The Ends Don't Justify the Means

Yet Another Analysis Fails to Support the OAS' Delegitimization of Bolivian Elections

By David Rosnick*

May 2020

Center for Economic and Policy Research
1611 Connecticut Ave. NW
Suite 400
Washington, DC 20009

Tel: 202-293-5380
Fax: 202-588-1356
https://cepr.net

*David Rosnick is an Economist at the Center for Economic and Policy Research.

Contents

Acknowledgements

The author thanks Dan Beeton, Karen Conner, Brett Heinz, and Jake Johnston.

Executive Summary

The Bolivian mission to the Organization of American States (OAS) *promoted* a new study purportedly proving that the results of the October election in Bolivia were fraudulent.[1] This study, "The OAS Conclusions about the election integrity of the Bolivian election are correct" by John Newman, primarily expresses concern that in the preliminary count the late votes are distributed differently than the votes counted prior to the interruption in reporting.[2] Newman argues that the unexplained difference is necessary in order to account for Evo Morales's victory. That is, if not for the unexplained increase in his votes late in the count, Morales's margin would have fallen short of the 10 percentage points necessary to avoid a runoff election.

Unfortunately, Newman's statistical tests are uninformative. By conditioning samples on the early vote margins, the tests ought to come back positive — with his split Newman creates a difference to be detected and a test should detect that difference. All negative results — failures to distinguish statistically between the distributions— likewise, are false negatives even in the absence of any fraud.

We suggest an alternative split of the data that suffers less of this conditioning problem while maintaining the spirit of the exercise. When we apply the tests on this alternative split, we expect to find that all results are negative and — following Newman's approach — the election result is entirely explained.

Further, Newman's choice of cutoff increases the relative likelihood of finding a false negative in cases where a positive result would lend counterfactual support to Morales over the runner-up, Carlos Mesa.

Finally, we note that Newman builds the final counterfactual based on municipality-level analysis. By ignoring intra-municipality variation, this necessarily overestimates the unexplained difference.

We therefore conclude that Newman's study is fatally flawed and should be retracted.

---

1 Bolivia en la OEA (2020).
2 Newman (2020).

Background

There are a number of ways in which researchers have attacked the question of whether the post-interruption results in Bolivia's October 2019 elections were predictable. For example, we projected directly the late votes using multiple imputations of tally sheets based on geographic indicators and the pre-interruption results; this led us to reject the idea that something inexplicable drove Morales's final margin above the 10 percentage point threshold required for a first-round victory.[3] That Morales was headed for a decisive victory appeared inevitable based on the early reporting.

Taking a different approach, Escobari and Hoover construct an implicit counterfactual using regression analysis. Seeking to quantify the degree to which the late vote is inexplicable, they find that incorporating full geographic information into their model leaves very little unexplained. In other words, once they take geography into full account, their model shows Morales's victory was predictable.

John Newman takes a third approach. Newman first breaks down the data into three sets: geographical areas where all tally sheets were counted pre-interruption, areas where all tally sheets were counted after the interruption, and areas which were partially but not completely counted after the interruption. Newman accepts the results from the first and second groups; for the third group, two questions are posed. First, do the late tally sheets look different than the early ones? Second, if they do look different, how much of this difference is accounted for by a shift in the mix of geographies within the group? Newman argues that absent non-geographical differences, Morales's margin would have fallen short of the threshold for a first-round victory.

As it happens, with a careful understanding of the first question, Newman's second question becomes moot.

3 Throughout this paper, for consistency, by "margin" we mean the number of votes recorded for the leading candidate, less those of the runner-up, as a share of the valid votes cast. The margin is normally expressed in percentage points. Margins may be aggregated at different levels, but by "final" margin we mean aggregated across the entire presidential election.

Understanding Newman's Test

Newman broadly seeks to answer the question of whether the early and late reported tally sheets are similar. One way of trying to answer this question is to aggregate the early and late votes by precinct.

Consistent with our broad expectations about the election, we assume that absent fraud, tally sheets in the same precinct should show statistically similar voting patterns irrespective of the time the sheet was counted. Specifically, whether or not a tally sheet was counted before or after the interruption in TREP reporting should not have any impact on the observed votes.[4]

If every precinct had some — but not all — votes counted prior to the interruption, then we would expect Morales's margin of victory not to differ much between the tally sheets counted early and those counted late. However, even if the population of residents is exactly split among those in support of Morales and those in support of Mesa, we are almost assured that there will be benign random differences between those voting early and those voting late.

In order to understand better these minor variations, we repeatedly simulate a single precinct represented by a populace with equal support for each candidate. In each simulation, we assume that some number of 1,200 voters are represented on tally sheets counted early. On average across simulations, 1,000 voters appear on these sheets. The remaining voters (200 on average) are represented on tally sheets counted late.[5]

Next, we assume that in each simulation every voter had a 50 percent chance of voting for either candidate, so that on average we expect that the total votes will break 600 to 600. Likewise, the average on early tally sheets will be something like 500 to 500 (but possibly 490 to 510 on some, and 510 to 490 on others). Regardless, the margin on early tally sheets should average zero. Likewise, the margin on the votes recorded on late tally sheets also should average zero.

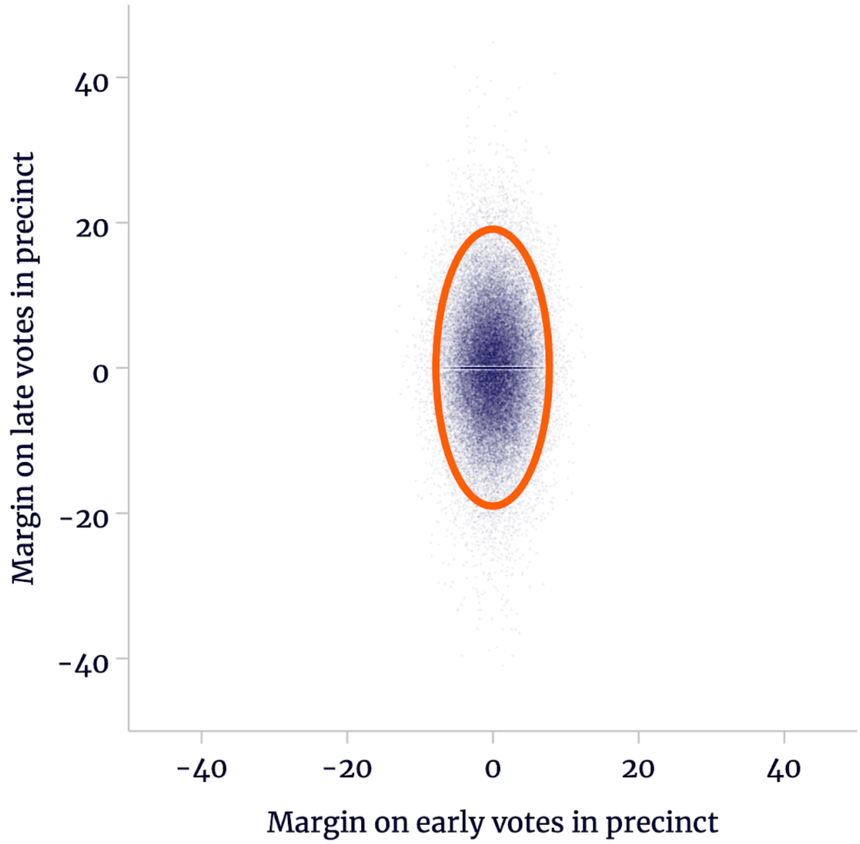
4 This is a conservative assumption (more likely to result in a finding of an unexplained change) because voters are assigned to specific tally sheets by name. As a family with a shared family name is probably more likely to vote in a like manner than will strangers, we expect some clustering of votes for one party or another on different sheets.

5 The number of voters assumed to vote on early sheets is assumed to be binomial with 1,200 trials (voters) each with probability drawn at the precinct level from a beta distribution p~Beta(25,5).

We see the results of 50,000 simulations in **Figure 1**. Most margins fall within an ellipse taller than it is wide. This is because the sample sizes for the late margins are in general considerably smaller than those for the early margins. If only three voters show up late (as in one simulation) then the margin has to be at least plus or minus 33 percentage points.

FIGURE 1

Early and Late Margins May Differ Greatly in Any Given Precinct



Source: Author's calculations.

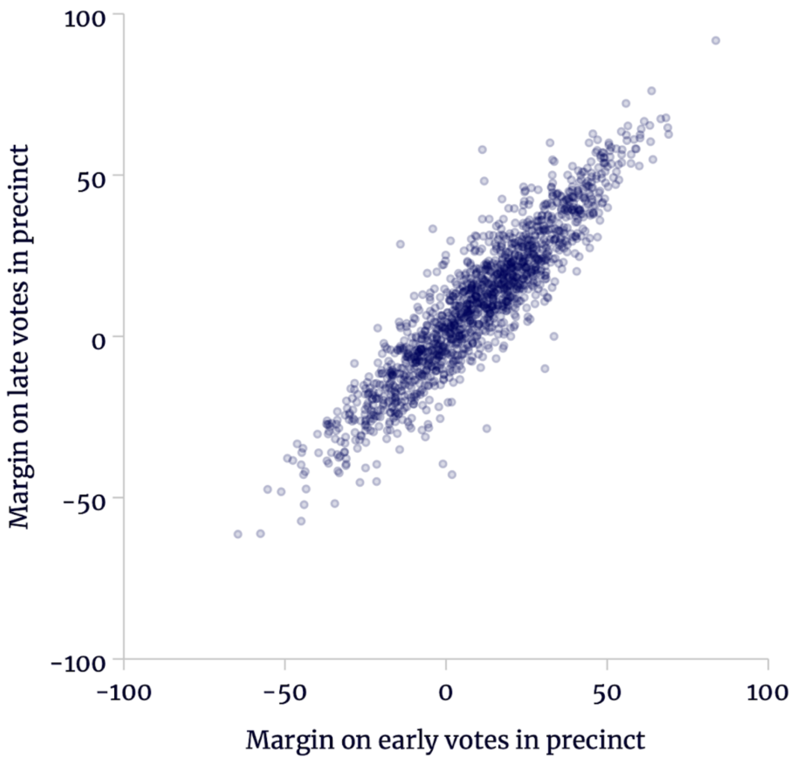
Importantly, the late margin can frequently swing by plus or minus 20 percentage points, even though every voter is exactly 50/50 in their voting. The early margins, however, are more closely distributed. The variations in observed results — both early and late — are entirely due to chance. However, the distribution of early and late margins are clearly different. No test is really required to detect this difference as it is inherent in the fraud-free simulations.

In this sense, Newman's premise is all wrong. Even absent fraud, we would not necessarily expect the early and late distributions to be similar.

Now, rather than simulating a single precinct 50,000 times, let us simulate 1,500 different precincts at once.[6] Rather than assuming each voter is equally likely to vote for each candidate, we now randomize that tendency at the precinct level.[7] We see in **Figure 2** that despite the randomization, there is a strong correlation between the early and late margins at each precinct.[8]

FIGURE 2

Simulated Election Results by Precinct



Source: Author's calculations.

Now let's add to this figure a histogram of the early vote margins for each precinct. Imagine all the blue points falling down to the horizontal axis. Note that the green pile (distribution) forms a nice bell curve centered to the right of zero. The average precinct preference is 11 to 9 in favor of the leader, or 10 percent of the total votes. This is not important at the moment, but will matter later.

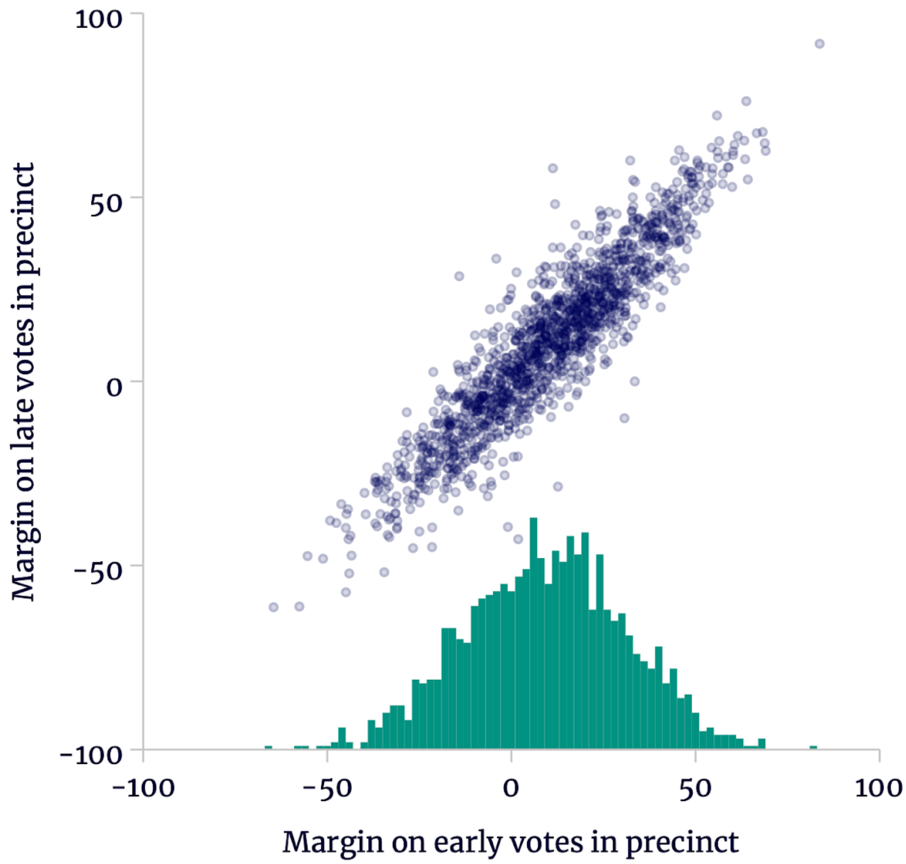
6 Newman reports 1,498 precinct observations in his precinct-level test.

7 Voters at precinct number *n* have a probability $v_n \sim \text{Beta}(11,9)$ of voting for the leader, which should result in an (11-9)/(11+9)=10 percent margin of victory. We have also increased the variance in early versus late turnout with $p_n \sim \text{Beta}(15,3)$.

8 The observed 92 percent correlation understates the true relationship between the early and late margins because of the "errors-in-variables" problem — that is, the late margin is related to the true preferences of voters in the precinct, rather than the approximation we observe in the early votes.

FIGURE 3

Distribution of Early Vote Margins

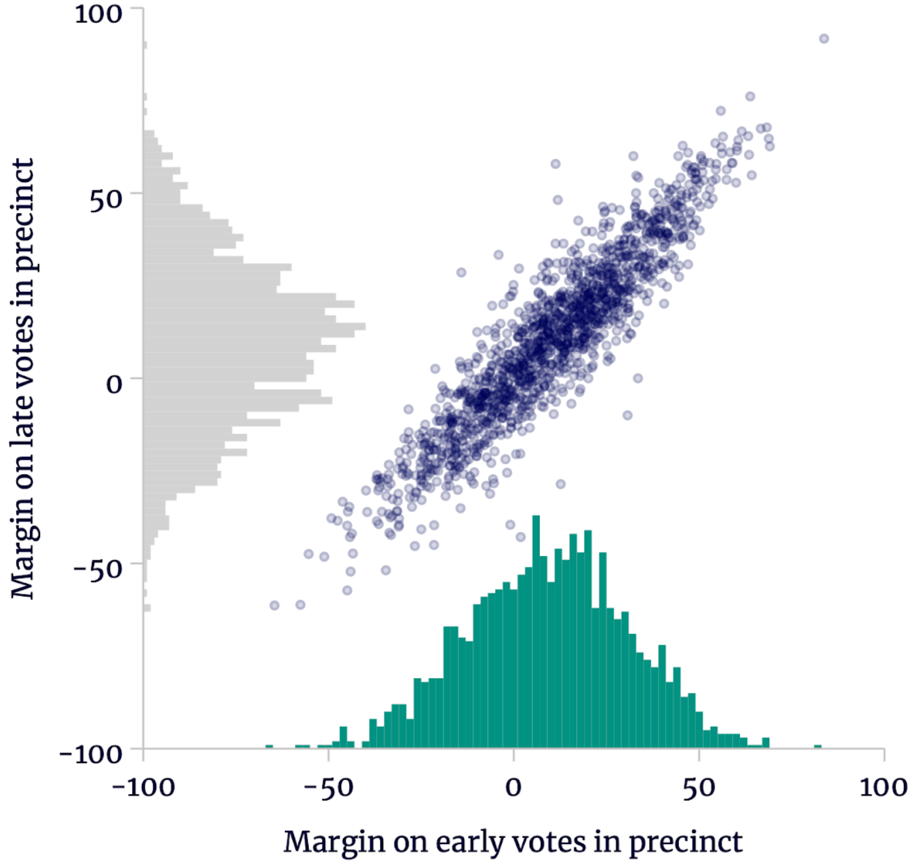


Source: Author's calculations.

We may also add a histogram of the late vote. Imagine all the blue points now sliding left into a gray pile on the vertical axis, and we get a nice bell curve centered just above zero in keeping with our assumed voter preferences.

FIGURE 4

Distribution of Late Margins

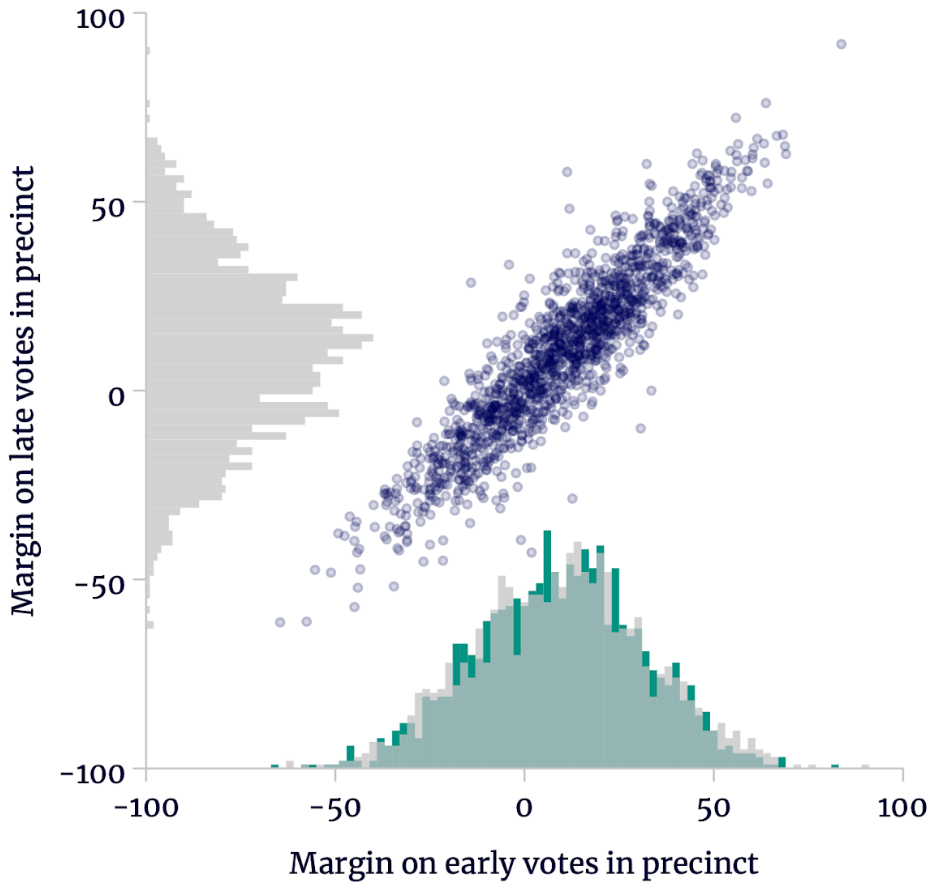


Source: Author's calculations.

While the second histogram makes sense in terms of the graph, it is not easy to compare the two. In **Figure 5**, then, we overlay the data for the late margins right over the data for the early margins. Clearly, there is nearly zero difference between the early and late distributions.

FIGURE 5

Comparison of Early and Late Margins



Source: Author's calculations.

Of course, as Newman points out, "clearly" is not a statistically scientific term. We may employ a Kolmogorov-Smirnov test for equality of distributions to get a more concrete description of how close or far the distributions are from one another.

TABLE 1

K-S Test for All Precincts

Hypothesis	Test Statistic (D)	P-value
Late < early (D^+)	0.0213	0.505
Early < late (D^-)	-0.0187	0.593
Combined K-S test (D)	0.0213	0.884

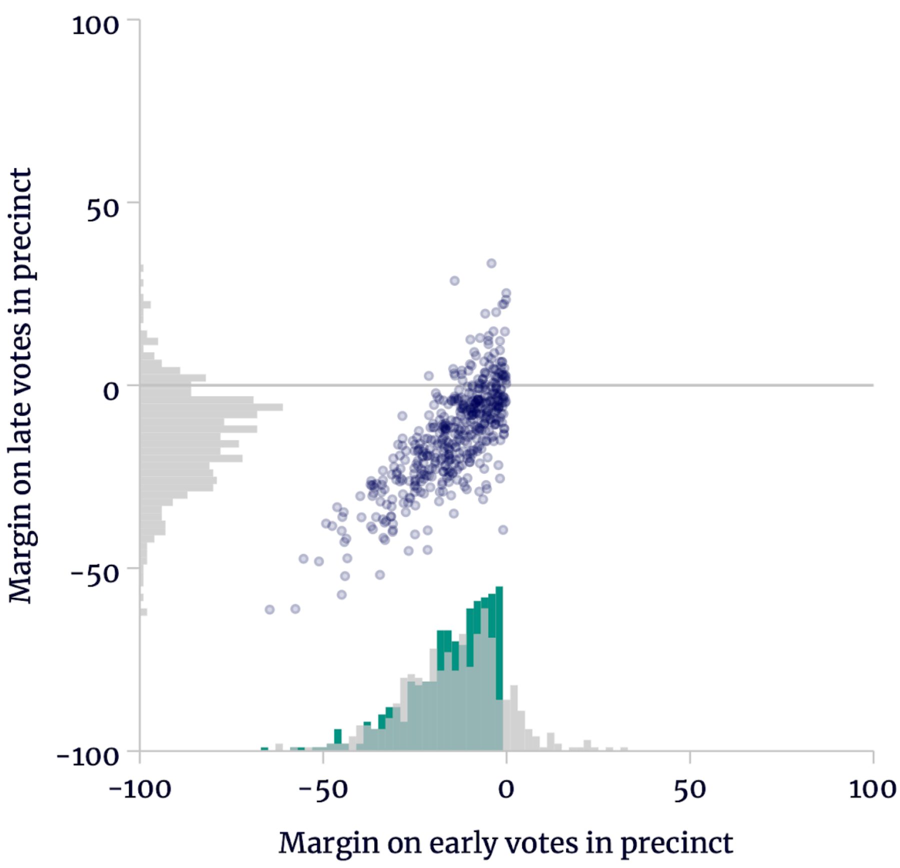
Source: Author's calculations.

Resoundingly, the K-S test fails to reject the hypothesis that the early and late margins differ. This is, of course, consistent with what we already know: the early and late margins are constructed identically apart from the sample sizes included in each precinct, so their margins should be very similar. The variance in the late margins is generally larger, conditioned on the overall margins (see Figure 1, for example), but this largely washes out across the distribution. If we increased sufficiently the number of precincts, the difference in the tails would be detectable. Thus, our negative result is a false negative. In fact, the wider the distribution, the more difficult the additional variance in the distribution of late margins is to detect.

Now here is where things get strange. Suppose we now split our data, analyzing separately the precincts which (in the early vote) favored one candidate over the other. In **Figure 6**, we examine only precincts favoring the runner-up.

FIGURE 6

Simulated Data for Precincts with Early Support for Runner-Up



Source: Author’s calculations.

Because we are looking only at precincts with early margins less than zero, the early (green) distribution falls off precipitously at zero. Obviously, we cannot have early margins to the right of zero because we are ignoring those by design. However, the early margins do not exactly reflect the underlying voter preferences. Some of the precincts with early margins less than zero are in precincts where the population favors the winning candidate (a positive margin) but it happens randomly that a majority of the early vote favors the runner-up. (Likewise, even when the precinct does not favor the leader, the late voters might randomly happen to favor the leader anyway). Consequently, in many of these precincts the late vote will favor the leader. The distribution of late vote margins therefore extends past zero. We say here that the late margins "bleed" past our intended cutoff.

What does this mean, then? If the early vote is cut off at zero, but the late vote has a tail that extends past zero, then we know for a fact that the distributions are different.[9] In fact, we know specifically that these early margins must tend to be smaller than the late margins. We may confirm this with a second K-S test.

TABLE 2

K-S Test for Runner-Up Precincts

Hypothesis	Test Statistic (D)	P-value
Late < early (D^+)	0.0580	0.197
Early < late (D^-)	-0.1429	<0.001
Combined K-S test (D)	0.1429	<0.001

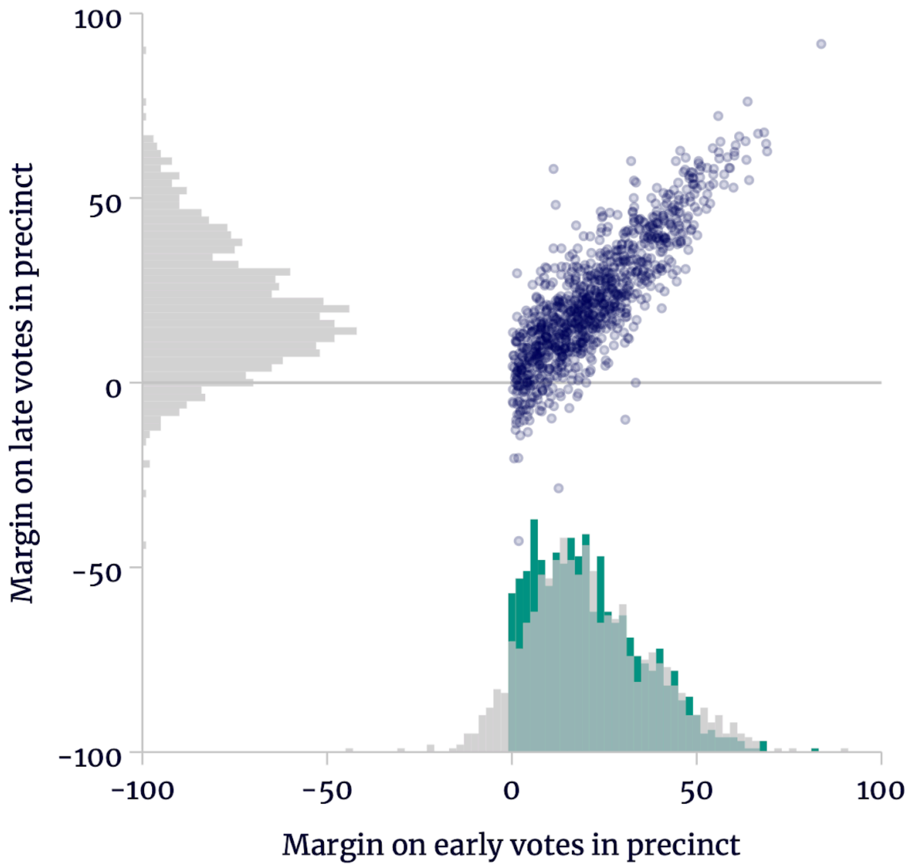
Source: Author's calculations.

Indeed, the K-S test rejects the hypothesis that the early margins are not smaller than the late margins. Similarly, we may perform the same analysis on the rest of the precincts — where the leader won early.

9 This is in addition to the added variance in the distribution of the late margins.

FIGURE 7

Simulated Data for Precincts with Early Support for Leader



Source: Author's calculations.

TABLE 3

K-S Test for Leader Precincts

Hypothesis	Test Statistic (D)	P-value
Late < early (D^+)	0.0806	0.001
Early < late (D^-)	-0.0275	0.463
Combined K-S test (D)	0.0806	0.003

Source: Author's calculations.

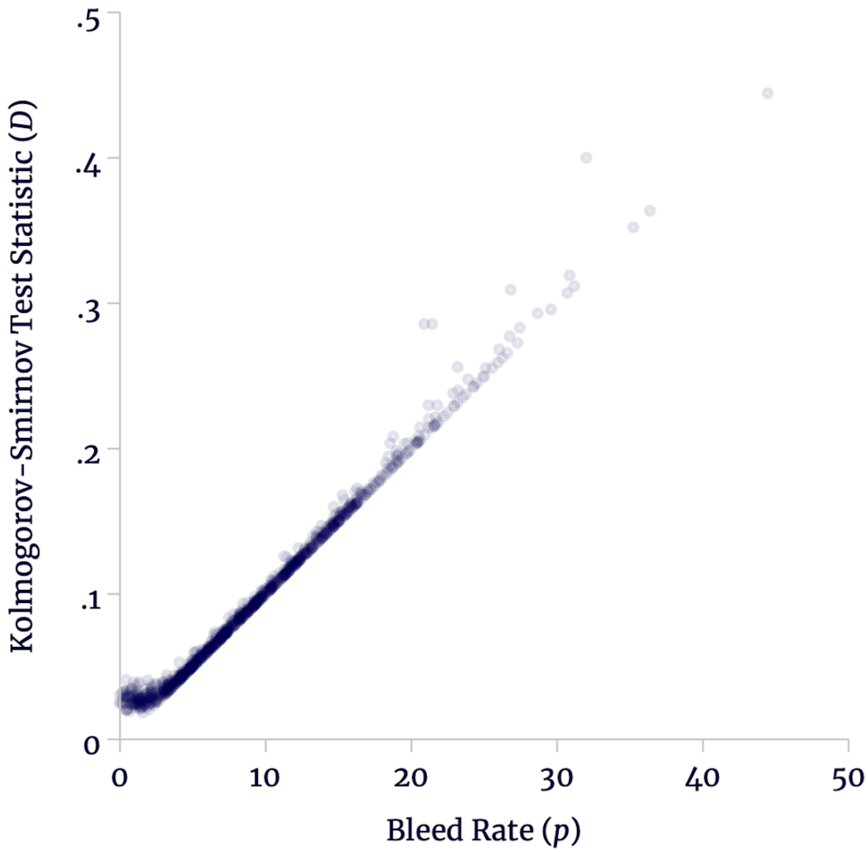
It appears that the sharp cutoff defining our data split is causing trouble. As it happens, there is a pretty simple way for us to confirm that the distributional bleed is driving the results. In such a case, the test statistic *D* can be no less than the bleed rate — the percent of observations where the late margin bleeds past the cutoff.[10]

10 See the Technical Appendix for the motivation behind this.

In **Figure 8**, we show that in our simulated data the K-S test statistic does in fact generally follow the bleed rate *p*. Each point represents a simulation with a randomly selected cutoff.

FIGURE 8

The K-S Test Statistic Follows the Bleed Rate



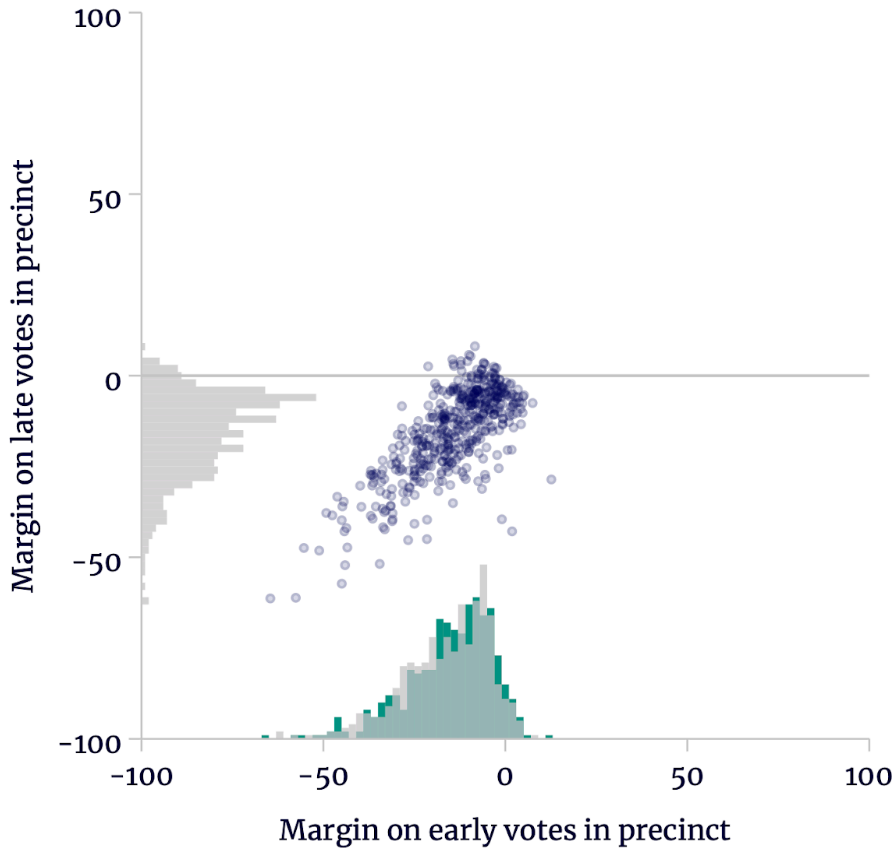
Source: Author's calculations.

Clearly, the sharp cutoff at zero is causing trouble. We are understandably rejecting the idea that the early and late margins are similarly distributed, and yet we know that the data here is random, so there is no fraud. Again, we must not be misled by the test into believing there is anything funny in this data.

We would like to separate precincts into two groups, each representing precincts favorable to each candidate, but we would like to avoid a sharp fall in the early vote margins. One possible fix is to soften the cutoff for the sample to zero in the average of the early and late votes. In theory, a precinct might have gone 100 percent for one candidate early, and 100 percent for the other late. But obviously, this is unlikely over 1,200 votes. So what might this look like?

FIGURE 9

Precincts with Early Support for Runner-Up (Soft Cutoff)



Source: Author's calculations.

Here, the spirit is maintained: the precincts are basically unfavorable to the leader candidate.[11] However, both the early and late margins bleed past zero. Once that flexibility is permitted, we can no longer argue based on the K-S test that the distributions are different.

TABLE 4

K-S Test for Runner-Up Precincts (Soft Cutoff)

Hypothesis	Test Statistic (D)	P-value
Late < early (D^+)	0.0674	0.116
Early < late (D^-)	-0.0105	0.949
Combined K-S test (D)	0.0674	0.231

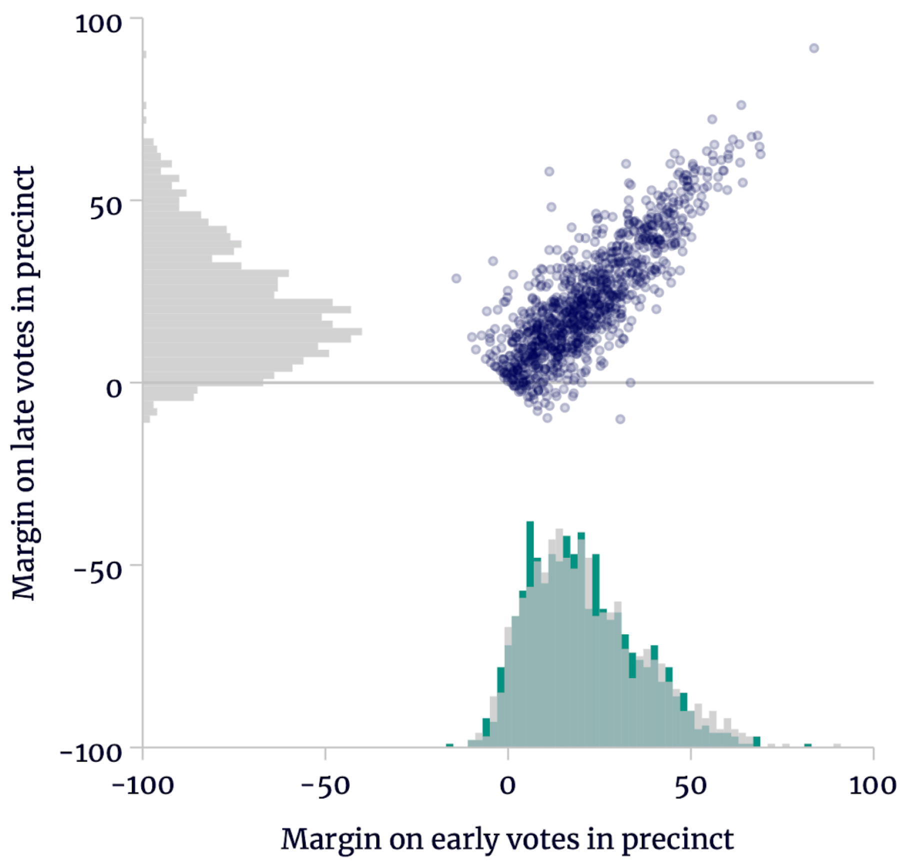
Source: Author's calculations.

11 Newman's cutoff excludes from the runner-up precincts some that likely favor the runner-up because the leader randomly happened to win early, and includes some precincts that likely favor the leader because the runner-up happened to win early. Likewise, our soft cutoff does not divide precincts by whether or not the data suggests that a precinct on the whole is likely to favor one candidate or the other; still it generally divides the data into groups that largely favor one or another without limiting the range of possible margins for either the early or late vote.

Likewise, to the right, we see a similar effect.

FIGURE 10

Precincts with Early Support for Leader (Soft Cutoff)



Source: Author's calculations.

Again, both distributions bleed past zero, but do so in a similar fashion so that they are not distinguishable, statistically. Again, this does not imply that the distributions are in fact the same. We know the opposite to be true: the late distribution has greater variance. Our test simply lacks the power to detect the difference.

TABLE 5

K-S Test for Leader Precincts (Soft Cutoff)

Hypothesis	Test Statistic (D)	P-value
Late < early (D^+)	0.0088	0.924
Early < late (D^-)	-0.0273	0.465
Combined K-S test (D)	0.0273	0.839

Source: Author's calculations.

Before moving on, we note one other subtle feature of the results. The p-values associated with the tests of precincts favorable to the leader are much larger than those for the other. This is because our cutoff does not evenly split the data. The precincts that bleed are relatively close to the cutoff, so the more the sample data lies far from that boundary, the closer the distributions will appear to be. This is true whether or not we use a hard (Newman) cutoff to divide the data or our softer one.

In the tables that follow, we reset the cutoff to 20 percentage points in favor of the leader — 10 points above the election-wide average, instead of 10 points below. This splits the precinct types much more evenly.

TABLE 6

K-S Test for Runner-Up Precincts (Newman Cutoff at +20)

Hypothesis	Test Statistic (D)	P-value
Late < early (D^{+})	0.0321	0.358
Early < late (D^{-})	-0.1002	<0.001
Combined K-S test (D)	0.1002	<0.001

Source: Author's calculations.

TABLE 7

K-S Test for Leader Precincts (Newman Cutoff at +20)

Hypothesis	Test Statistic (D)	P-value
Late < early (D^{+})	0.1713	<0.001
Early < late (D^{-})	-0.0538	0.234
Combined K-S test (D)	0.1713	<0.001

Source: Author's calculations.

TABLE 8

K-S Test for Runner-Up Precincts (Soft Cutoff at +20)

Hypothesis	Test Statistic (D)	P-value
Late < early (D^{+})	0.0451	0.236
Early < late (D^{-})	-0.0183	0.788
Combined K-S test (D)	0.0451	0.467

Source: Author's calculations.

TABLE 9

K-S Test for Leader Precincts (Soft Cutoff at +20)

Hypothesis	Test Statistic (D)	P-value
Late < early (D^{+})	0.0089	0.940
Early < late (D^{-})	-0.0354	0.371
Combined K-S test (D)	0.0354	0.704

Source: Author's calculations.

This cutoff effect, we believe, impacts Newman's results. However, the dominant problem is the choice of a hard cutoff which causes the late margins to bleed but — critically — not the early ones.

Newman's approach to discerning fraud fails when applied to clean data, so there is no reason to trust his test results when applied to the actual election results.

Replication and Reanalysis of Newman's Results

First, using publicly available data, we replicate closely Newman's Table 6 (our **Table 10**).[12]

TABLE 10

Replication of Newman (2020)'s Table 6

Kolmogrov-Smirnov Tests	All Observations Precincts (1,497 obs)	Precincts Where (MAS-CC)<0 before cut-off (538 obs)	Precincts Where (MAS-CC)>=0 before cut-off (959 obs)
Late < early			
D^+	0.0147	0.0353	0.037
P Value	0.724	0.511	0.279
Early < late			
D^-	-0.0227	-0.0967	-0.0355
P Value	0.462	0.007	0.300
Combined K-S			
D	0.0227	0.0967	0.0365
P Value	0.835	0.013	0.545

Source: OEP (2019) and author's calculations.

Although there are some unimportant differences, we come to the same conclusion: if taken as valid, these tests suggest that in precincts where the early vote margins went against Morales, the late votes were more favorable to the leader.

However, as the exercises in the previous section suggested, we believe this positive test result is an artifact of the choice of a hard cutoff at zero. Employing a soft cutoff instead, we find

12 We suspect the difference is due to use of nonpublic data and (possibly) inconsistent treatment of tally sheets with zero observations.

that none of the tests reject the null hypothesis that the distribution of margins is unchanged post-interruption.

TABLE 11

Reanalysis of Newman (2020)'s Table 6 (Soft Cutoff)

Kolmogrov-Smirnov Tests	All Observations Precincts (1,497 obs)	Precincts Where (MAS-CC)<0 before cut-off (532 obs)	Precincts Where (MAS-CC)>=0 before cut-off (965 obs)
Late < early			
D^+	0.0147	0.0414	0.0145
P Value	0.724	0.403	0.816
Early < late			
D^-	-0.0227	-0.0207	-0.0352
P Value	0.462	0.797	0.302
Combined K-S			
D	0.0227	0.0414	0.0352
P Value	0.835	0.753	0.587

Source: OEP (2019) and author's calculations.

This deals a fatal blow to Newman's proposed counterfactual analysis of the election. As he explains, effects on precincts where Morales performed well pre-interruption are

not factored into the counterfactual estimate, because the Kolmogorov-Smirnov test indicated that one could not reject the null hypothesis that the distributions were equal. Only where there was support for the proposition that the change in distribution was significant, was the change incorporated into the counterfactual estimate.

Now that the corrected K-S tests failed to reject all the nulls and there is no support for the proposition that any changes in the distributions are significant, there are no changes at all in the counterfactual estimate relative to the official results. If we were to use this information to correct Newman's Table 12, we find no evidence that Morales failed to win the first round of the election by 10.56 percentage points.

A Flawed Counterfactual

Even if we were to accept Newman's statistical tests as valid, his conclusions are still invalid because he performs his decisive analysis based on data at the municipality level. Specifically, he accounts for shifts in the voting patterns between municipalities, but ignores that changes in margin within municipalities are largely accounted for by the mix of smaller geographies within each municipality.

To see this, let us start with the precinct calculation. **Table 12** below is the precinct equivalent of Newman's Table 12 for municipalities. Rather than testing various groups, we simply assume for our counterfactual that the margins in precincts that were partially, but not completely, counted do not change from early to late.

TABLE 12

Counterfactual Estimate for Precincts Partially, But Not Completely, Counted Early

	Votes	Official Margin		Counterfactual	
		Difference	Share of Votes	Difference	Share of Votes
Total	3,915,993	372,491	9.51	367,914	9.40
Early	3,088,170	210,700	6.82	210,700	6.82
Late	827,823	161,791	19.54	157,214	18.99

Source: OEP (2019) and author's calculations.

These precincts accounted for nearly 4 million votes. As we see, Morales's margin on the 3,088,170 votes counted early was 210,700 — or 6.82 percent. In our counterfactual, we assume that this is correct. However, in the 827,823 votes in these precincts that were counted late, Morales won by 161,791 — or 19.54 percent. Nearly all of this increase in margin is due to a more favorable mix of precincts for Morales. If the early precinct margins simply carried over to the late, we can expect Morales to win the late precincts by 18.99 percent. This leaves 4,577 votes unexplained by the mix. Again, there may be benign reasons for this difference. It may be that for structural reasons more favorable tally sheets within a precinct tend to be counted late. It may also be completely random. Regardless, there are 4,577 net votes for Morales left unexplained by our counterfactual, out of nearly 4 million total votes.

Next, we repeat this exercise at the locality level. As we see in **Table 13**, there are more votes in these localities than there were in precincts covered in Table 12. This makes sense: localities are made of one or more precincts. A precinct partially but not completely counted early (covered in Table 12) can only exist in a locality that is also partially, but not completely, counted early (covered in Table 13). In a locality that is partially but not completely counted early, however, there may be one or more precincts that were completely counted early or completely counted late. In other words, Table 13 covers every precinct covered in Table 12, and then some.

Importantly, while Table 13 covers a larger share of the election, simply considering localities rather than precincts comes at a price. When the late margin in a locality differs from the early margin, we can no longer discern how much of the change is due to the shifting importance of precincts within that locality. Table 13 therefore understates the extent to which geographic mix explains the overall margin.

TABLE 13

Counterfactual Estimate for Localities Partially, But Not Completely, Counted Early

	Votes	Official Margin		Counterfactual	
		Difference	Share of Votes	Difference	Share of Votes
Total	5,189,142	181,383	3.50	125,220	2.41
Early	4,338,909	18,195	0.42	18,195	0.42
Late	850,233	163,188	19.19	107,025	12.59

Source: OEP (2019) and author's calculations.

Though there are only about one-third more votes reflected in Table 13, there are more than 12 times as many votes — 56,163 — left unexplained by the mix of localities. If we try to pin these down by using what learned in Table 12, we get into serious trouble. In **Table 14**, we simply take the results of Table 13 and subtract Table 12 — effectively accounting for those precincts we previously analyzed.

TABLE 14

Counterfactual Estimate for Precincts Completely Counted Early or Late Residing in Localities Partially, But Not Completely, Counted Early

	Votes	Official Margin		Counterfactual	
		Difference	Share of Votes	Difference	Share of Votes
Total	1,273,149	-191,108	-15.01	-242,694	-19.06
Entirely Early	1,250,739	-192,505	-15.39	-192,505	-15.39
Entirely Late	22,410	1,397	6.23	-50,189	-223.96

Source: OEP (2019) and author's calculations.

Clearly, these remaining precincts in the covered localities were generally counted early, and the early ones broke heavily for Mesa. However, we have to account for a total of 56,163 votes, of which we have already accounted for 4,577. This leaves 51,586 votes unaccounted for. If we assume that we cannot find them in the early votes, then they all must be in the remaining late precincts. Subtracting 51,189 net votes for Morales from the official 1,397 leaves 50,189 net votes for Mesa. Not only is this difference enormous, it ceases to be logical. We simply cannot expect 50,189 more votes for Mesa than Morales in precincts with only 22,410 total votes. Nor can we attribute this to disappearance of validly cast ballots; there were only 27,404 eligible voters in those precincts.

This exercise fails because we know that shifting the precinct mix accounts almost entirely for the overall shift in margin, yet the locality-level analysis fails to account for it. Table 13 must represent a severe overestimate of the unexplained margin.

However, Newman's Table 12 offers a second way forward. There, Newman analyzes at the municipality level — ignoring measurable shifts within precincts and localities. More importantly, he leaves unchanged in the counterfactual any municipality counted entirely early or entirely late. In effect, he assumes that the votes in such a municipality are fully explicable.[13] We may do likewise with our precinct-level analysis, as in **Table 15**, extending Table 12 to cover the precincts counted entirely early or entirely late.

13 Simply, in a municipality fully counted late, we have no way of knowing what direction to adjust, or by how much. All the localities within were also fully counted late, as are all the precincts. Newman's analysis offers no further basis for comparison. In our previous work, we assume that a fully late municipality should likely look something like any other municipality within the same province for which we have early results. Depending on the choice of comparison, the counterfactual varies, so we repeat our analysis hundreds of times, choosing comparisons at random as we go, and presenting a distribution of possible counterfactual results. We find that none of our counterfactuals diverge from the official results sufficiently enough to question the original, official outcome of the election.

TABLE 15

Full Counterfactual Estimate for All Precincts

	Votes	Official Margin		Counterfactual	
		Difference	Share of Votes	Difference	Share of Votes
Total	6,137,671	648,439	10.56	643,862	10.49
Entirely Early	2,067,788	195,738	9.47	195,738	9.47
Entirely Late	153,890	80,210	52.12	80,210	52.12
Others	3,915,993	372,491	9.51	367,914	9.40
Early	3,088,170	210,700	6.82	210,700	6.82
Late	827,823	161,791	19.54	157,214	18.99

Source: OEP (2019) and author's calculations.

Based on Newman's approach, but analyzing at the precinct level, we produce a counterfactual that leaves very little of the election results unexplained — only those 4,577 votes, or less than 0.075 percent of all valid votes. If early votes are not in question, then there are only 153,890 votes (with a net margin of 80,212) in precincts counted entirely late that may possibly even need further explanation.

Holding constant all other pieces of the counterfactual, how strong an assumption can we make regarding those 153,890 votes and still not change the apparent outcome of the election? We need to knock down the counterfactual difference there to 50,115 — a bit less than 33 percent of the total vote in precincts counted entirely late. This would represent an absolutely massive swing unsupported by any evidence presented to date.

In effect, there are simply too few late votes not covered by Table 12 to seriously accept the OAS claim that the late vote represents an "inexplicable change in trend that drastically modifies the fate of the election." Rather, it was the OAS statement itself that rubber-stamped the opposition's efforts to delegitimize the electoral process.

References

Bolivia en la OEA. 2020. "Estudio del Sr. John Newman, economista y líder estadístico que trabajó en el Banco Mundial confirmando los resultados del informe de auditoría de la OEA sobre las elecciones del 20 de octubre de 2019 https://drive.google.com/file/d/1bu8mMS..." Twitter, May 1. *https://twitter.com/BoliviaOEA/status/1256294524269314049*

Newman, John. 2020. "The OAS Conclusions about the election integrity of the Bolivian election are correct." April 29. *https://drive.google.com/file/d/1bu8mMS5EIyJVCkdmTtrCeMb38n1F-w8A/view*

Órgano Electoral Plurinacional (OEP). 2019. "Elecciones Generales Bolivia, 20 de octubre de 2019." Accessed November 2019. *https://www.oep.org.bo/elecciones-generales-2019/*.

Technical Appendix: Understanding the Effects of Tail Censoring of Data

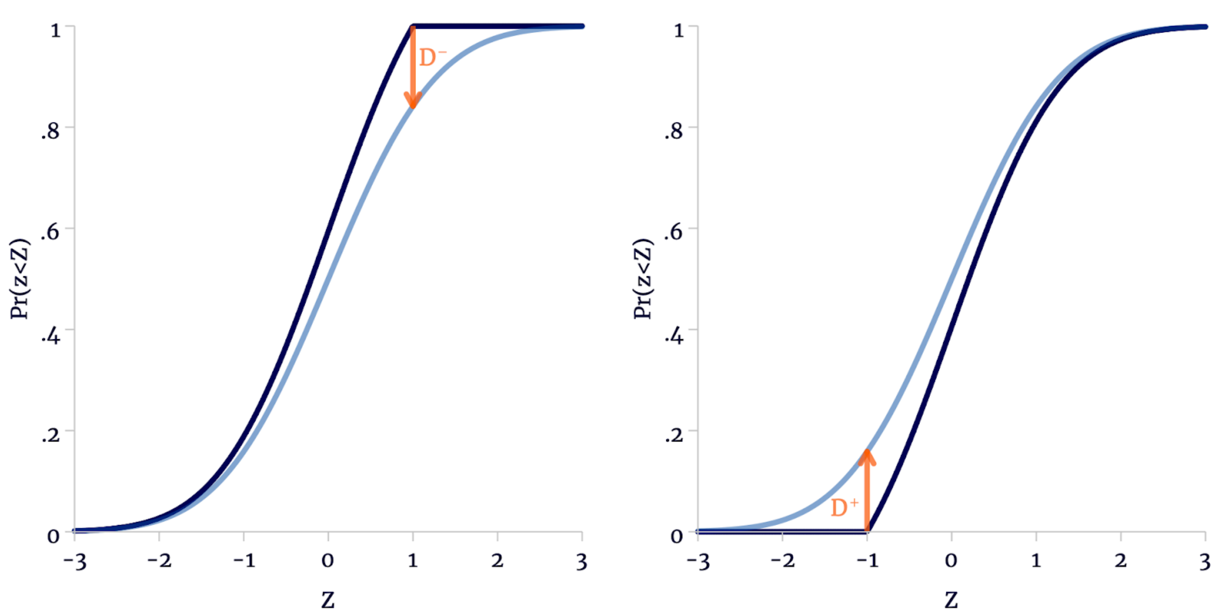
In **Figure A**, we show the schematics for Kolmogorov-Smirnov testing in the presence of censored data. On the left, the dark blue line shows the cumulative distribution function for data with a censored right tail, in contrast to the full data that bleeds to the right of 1. A large D^- indicates that the censored data is more likely to be less than 1 than the full sample.

Likewise, on the right, the dark blue line shows the cumulative distribution function for data with a censored left tail, while the full data bleeds to the left of -1. A large D^+ indicates that the censored data is more likely to be greater than -1 than the full sample.

In both cases, a sufficiently powerful test would reject the hypothesis that the censored distributions are identical to the uncensored.

FIGURE A

Schematic for the Kolmogorov-Smirnov Test with Censored Tail Data



Source: Author's calculations.

Conceptually, this is not identical to Newman's splitting of the data. The data split censors early margins, but he compares to the late margins — not the uncensored early margins.

Schematically, however, Figure A shows how censoring and bleeding will show up in the Kolmogorov-Smirnov tests. For runner-up precincts, we expect the final test statistic to be dominated by D^-; in leader precincts, D^+.

However, the censoring does not take place at the median precinct, but rather leader precincts dominate the full sample and tend to be more distant from the cutoff than do runner-up precincts. Even though precincts should bleed each way in equal numbers, the percentage of leader precincts that should bleed is about 56 percent that of the runner-up. The test statistic should be 56 percent smaller, but with a critical value for significance only 75 percent as large — meaning it is more likely that we reject the null that the margin in runner-up precincts did not rise than it is we reject the null that the margin in leader precincts did not fall.

Note again, though, that because the late distributions naturally bleed past the cutoffs even in the absence of fraud, we already know that the late and early distributions should differ. Any failure to reject the null is a false negative, but with a relatively low cutoff it is more likely that a K-S test will find a positive result in runner-up precincts than the leader.