:06:59 p.m. on election day, to continue the following morning (hence the second spike in the density of verification time in Figure A.1a). This is perhaps even more apparent in Figure A.1b, which plots the cumulative distributions of transmission time and of verification time.

The *preliminary* audit report of the OAS (OAS, November 10, 2019*b*) highlighted this overnight pause in the verification of tally sheets, presenting a graph similar to Figure A.1a. The preliminary report then noted—correctly—that the first tally sheets verified on October 21 (the day after the election) are significantly more pro-incumbent than the last tally sheets verified on election night (i.e., those verified right before 8:06:59 p.m.). In other words, in a graph of incumbent vote share against *percentile* of the vote verified in the preliminary system, there *does* appear to be a discontinuous jump in vote share at the point corresponding to 8:06:59 p.m. (see Figure A.3)—and a formal test rejects the null of continuity at this point. As we explain in Section 2.1 of the main text, there is an innocuous reason for this: verifiers view tally-sheet images drawn randomly from the pool of tally sheet images transmitted thus far. In other words, when the verification process resumed on the morning after the election, tally

[18] Specifically, at 95% of votes verified in the preliminary results system (TREP), which excludes the 4.4% of votes that were never verified in the preliminary system.

[19] 34,178 tally sheets have transmission times in the preliminary results system and 33,044 tally sheets were verified in the preliminary results system (there are 34,555 tally sheets total; 377 have missing transmission times, and 1,511 sheets were never verified).

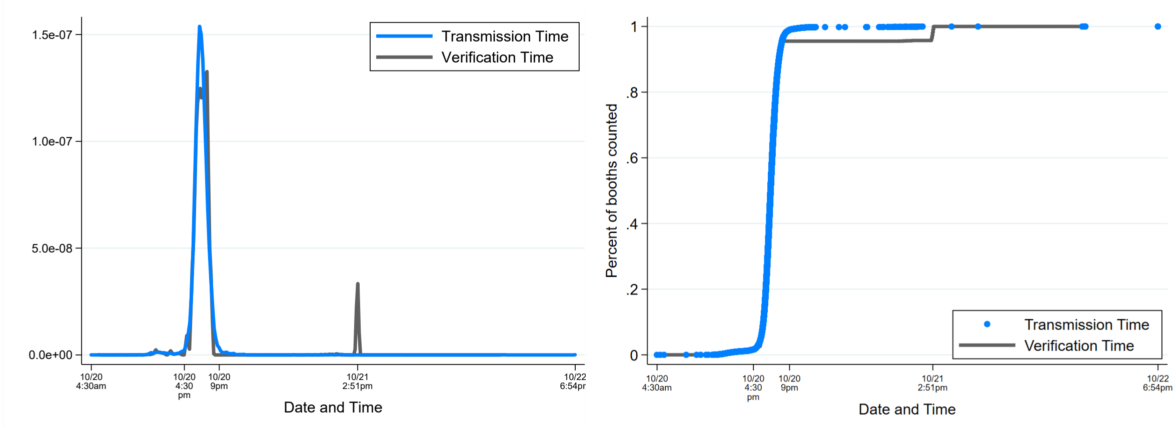
[20] As a side note, we observe that the Bolivian preliminary results system is extremely fast. For comparison, the Mexican preliminary results system typically collects a mere 50% of voting-booth tallies by midnight on election day (Garrido, 2021).

Figure A.1: Tally-Sheet Verification Paused Overnight

Fig. (a) plots the densities of tally-sheet transmission time (i.e. reporting time, in blue) and of verification time (in gray). Transmission of tally sheets continued overnight, while verification stopped abruptly at 8:06:59 p.m. and did not continue until the following morning. This pause is also visible in the CDFs, plotted in Fig. (b).

(a) Distribution of Transmission and Verification Times

(b) CDF of Transmission and Verification Times



The *transmission time* distributions are fit to the $N = 34,123$ tally sheets transmitted to the preliminary results system.

sheets were drawn randomly from all those transmitted after 8:06:59 p.m. the previous evening (and from those that were transmitted but not verified before 8:06:59 p.m.). To see this, consider Figure A.2a: among tally sheets verified on October 21, the day after the election, there is *no correlation* between transmission time and verification time. We would thus *expect* the first tally sheets verified on October 21 to look different from the last tally sheets verified on October 20 (we elaborate this point in Footnote 12 in the main text).

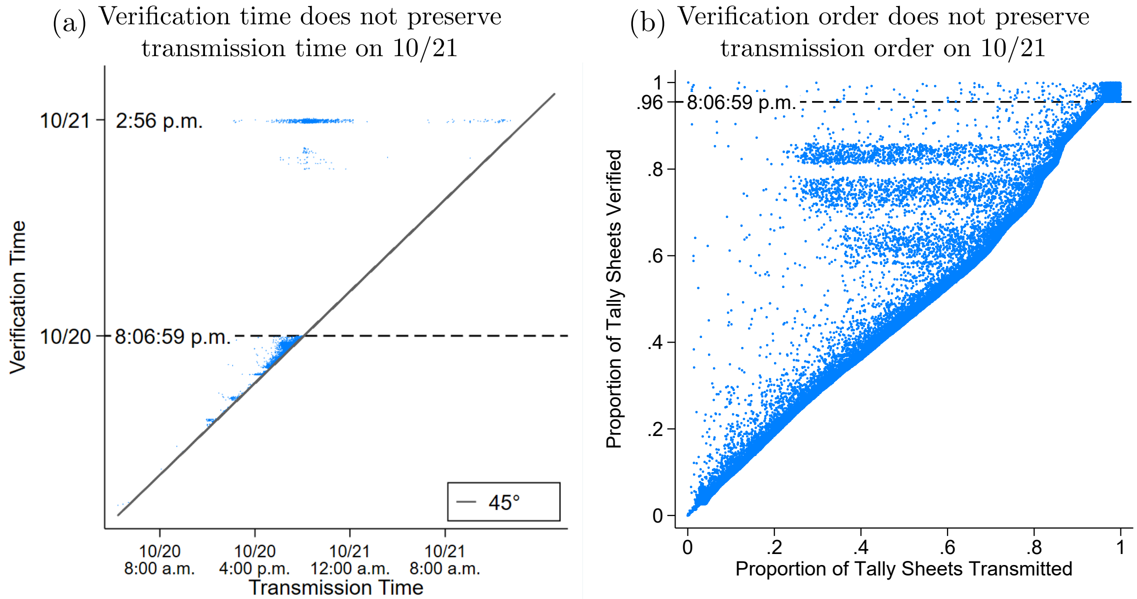
In sum, the preliminary audit report of the OAS correctly observed a discontinuous jump at this cutoff, but incorrectly considered it anomalous rather than predictable.

The *final* audit report of the OAS, in contrast, studied a different cutoff: rather than 8:06:59 p.m., which corresponds to 95**.6**% of the vote verified in the preliminary system, the final audit report chose 8:03:59 p.m., which corresponds to 95**.0**% of the vote verified.[21] We do not know why. The preliminary report did refer to the original

[21] It is also the case that the verification time series used in the OAS analysis differs subtly from the true *votes counted* time series. When the first attempt to verify a tally sheet failed—because the vote totals transcribed by verifiers at a central location did not match the totals transcribed

Figure A.2: Any Pause in Verification Scrambles the Tally-Sheet Sort Order

Fig. (a) plots tally-sheet *verification* time against transmission time, revealing that, when verification resumed on the day after the election, tally sheets were verified in a random order. This is also apparent in Fig. (b), which plots verification order against transmission order. Indeed, any pause or delay in the verification process scrambles the tally-sheet sort order, because verifiers view tally sheets drawn randomly from all those submitted thus far.



The scatter plots include the $N = 33,038$ tally sheets verified in the preliminary results system. See Footnote 19 for additional discussion.

cutoff as "a threshold that represents *about the last 5% of the cumulative vote count*" (OAS, November 10, 2019*b*, p. 9, emphasis added); perhaps the final report then adopted this "last 5%" cutoff without heeding the "about" qualifier in the preliminary report. Paradoxically, correcting this imprecision would have resulted in the discovery of an actual (rather than artificial) discontinuous jump at the cutoff—albeit for the innocuous reason summarized above.

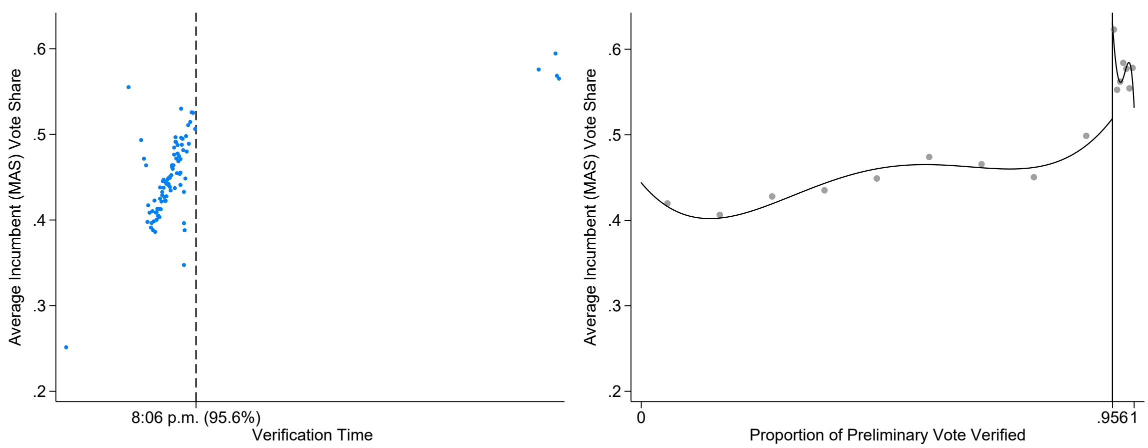
Figure A.3 also highlights a potential problem with testing for discontinuities using *percentiles* of verification time as the running variable. In these data, *percentile* is

into the cell-phone app at each precinct—the time stamp of that *failed* verification nevertheless appears in the *verification date* column. If/when that tally sheet is later approved with corrected vote totals, the approval time stamp appears in a column called *approval date*. The latter time stamp corresponds to the moment when the tally sheet was added to the count; a correct "counted" time series would thus use *approval* dates in place of *failed verification* dates whenever possible. The OAS instead used the series that includes failed verification times; we do the same in order to replicate the OAS analysis.

Figure A.3: A Predictable Jump in Vote Share after 8:06:59 p.m. (95.**6**%)

Figure (a) plots average incumbent vote share in bins of the preliminary results system verification times. Figure (b) instead plots average incumbent vote share in bins of *percentiles* of the preliminary results system transmission times, marking 8:06:59 p.m. This transformation of the x axis visualizes a discontinuous change in vote share at a moment when the underlying running variable—verification time—is itself discontinuous.

(a) Vote share v. verification *time* (b) Vote share v. verification *percentile*



Both figures use optimal (data-driven) bins (Cattaneo et al., 2019) fit to data from the $N = 33,038$ tally sheets verified in Bolivia's preliminary results system. The black line in (b) marks a fourth-order polynomial fit on each side of the cutoff (Calonico, Cattaneo and Titiunik, 2015).

a transformation of verification time that places the 8:06:59 p.m. time stamps right next to the 10:37 a.m. (next day) time stamps, closing the long gap in the actual time series. Thus testing for a discontinuity at 95.6% of the vote verified implies testing for a discontinuous change in incumbent vote share at a moment when the running variable (time) is itself discontinuous.

In sum, the preliminary and final audit reports of the OAS, respectively, commit the two types of errors outlined in Section 1 of this paper. Again adopting the language of Eggers, Garro and Grimmer (2021), the preliminary audit report presents an allegedly anomalous fact—a discontinuity at 95.6%—that is a fact but is not anomalous; the final audit report presents an allegedly anomalous fact—a discontinuity at 95%—that is not a fact.

B Semiparametric results

Section 2.1 in the main text describes the features of the Bolivian vote-counting system that would lead us to *expect* the incumbent's margin to increase over time. To review:

1. Voting-booth jurors are randomly selected from voters registered *at that booth*.
2. Jurors are responsible for counting ballots and filling out a paper tally sheet, among other tasks.
3. The speed with which jurors complete these tasks is likely correlated with their education level. The distribution of reporting time is so compressed that even a ten-minute delay is significant.
4. Education is correlated with political preferences.
5. We would therefore expect a time-trend in vote shares.

If this hypothesis is correct, we would expect that controlling for observable precinct characteristics would reduce the slope of the time trend in vote margin. In the main text, we investigate this hypothesis by estimating an equation in which precinct characteristics enter linearly and $Time$ enters flexibly (Equation 1). The resulting estimate of $f\left(\text{Time}_{bpm}\right)$, plotted in Figure 1a, is approximately linear. This allows us to estimate a fully parametric version of Eq. 1 in which $f\left(\text{Time}_{bpm}\right) = \beta_1 \text{Time}_{bpm}$, i.e.,

$$\text{Margin}_{bpm} = \beta_1 \text{Time}_{bpm} + \boldsymbol{\beta X}_{bpm} + \epsilon_{bpm} \tag{2}$$

We standardize the time variable so that a one-unit increase corresponds to a one-standard-deviation increase (approximately 45 minutes). Table B.1 reports the results. When we omit any controls, $\widehat{\beta}_1 \approx 0.074$, which is to say that Morales's average margin (not cumulative margin) increases by 7.4 percentage points every 45 minutes (Column 1). Including one indicator for all precincts in the lowland departments Pando, Beni, Santa Cruz, and Tarija reduces this slope by more than 40%, to 4.3 percentage points (Column 2). Crude proxies for education (at the municipality level) and rurality further reduce the slope, to 3.7 percentage points (Column 3). This is remarkable given the relative parsimony of Equation 2.

We also consider the previous poll, in 2016, when voters defeated Morales's proposed

Table B.1: Education, Region, and Rurality Explain Most of the Time Trend

Estimates of a fully parametric version of Equation 1, in which $f(\text{Time}) = \beta_1\text{Time}$. We standardize time so that a one-unit increase corresponds to one standard deviation, or $\approx$ 45 minutes. Column (1) presents the bivariate specification; Columns (2)–(5) sequentially add controls.

	(1) No controls	(2) +$\mathbb{1}$(Lowlands)	(3) +Rural, Education	(4) +2016	(5) +Department +2016×Dep't
Transmission Time	0.074	0.043	0.037	0.019	0.008
	(0.035)	(0.024)	(0.013)	(0.003)	(0.002)
Observations	31,724	31,724	31,599	30,932	30,932
R-squared	0.028	0.113	0.533	0.888	0.896

Standard errors, clustered at the municipality level, in parentheses. Column (2) includes an indicator for the lowland departments of Pando, Beni, Santa Cruz, and Tarija; Column (3) adds measures of education and a vector of proxies for rurality; Column (4) adds average precinct MAS margin in the previous election (2016); Column (5) adds indicators for Bolivia's nine departments as well as interactions between each department indicator and the 2016 margin.

constitutional amendments in a yes-or-no referendum. That result was not contested; indeed, the OAS electoral observation mission made no reference to malpractice in its reports (though in 2016 the OAS did not conduct an audit, as it did in 2019) (OAS, 2016*a*,*b*). When we control for the mean margin in each precinct in 2016 (we cannot match voting booths across elections), our estimate of β_1 falls further, to 0.02 (Column 4).

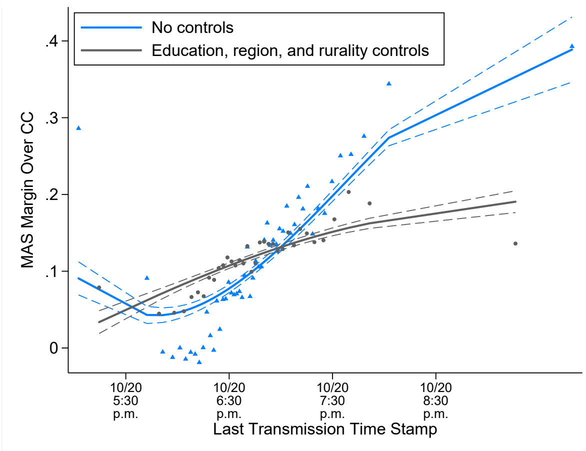
When we further include indicators for Bolivia's nine departments and interactions between those indicators and the 2016 precinct-level vote share (Column 5), β_1 falls an additional 55%, to 0.008, which is to say that every forty-five minutes Morales's average margin increases by less than one percentage point more than we would expect given a minimal set of controls (none of which are at the level of the *voting booth*). We can think of this as an upper bound on the contribution of unobserved factors to the time-trend in vote margin; again, these unobserved factors include (a) all voting-booth-level characteristics (b) all precinct-level characteristics except 2016 vote margin, among many others.

Figure B.1 establishes the robustness of our semiparametric estimate to various specifications discussed in the main text.

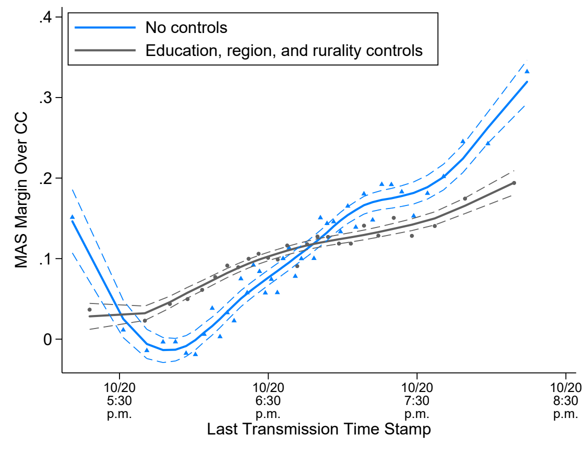
Figure B.1: Additional Semiparametric Results

This figure shows the robustness of our semiparametric results to (a) including all time outliers in the sample; (b) dropping the observations without time stamps in the preliminary results system; (c) treating them as early reporters in their respective municipalities; and (d) treating them as late (maximum) reporters in their respective municipalities. In Figure (e), we show the overall bivariate time trend in the full sample of voting booths (the other figures are restricted to the sample for which we have covariates; this excludes voting booths abroad).

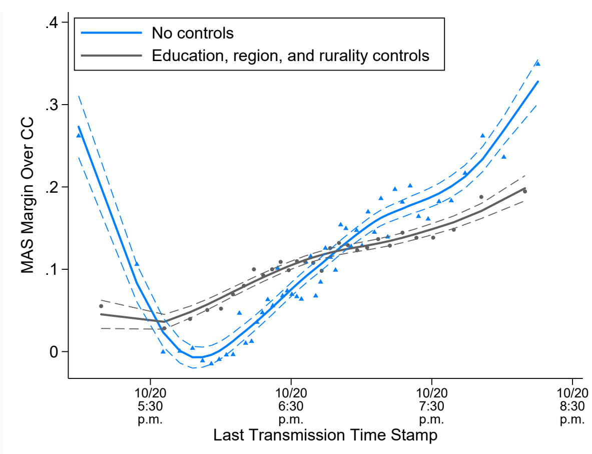
(a) No trimming



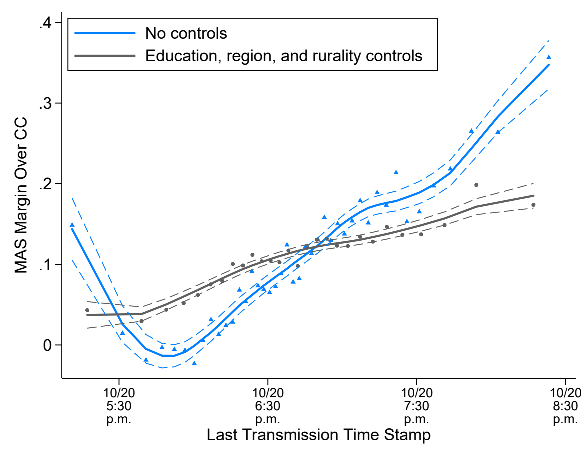
(b) Dropping missings



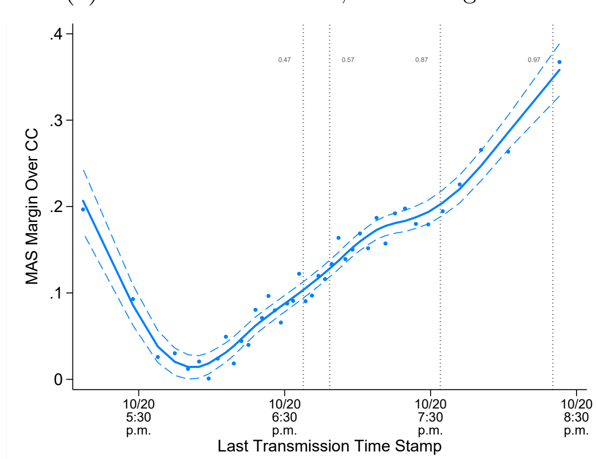
(c) Replacing at the minimum



(d) Replacing at the maximum



(e) Overall Time Trend, All Voting Booths



Points mark the average MAS (incumbent) margin over Civic Community (runner-up) in optimal (data-driven) bins of the transmission time (Cattaneo et al., 2019), using data from Bolivia's $N = 34,555$ tally sheets. The solid blue fits mark a local-polynomial estimation that follows Calonico, Cattaneo and Farrell (2018). The grey lines mark an estimate of $f(\text{Time})$ from Equation 1, using the semi-parametric estimator proposed in Robinson (1988).

C The 4.4% of observations excluded from the preliminary results

The analysis in OAS (2019*a*) focuses on a data set that merges the preliminary results system (TREP) *verification* time stamps with the definitive-system (*cómputo*) vote tallies, at the level of the voting booth (our replication does the same).

Using the preliminary-system *verification* time stamps entails a challenge: how to treat the set of voting booths whose tallies were never verified in the preliminary system (TREP). These 1,513 voting booths account for 4.4% of all observations, and they are excluded from the preliminary system for diverse reasons (including lack of cellular service and tally sheet illegibility). Regardless, their verification stamps are unobserved (perhaps even undefined).

The text of the OAS audit report claims to treat these voting booths as "late reporters" (p. 86), under the assumption that they finished tallying only after the preliminary results system closed. The report states: "All the analysis conducted below include these additional polling stations. Since they were not included in the TREP [preliminary system], they are treated as being late reporters" (p. 86). We interpreted this to mean that OAS (2019*a*) sorted the first 33,038 booths by their preliminary results system time stamps, and then appended the remaining 1,513 voting booths (4.4%) at the end, presumably in a random order.

In the first draft of this paper, we alleged that, rather than append the "late-reporting" voting booths to the end of the data set as claimed, the OAS dropped them when creating Figure 2a. This is a consequential exclusion. The "late-reporters" account for 4.4% of tally sheets and 4.1% of votes, which is to say, the vast majority (82%) of the last 5% of votes counted (if we assume, as the OAS does, that they were late reporters). Any analysis focused on the last 5% of votes counted will therefore be quite sensitive to the treatment of the booths without preliminary results system time stamps.

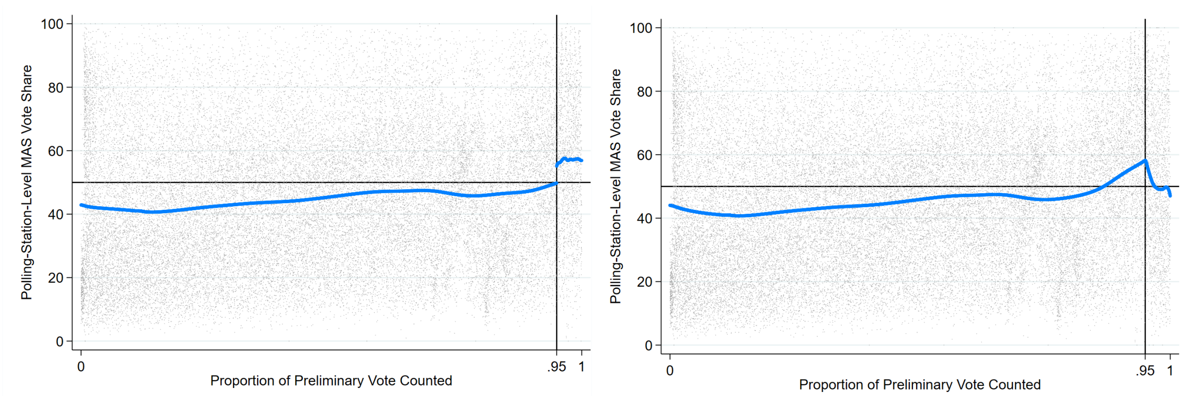
The OAS replication materials (Nooruddin, 2020*b*), posted in response to our earlier draft, confirm that the "late-reporting" voting booths were in fact dropped in creating Figure 2a (which, recall, was the graph presented as evidence of a discontinuous jump in MAS's vote share). If we include the "late reporters" at the end, as the OAS audit report claimed to do, we obtain Figure C.1b. In this case, there is neither a jump nor an uptick in the trend of MAS's vote share in the final 5% of the count.

Figure C.1: Exclusion of "Late Reporters" and the Jump at 95%

Figure (a) reproduces OAS (2019*a*) (p. 88). Figure (b) shows that the apparent jump disappears when we append the observations without preliminary-system time stamps.

(a) Dropping "Late Reporters" (Replication of OAS, 2019*a*, p. 88)

(b) Including "Late Reporters" (No other changes)



The gray dots mark the underlying raw data, with the preliminary results system *verification time* on the x axis and the final (definitive results system) incumbent vote shares on the y axis, following Nooruddin (2020*b*). The lines mark nonparametric fits using the tricube weighting function and the bandwidths handpicked in Nooruddin (2020*b*), namely, 0.3 to the left of the cutoff and 0.6 to the right of the cutoff. Figure (a) uses only the $N = 33,038$ tally sheets verified in the preliminary results system, excluding the 4.4% of tally sheets that were never verified; Figure (b) appends these tally sheets at the end, in a random order.

In his response to our earlier draft, Nooruddin (2020*a*) argued that the OAS audit report never *claimed* to include the "late-reporting" voting booths in this key results figure. We maintain that the language of the report implies otherwise. Regardless, excluding the "late-reporters" from the key results figure in an analysis of late-counted votes strikes us as unfortunate—whether by choice or by mistake.

D More on the within-precinct trend

Documenting the trend. Before studying within-precinct variation in vote shares, we note that the within-precinct variation in reporting time is substantial. 70% of precincts have more than one voting booth; among these precincts, the median within-precinct standard deviation in reporting time is 35 minutes—more than one fourth of the active reporting window (see Appendix Figure A.2). Moreover, 26% of precincts—and 37% of precincts with more than one voting booth—contain booths reporting before *and* after the public information blackout.

Figure D.1a presents an example of within-precinct variation; the blue diamonds mark MAS's margin in each of the 40 voting booths in a single precinct in the town of Llallagua, Potosí. In this example, MAS's margin increases with verification time even before the government stopped publishing updated results (at 7:40 p.m.).[22] This is not an isolated case. Let m_{bp} denote MAS's margin in voting booth b in precinct p, and $\overline{\mathrm{m}}_p$ denote the average margin in precinct p. Then Figure D.1b reveals that the residual MAS vote margin $\mathrm{m}_{bp} - \overline{\mathrm{m}}_p$ increases with verification time.[23]

Critically, the within-precinct divergence between MAS and CC does not accelerate after the shutdown of the public preliminary results system. If anything, the candidates' fortunes diverge more slowly after 7:40 p.m. (This fact is robust to bandwidth choice, as we show in Appendix Figure H.3).

The time trend in Figure D.1b highlights a problem with the interpretation of results in Escobari and Hoover (2019). They regress MAS's margin on an indicator for *post-shutdown* and precinct fixed effects, finding that the coefficient on *post-shutdown* is positive and significant even with precinct fixed effects included. The magnitude

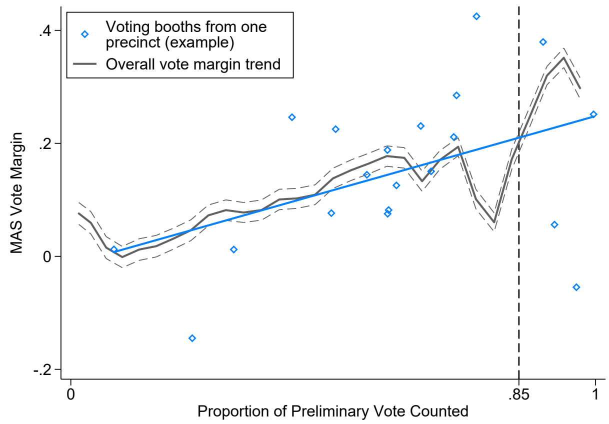
[22] The gray line in Figure D.1a, which marks the overall time trend in MAS margin, differs from the time trend in Figure 1 for two reasons. First, Figure D.1a uses the *verification* time stamp (following the work we replicate in this section), while Figure 1 uses the *transmission* time stamp, which better captures reporting time (Appendix A). The sharp non-monotonicity in Figure D.1a is caused by a server backup that produced a burst of verifications of tally sheets from the anti-Morales department of Santa Cruz. Second, Figure D.1a uses *percentiles* of reporting time on the x-axis (again following other work), while Figure 1 uses clock time. Using percentiles has the effect of visually compressing the long tails of the distribution of clock time: many more minutes elapsed between the 95th and the 96th percentiles than between the 65th and 66th percentiles; Figure D.1a obscures this fact.

[23] This trend differs from Figure 1b in the main text for two reasons: here, following Escobari and Hoover, we use *verification* time stamps rather than *transmission* time stamps; and, again following Escobari and Hoover, we use proportion of vote counted on the x axis, rather than clock time.

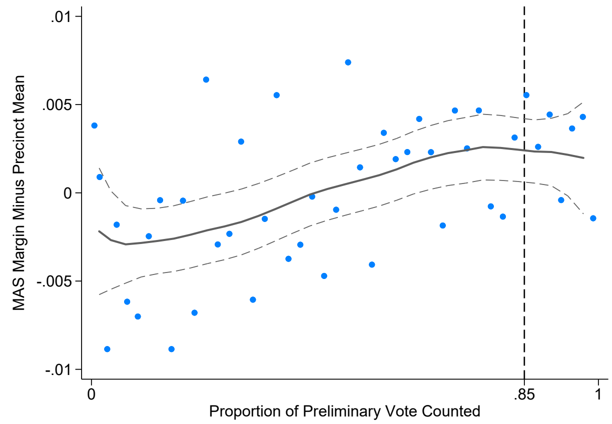
Figure D.1: Within Precincts, MAS Vote Margin Increases with Reporting Time

Figure (a) provides an example of within-precinct variation; the blue diamonds mark MAS's vote margin in each of the 20 voting booths in a single precinct in the town of Llallagua. The gray line marks the overall margin trend; it differs from the trend in Figure 1 for two reasons, explained in Footnote 22. Figure (b) plots the voting-booth-level MAS margin after subtracting the precinct mean (i.e., $\mathrm{m}_{bp} - \overline{\mathrm{m}}_p$).

(a) Within-Precinct Variation: An Example



(b) Residual MAS Margin ($\mathrm{m}_{bp} - \overline{\mathrm{m}}_p$)



Lines mark local linear fits with the rule-of-thumb bandwidth from Calonico, Cattaneo and Farrell (2018). Figure (b) uses data from the $N = 33,038$ tally sheets verified in the preliminary results system. Note that this figure looks slightly different from Figure 1b because the latter uses transmission time on the x axis, while this figure uses verification time (following Escobari and Hoover, 2019).

of the coefficient is consistent with our Figure D.1b; it reveals that MAS's post-shutdown vote margin was approximately four tenths of a percentage point larger than MAS's pre-shutdown margin in the same precincts. But the Escobari and Hoover (2019) specification does not account for the secular trend in Figure D.1b: even within precinct, voting booths that report later favor MAS, even before the shutdown. Adding a time trend to the regression in Escobari and Hoover (2019) reduces the estimate of the post-shutdown increase to zero.

To see this, consider a regression of the form:

$$\mathrm{M}_{bp} = \gamma_p + \beta_1(\text{Time percentile})_{bp} + \beta_2 \mathbb{1}(\text{Post shutdown})_{bp} + \beta_3(\text{Percentile} \times \text{Post})_{bp} + \epsilon_{bp} \quad (3)$$

where M_{bp} is MAS's margin over CC in voting booth b in precinct p; γ_p are precinct fixed effects; $(\text{Time percentile})_{bp}$ is the percent of the vote counted when voting booth b was verified in the preliminary results system (TREP); $(\text{Post shutdown})_{bp}$ takes a value of 1 if voting booth b reported after the government stopped publishing updated

Table D.1: Within Precinct, MAS Margin Does Not Grow Faster Post-Shutdown

Estimates of Equation 2. The dependent variable is MAS's margin over Civic Community (scaled −1 to 1). Column (1) reveals that the (linearized) growth in MAS's margin does accelerate after the shutdown; Column (2) shows that this is not true of within-precinct variation; Column (3) replicates Escobari and Hoover (2019, Table 3, Col. 3), showing that omitting the within-precinct secular trend in MAS margin produces a positive and (marginally) significant coefficient on the post-shutdown dummy; and Column (4) adds the time trend, revealing that, in this specification, the coefficient on *post-shutdown* is estimated at zero.

	(1) No Precinct FEs	(2) + Precinct FEs	(3) No time trend§	(4) + time trend
$\hat{\beta}_1$: Reporting time percentile†	0.173 (0.02)	0.014 (0.003)		0.013 (0.003)
$\hat{\beta}_2$: Post shutdown (0/1)	0.102 (0.02)	0.005 (0.003)	0.006 (0.002)	0.000 (0.002)
$\hat{\beta}_3$: Percentile × Post	-0.026 (0.2)	-0.065 (0.04)		
Observations	34,551	32,946	32,946	32,946
Precinct FEs		✓	✓	✓

Standard errors, clustered by precinct, in parentheses. §This is the specification in Escobari and Hoover (2019); see Appendix E for discussion. †For ease of interpretation of the coefficients, we center the reporting time percentile at the moment of the shutdown (7:40 p.m. on election night). Thus the coefficient on *reporting time percentile* can be interpreted as the slope of MAS's vote share before the shutdown, the coefficient on *Post* is the estimated jump (new intercept) after the shutdown, and the coefficient on the interaction term is the increase in slope after the shutdown.

results (7:40 p.m.) and 0 otherwise; $(\text{Percentile} \times \text{Post})_{bp}$ interacts $(\text{Time percentile})_{bp}$ with $(\text{Post shutdown})_{bp}$; and ϵ_{bp} is a voting-booth-specific error term.

Column (1) of Table D.1 reports estimates of a version of Equation 3 that *excludes* precinct fixed effects; in this specification, MAS's margin grows faster after the government stopped publishing updated results. But when we include precinct fixed effects, in Column (2), MAS's margin grows no faster after than before the shutdown. If anything, and again consistent with Figure D.1b, the growth in MAS's margin slows after the shutdown ($\widehat{\beta}_3$ is negative but imprecisely estimated).

Column (2) of Table D.1 also reveals that, *even within precinct*, there is a secular increase in MAS's margin over the reporting window. This is captured in the positive and significant coefficient on β_1. And this is the problem with the conclusions Escobari and Hoover (2019): if we omit that secular trend, as in Column (3), then the coefficient on the *post shutdown* is positive and significant.[24] When we include the secular trend, as in Column (4), the coefficient on *post shutdown* is estimated at zero. The same would be true of an indicator for any artificial *post* period: post-50% of the count, post-70% of the count, et cetera. In other words, because of the within-precinct secular trend in MAS margin, the specification that Escobari and Hoover propose as a "natural experiment" is not, in fact, a natural experiment.

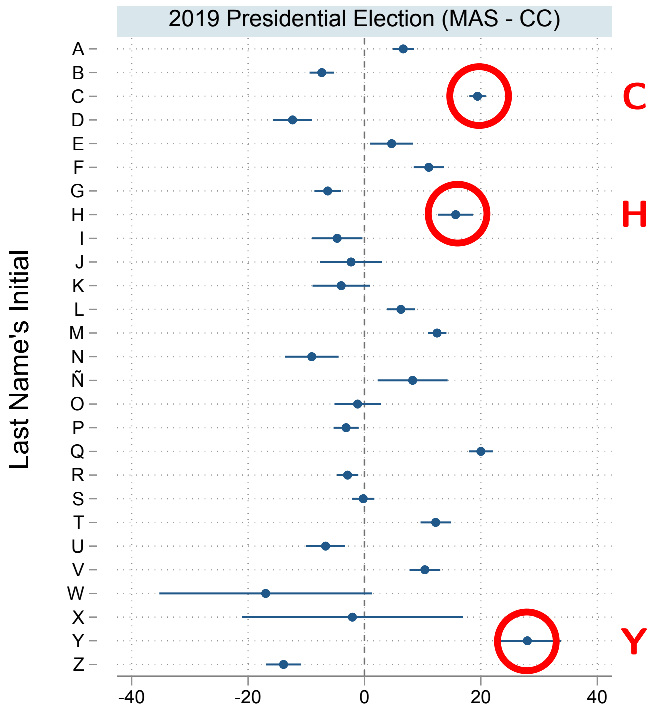
Possible explanations. As noted in the main text, voting-booth jurors (*jurados*) are chosen randomly from among each voting booth's voters—not from among voters in the whole precinct. At the close of voting, the jurors count the ballots and fill out a paper tally sheet (*acta*). This aspect of electoral administration in Bolivia could easily generate a correlation between MAS vote margins and verification time. Voters' socio-economic status is unlikely to be exactly identical across voting booths within a precinct. Booths with voters of lower socio-economic status and lower levels of education are more likely to vote MAS (Madrid, 2012, p. 69–72). It is easy to imagine why those booths might also report later: voters with lower levels of education may take more time to vote; moreover, jurors with lower levels of education would likely take more time to count votes and fill out the tally sheet. It is therefore unsurprising that we find a positive within-precinct correlation between MAS margin and time.

These differences across voting booths within a precinct are likely greater because voters are assigned alphabetically—not randomly—to voting booths within precincts, as in much of the United States (Exeni Rodríguez, 2020). Of course, surname is related to ethnicity, which is related to socio-economic status in Bolivia (including education, see UNICEF, 2014, p. 30)—and indigenous surnames are distributed differently throughout the alphabet than non-indigenous surnames. Indigenous surnames are more likely to begin with C, H, or Y, for example, while non-indigenous names are more likely to begin with F, R, or S (Forebears.io, 2020). For that reason, different voting booths likely have different proportions of indigenous voters.

To illustrate, consider a hypothetical precinct with the mean number of voting booths (6.5). Each voting booth has approximately 15% of the precinct's voters. Consider two clusters of last names: those that begin with the letter C, which includes 15.9%

[24] The estimate in Column (3) of Table D.1 is larger than the corresponding estimate in Escobari and Hoover (2019), because we use slightly different time stamps to construct the *post* variable. When we use the same time stamps, we can replicate Escobari and Hoover's estimate, as we show in Appendix E.

Figure D.2: Figure from Escobari and Hoover (2020), Our Annotations in Red



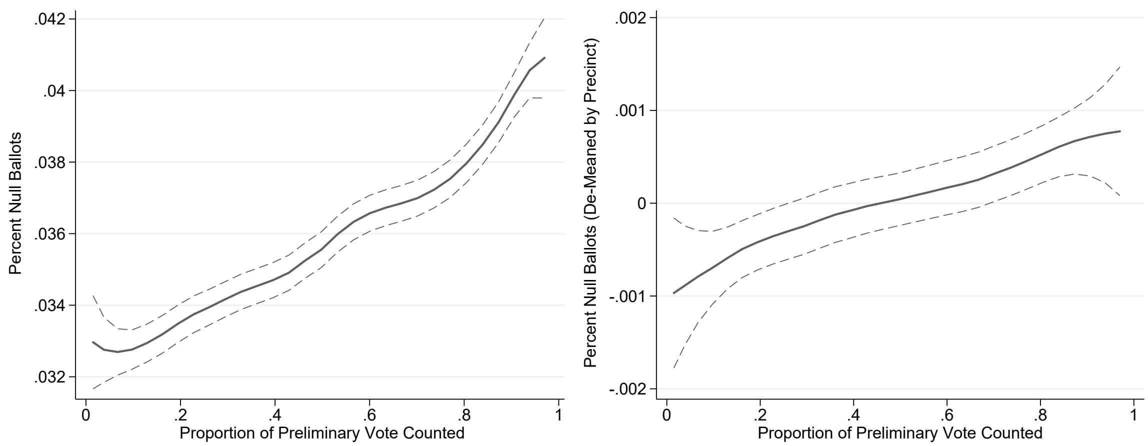
of the population, and those that begin with R or S, which together cover 14% (Forebears.io; see also Rodriguez-Larralde et al., 2011). This hypothetical precinct could then have one voting booth in which all voters' surnames begin with C, and another in which all voters' surnames begin with R or S. These booths would likely have very different proportions of indigenous voters: among the 911 most common surnames (which account for 88% of the population), 33.1% of people with C surnames have indigenous surnames, while 1.4% of the people with R or S surnames have indigenous surnames. It would therefore be unsurprising if MAS performed better in the C voting booth than in the $R + S$ voting booth; nor would it be surprising if the C voting booth reported later than the $R + S$ voting booth.

In a subsequent paper, Escobari and Hoover (2020) obtained data that allow them to observe which surname first letters correspond to which voting booths. Regressing voting-booth-level incumbent margin on indicators for the first letters voting at each booth, they estimate the coefficients on those indicators; we copy-paste the figure from their paper in Figure D.2, adding our own annotations in red. Consistent with our conjecture prior to seeing these results, voting booths with voters whose last

Figure D.3: Preliminary Results System Time is Correlated with % Null Ballots

Less-educated voters are more likely to cast null ballots. Consistent with the hypothesis that voting booths with less-educated voters were more likely to report later, the share of null ballots rises over the reporting window (a). And consistent with the hypothesis that within-precinct variation in socio-economic status drives within-precinct variation in reporting time, within-precinct variation in null ballot share is correlated with reporting time (b).

(a) % Null Rises Over Time (b) Residual % Null Rises Over Time



Grey lines mark local linear fits using the rule-of-thumb bandwidth from Calonico, Cattaneo and Farrell (2018); the dashed lines mark 95% confidence intervals. Top and bottom 1% of de-meaned reporting times are excluded.

names begin with C, H, and Y support the incumbent at much higher rates. (The same is true of voting booths with Q voters; we did not mention Q in our previous draft, but it also has a high concentration of indigenous last names.)

While Escobari and Hoover (2020) acknowledge that, as we suggested, first-letter-of-surname is highly predictive of vote margin, they also claim that the first-letter fixed effects do not explain the entire within-precinct trend. If true, this would not be inconsistent with our proposed explanation: first-letter-of-surname should be *correlated* with one aspect of socio-economic status—the source of cross-booth heterogeneity in our account—but they need not capture all of the relevant variation. Moreover, Escobari and Hoover (2020) do not include the first-letter fixed effects in the specification used in their previous paper. Instead, they include these fixed effects in a specification in which they pool data from two elections (2016 and 2019) and restrict the secular time trend to be the same in both election years (Table 4).

One implication of our hypothesis is that, even *within* precinct, the proportion of null ballots would be correlated with reporting time. While *blank* ballots might be interpreted as protest votes, null ballots occur when the voter makes a mistake (for example, marking two candidates instead of one). Less-educated voters are more likely to cast these ballots (Fujiwara, 2015). Thus, if within-precinct variation in voters' socioeconomic characteristics is correlated with within-precinct variation in verification time, we would also expect within-precinct variation in null ballots to be correlated with within-precinct variation in verification time. We show graphically that it is (Appendix Figure D.3).

Another possible explanation for the within-precinct trends in MAS margin and in null ballots is that pro-MAS jurors strategically invalidate ballots cast for the opposition, and that doing so takes time. Writing and estimating a model to adjudicate between these explanations strikes us as a worthy objective for future work. In any case, decentralized invalidation of opposition votes throughout election night does not resemble mechanics implicitly alleged by Escobari and Hoover (2019) and Newman (2020), in which the government stopped publishing results in order to enable centralized tampering with vote tallies in late-counted voting booths.

E Escobari and Hoover (2019) Replication

In the main text, we note a problem with the specification in Escobari and Hoover (2019): it includes an indicator for *post* without accounting for a secular (within-precinct) trend in MAS's vote margin. We show that when we account for this trend, the coefficient on *post* is estimated at zero.

The results presented in Table D.1, Column (3)—reproduced in Column (3) of Table E.1 below—do not exactly replicate Escobari and Hoover (2019). Our coefficient on *post* is estimated at 0.0057 (about half of one percentage point), whereas theirs is estimated at 0.0037. The principal difference is that Escobari and Hoover use what we call the *website* time stamps (see previous section, Appendix A), whereas we use the internal *verification* time stamps. When we use the *website* time stamps, as in Column (5) of Table E.1, we can replicate their result almost exactly.

Table E.1: Replication of Escobari and Hoover (2019)

Estimates of Eqn. 2. The D.V. is MAS's margin over Civic Community (scaled 0–1).

	Last Transmission		Verification		Website	
	(1)	(2)	(3)	(4)	(5)	(6)
Post shutdown (0/1)	0.0052	-0.0007	0.0057	0.0003	0.0038	0.0036
	(0.0021)	(0.0023)	(0.0019)	(0.0023)	(0.0018)	(0.0018)
Reporting time percentile		0.0172		0.0129		
		(0.0033)		(0.0033)		
Observations	32,946	32,946	32,946	32,946	32,925	32,946
Precinct FEs	✓	✓	✓	✓	✓	✓

Standard errors, clustered by precinct, in parentheses. Column (5) uses MAS's margin as Escobari and Hoover (2019) calculated it; Column (6) uses MAS's margin as it appears in the final tally.

A secondary difference is that Escobari and Hoover calculate MAS's margin based on a preliminary count of valid votes (the one published on the website), whereas we calculate MAS's margin based on the final count of valid votes. Because the number of valid votes differs only for 2.75% of observations, and because these differences are quite small, this alone makes little difference for the final estimates: Column (5) of Table E.1 uses the website count of valid votes; Column (6) uses the final count of valid votes. The point estimate changes by 0.0002.

F RD estimate: Robustness

Sort order. As noted in the main text, only 8% of observations have unique time stamps. This is not surprising given the number of tally sheets and the length of the reporting window: there are 34,555 tally sheets, almost all of which were verified within a two-hour window, or 7,200 seconds (the time stamps include seconds, but not milliseconds). In the main text, we present results based on sorting the observations first by time stamp and then by a random number.

Of course, the sort order could affect our regression discontinuity (RD) results. To investigate whether our main RD result—failure to reject the null of continuity—is robust to different possible sort orders, we repeat the analysis 1,000 times, each time sorting (within time stamp) according to a different random draw. This exercise reveals that our failure to reject continuity is not the artifact of a specific sorting.

Figure F.1: No Evidence of Discontinuities, Regardless of Sort Order

Each figure plots the magnitude of the RD estimate (Calonico, Cattaneo and Titiunik, 2014) against the corresponding p-value, for each of 1,000 draws of the random variable used to sort observations within time stamps.

(a) 95% (truncated sample) (b) 7:40 p.m.

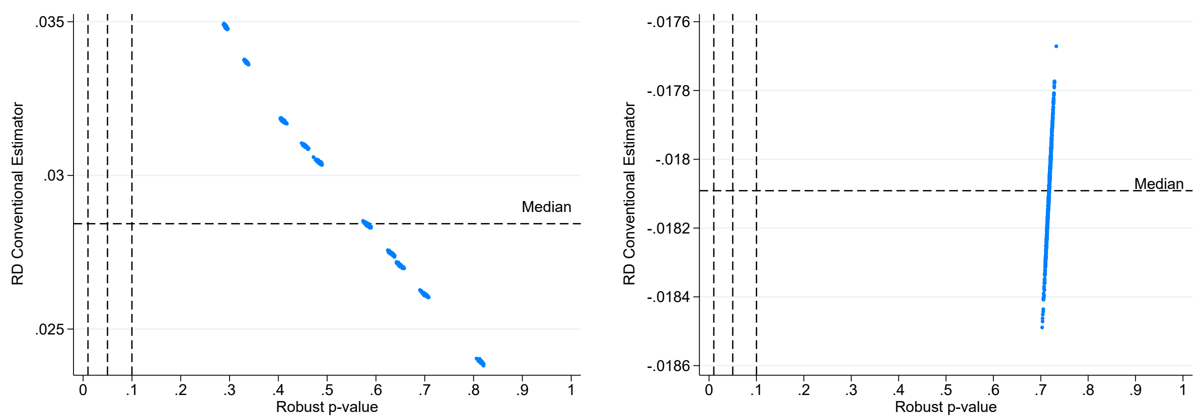


Figure F.1 plots the magnitude of the RD estimates against the corresponding p-values for each of the 1,000 draws, for each of the two cutoffs studied in the paper. Table F.1 summarizes the results. The mean and median robust p-values are above 0.5, implying that the results presented in the main text are not anomalous: there is no evidence of a statistical discontinuity in MAS vote share at those cutoffs.

Table F.1: No Evidence of Discontinuities, Regardless of Sort Order

Cutoff	Date & Time	Robust p-value		RD Estimate		N Sortings
		Mean	Median	Mean	Median	
0.950	10/20/2019 20:03:59	0.541	0.580	0.029	0.028	1,000
0.889	10/20/2019 19:40:57	0.717	0.717	-0.018	-0.018	1,000

Polynomial degree and bandwidth. The results in the main text show that we cannot reject the null of continuity at the three cutoffs using a degree-one local polynomial with the MSE-optimal bandwidth. Table F.2 shows that, indeed, we cannot reject the null of continuity for other combinations of polynomial degree and bandwidth. Specifically, for each polynomial degree $p \in \{1, 2, 3\}$, we estimate the treatment effect using bandwidths selected with and without the regularization term (the regularization term shrinks the optimal bandwidth, Cattaneo, Idrobo and Titiunik, 2020, Section 4.4.2).

Table F.2: Robustness to Polynomial Degree and Bandwidth Choices

Cutoff	Date	Reg.	Deg.	Estimate	BW	p-val.	Robust C.I.	N Left	N Right
0.950	10/20/2019 20:03:59	1	1	0.031	0.041	0.456	[-0.036, 0.080]	1,325	1,379
0.950	10/20/2019 20:03:59	1	2	0.023	0.048	0.917	[-0.066, 0.074]	1,538	1,602
0.950	10/20/2019 20:03:59	1	3	0.006	0.059	0.465	[-0.110, 0.050]	1,872	1,663
0.950	10/20/2019 20:03:59	0	1	0.032	0.096	0.218	[-0.025, 0.108]	3,037	1,663
0.950	10/20/2019 20:03:59	0	2	-0.001	0.180	0.625	[-0.523, 0.871]	5,667	1,663
0.950	10/20/2019 20:03:59	0	3	0.007	0.076	0.442	[-0.103, 0.045]	2,415	1,663
0.889	10/20/2019 19:40:57	1	1	-0.018	0.039	0.721	[-0.065, 0.045]	1,248	1,245
0.889	10/20/2019 19:40:57	1	2	-0.006	0.047	0.918	[-0.062, 0.069]	1,496	1,496
0.889	10/20/2019 19:40:57	1	3	-0.051	0.045	0.125	[-0.143, 0.018]	1,434	1,434
0.889	10/20/2019 19:40:57	0	1	-0.025	0.095	0.665	[-0.087, 0.055]	2,996	3,078
0.889	10/20/2019 19:40:57	0	2	-0.048	0.117	0.174	[-0.138, 0.025]	3,661	3,612
0.889	10/20/2019 19:40:57	0	3	-0.019	0.076	0.840	[-0.080, 0.065]	2,395	2,440

"Reg." reports whether we choose the bandwidth with or without the regularization term; "Deg." reports the degree of the local polynomial.

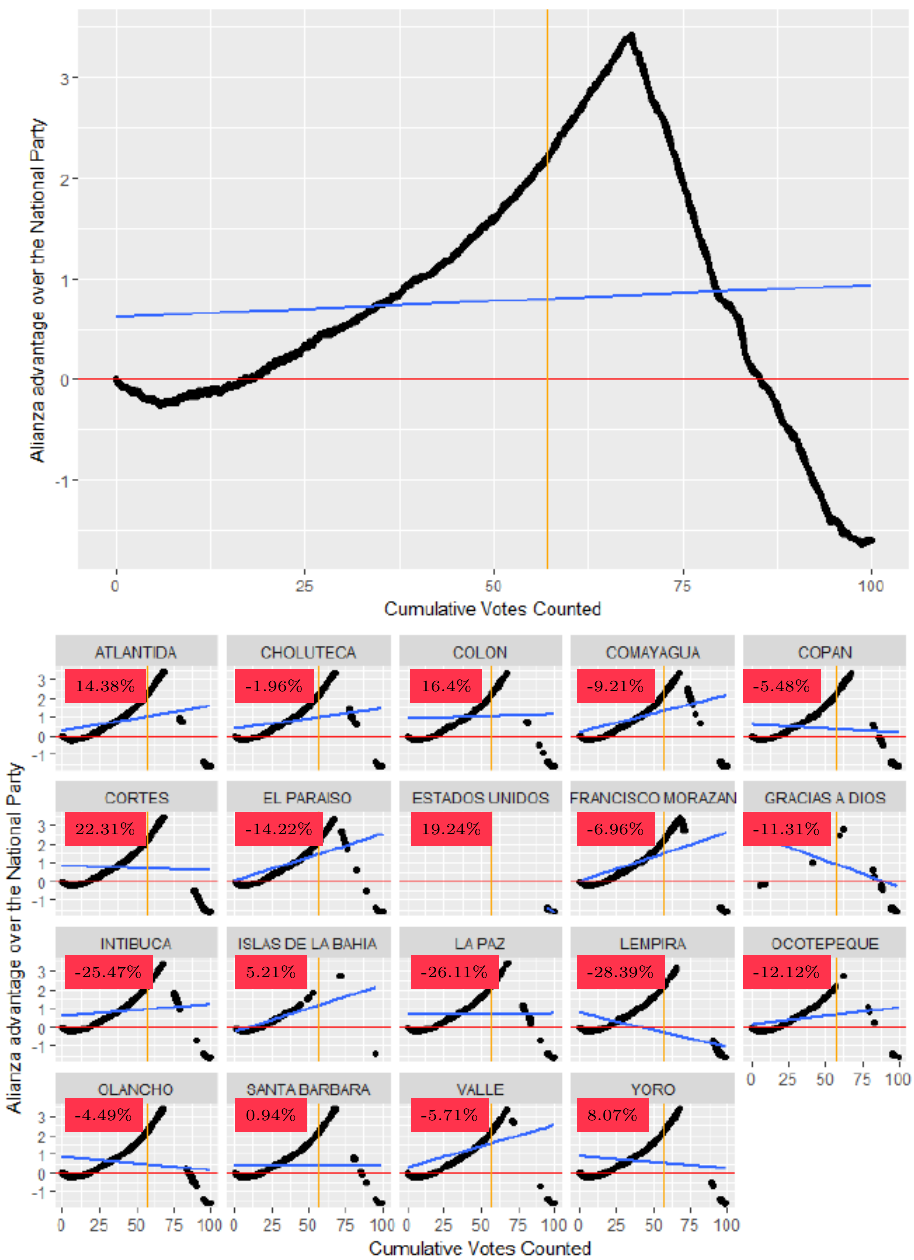
G The case of Honduras in 2017

The OAS report on the Honduran presidential election of 2017 raised concerns about electoral integrity, based in part on the observation that the same non-monotonic time trend in vote share appeared in every single department (OAS, 2017; Nooruddin, 2017). This assertion was based on the figure reproduced below (Figure G.1), over which we superimposed red text boxes reporting the final margin in each department.

The department-specific graphs are clearly inconsistent with the overall department margins. Consider, for example, the coastal department of Atlántida, one of the most opposition-leaning departments in the country. In Atlántida, the opposition ended up with a 14-point margin over the incumbent, according to official data (CNE Honduras, 2017), but the graph shows the cumulative margin varying between -1 point and 3 points, and ending around -1 point. Similarly, consider La Paz, one of the most pro-incumbent departments, where the incumbent ended up with a 26-point margin over the opposition. There, too, the graph shows the final cumulative margin at -1. In short, it may be the case that there was a swing away from the opposition across all departments, but these graphs do not establish that fact. They appear, rather, to be the artifact of projecting the national trend onto time stamps from each department.

Figure G.1: Coding Error Produces Appearance of Identical Trends?

Our annotations in red report the actual final Alianza margin in each department.

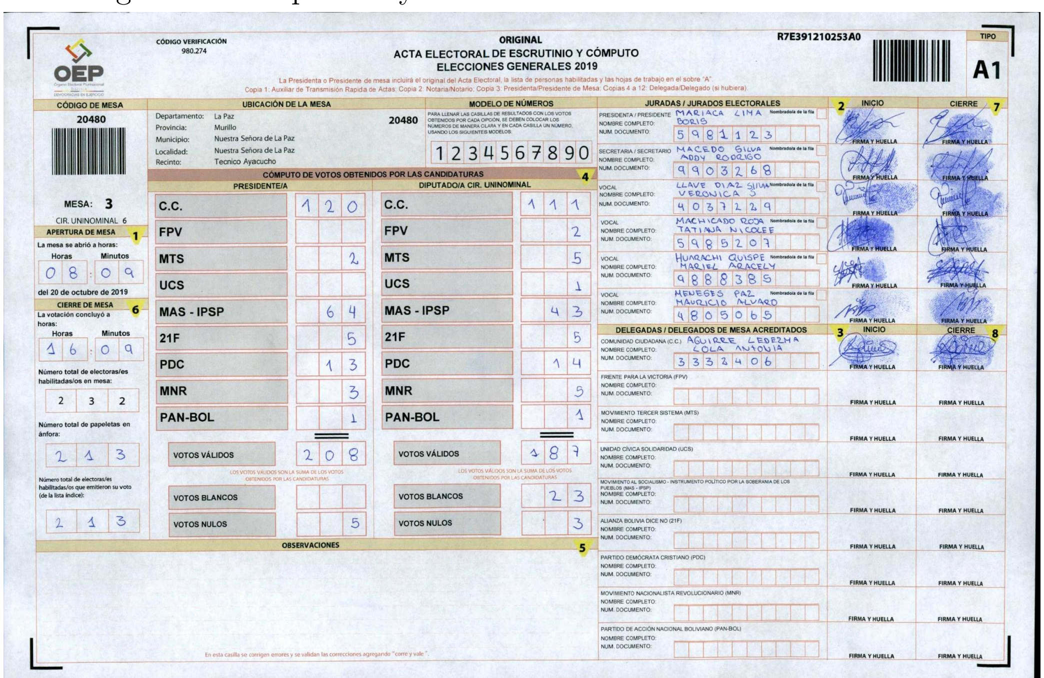


H Additional tables and figures

Figure H.1: Paper Ballot in Bolivia's Presidential Election

Source: Jorge Bernal

Figure H.2: Paper Tally Sheet in Bolivia's Presidential Election



CÓDIGO VERIFICACIÓN 980.274

ORIGINAL

ACTA ELECTORAL DE ESCRUTINIO Y CÓMPUTO
ELECCIONES GENERALES 2019

R7E391210253A0

TIPO A1

La Presidenta o Presidente de mesa incluirá el original del Acta Electoral, la lista de personas habilitadas y las hojas de trabajo en el sobre "A".
Copia 1: Auxiliar de Transmisión Rapida de Actas; Copia 2: Notaria/Notario; Copia 3: Presidenta/Presidente de Mesa; Copias 4 a 12: Delegada/Delegado (si hubiera).

CÓDIGO DE MESA: 20480

UBICACIÓN DE LA MESA: Departamento: La Paz; Provincia: Murillo; Municipio: Nuestra Señora de La Paz; Localidad: Nuestra Señora de La Paz; Recinto: Tecnico Ayacucho — 20480

MODELO DE NÚMEROS: PARA LLENAR LAS CASILLAS DE RESULTADOS CON LOS VOTOS OBTENIDOS POR CADA OPCIÓN, SE DEBEN COLOCAR LOS NÚMEROS DE MANERA CLARA Y EN CADA CASILLA UN NÚMERO, USANDO LOS SIGUIENTES MODELOS: 1 2 3 4 5 6 7 8 9 0

MESA: 3

CIR. UNINOMINAL 6

APERTURA DE MESA (1): La mesa se abrió a horas: Horas 08 : Minutos 09 del 20 de octubre de 2019

CIERRE DE MESA (6): La votación concluyó a horas: Horas 16 : Minutos 09

Número total de electoras/es habilitadas/os en mesa: 232

Número total de papeletas en ánfora: 213

Número total de electoras/es habilitadas/os que emitieron su voto (de la lista índice): 213

CÓMPUTO DE VOTOS OBTENIDOS POR LAS CANDIDATURAS (4)

	PRESIDENTE/A	DIPUTADO/A CIR. UNINOMINAL
C.C.	120	111
FPV		2
MTS	2	5
UCS		1
MAS - IPSP	64	43
21F	5	5
PDC	13	14
MNR	3	5
PAN-BOL	1	1
VOTOS VÁLIDOS	208	187
VOTOS BLANCOS		23
VOTOS NULOS	5	3

LOS VOTOS VÁLIDOS SON LA SUMA DE LOS VOTOS OBTENIDOS POR LAS CANDIDATURAS

OBSERVACIONES (5)

En esta casilla se corrigen errores y se validan las correcciones agregando "corre y vale".

JURADAS / JURADOS ELECTORALES — INICIO (2) / CIERRE (7)

Cargo	NOMBRE COMPLETO	NUM. DOCUMENTO	INICIO	CIERRE
PRESIDENTA / PRESIDENTE	MARIACA LIMA BORIS	5981123	FIRMA Y HUELLA	FIRMA Y HUELLA
SECRETARIA / SECRETARIO	MACEDO SILVA ADDY RODRIGO	9903268	FIRMA Y HUELLA	FIRMA Y HUELLA
VOCAL	LLAVE DIAZ SILVA VERONICA S	4037229	FIRMA Y HUELLA	FIRMA Y HUELLA
VOCAL	MACHICADO ROJA TATIANA NICOLEE	5985207	FIRMA Y HUELLA	FIRMA Y HUELLA
VOCAL	HUARACHI QUISPE MARIEL ARACELY	9888385	FIRMA Y HUELLA	FIRMA Y HUELLA
VOCAL	MENESES PAZ MAURICIO ALVARO	4805065	FIRMA Y HUELLA	FIRMA Y HUELLA

Nombrado/a de la fila

DELEGADAS / DELEGADOS DE MESA ACREDITADOS — INICIO (3) / CIERRE (8)

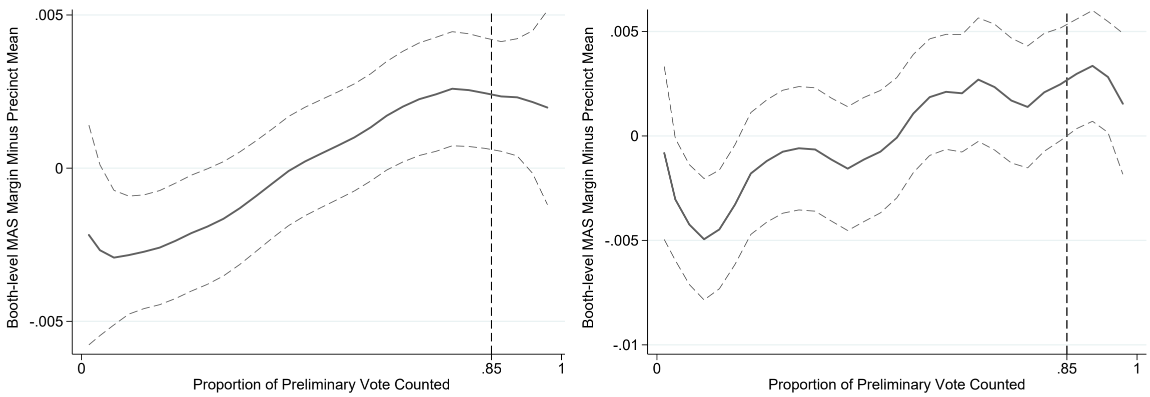
Organización	NOMBRE COMPLETO	NUM. DOCUMENTO	INICIO	CIERRE
COMUNIDAD CIUDADANA (C.C.)	AGUIRRE LEDEZMA LOLA ANTONIA	3332406	FIRMA Y HUELLA	FIRMA Y HUELLA
FRENTE PARA LA VICTORIA (FPV)			FIRMA Y HUELLA	FIRMA Y HUELLA
MOVIMIENTO TERCER SISTEMA (MTS)			FIRMA Y HUELLA	FIRMA Y HUELLA
UNIDAD CÍVICA SOLIDARIDAD (UCS)			FIRMA Y HUELLA	FIRMA Y HUELLA
MOVIMIENTO AL SOCIALISMO - INSTRUMENTO POLÍTICO POR LA SOBERANÍA DE LOS PUEBLOS (MAS - IPSP)			FIRMA Y HUELLA	FIRMA Y HUELLA
ALIANZA BOLIVIA DICE NO (21F)			FIRMA Y HUELLA	FIRMA Y HUELLA
PARTIDO DEMÓCRATA CRISTIANO (PDC)			FIRMA Y HUELLA	FIRMA Y HUELLA
MOVIMIENTO NACIONALISTA REVOLUCIONARIO (MNR)			FIRMA Y HUELLA	FIRMA Y HUELLA
PARTIDO DE ACCIÓN NACIONAL BOLIVIANO (PAN-BOL)			FIRMA Y HUELLA	FIRMA Y HUELLA

Source: Plurinational Electoral Organ of Bolivia

Figure H.3: Within-Precinct Variation Trend, Smaller Bandwidth

Figure (a) repeats Figure D.1b; the takeaway is that, after accounting for precinct characteristics, the growth in MAS's margin does not accelerate after the public information blackout. Figure (b) shows that this result is not an artifact of bandwidth choice.

(a) Rule-of-thumb bandwidth (Calonico, Cattaneo and Farrell, 2018) (b) Arbitrary small bandwidth (0.15)

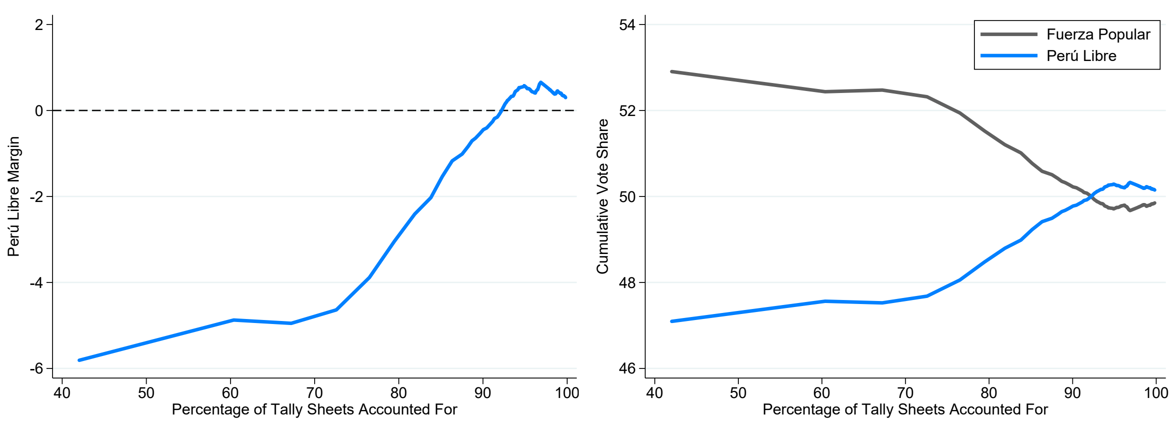


Grey lines mark local linear fits using the rule-of-thumb bandwidth from Calonico, Cattaneo and Farrell (2018); the dashed lines mark 95% confidence intervals.

Figure H.4: Example of a Non-Monotonic Time Trend in Vote Share

These figures plot the winner's margin (a) and vote shares of the two candidates (b) in Peru's 2021 runoff presidential election. The count shifted toward the left candidate, Pedro Castillo, as the share of rural votes increased; in the final 5% of the count, the trend reversed, as votes from abroad were finally tallied (votes at embassies are transmitted physically via diplomatic pouch) (Infobae, 2021).

(a) Margin vs. Time Percentiles (b) Vote Shares vs. Time Percentiles



Authors' elaboration with data from ONPE