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

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

June 7, 2020

Abstract

Surprising trends in late-counted votes can spark conflict. In Bolivia, electoral observers recently sounded alarms about trends in late-counted votes—with dramatic political consequences. We revisit the quantitative evidence, finding that (a) an apparent jump in the incumbent's vote share was actually an artifact of the analysts' error; (b) analysis of within-precinct variation mistakenly ignored a strong secular trend; and (c) nearly identical patterns appear in data from the previous election, which was not contested. In short, we examine the patterns that the observers deemed "inexplicable," finding that we can explain them without invoking fraud.

[†]We are grateful to Santiago Anria and Marc Meredith for detailed guidance. For research assistance, we thank Mateo Arbeláez, Estefanía Bolívar Méndez, and Juan Vera. For comments, we thank María Eugenia Boza, Matias D. Cattaneo, Tulia Falleti, Camilo García-Jimeno, Etan Green, 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 Rocío Titiunik. All errors are our own.

[‡]Department of Political Science, University of Pennsylvania.

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

[¶]Greenleaf Visiting Professor, Center for Inter-American Policy and Research, and Stone Center for Latin American Studies, Tulane University.

Democracy requires elections you can believe in. Doubts about the legitimacy of the electoral process can demoralize and demobilize voters (Birch, 2010; Norris, 2014; Alvarez, Hall and Llewellyn, 2008; Simpser, 2012)—or even spark violence (Tucker, 2007). One obvious threat to legitimacy is electoral malpractice. Another is unsubstantiated allegations of fraud. Politicians often sound alarm bells about ballot-stuffing or double voting, for instance, even in the absence of such abuses (Goel et al., 2020; Foley and Stewart, 2020; Norris, Garnett and Grömping, 2020).

One threat to the perceived legitimacy of the electoral process comes from the fact that votes are typically counted in a non-random order. When a candidate leads on election night but ultimately loses the next day, her supporters often cry foul. In the 2007 presidential election in Kenya, for example, the opposition candidate suffered a narrow loss after holding an early lead (Kanyinga, 2009). His party accused the government of fraud. Hundreds were killed in the ensuing crisis; hundreds of thousands were displaced.

Researchers understand why late-counted votes disproportionately favor the Democrats in the United States: young and nonwhite voters are more likely to cast mail-in and provisional ballots, which are more likely to be counted after election day (Foley, 2013; Foley and Stewart, 2020; Li, Hyun and Alvarez, 2020). This finding may constrain politicians who would otherwise decry the "blue shift" as evidence of fraud. But other countries' shifts in late-counted votes, while common, are less well understood—leaving them open to politicized interpretation.

We revisit the controversial Bolivian presidential election of October, 2019. On election night, electoral authorities announced that, with more than 80% of the vote counted, incumbent Evo Morales had a 7.9-point lead over the runner-up—less than the 10 points he needed to avoid a runoff. But the following evening, with nearly all of the vote counted, Morales's margin narrowly exceeded 10 points. The runner-up cried fraud (Mesa, 2019). And critically, the Organization of American States (OAS) 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*d*).

The political consequences were dramatic. In part because of allegations of fraud, the Bolivian military asked Morales to resign; he fled to Mexico. An opposition-party senator took office as interim president. At this writing, she remains in office.

We revisit the quantitative patterns that the OAS and other researchers presented as "inexplicable" (OAS, 2019*a*, p. 8; Escobari and Hoover, 2019; Newman, 2020).[1] We find instead that we can explain these patterns without invoking fraud.[2] On their own, these results do not "call into question the credibility of the [electoral] process" (OAS, 2019*a*, p. 8). We do not assess the integrity of the election overall. As noted below, the OAS presented many qualitative indicators of electoral malpractice; we study only the quantitative evidence.

The final report of the OAS emphasized a discontinuous jump in the incumbent's vote share after 95% of the vote had been counted (OAS, 2019*a*, p. 8, 88). We find that this apparent jump was likely the artifact of two errors: first, the mistaken exclusion of 4.4% of observations; second, the use of an estimator not designed for regression discontinuity analysis.[3] Correcting either error eliminates the appearance of a jump. Moreover, when we implement a formal test following best practices (Calonico, Cattaneo and Titiunik, 2014), we cannot reject the null that the incumbent's vote share is continuous at the cutoff.

In related work that echoes the OAS's concern about the integrity of the Bolivian election, Escobari and Hoover (2019) find what they deem a suspicious pattern. Voting booths counted after 7:40 p.m., when the government suspiciously stopped publishing updated results, favor the incumbent more than voting booths *from the same polling place* that were counted before 7:40 p.m. Newman (2020) presents a related result, also citing it as suggestive of fraud. We show that these pre-post differences are the product of a secular trend: within precincts, the incumbent's vote share increased with time all evening—even before 7:40 p.m. Accounting for this secular trend eliminates the appearance of an anomalous within-precinct pre-post difference in vote shares. We offer two possible explanations for the secular within-precinct trend, neither of which involves centralized tampering with the tally.

[1]Escobari and Hoover (2019, p. 1) claim to find "evidence that electoral fraud was highly statistically significant;" Newman (2020, p. 1) writes that "the OAS findings were correct."

[2]The OAS declined our request for replication materials. We plan to post our replication materials upon publication; in the meantime, they are available by request.

[3]The literature uses the term "discontinuity" in reference to the probability of receiving treatment at the cutoff; in this case, the probability of being subject to electoral manipulation after 95% of the vote was counted. OAS (2019*a*) instead use the term "discontinuity" to refer to a jump in the outcome (in this case, the incumbent's vote share) at the cutoff. We follow the literature, using *discontinuity* to refer to the probability of receiving treatment and *treatment effect* to refer to the difference in the limits of the incumbent's vote share from below and above the cutoff.

The OAS also expressed concern about an acceleration in the growth of the incumbent's lead after 7:40 p.m. on election night (OAS, 2019*a*, p. 87). We find that this change in trend also appears in analysis of data from the previous poll, in 2016. In fact, we can predict the contested 2019 vote margin within three hundredths of one percentage point using data from (a) 2016 and (b) before 7:40 p.m. on the election night in question (2019). The OAS observed the 2016 election, too, and raised no concerns about electoral malpractice (though there was no audit in 2016, because the result was not contested OAS, 2016*a*,*b*). These results suggest that the incumbent's lead grew faster later in the evening because of a change in the composition of voting booths entering the count. In other words, we can explain the change in trend without invoking electoral malpractice.

In sum, we offer a different interpretation of the quantitative evidence that led the OAS and other researchers to question the integrity of the Bolivian election. We do not establish the absence of fraud; we did not observe the election and make no claim of comprehensive evaluation. Rather, we find that we do not *require* fraud in order to explain the quantitative patterns used to help indict Evo Morales.

Alongside the quantitative results, the OAS emphasized many other indicia of electoral manipulation: secret servers, falsified tally sheets, undisclosed late-night software modifications, and a fragile chain of custody for voter rolls and ballots, among other problems (OAS, 2019*a*).[4] We assess only the quantitative evidence, not the integrity of the election overall. The quantitative results alone merit attention because they played an important role in the evolution of Bolivia's political crisis (see Context section). The OAS drew an explicit connection between their quantitative findings and the outcome, stating that Morales's 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" (p. 94). In contrast, the OAS presented other irregularities (secret servers, etc.) as evidence that the government "*sought* to manipulate the results," not that the government *actually* manipulated the results (p. 4, emphasis added).

These findings contribute to an ongoing debate over quantitative patterns in the Bolivian electoral returns (OAS, 2019*a*; Escobari and Hoover, 2019; Johnston and Rosnick, 2020; Williams and Curiel, 2020; Mebane, 2019; Noorudin, 2020; Minoldo

[4]Other authors claim that these findings do not reveal intentional electoral manipulation (Johnston and Rosnick, 2020). We restrict our analysis to the statistical evidence.

and Quiroga, 2020; Newman, 2020; Rosnick, 2020). We use more fine-grained data than previous critics of the OAS report (though Newman, 2020, who endorses the OAS's conclusions, uses the same data, to the best of our knowledge). The New York *Times* obtained these data from Bolivian electoral authorities and shared them with us. These data allow us to (a) identify the coding error referenced above, (b) estimate discontinuities, and (c) study the shape of the trend around 7:40 p.m., when the government stopped publishing updated results.

Beyond Bolivia, we contribute to three literatures. First, our results echo work in American Politics about the "blue shift:" votes counted after election day disproportionately favor the Democrats (Foley, 2013; Foley and Stewart, 2020; Li, Hyun and Alvarez, 2020). While politicians and pundits often point to the blue shift as evidence of fraud, scholars find that it is predictable. In Bolivia, too, compositional changes likely explain the shift in late-counted votes.

Second, we contribute to literature on the role of international electoral observers (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). One central finding of previous work is that intergovernmental organizations (such as the OAS) are *less* likely to question electoral integrity than nongovernmental organizations (Kelley, 2009, 2012), perhaps because the former are beholden to member states, who may push for leniency. Indeed, in Kelley's data, the OAS itself—one of "a small core of organizations with a serious commitment to high-quality election observation" (Carothers, 1997, p. 21)—ranks among the observers least likely to criticize or condemn electoral integrity (2009, p. 779). In that sense, the Bolivian case constitutes something of an exception. On the other hand, the Bolivian case is consistent with Bush and Prather (2017), who find that third-party monitors can powerfully shape local perceptions of electoral credibility—especially those of political losers inclined to discredit the election anyway.

Finally, our results underscore the importance of quality electoral administration for democratic representation across the Americas (Alvarez et al., 2013; Berger, Meredith and Wheeler, 2008; Meredith and Salant, 2013; Hopkins et al., 2017; Goel et al., 2020; Fujiwara, 2015).

1 Context: Chronicle of a Crisis Foretold

On October 20, 2019, Bolivian voters cast ballots in the first round of a presidential election. The contest pitted incumbent Evo Morales against eight challengers. Morales, first elected in 2005 as part of Latin America's pink tide (Falleti and Parrado, 2018), was seeking a fourth term in office.

This alone was controversial. Bolivia's 2009 constitution imposed a two-term limit, but in 2013 courts had 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. Then, in 2016, Morales held a referendum on his proposal to eliminate term limits all together—and voters defeated it, 51% to 49%.[5] Morales was able to run in 2019 only because Bolivia's highest court later ruled that term limits violated the American Convention on Human Rights (Anria and Cyr, 2019). The president of the electoral tribunal resigned in protest (Aguilar, 2018).

To avoid a runoff, Morales needed more than 40 percent of the vote and a 10-point margin over the second-place candidate (Bolivian Constitution, Article 166).[6] After the polls closed at 7:00 p.m., Bolivia's electoral authority began posting online preliminary results from the preliminary results system (see the following section for details on this system). At 7:40 p.m., the electoral authority initiated a planned pause in the public transmission of results, in advance of a scheduled press conference. The idea was to freeze website updates during the televised announcement, to avoid confusion (NEOTEC, October 28, 2019, p. 3). Just minutes earlier, the Panamanian cybersecurity company that the Bolivian government had hired to monitor the election issued a "maximum alert" about a burst of activity from one of the secret servers (Ethical Hacking, 2019, p. 35). At the press conference, which began at 7:50 p.m., authorities reported that, with 83% of voting booths reporting,[7] Morales had 45.71% of the vote to Carlos Mesa's 37.84%, a gap of 7.87 points (Bolivia tv).

Trouble began when the electoral authority did not resume the public transmission

[5] The two-term limit in the 2009 constitution was itself more favorable to the incumbent than the previous rule, which forbade immediate reelection, allowing reelection only after sitting out at least one term (Corrales, 2016, p. 8).

[6] Or an outright majority.

[7] The OAS later noted that 89%—not 83%—of tally sheets had been transmitted at this point, and that the electoral authorities "deliberately hid from citizens 6% of the tally sheets that were already in the [preliminary results system] but not published" (p. 4).

of the results. The reason is disputed. Critics charge that the government used the shutdown in order to tamper with the electoral results. The government claimed that they never intended to tally 100% of the vote in the preliminary results system (Los Tiempos, 2019*b*). Other accounts attribute the shutdown to an "enormity of technical fuck-ups" and "lack of expertise" (*impericia*) (Cambara Ferrufino, 2019). At 10:30 p.m. on election night, the Organization of American States (OAS) publicly urged electoral authorities to explain why updated results had not been published (OAS, October 23, 2019*c*, p. 3). At 11:23 p.m., opposition candidate Carlos Mesa tweeted a video in which he said, "we cannot accept the manipulation of a result that obviously leads us to a second round."

Electoral authorities did not update the public results until the evening of the following day. By then, Morales had gained a 10.15% lead over Carlos Mesa (Los Tiempos, 2019*a*). Three days later, on October 24, the Plurinational Electoral Organ (OEP) published final near-final results in which Morales won 47.05% to Mesas's 36.53%—a margin of 10.52 points, large enough for Morales to avoid a runoff.[8]

Opposition leaders cried fraud (AFP, 2019). Bolivians "exploded in protest" (Kurmaneav and Castillo, 2019); two protesters were killed, many were injured, offices of the electoral authority were vandalized, and a local MAS building was burned (MAS is Morales's party). Polls suggested that a run-off election would have been close, because opposition votes may have coalesced around Carlos Mesa (ANF, 2019).

Statements from the OAS played an important role in the evolution of Bolivia's political crisis. Together with the European Union, the OAS's Department of Electoral Cooperation and Observation sent a mission to observe the elections. On the evening of October 21, one day after the election, the 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," and "urg[ing] the electoral authority to firmly defend the will of the Bolivian citizenry" (OAS, October 21, 2019*d*). Two days later, on October 23, the mission published a preliminary bulletin recommending that Bolivia hold a runoff election even if Morales were to earn a margin greater than ten points in the final tally (OAS, October 23, 2019*c*, p. 5). The OAS also repeatedly called for all actors to abstain from violence.

[8] The results announced on October 24 included more than 99.5% of the vote and thus Morales's first-round victory was irreversible. The final results, announced the following day (October 25), gave Morales 47.08% to Mesa's 36.51%, a margin of 10.57 points.

Key political actors within Bolivia cited the OAS in calls for new elections and for Morales's resignation. For example, in a statement requesting the resignation of all electoral authorities and the convening of a new electoral process, Carlos Mesa's party 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).

Amid continuing unrest over the disputed result, Morales's government signed an agreement with the OAS to conduct a formal audit (Flores and Valdez, 2019). The audit team, which was separate and independent from the OAS's electoral observation mission, began work on November 1.

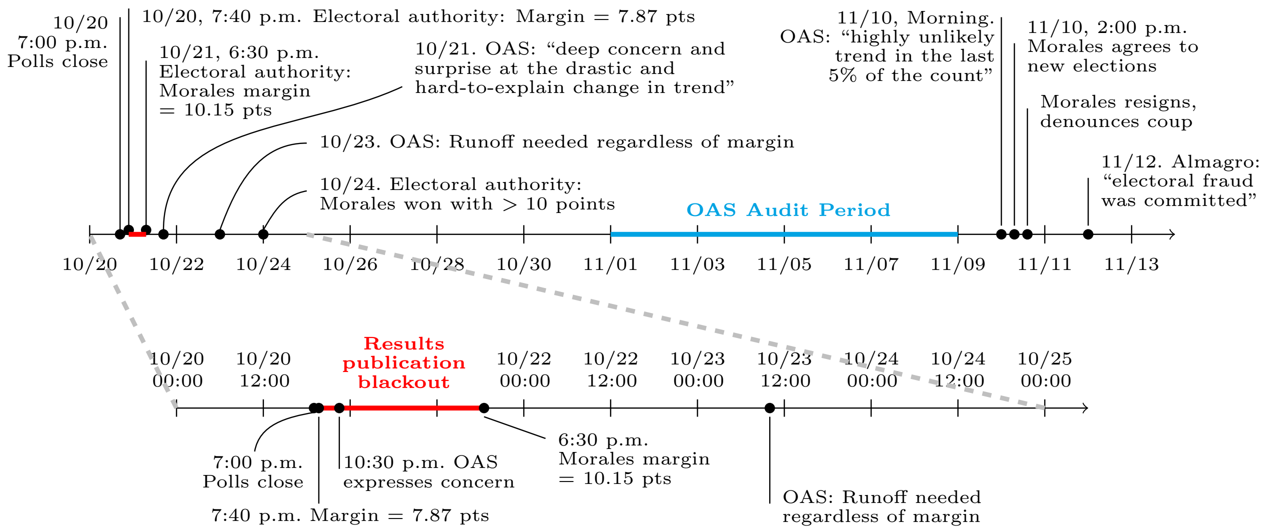
The audit team published their preliminary report on the morning of Sunday, November 10. The report found secret servers, falsified tally sheets, and a deficient chain of custody for critical electoral material—as well as a "highly unlikely trend in the last 5% of the vote count" (OAS, November 10, 2019*b*, p. 9). That afternoon, Morales responded by announcing that the government would convene new elections, under new electoral authorities (Collyns, 2019). But just hours later, under intense public pressure, Bolivia's military chief and police chief asked Morales to resign (Kurmanaev, Machicao and Londoño, 2019). He stepped down that evening and flew to political asylum in Mexico, claiming that he had been ousted in a coup. Two days later, in a speech to the Permanent Council of the OAS, OAS Secretary General Luis Almagro said that "yes, there was a coup d'etat in Bolivia: it happened on October 20, when electoral fraud was committed."

Figure 1 plots several of these events on a timeline. One key moment for the quantitative analysis is 7:40 p.m. on election night (October 20). As noted above, this was when the government stopped publishing updated results from the preliminary results system (see inset timeline in Figure 1)—and it was around this same time that Morales's margin over the runner-up began to grow faster.

On December 4, the OAS published its final report on the election (OAS, 2019*a*), including quantitative analysis that appeared to reveal (a) a jump in Morales's vote

Figure 1: Key Announcements from the Bolivian Government and the OAS

For space reasons, we restrict the events on this timeline to the announcements key to understanding the statistical analysis. However, we note that protests began on election night and escalated as soon as electoral authorities announced the reversal the next evening.



share after 95% of the vote had been counted and (b) a suspicious acceleration of the growth in Morales's margin after electoral authorities suspended the publication of results from the preliminary results system (TREP). We investigate these results. We also investigate results that other researchers (Escobari and Hoover, 2019; Newman, 2020) have presented as evidence of fraud in the Bolivian election.

2 Data

Bolivian voters cast paper ballots at one of 34,555 voting booths (*mesas*) located within 5,296 precincts, or polling places (*recintos*). The ballot uses colors and photos as well as text to communicate voters' choices (see Appendix Figure E.1 for an image).

Three types of poll workers administer the election. First, each voting booth has six "jurors" (*jurados*), who are (a) randomly selected from among each booth's registered voters and (b) legally required to serve (Exeni Rodríguez, 2020). The jurors attend training in advance of election day. They are responsible for checking voters' names against the registration list, distributing and receiving ballots, and helping voters who need assistance. Most importantly, at the close of voting, the jurors count the paper ballots, tally them, write the totals on a paper tally sheet (*acta*), and sign the sheet. Any citizen or party representative may observe this process. Second, an *electoral notary*, hired by the electoral authority, checks the tally sheet for obvious errors (TSE, 2019); there is one notary per precinct. Finally, a *preliminary results system operator*, also hired by the electoral authority, takes a photo of the tally sheet and transmits it via an app. The preliminary results system operator also types the vote totals into the app.

Two systems aggregate the tally sheets. The Transmission of Preliminary Electoral Results, or TREP (*Transmisión de Resultados Electorales Preliminares*), provides a preliminary count. After the preliminary results system operator transmits a tally sheet image and tally sheet data through the app, a team of verifiers look at the image and re-type the totals into the system. If these re-typed figures match the figures typed by the on-site operator, the tally sheet is recorded as *verified* and the numbers are added to the preliminary results system.

The second aggregation is the *cómputo*, or calculation, which is the legally binding official count. This count is much slower and more accurate than the preliminary results system. The paper tally sheets are delivered to electoral authorities in each of Bolivia's nine states (*departamentos*), where they are scanned and transmitted to the national electoral authority. Two separate teams independently transcribe the tally sheets. If the transcriptions match, the totals are added to the count (OEP, 2019, p. 5); otherwise, a third operator checks the transcription. These figures—not the preliminary results (TREP) numbers—determine the outcome of the election. In principle, the official *cómputo* count and the preliminary results system are separate; in practice, the OAS found evidence of contamination, with some preliminary figures funneled directly into the official count (OAS, 2019*a*, p. 6).

The analysis in OAS (2019*a*) focuses on a data set that merges the preliminary results system (TREP) time stamps with the *cómputo* vote tallies, at the level of the voting booth.[9] In our view, this makes sense. The preliminary results system time stamps capture when each voting booth's tallies were verified, which is the relevant time series for investigating the shift in late-counted votes. The various *cómputo* time stamps record when tally sheets physically arrived at the state electoral authorities, as well

[9] The preliminary results system data actually contain five different time stamps corresponding to different stages of the process. In the main text, we use the last of these time stamps—the *verification* time—because it allows us to most closely replicate the figures in the OAS report, even though *verification* time does not always reflect *reporting* time. See Appendix A for details.

as when they were finally verified; this timing has little relevance for election-night dynamics. But the *cómputo* vote tallies are those that determine the final margin.

Even though the preliminary results system time stamps include minutes and seconds, only 8% of tally sheets have unique time stamps. This makes sense given that there are 34,555 tally sheets, almost all of which arrived within two hours—7,200 seconds—of the polls closing.[10] Within each time stamp, we sort tally sheets in a random order.

Previous critiques of the OAS report (2019*a*) used a data set scraped from the website of Bolivia's electoral authority (Johnston and Rosnick, 2020; Williams and Curiel, 2020; Minoldo and Quiroga, 2020; Escobari and Hoover, 2019). These data have two limitations. First, the preliminary results system time stamps in the public data were rounded to the nearest three minutes, precluding other analysts from directly studying discontinuities.[11] Second, and perhaps more importantly, all tally sheets reported after the election-night preliminary results system shutdown were assigned a single reporting time, on the evening of the day after the election. Thus, in the public data, 11.6% of the tally sheets share a single time stamp (83.8% arrived before the shutdown, and 4.1% have no preliminary results system time stamps at all; we discuss the latter group in detail in the following section). See Williams and Curiel (2020, p. 2) for a figure that visualizes this bunching. Our data, obtained from the electoral authority by the New York *Times*, include all of the preliminary results system time stamps (even those for voting booths verified after the government stopped publishing results). To the best of our knowledge, these are the same data used by Newman (2020) and by the OAS (the OAS declined our request for replication materials).

The principal time measure used in the OAS analysis is not the time stamp itself but rather *percent of vote counted when a given voting booth's numbers were verified in the preliminary results system*. This transformation of the underlying time variable conveys an important advantage: while the time stamps themselves have long tails, the percent of vote counted is distributed nearly uniformly between zero and one.[12]

[10] Actually, as discussed in detail in the following section, only 33,038 tally sheets made it into the preliminary results system; the remaining 1,513 have no preliminary results system time stamps.

[11] 99.4% of tally sheets that appear before the shutdown in the public data have time stamps within three minutes (that is, within rounding) of our private time stamps. However, there are also 453 more tally sheets before the shutdown in our data than in the public (website) data: in the public data, there are 28,975; in our data, there are 29,428. See Appendix Figure A.1.

[12] It is not *exactly* uniform because voting booths are not all the same size. A voting booth verified after 10% of *voting booths* were verified is not the same as a voting booth verified after 10% of the *vote* verified.

The graphs therefore visualize how vote shares change as the overall preliminary results tally progressed—not as time itself progressed. However, this transformation also entails an important drawback: while the underlying time variable is discontinuous, the transformed variable is continuous, effectively smoothing over gaps in the actual time series. We discuss this issue in Appendix A.

3 Results

OAS (2019*a*), Escobari and Hoover (2019), and Newman (2020) interpret certain patterns in Bolivia's electoral returns as suggestive of electoral malpractice. We do not. This disagreement stems from three key differences:

(1) The OAS claimed to find a suspicious jump in MAS's vote share after 95% of the vote had been counted (p. 88).

   (a) We believe that a coding error led the OAS to mistakenly drop observations corresponding to the last 4.1% of the vote. When these observations are included, there is no jump at 95%.

   (b) Moreover, the apparent jump in the truncated sample—that is, the sample excluding the last 4.1% of the vote—is the artifact of using an estimator not designed for regression discontinuity analysis. When we follow best practices (Calonico, Cattaneo and Titiunik, 2014), we find no jump at that point (or, for that matter, at other potentially concerning moments in the count).

(2) Escobari and Hoover (2019) find what they deem a suspicious pattern: late-reporting voting booths favor MAS more than early-reporting voting booths *from the same precinct* (that is, from the same polling place); Newman (2020) presents a related result, also interpreting it as suggestive of fraud. We show that these pre-post differences are in fact the product of a secular within-precinct trend. We offer two possible explanations for this secular trend; neither of them involves centralized tampering with the tally.

(3) The OAS also noted (p. 87) that the shift in vote share accelerated after 7:40 p.m. on election night, when the government suspiciously stopped publishing results from the preliminary results system. We find that we can project the overall shape of the shift—including changes in slope—using a combination of (a) electoral returns from the previous poll (2016), which the OAS endorsed, and

(b) results reported early on election night, before the preliminary results system interruption. It is therefore possible to explain the shift without invoking electoral manipulation.

3.1 The jump at 95%: Artifact of mistakes?

It is not obvious to us whether—or under what conditions—observers should interpret a discontinuous change in vote share as evidence of fraud. On the one hand, we can easily generate innocuous explanations for these jumps; for example, if all of Philadelphia were to submit results at the same moment, the trend in Democratic vote share in Pennsylvania would undoubtedly be discontinuous at that point. On the other hand, it is at least as easy to construct theories of fraud that would produce jumps in the vote-share trend. Key references on election forensics do not mention discontinuous changes (e.g. Hicken and Mebane, 2017; Alvarez, Hall and Hyde, 2009). Regardless, in the election we study, we find that vote-share trends were in fact continuous at the point emphasized by the OAS.

The first difference between our analysis and that of OAS (2019*a*) stems from how we treat the set of voting booths that never appeared in the preliminary results system. There are 34,555 voting booths in the *Cómputo*, which is to say, 34,555 booths in the final, verified, legally binding tally of votes. Of these, 1,513 booths (4.4%) do not have time stamps in the preliminary results system (TREP).[13] The OAS treats these booths as "late reporters" (p. 86), under the assumption that they finished tallying only after the preliminary results system closed.[14] Operationally, this means 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 at the end, presumably in a random order.

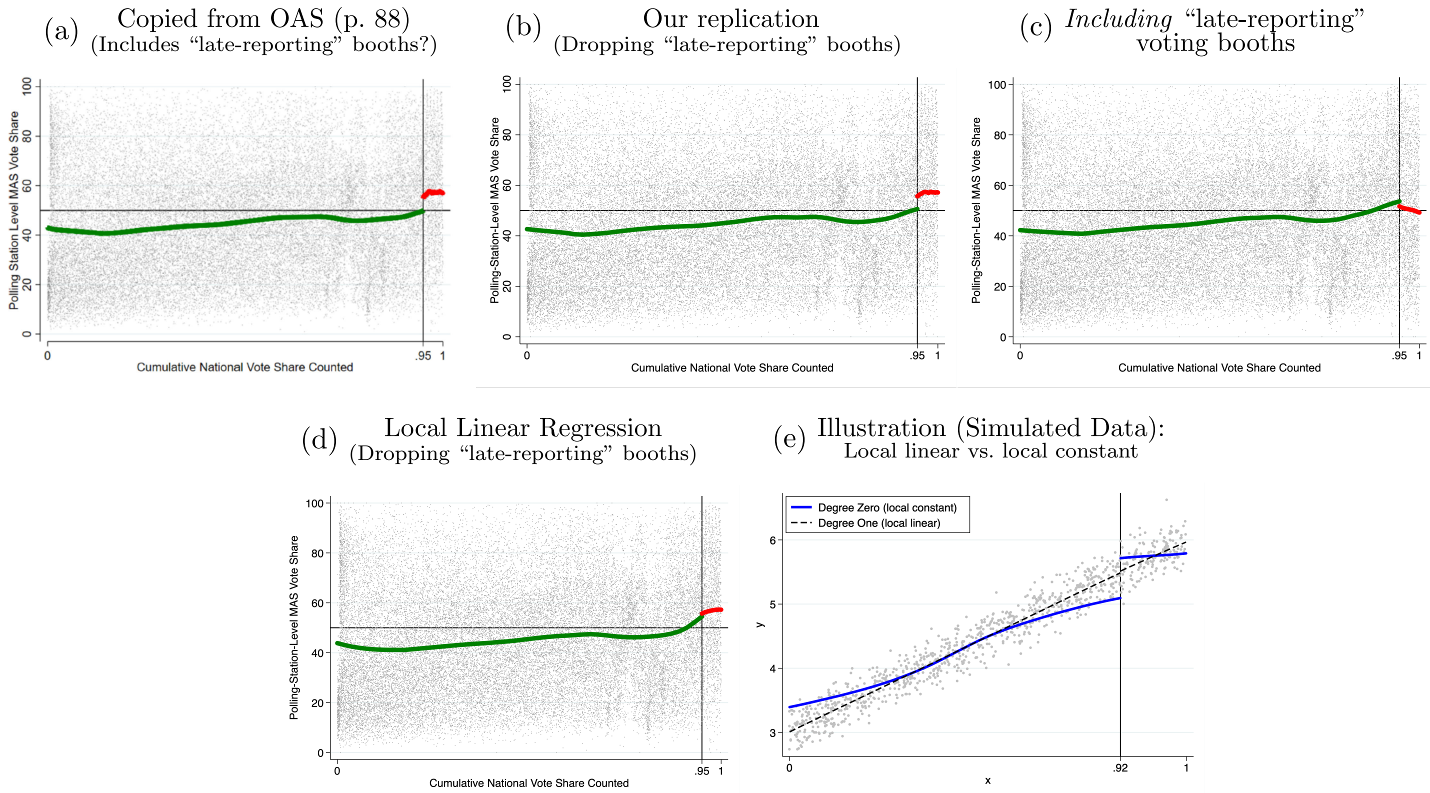
When we follow this approach, we cannot replicate the OAS results. We can only replicate the OAS results when we instead *drop* the 1,513 booths without preliminary results system time stamps (for brevity, we refer to these as "late-reporting" voting booths in Figure 2). The difference is significant: these booths account for 4.4% of tally sheets and 4.1% of votes, which is to say, the vast majority (82%) of the last

[13] OAS (2019*a*) report 1,511 precincts without time stamps in the TREP data.

[14] The report states: "All the analysis conducted below include these additional polling stations. Since they were not included in the TREP, they are treated as being late reporters" (p. 86). Irfan Noorudin confirmed over the phone that they meant to append the "late-reporting" booths to the end of the preliminary results system data.

Figure 2: Analytic mistakes and the jump at 95%

Figure (a) is taken directly from OAS (2019*a*) (p. 88). Figure (b) presents our replication, dropping the voting booths without preliminary results system time stamps. When we include them as late reporters, as the OAS text claims to do, we obtain Figure (c). Figure (d) shows that the apparent jump in the truncated sample disappears when we use a degree-one local polynomial rather than degree-zero; Figure (e) illustrates why, using simulated data.



The gray dots mark the underlying raw data. The lines mark nonparametric fits using the Epanechnikov kernel and a grid of equally spaced points; Figures (b)–(c) use local constant regression (degree-zero local polynomials), while Figure (d) uses local linear regression (degree-one local polynomials). In Figure (b), we selected bandwidths arbitrarily to visually match the OAS (Figure (a)); the bandwidths are 0.0475 before the cutoff and 0.0075 after. In (c)–(e), we use the rule-of-thumb bandwidths from Fan and Gijbels (1996, p. 110–113).

5% of votes counted (if we assume, as the OAS does, that they were late reporters). Any study focused on vote share trends in the last 5% of votes counted will therefore be quite sensitive to the treatment of the booths without preliminary results system time stamps.

Figure 2 describes the consequences of these voting booths for a key piece of evidence in the OAS report: the apparent jump in MAS's vote share after 95% of the votes were counted. Consider first Figure 2a, which copy-pastes the image published in the OAS report (p. 88). The OAS presented this jump as anomalous; Irfan Noorudin, the analyst who conducted the quantitative analysis for the report, wrote in subsequent commentary that "a sharp discontinuity around an arbitrary point such as the 95 percent threshold demands explanation" (2020).

In Figure 2b, we nearly replicate this key figure—but only by *dropping* the booths without preliminary results system time stamps.[15] Figure 2c shows what happens when we include them at the end, as the OAS claims to do. 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.

Numbers presented in the text of the OAS report itself also suggest that the "late-reporting" voting booths were mistakenly excluded from the discontinuity analysis. They note (p. 89) that Morales obtained 128,025 of 247,025 votes in these voting booths which is to say, 51.8%.[16] But their graph (reproduced in Figure 2a) shows Morales's vote share in the last 5% of votes somewhere clearly above 55% (indeed, in our near-replication in Figure 2b, Morales earns 56.8% of the last 5% of the vote). These two facts are irreconcilable. If Morales earned 51.8% of the vote in the "late-reporting" booths (as the OAS text reports), and given that these booths represent the vast majority (82%) of the last 5% of the vote (as the OAS text also reports), he would have had to obtain 79.6% of the vote in the other last-5% voting booths in order to earn 56.8% in that last 5% overall. He did not. (He earned 56.6%, on average, in these other booths.)

It is far from obvious that the voting booths that did not appear in the preliminary results system *should* be appended to the end of the preliminary results data. We therefore also study the apparent jump at 95% in the truncated sample (that is, the sample without the non-preliminary-results-system booths).

This jump is itself the artifact of using an estimator not designed for estimating discontinuities. We suspect that the OAS created Figure 2a using a Stata command called `lpoly`, separately for the data to the left and right of 0.95. This is how we nearly replicate their graph (Figure 2b). This approach is not appropriate for the OAS's purpose, for two reasons. First, `lpoly` by default estimates local regressions

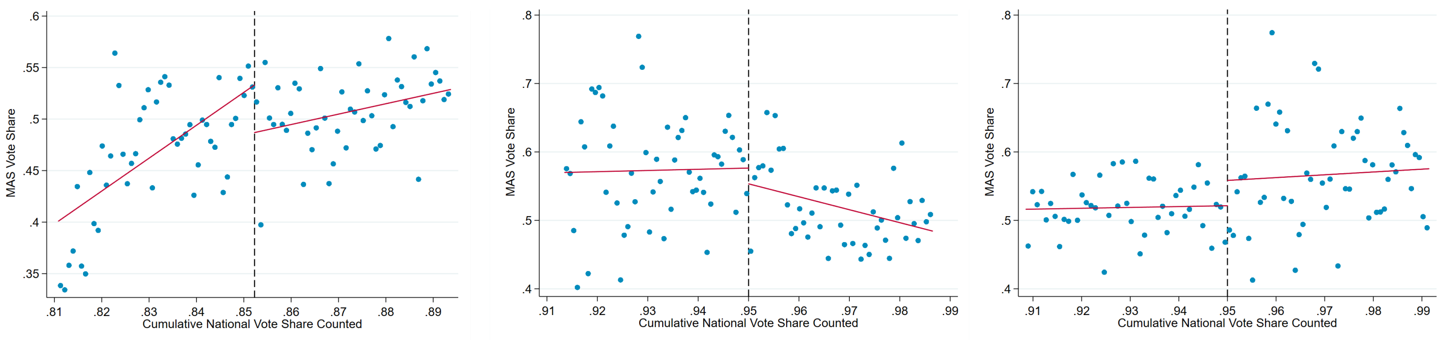
[15] In Figure 2, we include the blank and null (spoiled) ballots in the denominator, because this allows us to most closely replicate the figures in the OAS report. However, our own analysis of the vote margin (below) excludes blank and null ballots, because the Bolivian government excludes them when calculating the final margin.

[16] Our data agree; in our data, Morales obtains 128,315 of 248,473 votes in the voting booths that do not appear in the preliminary results system. The slight differences are likely due to the fact that we have 1,513 such voting booths, while the OAS reports 1,511.

Figure 3: The Absence of Discontinuities at Three Points

The points mark means of MAS's vote share in bins of 0.1 points (one tenth of one percent); the lines mark local polynomial fits with a triangular kernel and the bandwidth proposed by Calonico, Cattaneo and Titiunik (2014).

(a) At the shutdown (10/20, 7:40 p.m.) (b) At the true 95% (95% of full sample) (c) At the OAS cutoff (95% of truncated sample)



at each of fifty evenly spaced grid points—but (again, by default) the first and last grid points are not located at the boundaries of the support of the data. In other words, Figure 2a (and Figures 2b, 2c, 2d, and 2e) estimate local regressions at points far from the cutoff of 0.95, visually extrapolating those estimates to the edges. This is problematic for studying what happens precisely at 0.95, which is the objective of regression discontinuity analysis.

Second, the default settings for Stata's `lpoly` use local polynomials of degree zero (i.e., local constant regression). Degree-zero local polynomials often fail to fit the data well at boundary points (that is, at the edges). This "boundary bias" problem is well understood in statistics: "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, 2019, p. 38).[17] In Figure 2d, we use a local polynomial of degree one (i.e., local linear regression); this change alone is sufficient to eliminate the appearance of a jump. Figure 2e illustrates why, using simulated data in which there is no discontinuous change in the outcome. By construction, degree-zero local polynomials artificially flatten the slope of data that trend (linearly) upward or downward.[18]

[17] 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.

[18] Applying a local constant estimator (such as that used by the OAS) to data from other elections

Table 1: Non-Parametric Regression Discontinuity Estimates

Estimates of discontinuities at three points (Calonico, Cattaneo and Titiunik, 2014).

					Robust		Observations	
Cutoff	Date & Time	Sample*	RD Estimate	BW	p-val	C.I.	Left	Right
0.852	10/20/2019 19:40:57	Full	-0.024	0.041	0.511	[-0.072, 0.036]	1,368	1,367
0.950	10/21/2019 14:54:25	Full	-0.008	0.036	0.675	[-0.039, 0.061]	1,271	1,332
0.950	10/20/2019 20:03:59	Truncated	0.031	0.041	0.467	[-0.036, 0.079]	1,325	1,378

* *Truncated* refers to the sample used by the OAS, which excludes the voting booths without time stamps in the preliminary results system. This is thus the threshold analyzed by OAS (2019*a*).

Calonico, Cattaneo and Titiunik (2014) propose a now widely used data-driven regression discontinuity estimator to address these and other problems. This approach estimates the treatment effect by running two local linear regressions precisely at the cutoff (one to the left, one to the right).[19] We use this estimator to formally test for discontinuities at three points: (1) 7:40 p.m. on election night, when the government stopped publishing updated results; (2) the point originally studied by the OAS (95% of the truncated sample); and (3) 95% of the full sample. We cannot reject the null of continuity at any of these points.

Figure 3 presents graphs of MAS's vote share at these three moments. The dots mark average MAS vote share in 0.1-point bins; the lines plot the estimated local polynomial of degree one, with optimal bandwidth and a triangular kernel (for a comprehensive discussion, see Cattaneo, Idrobo and Titiunik, 2019). None of the three graphs presents visual evidence of a treatment effect.

To estimate the size of the treatment effects in Figure 3, and to test whether they are statistically distinguishable from zero, we use the estimator proposed by Calonico, Cattaneo and Titiunik (2014). Table 1 reports the results. At all three cutoffs, the estimated treatment effect is statistically indistinguishable from zero. In Appendix

in the region would also create the false appearance of jumps in the vote-share trend. In Brazil, for example, in the first-round presidential poll held on October 7, 2019, the vote share of second-place candidate Fernando Haddad increased as the count progressed. Using a local constant estimator, we show in Appendix Figure E.2a that his vote share would appear to jump discontinuously after 97% of the vote had been counted. Again, using a local linear estimator eliminates the appearance of a jump (Appendix Figure E.2b). The OAS mission observing this election praised Brazilian electoral authorities: "The Mission congratulates the Brazilian electoral authorities for their results transmission system, which gives citizens fast access to official information, contributing to the certainty of the process" (OAS, 2018, p. 3).

[19] Calonico, Cattaneo and Titiunik (2014) also use a triangular kernel rather than Epanechnikov, which is the default in Stata's `lpoly` (see also Calonico, Cattaneo and Farrell, 2018).

Table D.2, we show that this result is robust to various choices of polynomial degree and bandwidth.

3.2 Within-precinct variation

In papers that echo the OAS's concerns about the 2019 election in Bolivia, Escobari and Hoover (2019) and Newman (2020) study within-precinct variation in MAS's vote share. Specifically, they note that MAS performed better in voting booths reporting after the government stopped publishing updated results (*post-shutdown*) than in voting booths *from the same precinct* that reported earlier (*pre-shutdown*). We replicate this finding, but interpret it differently. Escobari and Hoover view the within-precinct variation as evidence of "a statistically significant case of electoral fraud" (p. 1); Newman (2020) interprets a related pattern as evidence that "the OAS findings were correct" (p. 1).

In our view, these inferences are incorrect. The analysis in Escobari and Hoover (2019) and Newman (2020) compares two periods (*pre* and *post*) without accounting for a secular trend. We show in this section that the within-precinct increase in MAS's vote margin begins early on election night—well before the 7:40 p.m. suspension of the publication of electoral results. Accounting for this secular trend eliminates the appearance of an anomalous within-precinct pre-post difference in vote shares. We first demonstrate this, and then turn to two possible explanations for the secular within-precinct trend—neither of which requires centralized tampering with the tally.

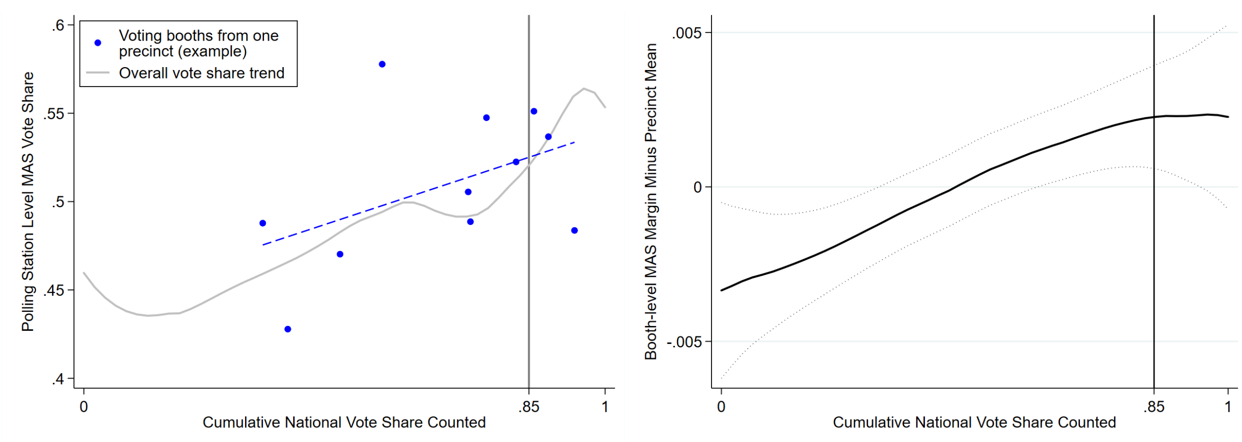
The secular within-precinct trend. Before studying the within-precinct variation in reporting time, we note that this variation is substantial relative to the overall reporting window. Almost all voting booths transmit preliminary results system figures between 7:00 p.m. and 9:00 p.m. on election night (see Appendix Figure A.2). 70% of precincts have more than one voting booth; among these precincts, the median within-precinct standard deviation in reporting time is 35 minutes (i.e., more than one fourth of the active reporting window). 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 4a presents an example of within-precinct variation; the blue dots mark MAS's vote share in each of the 11 voting booths in a single precinct in the town of Guanay. In this example, MAS's vote share increases with reporting time even before the

Figure 4: Within Precincts, MAS Vote Share Increases with Reporting Time

Figure (a) provides an example of within-precinct variation; the blue dots mark MAS's vote share in each of the 11 voting booths in a single precinct in the town of Guanay. Figure (b) plots the voting-booth-level MAS margin after subtracting the precinct mean (i.e., the residual margin $\mathrm{m}_{bp} - \overline{\mathrm{m}}_p$).

(a) Within-Precinct Variation: An Example (b) Residual MAS Margin ($\mathrm{m}_{bp} - \overline{\mathrm{m}}_p$)



Local linear fits with the rule-of-thumb bandwidth from Fan and Gijbels (1996, p. 110–113), Epanechnikov kernel.

government stopped transmitting updated results (at 7:40 p.m., with 85.2% of the vote verified). 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 4b reveals that the residual MAS vote margin $\mathrm{m}_{bp} - \overline{\mathrm{m}}_p$ increases with reporting time.

Critically, the within-precinct divergence between MAS and CC does not accelerate either after the shutdown of the public preliminary results system (at 7:40 p.m.) or after 95% of the votes were counted (Figure 4b). 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 E.4).

The time trend in Figure 4b highlights a problem with the interpretation of results in Newman (2020) and Escobari and Hoover (2019). Newman (2020) shows that MAS's margin was slightly higher in voting booths reporting after the shutdown, compared to those reporting before the shutdown, even when restricting the sample to precincts with voting booths both before and after (though the difference in distributions is not statistically distinguishable from zero; p. 12, 14). But Figure 4b shows that this within-precinct growth in MAS's margin began *prior* to the shutdown, suggesting that Newman's test would detect non-zero differences at many other cut points, too.[20]

[20]Newman (2020) also shows that the pre-post difference in distributions *is* statistically significant

Escobari and Hoover regress MAS's vote 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 of the coefficient is consistent with our Figure 4b; it reveals that MAS's post-shutdown vote margin was approximately four tenths of a percentage point larger than MAS's pre-shutdown margin. But the Escobari and Hoover (2019) specification does not account for the secular trend in Figure 4b: 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:

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

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 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 2 reports estimates of a version of Equation 1 that excludes precinct fixed effects. Consistent with Figure 4a, when we *ignore* precinct characteristics, 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 4b, the growth in MAS's margin slows after the shutdown (the estimate of β_3 is negative but statistically indistinguishable from zero).

Column (2) of Table 2 also reveals that, *even within precint*, there is a secular increase in MAS's margin over the reporting window (see also Figure 4b). This is captured in the positive and significant coefficient on β_1. And this is the problem with the conclusions of Newman (2020) and Escobari and Hoover (2019): if we omit that secular trend, as in Column (3), then the coefficient on the *post shutdown* is naturally

in a specific subsample. But as Rosnick (2020) points out, the subsample was defined in a way that creates pre-post differences *by construction*.

Table 2: Within Precinct, MAS Margin Does Not Grow Faster Post-Shutdown

Estimates of Equation 1. The dependent variable is MAS's margin over Civic Community (scaled −1–1). 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
β_1: Reporting time percentile[†]	0.173	0.014		0.013
	(0.02)	(0.003)		(0.003)
β_2: Post shutdown (0/1)	0.102	0.004	0.006	0.000
	(0.02)	(0.003)	(0.002)	(0.002)
β_3: Percentile $\times$ Post	-0.019	-0.052		
	(0.2)	(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 B 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.

positive and significant.[21] 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.

Two possible explanations. This analysis raises a question: why is there a secular within-precinct trend in MAS's vote margin? Why do voting booths counted later favor MAS more than voting booths *from the same precinct* that were counted earlier? We offer two possible explanations.

[21]The estimate in Column (3) of Table 2 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 B.

As noted above in the Context section, voting-booth jurors (*jurados*) are chosen randomly from among that 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 the same 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 (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, for example, two clusters of last names: those that begin with the letter C, which includes 15.9% 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 completely unsurprising if MAS performed better in the C voting booth than in the $R + S$ voting booth (ethnicity is correlated with political preferences, Madrid, 2012, p. 69–72); nor would it be surprising if the C voting booth reported later than the $R + S$ voting booth.

One implication of this 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 E.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.

3.3 The trend around the information blackout

Even in the absence of discontinuities in MAS's vote share, Figure 4a does seem to reveal a strange nonmonotonicity in the vote-share trend. Sometime after 7:00 p.m. on the evening of the election, the preliminary results system reported that MAS's lead over Civic Community (CC) had begun to fall after steadily climbing all afternoon. Later, the trend reversed, and MAS's margin began to rise again. Right around this time, the government suspended the publication of figures from the preliminary results system (see the Context section for details).

OAS (2019*a*, p. 86) express concern about "the steep slope of that line" (that is, the slope of the trend MAS's margin) after 7:40 p.m. on election night. We show in this section that we can predict the shape of the trend—and the final margin—using data from the previous poll (2016), together with data from early-reporting voting booths in 2019. While this does not establish the absence of fraud, it does imply that we can explain key features of the vote-share trend without invoking fraud. (Unless, of course, one were to conjecture that the same type of fraud occurred in the 2016 referendum, a yes-or-no vote on Evo Morales's proposed changes to the constitution. The OAS electoral observation mission made no reference to electoral manipulation in its reports on the 2016 election (2016*a*; 2016*b*), though in 2016 the OAS did not conduct an audit—and the composition of the electoral tribunal changed between the two elections. Incidentally, MAS lost the 2016 referendum: voters defeated the proposed constitutional amendments, 51% to 49%.)

Before developing our projection of the vote margin in late-reporting precincts, we present a simple exercise that illustrates how the 2016 data can help us draw inferences about 2019. It is not possible to match *voting booths* across elections, because of how the booth identifiers changed. However, we can match *precincts* across elections (Minoldo and Quiroga, 2020, show a high correlation between 2016 and 2019 precinct-level vote shares). We then calculate *average precinct-level MAS vote margin* for each voting booth in each election ($\overline{\mathrm{m}}_p$, using the notation of the previous section), and plot these average precinct-level margins against each voting booth's 2019 reporting time percentile.[22]

Figure 5a presents the result: the shape of the vote-share trend appears nearly identical if we use 2016 vote margins rather than 2019 vote margins. In other words, features that the OAS flagged as anomalous in 2019 also emerge in analysis of data from 2016, an election for which the OAS congratulated Bolivia and praised the leadership of the electoral authority (OAS, 2016*a*,*b*).

While Figure 5a is suggestive, it does not allow us to evaluate whether we can predict MAS's final 2019 margin—10.56%—on the basis of (a) the 2016 electoral returns, and (b) early-reported results on election night in 2019. By "early-reported results," we mean voting booths that reported before the government stopped publishing updated results, at 7:40 p.m.; we refer to these observations as *pre-shutdown*, and the rest as *post-shutdown*. We predict the MAS margin M_{bp} in each post-shutdown voting booth b in precinct p as follows:

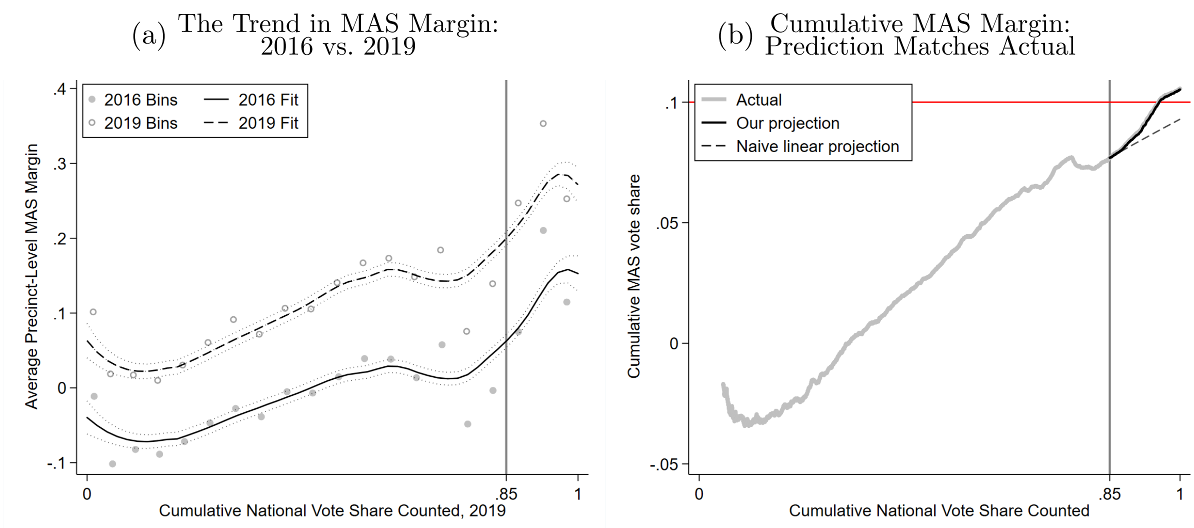
$$\widehat{\mathrm{M}}_{bp} = \widehat{\mathrm{m}}_p + \widehat{\mathrm{e}}_{bp} \tag{2}$$

where $\widehat{\mathrm{m}}_p$ is a prediction of the average margin in precinct p, and $\widehat{\mathrm{e}}_{bp}$ is a prediction

[22] To be clear, the y-axis values in Figure 5a are not the voting-booth-specific MAS margins m_{bp}, but rather the average precinct-level MAS margin $\overline{\mathrm{m}}_p$. In other words, all booths b in a given precinct p have the same y-axis value in Figure 5a. This is not the case in the prediction exercise below.

Figure 5: 2016 Electoral Returns Reveal Similar Patterns

Figure (a) plots average precinct-level MAS vote margins $\overline{m}_p$ in two elections against each voting booth's percentile of reporting time in 2019 (i.e., the x-axis values are the same). Figure (b) plots the actual cumulative MAS margin in 2019 together with our predictions (Equation 2), as well as predictions from a naive linear regression.



In (a), we only include observations corresponding to voting booths present in both 2016 and 2019 (i.e., the samples are the same). In (a), lines mark local linear fits using the rule-of-thumb bandwidth from Fan and Gijbels (1996, p. 110–113); the dotted lines mark 95% confidence intervals, and the bins are obtained using Cattaneo et al. (2019).

of voting booth b's deviation from that precinct mean. We predict the former (m_p) based on precinct-level returns from 2016, and we predict the latter (e_{bp}) based on pre-shutdown data from 2019.[23]

To generate $\widehat{\mathrm{m}}_p$, our prediction of MAS's average margin in precinct p, we first divide the 5,296 precincts into fifty strata based on geography and precinct size.[24] Within each stratum, we *use only the precincts in which all booths reported before 7:40 p.m.* to estimate:

$$\mathrm{m}_p = \alpha + \beta \mathrm{m2016}_p + \epsilon_p \tag{3}$$

[23]As noted above, we can not match observations across elections at the voting booth level; only at the precinct level. Otherwise we could use 2016 data to predict M_{bp} directly. Nor do we have time stamps for the 2016 data.

[24]There are ten geographic groups and five precinct-size groups (50 strata total). Nine of the geographic groups correspond to Bolivia's nine departments; the tenth includes all precincts abroad (embassies, consulates, etc.). The five precinct-size groups (based on 2016) are: (1) precincts with only one voting booth, (2) precincts with two voting booths, (3) precincts with 3–5 voting booths, (4) precincts with 6–9 voting booths, and (5) precincts with 10+ voting booths.

where m_p is MAS's margin in precinct p in 2019, and $\mathrm{m2016}_p$ is MAS's margin in precinct p in 2016. We then calculate:

$$\widehat{\mathrm{m}}_p = \widehat{\alpha} + \widehat{\beta}\mathrm{m2016}_p \tag{4}$$

where, again, the coefficients $[\widehat{\alpha}, \widehat{\beta}]$ are stratum-specific.[25]

These predictions alone—$\widehat{\mathrm{m}}_p$, the precinct-level average MAS margin—produce a fairly good projection of the cumulative final margin, as we show below. However, as noted in the previous section, within-precinct variation also contributes (a bit) to the final margin. For completeness, we therefore also generate $\widehat{\mathrm{e}}_{bp}$, predictions of voting booth b's deviation from the within-precinct mean m_p. Using *only the pre-shutdown 2019 data*, we estimate:

$$\mathrm{m}_{bp} = \gamma_p + \phi\,(\text{Time Percentile})_{bp} + \eta_{bp} \tag{5}$$

where m_{bp} is MAS's margin in booth b in precinct p, γ_p are precinct fixed effects, and $(\text{Time Percentile})_{bp}$ is voting booth b's reporting time percentile. Again, we only estimate Equation 6 using *pre-shutdown* observations. Note that this equivalent to estimating the de-meaned equation:

$$\left[\mathrm{m}_{bp} - \overline{m}_p\right] = \phi \times \left[\,(\text{Time Percentile})_{bp} - \overline{\text{Time Percentile}}_p\right] + \nu_{bp} \tag{6}$$

We then calculate:

$$\widehat{\mathrm{e}}_{bp} = \widehat{\phi} \times \left[\,(\text{Time Percentile})_{bp} - \overline{\text{Time Percentile}}_p\right] \tag{7}$$

That is, we predict voting booth b's deviation from the precinct average m_p based on (a) booth b's deviation from the precinct average time reporting percentile and (b) the within-precinct time trend in the pre-shutdown period. Of course, Equation 6 imposes linearity on the within-precinct time trend; we show in Appendix Figure C.1 that this restriction is not unreasonable. In any case, as noted above, the vote-share trend is almost entirely driven by cross-precinct variation rather than within-precinct variation; when we conduct this exercise using only $\widehat{\mathrm{m}}_p$ and ignoring $\widehat{\mathrm{e}}_{bp}$, the results

[25] For 363 post-shutdown observations (7%), we do not observe 2016 data. For these new precincts, we obviously cannot generate $\widehat{m}_p$ according Equation 4. Instead, we generate predicted values using (a) the time trend in $\widehat{m}_p$ for the other precincts (i.e., old precincts), and (b) the pre-shutdown vote margins in the new precincts. See Appendix C for details.

Table 3: Predicting the Final Margin with 2016 Returns & Early-Counted 2019 Results

Description	Estimate
Actual cumulative margin	10.56
Predicted cumulative margin, using $\widehat{\mathrm{M}}_{bp} = \widehat{\mathrm{m}}_p + \widehat{\mathrm{e}}_{bp}$	10.53
Predicted cumulative margin, using only predicted precinct means $\widehat{\mathrm{m}}_p$	10.49
Predicted cumulative margin, naive linear extrapolation	9.29

are quite similar.

To calculate our projection of the cumulative final margin, we weight the predictions $\widehat{\mathrm{M}}_{bp}$ (or $\widehat{\mathrm{m}}_p$) by the actual, observed number of votes in each voting booth. This approach allows us to predict MAS's final vote margin almost exactly, as Figure 5b reveals. Our predicted booth-level margins $\widehat{\mathrm{M}}_{bp}$, weighted by turnout, cumulate to produce a projected final margin of 10.53, within three hundredths of a point of the actual final margin (10.56). If we use only the cross-precinct variation $\widehat{\mathrm{m}}_p$, ignoring the within-precinct residual predictions $\widehat{\mathrm{e}}_{bp}$, we obtain a predicted final margin of 10.49. Table 3 summarizes these results.

This is all to say that the 2016 electoral returns, together with early-reporting voting booths in 2019, are sufficient to predict the final outcome. This result does not establish the absence of manipulation of vote shares in the late-reporting booths in 2019; rather, it implies that we do not *need* electoral manipulation in order to explain MAS's first-round victory.

4 Conclusion

The OAS and other researchers have used three quantitative results to question the integrity of the Bolivian presidential election of October, 2019: (1) an apparent jump in the incumbent's vote share after 95% of the vote had been counted, (2) comparisons across voting booths within the same precinct, and (3) acceleration in the growth of the incumbent's lead after 7:40 p.m. on election night, when the government stopped publishing updated results. We revisit the evidence, finding that: (1) the jump does not exist; (2) a secular trend explains the within-precinct results; and (3) we can predict the post-7:40-p.m. results almost exactly using data from the previous poll, which the OAS endorsed.

Our analysis does not establish the absence of fraud in this election; that could never be determined on the basis of quantitative analysis alone. The quantitative results that we revisit formed just one part of the OAS's case against the integrity of the Bolivian election. Their team presented evidence of secret servers, improperly completed tally sheets, undisclosed late-night software modifications, and myriad other reasons for suspicion.

But while quantitative evidence was merely one of the findings of the OAS audit report, it played—and continues to play—an outsize role in Bolivia's political crisis. It helped convict Morales of fraud in the court of public opinion. We find that this key piece of evidence is faulty and should be excluded.

Our findings also speak to a general problem in election administration. Governments rarely announce election results all at once; instead, they release partial tallies as they trickle in, telling the public how things stand with (e.g.) 30% of precincts reporting, 70%, 90%, and so on. These updates create transparency and respond to the public's demand for information. But they also entail an important and seldom-studied cost: raising false hope. This is dangerous, because dashed hopes can spark conflict.

Incremental reporting of results thus creates a tradeoff between transparency and certainty. One way to lower the costs of transparency is to study shifts in late-counted votes, weighing fraud against more innocuous explanations. Researchers have done this work for the United States (Foley, 2013; Foley and Stewart, 2020; Li, Hyun and Alvarez, 2020), but, to the best of our knowledge, we are the first to do so elsewhere. In Brazil, for example, the left candidate in the 2018 presidential election earned just 25% of votes counted early but more than 40% of votes counted late. In the Colombian presidential election that same year, Gustavo Petro fared far better as election night progressed. Do these trends stem from regional variation in the order in which votes are counted? Or from changes in the mix of urban and rural ballots? Distinguishing these mechanisms can help protect the legitimacy of the electoral process.

Our findings suggest opportunities for future work. First, future studies could investigate the conditions under which electoral observers use quantitative analysis to study electoral integrity; as we note, the quantitative indicators applied to the Bolivian case would have revealed similar patterns in (e.g.) Brazil, or in the previous poll in Bolivia, both of which were endorsed by OAS missions. Second, voting technology in many countries is better suited to documenting shifts in late-counted votes than voting technology in the United States; comparative evidence on the magnitude of these shifts would provide important perspective on the Bolivian and U.S. cases. Finally, comparative work could assess which (if any) characteristics of shifts in late-counted votes *should* be interpreted as evidence of possible fraud.

References

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

Aguilar, Wilson. 2018. "Renuncia de Katia Uriona deja más sombras en el TSE y las primarias." *Los Tiempos* .

Alvarez, Michael R., Thad E. Hall and Morgan H. Llewellyn. 2008. "Are Americans Confident Their Ballots Are Counted?" *Journal of Politics* .

Alvarez, R. Michael, Ines Levin, Julia Pomares and Marcelo Leiras. 2013. "Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions." *Political Science Research and Methods* 1(1):117–137.

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.
URL: *https://books.google.com/books?id=HeUyo5RcI7wC*

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

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

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.

Berger, Jonah, Marc Meredith and S. Christian Wheeler. 2008. "Contextual priming: Where people vote affects how they vote." *Proceedings of the National Academy of Sciences* 105(26):8846–8849.

Birch, Sarah. 2010. "Perceptions of Electoral Fairness and Voter Turnout." *Comparative Political Studies* .

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.

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.
URL: *https://doi.org/10.1080/01621459.2017.1285776*

Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* .

Cambara Ferrufino, Pablo Cesar. 2019. "Costas atribuye a una 'impericia' la paralización de la transmisión de votos del TREP." *El Deber* .

Carothers, Thomas. 1997. *The Rise of Election Monitoring: The Observers Observed*. Journal of Democracy.

Cattaneo, Matias D., Nicolás Idrobo and Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs*. Cambridge University Press.

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

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* .

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

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.

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

Ethical Hacking. 2019. "Informe Consolidado de Producto 1 y Producto 2: Línea de Tiempo de la Consultoría al OEP." *Ethical Hacking* .

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.

Flores, Paola and Carlos Valdez. 2019. "Bolivia says OAS team to audit disputed election." *AP News* .

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.".
URL: *https://forebears.io/bolivia/surnames*

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

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* .

Hopkins, Daniel J., Marc Meredith, Michael Morse, Sarah Smith and Jesse Yoder. 2017. "Voting But for the Law: Evidence from Virginia on Photo Identification Requirements." *Journal of Empirical Legal Studies* 14(1):79–128.

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.
URL: *https://doi.org/10.1093/restud/rdr043*

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

Kanyinga, Karuti. 2009. "The legacy of the white highlands: Land rights, ethnicity and the post-2007 election violence in Kenya." *Journal of Contemporary African Studies* 27(3):325–344.

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.

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. 2020. “Why Do Election Results Change After Election Day? The ”Blue Shift” in California Elections.” *American Government and Politics* .

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* .

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

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

Meredith, Marc and Yuval Salant. 2013. “On the Causes and Consequences of Ballot Order Effects.” *Political Behavior* 35(175–197).

Mesa, Carlos D. 2019. “Post-election thread.” *Twitter* .

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

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* .

Noorudin, Irfan. 2020. “Yes, Bolivia’s 2019 election was problematic. Here’s why.” *Washington Post Monkey Cage* .

Norris, Pippa. 2014. *Why Electoral Integrity Matters.* Cambridge University Press.

Norris, Pippa, Holly Ann Garnett and Max Grömping. 2020. "The paranoid style of American elections: explaining perceptions of electoral integrity in an age of populism." *Journal of Elections, Public Opinion and Parties* 30(1):105–125.

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. 2018. "OAS Electoral Observation Mission Congratulates Brazil for Holding its General Elections and Announces that It Will Observe the Second Round." *Organization of American States* .

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*d*. "Statement of the OAS Electoral Observation Mission in Bolivia." *Organization of American States* .

OAS. October 23, 2019*c*. "Informe Preliminar de la Misión de Observación Electoral en Bolivia." *Organization of American States* .

OEP. 2019. "Cartilla de información pública: ¿Cómo funcionan los sistemas del TREP y del cómputo oficial?" *Órgano Electoral Plurinacional* .

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.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.21379*

Rosnick, David. 2020. "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* .

Simpser, Alberto. 2012. "Does Electoral Manipulation Discourage Voter Turnout? Evidence from Mexico." *The Journal of Politics* 74(3):782–795.

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

TSE. 2019. "Verificación del llenado del acta." *Tribunal Supremo Electoral de Bolivia Video* .

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.

Appendix

A Data: Details on time stamps

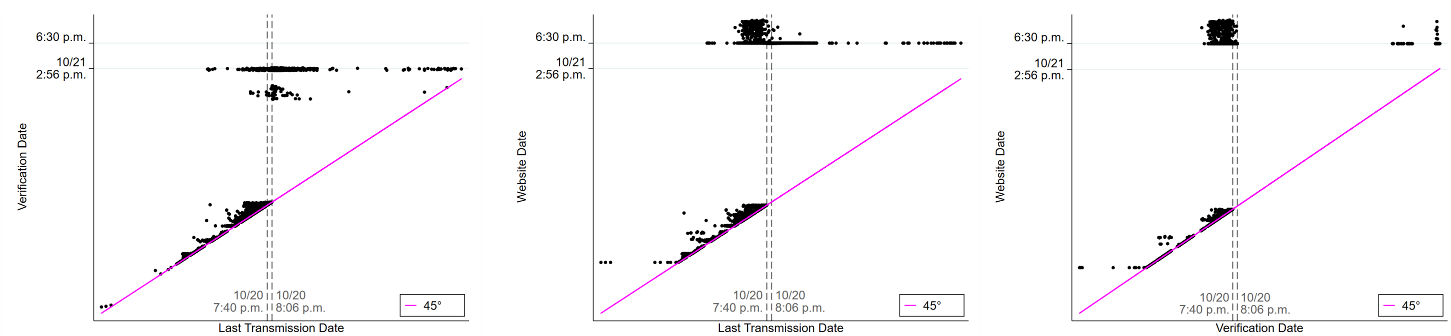
The preliminary results system (TREP) data contain five different time stamp variables: first registration, last registration, first transmission, last transmission, and verification.[26] (See the Data section for details on the preliminary results system.)

In the main text, we use *percentiles of verification date* as the time variable. We do this because it allows us to most closely replicate the figures in OAS (2019*a*); none of the other transformed time stamps generate graphs that look like those of the OAS. But *verification date* has two related drawbacks. First, it is itself discontinuous: the verification time stamps stop at 8:06 p.m. on election night and pick up again the following morning at 10:37 a.m. (October 21). Second, while the penultimate time stamp—last transmission date—is highly predictive of verification time prior to 8:06 p.m. on election night, this correlation breaks down after 8:06 p.m. That is, the tally sheets transmitted at 8:07 p.m., 8:08 p.m., etc., were not necessarily the first tally sheets verified on the morning of October 21.

Figure A.1: Website Times and Verification Times vs. Reporting Time

Figure (a) plots the *verification* time stamp—the one used in the main text—against the *last transmission* time stamp. The latter is continuous, whereas verification stops at 8:06 p.m. on election night and continues at 10:37 a.m. the next morning. Figure (b) plots the public (website) time stamp against the last transmission time stamp, revealing that all tally sheets transmitted after 7:40 p.m.—and many transmitted before—were published online at 6:30 p.m. the next day. Figure (c) plots the website time stamp against the verification time stamp, again revealing that many tally sheets *verified* soon after the polls closed were published online on the evening of the next day.

(a) Verification vs. Transmission (b) Website vs. Transmission (c) Website vs. Verification



All figures exclude one outlier, an observation that was not *verified* until October 22.

Figure A.1a visualizes both of these issues, plotting the *verification* time stamp against the *last transmission* time stamp. Again, the latter is continuous while the

[26] There is also a sixth, *approval*, which is missing for almost all observations.

former stops after 8:06 p.m. on election night. Moreover, while *last transmission* time strongly predicts *verification* time prior to 8:06 p.m., it does not predict verification time for those tally sheets verified the next day. Likewise, Figures A.1b and A.1c reveal that the order in which tally sheets were published online reflects the order in which they were transmitted only for a subset of early-reporting voting booths. Finally, A.2 clarifies that more than 85% of the vote was counted before the 8:06 p.m. interruption in verification.

Figure A.2: Progress of the Preliminary Results System (TREP)

This figure excludes the 4.4% of voting booths (4.1% of the vote) without TREP time stamps.

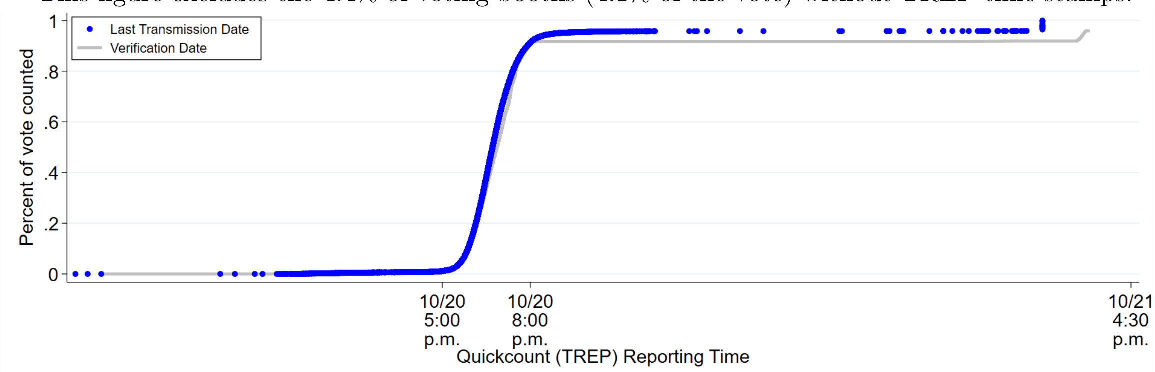
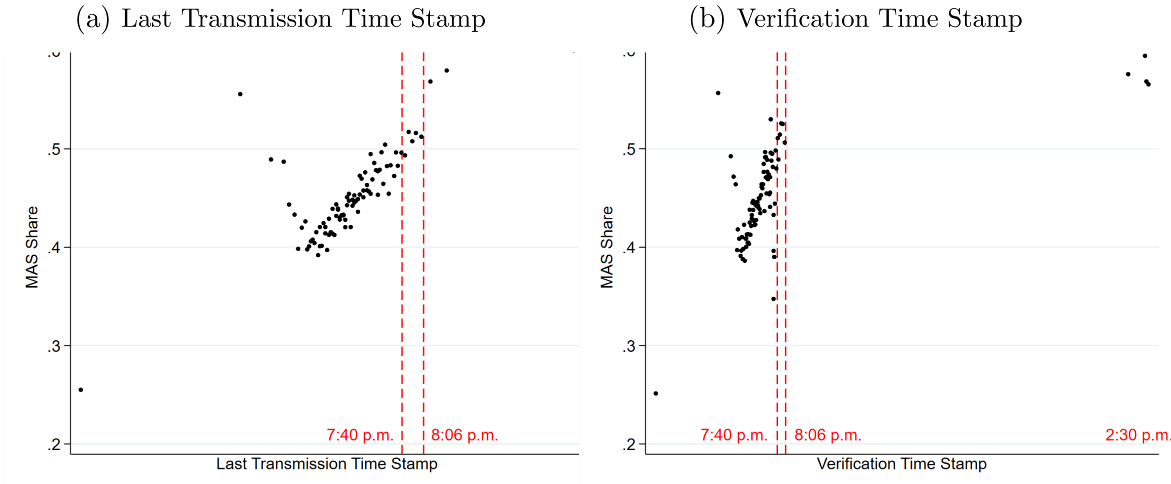


Figure A.3 shows how MAS's vote share changes as a function of clock time (rather than as a function of *percentile of vote verified*, which is what we study in the main text). The points mark average MAS vote share in optimal (data-driven) bins of the timestamp variables (Cattaneo et al., 2019). The vertical lines mark two times of interest: 7:40 p.m., when the government stopped publishing updated results; and 8:06 p.m., when the verification time stamps stop until the following morning. (As noted above, transmission time stamps continue through the night).

Figure A.3b highlights a potential problem with testing for discontinuities in these data. The discontinuity analysis uses *percentiles* of the verification time stamp, which is to say, a transformation of verification date that places the 8:06 p.m. time stamps right next to the 10:37 a.m. (next day) time stamps, effectively closing the long gap in the actual time series. Thus if we were to test for a discontinuity at 8:06 p.m. using percentiles of verification date, we would be testing for a discontinuity in MAS vote share at a moment when the running variable (time) is itself discontinuous. Worse, as noted above, the order in which tally sheets were verified on the morning of October 21 is only loosely related to the order in which they were transmitted the night before. As it happens, neither of the moments studied by the OAS—7:40 p.m., and 95% of

Figure A.3: Last Transmission Date, Verification Date, and MAS Vote Share

Figure (a) plots average MAS vote share in bins of the preliminary results system *last transmission date*, using the optimal (data-driven) bins (Cattaneo et al., 2019). Figure (b) plots average MAS vote share in optimal bins of the preliminary results system *verification date*.



the vote counted—coincide with 8:06 p.m. on election night (see Table 1), so this problem does not arise.

Using the last transmission time stamp, Figure A.3a reveals an apparently smooth trend in MAS vote share before and after 8:06 p.m. When we test for a discontinuity in MAS vote share at 8:06 p.m., using last transmission date and again following Calonico, Cattaneo and Titiunik (2014), the point estimate is positive (4.8 percentage points) but statistically indistinguishable from zero.

Overall, it strikes us that the last transmission time stamp better captures the reporting sequence, while the verification time stamp perhaps better captures the counting/tabulation sequence. Again, we focus on the verification time stamp in the main text because it allows us to replicate the OAS results. However, the substantive takeaway from our own analysis in Section 3.3 remains unchanged when we use the last transmission time stamp: MAS performs predictably well, *not* surprisingly well, in the period after the government stopped publishing updated results.

B 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 2, Column (3) in the main text—reproduced in Column (3) of Table B.1 below—do not exactly replicate Escobari and Hoover (2019). Our coefficient on *post* is estimated at 0.0056 (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 B.1, we can replicate their result almost exactly.

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

Estimates of Eqn. 1. 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)	0.0048 (0.0019)	-0.0018 (0.0023)	0.0056 (0.0019)	0.0001 (0.0023)	0.0038 (0.0018)	0.0036 (0.0018)
Reporting time percentile		0.0164 (0.0034)		0.0130 (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 B.1 uses the website count of valid votes; Column (6) uses the final count of valid votes. The point estimate changes by 0.0002.

C Projection: Additional details

Treatment of new precincts. For 363 of the post-shutdown observations (7%), we do not observe 2016 data. For these new precincts (new because they did not exist in 2016), we obviously cannot generate $\widehat{m}_p$ according Equation 4, which uses 2016 data. Instead, we generate predicted values using (a) the time trend in $\widehat{M}_{bp}$ for the other precincts (i.e., old precincts), and (b) the pre-shutdown vote margins in the new precincts. Specifically, for the 363 new precincts, we define:

$$\widehat{\mathrm{M}}_{bp}^{new} = \widehat{\alpha} + \widehat{f}(\text{Time Percentile})_{bp} \quad (8)$$

where $\widehat{f}(\text{Time Percentile})_{bp}$ is a nonparametric fit to the predicted values for voting booths in the other (i.e., old) precincts, $\widehat{\mathrm{M}}_{bp}$ (Equation 2); and $\widehat{\alpha}$ is the average difference in vote share between (i) voting booths in old precincts and (ii) voting booths in new precincts, *in the pre-shutdown period*. The inclusion of this term is important, because MAS's margin is considerably higher, on average, in new voting booths than in old ones. (The difference does not appear to change much over the reporting window.) This makes sense if, for example, new precincts were established in places with high population growth.

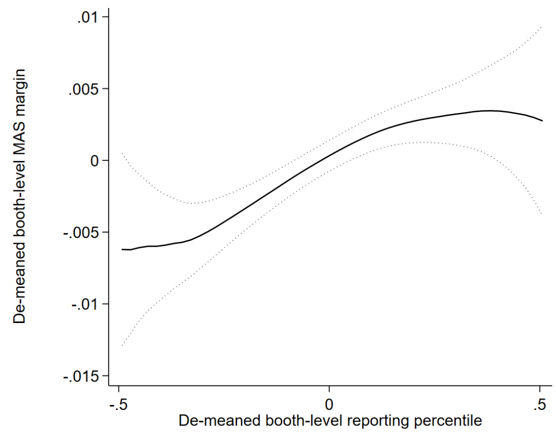
Functional form of the within-precinct variation. Our approach to predicting values for the within-precinct variation (Equation 7) imposes linearity on the (within-precinct) relationship between MAS margin and reporting time. Figure C.1 reveals that this restriction is not unreasonable. Of course, we could choose a functional form that more closely approximates the relationship in Figure C.1, but doing so would add complexity without meaningfully affecting the results; as noted in the main text, most of the variation is cross-precinct rather than within-precinct.

D 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

Figure C.1: Monotonicity of within-precinct variation

This plot shows the non-parametric relationship between de-meaned MAS vote margin ($\mathrm{m}_{bp} - \overline{m}_p$) and de-meaned reporting time percentile ($\text{Time Percentile}_{bp} - \overline{TimePercentile_p}$).



The black line marks a local linear fit using the rule-of-thumb bandwidth from Fan and Gijbels (1996, p. 110–113); the dashed lines mark 95% confidence intervals. Top and bottom 1% of de-meaned reporting times are excluded.

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 D.1: No evidence of discontinuities, regardless of sort order

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

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

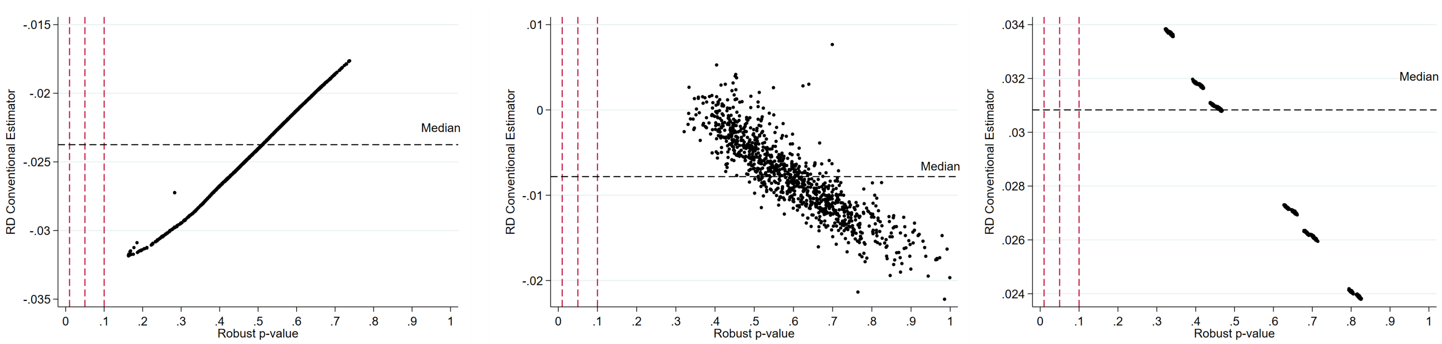


Figure D.1 plots the magnitude of the RD estimates against the corresponding p-values for each of the 1,000 draws, for each of the three cutoffs studied in the paper. Table D.1 summarizes the results. The mean and median robust p-values are above 0.5, implying that the results presented in the main text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

Table D.1: No evidence of discontinuities, regardless of sort order

			Robust p-value		RD Estimate		
Cutoff	Date & Time	Sample*	Mean	Median	Mean	Median	N Sortings
0.852	10/20/2019 19:40:57	Full	0.503	0.511	-0.024	-0.024	1,000
0.950	10/21/2019 14:54:25	Full	0.596	0.590	-0.008	-0.008	1,000
0.950	10/20/2019 20:03:59	Truncated	0.553	0.465	0.029	0.031	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 D.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, 2019, Section 4.4.2).

Table D.2: Robustness to polynomial degree and bandwidth choices

Cutoff	Date	Sample	Reg.	Deg.	Estimate	BW	p-val.	Robust C.I.	N Left	N Right
0.852	10/20/2019 19:40:57	Full	1	1	-0.024	0.041	0.511	[-0.072, 0.036]	1,368	1,367
0.852	10/20/2019 19:40:57	Full	1	2	-0.008	0.048	0.957	[-0.062, 0.065]	1,588	1,589
0.852	10/20/2019 19:40:57	Full	1	3	-0.004	0.062	0.894	[-0.065, 0.074]	2,043	2,059
0.852	10/20/2019 19:40:57	Full	0	1	-0.025	0.072	0.615	[-0.073, 0.043]	2,337	2,372
0.852	10/20/2019 19:40:57	Full	0	2	-0.027	0.077	0.554	[-0.071, 0.038]	2,532	2,580
0.852	10/20/2019 19:40:57	Full	0	3	-0.020	0.125	0.851	[-0.124, 0.102]	4,082	4,281
0.950	10/21/2019 14:54:25	Full	1	1	-0.008	0.036	0.675	[-0.039, 0.061]	1,271	1,332
0.950	10/21/2019 14:54:25	Full	1	2	0.008	0.049	0.765	[-0.051, 0.070]	1,698	1,804
0.950	10/21/2019 14:54:25	Full	1	3	-0.002	0.048	0.471	[-0.097, 0.045]	1,645	1,741
0.950	10/21/2019 14:54:25	Full	0	1	-0.033	0.075	0.508	[-0.063, 0.031]	2,547	1,828
0.950	10/21/2019 14:54:25	Full	0	2	-0.002	0.065	0.567	[-0.040, 0.073]	2,210	1,828
0.950	10/21/2019 14:54:25	Full	0	3	0.006	0.053	0.471	[-0.096, 0.045]	1,809	1,828
0.950	10/20/2019 20:03:59	Truncated	1	1	0.031	0.041	0.467	[-0.036, 0.079]	1,325	1,378
0.950	10/20/2019 20:03:59	Truncated	1	2	0.022	0.048	0.932	[-0.067, 0.073]	1,542	1,605
0.950	10/20/2019 20:03:59	Truncated	1	3	0.006	0.059	0.451	[-0.111, 0.049]	1,874	1,662
0.950	10/20/2019 20:03:59	Truncated	0	1	0.032	0.093	0.227	[-0.025, 0.105]	2,944	1,662
0.950	10/20/2019 20:03:59	Truncated	0	2	0.011	0.163	0.617	[-0.386, 0.650]	5,152	1,662
0.950	10/20/2019 20:03:59	Truncated	0	3	0.006	0.077	0.436	[-0.103, 0.044]	2,455	1,662

"Truncated" denotes the sample that excludes the voting booths without time stamps in the preliminary results system. "Reg." reports whether we choose the bandwidth with or without the regularization term; "Deg." reports the degree of the local polynomial.

E Additional tables and figures

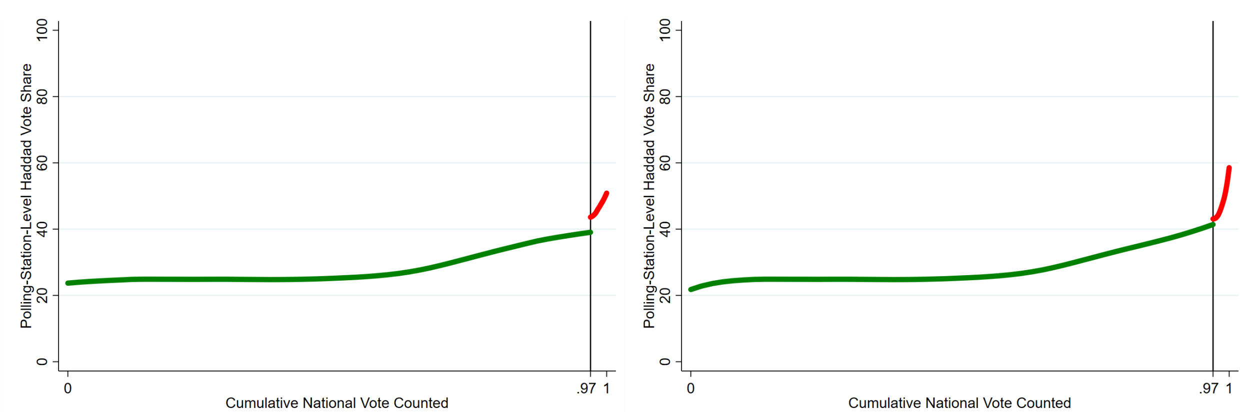
Figure E.1: Paper Ballot in Bolivia's Presidential Election

Source: Jorge Bernal

Figure E.2: Boundary Bias and Artificial Jumps in Haddad's Vote Share in Brazil

Figure (a) reveals that using a local constant fit creates the artificial appearance of a jump in Haddad's vote share at 97%. Figure (b) reveals that using a local linear fit corrects this.

(a) Degree-Zero Local Polynomial (b) Degree-One Local Polynomial



Rule-of-thumb bandwidth from Fan and Gijbels (1996, p. 110–113), Epanechnikov kernel.

Figure E.3: Preliminary Results System Time is Correlated with Share of 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

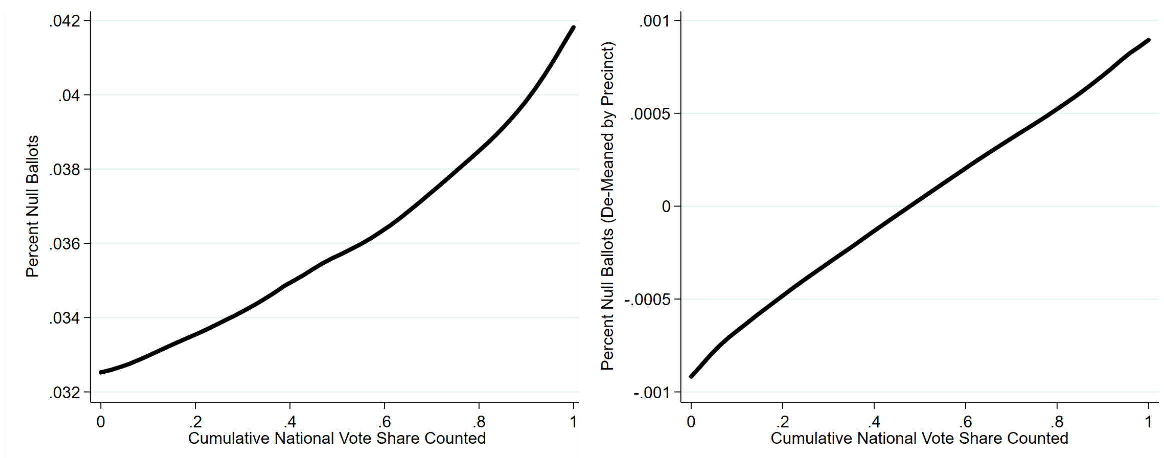


Figure E.4: Within-Precinct Variation Trend, Smaller Bandwidth

Figure (a) repeats Figure 4b from the main text; 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 (Fan and Gijbels, 1996, p. 110–113)

(b) Arbitrary small bandwidth (0.02)