Unnatural Claims in a ‘Natural Experiment’: Escobari and Hoover on the 2019 Bolivian Elections

By David Rosnick*

December 2019

* David Rosnick is an Economist at the Center for Economic and Policy Research (CEPR).
Acknowledgements

The author wishes to thank Andrés Arauz, Dan Beeton, Kevin Cashman, Brett Heinz, Jake Johnston, Alex Main and Rebecca Ray.
Introduction

On November 25, a short paper was released purporting to show that there was fraud in the October 20 Bolivian presidential election — and that absent this fraud, the election would have required a second round.¹ According to Bolivian election law, in order to win in a single round, a candidate must obtain a majority of the valid votes cast, or barring that, a plurality of at least 40 percent and a minimum 10 percentage point margin over the runner-up.

Every 180 seconds, the electoral authorities provided an update to the TREP — Transmisión de Resultados Electorales Preliminares, an unofficial preliminary count of the actas, or tally sheets — on their website as results came in. The purpose of the TREP is to provide a rough idea of how the election results look without waiting for the slow, official count (or cómputo) wherein tally sheets are actually examined.

Controversially, the electoral authorities ceased reporting regular updates to the TREP with 5,580 tally sheets (out of 34,555) remaining. At that point, the TREP reported 45.7 percent of the valid votes for incumbent president Evo Morales and a lead of 7.9 percentage points over runner-up Carlos Mesa — short of the necessary margin to avoid a runoff election. The next time the TREP was updated on the website, only 1,749 tally sheets remained uncounted and Morales’s lead had increased to 10.1 percent of the valid vote. The Organization of American States (OAS) cried foul, arguing that this change in Morales’s lead was “drastic” and “hard to explain.”² Though we found little difficulty predicting this result based on the first 83.85 percent of tally sheets counted in the TREP in our paper, “What Happened in Bolivia’s 2019 Vote Count?” the alarmism of the OAS gave free rein to the opposition to openly and chaotically contest the election.³ This led to a military coup d’etat and Morales’s flight to Mexico.⁴

To Diego Escobari, associate professor of economics at the University of Texas Rio Grande Valley, and to Gary Hoover, chair of the Department of Economics at the University of Oklahoma, this break in the quick count provided a “natural experiment” by which “fraud” in the election might be measured. At its most basic level, Escobari and Hoover assume that the tally sheets counted prior to the interruption were effectively “clean.” To the extent that the later tally sheets look inexplicably different, the authors attribute the difference to fraud. “Fraud,” in this case, is anything the model does not explain.

¹ Escobari and Hoover (2019a).
² OAS (2019).
³ Long et. al. (2019).
It is important to understand that because “fraud” here is just a label for any unexplained change, “fraud” in favor of Morales’s Movement Toward Socialism (MAS) party in the last 16 percent of tally sheets is completely indistinguishable from “fraud” in favor of the opposition in the first 84 percent. The analysis cannot tell us whether a finding of “fraud” should lead us to subtract from the final MAS margin or to add to it. If there is a large amount of “fraud,” then the statistics are ambiguous with respect to the outcome of a potentially “clean” election. If, however, the unexplained change is sufficiently small, we can agree that the result is consistent with a Morales lead sufficient to avoid a runoff election based on the “clean” votes.

Unfortunately, the model the authors highlight does not even attempt to explain the differences between groups of tally sheets. They simply take the observed difference itself as their measure of “fraud.”

We believe this approach is misguided. There is no reason to believe that tally sheets were counted completely at random.

 Appropriately, the authors do present results based on models with additional detail. These models better explain the difference in MAS margins before and after the interruption of the regular TREP reports and therefore drastically reduce the estimated extent of “fraud” in the election, but the authors quickly discount these additional results.

In this brief, we replicate the results of Escobari and Hoover, discover errors in their analysis, and eventually find that the approach essentially confirms our own findings that the change in margin was predictable.  

---

5 Long et. al. (2019).
Escobari and Hoover: Basic Points of Agreement and Disagreement

First, we successfully replicated precisely the results of Escobari and Hoover with data at our disposal, leading us to believe that we are basing our analyses on exactly the same data. Second, there is no question that — officially — MAS outperformed Mesa’s party, Civic Community (CC), more strongly on the 5,580 tally sheets that were not counted in the TREP by the time it reached 84 percent. The question is whether or not this overperformance has a benign explanation.

We believe this difference is overwhelmingly benign. Specifically, areas that favored MAS were — on average — less likely to be counted in the 84 percent. By projecting the remaining 16 percent of tally sheets based on nearby tally sheets already included in the count, we find that the change in trend was very predictable.

How is it that we find this, but Escobari and Hoover — with their “Placebo Analysis” — find otherwise? In theory, we ought to be doing very similar things with one very significant difference: the role of geography in informing our projections.

Let us turn now to the actual analysis of Escobari and Hoover. Though we are able to fully and precisely replicate their results, we find that this requires three errors of varying importance.

Problem 1: Vote Shares by Party are Not Computed as a Share of the Actual Valid Votes on Each Tally Sheet

There are a fair number of tally sheets in the official data where the “Votos Válidos” does not match the sum of valid votes. In some cases, “Votos Válidos” reported on tally sheets are irregularly missing or failing to match the sum of valid votes of the individual parties. According to Article 49 of the electoral regulations and the election guide, the reported “Votos Válidos” are intended only as an aid and are in no way binding on the actual count. Rather, the official valid vote count is the sum of the valid votes recorded for each party.

---

6 Throughout their paper, votes for each party by tally sheet come from the official final count. The tally sheets are grouped according to whether or not they were included in the 84 percent of tally sheets counted in the TREP prior to the interruption in reporting.

7 OEP (2019a).
The valid vote in the election came to 6,137,671 — both officially reported and noted in **Table 1** of Escobari and Hoover. However, the total “Votos Válidos” come to 6,137,778. While the aggregate difference is small, the difference may be very large on any given tally sheet.

**TABLE 1**

Table 1 from Escobari and Hoover

<table>
<thead>
<tr>
<th>Parties</th>
<th>Final Comupto</th>
<th>After the Shutdown*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Votes (1)</td>
<td>Percentage (2)</td>
</tr>
<tr>
<td>MAS</td>
<td>2,889,359</td>
<td>47.08%</td>
</tr>
<tr>
<td>CC</td>
<td>2,240,920</td>
<td>36.51%</td>
</tr>
<tr>
<td>PDC</td>
<td>539,081</td>
<td>8.78%</td>
</tr>
<tr>
<td>21F</td>
<td>260,316</td>
<td>4.24%</td>
</tr>
<tr>
<td>MTS</td>
<td>76,827</td>
<td>1.25%</td>
</tr>
<tr>
<td>MNR</td>
<td>42,334</td>
<td>0.69%</td>
</tr>
<tr>
<td>PAN</td>
<td>39,826</td>
<td>0.65%</td>
</tr>
<tr>
<td>UCS</td>
<td>25,283</td>
<td>0.41%</td>
</tr>
<tr>
<td>FPV</td>
<td>23,725</td>
<td>0.39%</td>
</tr>
<tr>
<td>Blank</td>
<td>93,507</td>
<td>1.52%</td>
</tr>
<tr>
<td>Null</td>
<td>229,337</td>
<td>3.74%</td>
</tr>
<tr>
<td>Valid Votes</td>
<td>6,137,671</td>
<td></td>
</tr>
<tr>
<td>Total Votes</td>
<td>6,460,515</td>
<td></td>
</tr>
<tr>
<td>polling Stations</td>
<td>34,555</td>
<td></td>
</tr>
</tbody>
</table>

Notes:  
* The numbers in columns 3 and 4 are calculated with final Comupto votes, but from the polling stations that were not included in the TREP prior to the shutdown.  
* Valid votes do not include blank and null votes, while total votes do.

Source: Escobari and Hoover (2019a).

Thus, when the authors chose to use “Votos Válidos” as their denominator when computing party shares or margins on individual tally sheets, their results failed to add up, as we observe in **Table 2**.

Each column of Table 2 represents a regression estimate of the vote share for a given party — the share for tally sheets included in the first 84 percent of the TREP count (the coefficient labeled “Constant”), and the excess share seen in the last 16 percent of tally sheets (labeled “SHUTDOWN”).

---

8 This table was reproduced exactly.
The “Constant” coefficients are meant to represent each party’s share of the vote in the “clean” (or “control”) 84 percent as opposed to the hypothetically contaminated (“treatment”) group. The sum of shares for the parties ought to equal 100, but Escobari and Hoover’s approach only gets to 99.95. This is not a rounding error. Repeating the analysis using the sum of valid votes on each tally sheet instead of “Votos Válidos” yields a sum of exactly 100. Likewise, the excess shares should total zero, rather than Escobari and Hoover’s -0.0199. Correcting the share denominator to the actual sum of valid votes increases very slightly the estimated MAS share in the final 16 percent from 54.67 to 54.71 and the overall MAS share from 47.98 to 48.01.

Neither result matches well the observed data. As agreed in Table 1, the shares after the shutdown, and overall for MAS should be 54.09 and 47.08, respectively. Even uncorrected, Escobari and Hoover’s approach overestimates the early MAS shares. Why?

Problem 2: Vote Shares Are Improperly Weighted

The main problem here is that Escobari and Hoover’s regressions are unweighted. Instead, shares should be weighted by the number of valid votes in order to produce results consistent with the data.

Let’s take an example. Suppose there were only two polling stations in the election. At one station, MAS received two valid votes, CC received only one, and other parties combined for one vote. Thus,
the MAS share at this station would be 50 percent. At the other station, MAS received only a single vote, CC received 90 valid votes, and other parties received nine. The MAS share at this station would be 1 percent.

Now, if we take the unweighted average share, we estimate MAS to have 25.5 percent of the valid vote overall. Yet MAS received only three votes out of 104 cast. Only when we weight the first station by its four valid votes and the other by its 100 do we correctly infer that MAS received 2.88 percent of the vote.

Turning back to the actual election data, we find that by properly weighting the shares by the sum of valid votes on each tally sheet, we get the correct shares for each party and in each group as observed in Table 1. While this weighting does not have a large effect on the estimated difference between treatment and control (the difference is slightly larger than what Escobari and Hoover report) proper weighting does allow us to interpret the results naturally.

The corrected “SHUTDOWN” coefficient of 8.351 tells us that Morales received 82,000 votes (or 1.33 percent of the total valid votes cast) more than would be expected if the MAS vote share had remained unchanged from the first 84 percent to the last 16 percent. However, the geographic mix of tally sheets in the last 16 percent was dissimilar from that of the first 84 percent. If we simply look at the location of the last 16 percent of tally sheets, we see that they are disproportionately in areas where MAS performed well in the initial 84 percent. Thus, we would expect MAS to perform better on the last 16 percent relative to the first 84 percent.

In technical terms, there was a significant selection bias in the natural experiment: the “treatment” group was not representative of the overall population. Escobari and Hoover attempt to correct for this selection bias by adding geographic controls to their regression. Their original results are shown in Table 3. They report that rather than an “expected” 16.26 percentage point increase in the MAS share (relative to CC), as in column (1), that precinct level controls reduce the estimated difference to only 0.954 percentage points — or 0.153 percent of the total valid vote.

In other words, when controlling for geography, the unexplained increase in the MAS margin, according to Escobari and Hoover, was not sufficient to change the outcome of the election. Morales’s lead grew beyond the necessary 10 percentage points over Mesa in the last 16 percent of the count for the simple reason that the as-of-yet uncounted tally sheets were disproportionately in precincts that already had shown a tendency to favor Morales and MAS over Mesa and CC.

8.3508 percent of 981,713 valid votes cast after the shutdown equals 81,981.
Once we use correct party shares and proper tally sheet weights in the analysis, the explained increase falls even further — to 0.927 percentage points, representing only 0.148 percent of the total valid vote.

But even this estimate overstates the extent of the change in margin.

**Problem 3: The Geographic Controls (Were) Incorrectly Coded**

To reproduce Escobari and Hoover’s original results, we coded fixed effects based on the name of the municipality or precinct. However, the names did not uniquely identify the municipality or precinct. To take the most obvious examples there are 17 precincts with the name “Embajada.” There are four municipalities called “Santa Rosa” located in four different departments. We alerted Escobari to the problem, and the paper was updated with a revised Table 3 (Table 4, here).\(^\text{10}\)

As we can see in column (3), once precinct-level controls were introduced, Escobari and Hoover reported an unexplained increase in the MAS margin of 0.365 percentage points, or about 0.0602 percentage points.
percent of the total valid vote, compared to 2.60 percent without any geographic controls. In other words, the selection bias in the mix of geography explains almost completely the change in margin. Escobari and Hoover’s results, when controlling for geography, show no “fraud” on a scale sufficient to challenge the official outcome of the election.

TABLE 4
Table 3 (Revised) from Escobari and Hoover

<table>
<thead>
<tr>
<th>Variables</th>
<th>Share of Votes</th>
<th>Number of Votes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>SHUTDOWN</td>
<td>16.26***</td>
<td>7.243***</td>
</tr>
<tr>
<td></td>
<td>(0.653)</td>
<td>(0.437)</td>
</tr>
<tr>
<td>Constant</td>
<td>9.830***</td>
<td>11.28***</td>
</tr>
<tr>
<td></td>
<td>(0.266)</td>
<td>(0.162)</td>
</tr>
<tr>
<td>Municipality FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Precinct FE</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>34,529</td>
<td>34,529</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.017</td>
<td>0.640</td>
</tr>
</tbody>
</table>

Notes: The dependent variable in columns 1 through 3 is the difference in the share of the votes (from 0 to 100), while in columns 4 through 6 is the difference in the number of votes. α Municipality and precinct fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Source: Escobari and Hoover (2019b).

We successfully replicated the results of the corrected Table 3 with one exception. In column (3) we find a “SHUTDOWN” coefficient of 0.377 — slightly, but not meaningfully, larger than that reported by Escobari and Hoover. As all other numbers agree exactly, we suspect this is simply a reporting error on their part.

However, when we correctly compute the vote shares and weight the tally sheets in the analysis, the unexplained increase in margin of the final 16 percent of votes falls all the way to 0.287 percentage points — not even statistically significant (Table 5). The 95 percent confidence interval on the “clean” MAS margin is 10.40–10.64 percent of the valid vote.
TABLE 5  
Corrected Regression Results

<table>
<thead>
<tr>
<th>Variables:</th>
<th>Share of Votes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>SHUTDOWN</td>
<td>16.77***</td>
<td>7.766***</td>
</tr>
<tr>
<td></td>
<td>(0.663)</td>
<td>(0.446)</td>
</tr>
<tr>
<td>Constant</td>
<td>7.883***</td>
<td>9.323***</td>
</tr>
<tr>
<td></td>
<td>(0.264)</td>
<td>(0.166)</td>
</tr>
<tr>
<td>Municipality FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Precinct FE</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>34,551</td>
<td>34,551</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.019</td>
<td>0.627</td>
</tr>
</tbody>
</table>

Source and notes: OEP (2019b) and author’s calculations.

Thus, we take Escobari and Hoover’s analysis as confirmation of our previous analysis of the election.

Spin

We strongly question the characterization of Escobari and Hoover’s finding, “the extent of the fraud is at least 2.67 percent of the valid votes” (emphasis added). Given that “fraud” is defined here as the unexplained increase in the MAS margin, the model without any other controls must produce an initial upper bound on what is left unexplained. When precinct-level controls are employed, that “upper” bound falls below 0.05 percent of the valid votes. Everything else is explained by the extent to which precincts were counted in the initial 84 percent TREP.

Escobari and Hoover dismiss their own findings of lesser “fraud,” arguing:

One drawback from the fixed effects estimates is that as we move to control for smaller geographical units, some of the variation in the dependent variable that is due to electoral fraud will be erased if, for example, electoral fraud affects all polling stations within the same precinct.

But their paper is predicated on the idea that there are two distinct classes of tally sheets — those included in the TREP at 84 percent and those (assumed potentially “fraudulent”) tally sheets not included in the TREP at 84 percent. This may permit detection of a specific type of “fraud” — specifically that applied to the later tally sheets, but not the former.

While it is true that a different sort of “fraud” could contaminate all tally sheets at a particular precinct, this is simply not the type of “fraud” suitable for detection by a natural experiment of this sort.
Obviously, an unexplained factor that applies steadily throughout the count cannot be detected. This “drawback” applies to the model without geographic fixed effects as well — the fixed effects have nothing to do with it. Rather, the results with fixed effects serve as direct evidence against the interpretation that the baseline estimate captures “fraud” at all.

Escobari and Hoover are free to categorize their estimates as lower bounds insofar as there could be fraud undetectable by exploiting the natural experiment, but they cannot take the highest estimate of “fraud” from the model with the least explanatory power — the model which amounts to nothing more than simply declaring the difference in margins “fraud” — and call that the “lower bound.” That is — bluntly — not what they found.

Finally, we express very deep concern with Escobari and Hoover’s characterization of their findings in Section 4.3. They observe that the tally sheets not included in the initial 84 percent TREP were reported more or less uniformly through the official count. They find this “puzzling,” saying:

We have no explanation for this other than the TSE was trying to deliberately hide these polling stations and avoid jumps. We read this as additional evidence that identifies these group [sic] of polling stations as fraudulent.

Like the OAS in failing to explain the change in trend, Escobari and Hoover here simply throw up their hands. But there is a very good explanation for what they observe. The TREP results are transmitted electronically when the tally sheets are filled, so that geography _per se_ (as opposed to factors correlated with geography) should have no impact on the order the results are received — and presumably are reported. On the other hand, the official cómputo results require delivery of the physical tally sheets to the nine departmental tribunals. Though tally sheets from the nine capital cities where the departmental tribunals are located were more likely to be reported early in the TREP, this is far truer in the official count as they came from precincts located nearby.

This means that the cómputo order is strongly related to the time required to deliver the tally sheets, rather than order of inclusion in the TREP. Because precincts represented in the final 16 percent — despite the tendency to favor MAS — are nevertheless geographically diverse they are correspondingly scattered throughout the official count.

It is important to recognize, however, that Escobari and Hoover’s Figure 3 (Figure 1, here) is misleading for several reasons:

- First, the presentation is arbitrarily selective — there is no reason that other time slices of the TREP could not be shown as placebo “controls.”
• Second, the choice to include a line indicating progress expands the vertical axis, making it hard to discern the extent to which the vote share varies.

• Third, as the shares shown are cumulative, this has the effect of increasingly masking any variation in the trend over time; the latter part of the graph contains effectively no information.\footnote{While in our original analysis we also showed cumulative shares of the margin over time, our point was the opposite — that the margins were changing over time with sufficient strength to impact the overall vote, and this was plainly visible in the figures. If anything, our choice understated our case. Here, Escobari and Hoover argue that the shares are so evenly distributed throughout that only fraud can explain, so masking the variation exaggerates.}

• Fourth, the choice of time as the horizontal axis unnecessarily stretches out the latter portions of the data so that a small fraction of the data consumes 60 percent of the figure.

**FIGURE 1**

Figure 3 from Escobari and Hoover

![Reported CompuTo Timing](image)

**Fig. 3:** Timeline of the reporting of the polling stations by CompuTo.

*Source: Escobari and Hoover (2019a).*
All four of these reasons contribute to exaggerating the extent to which the 16 percent of tally sheets considered appear to be uniformly distributed throughout the cómputo. **Figure 2** shows the distribution of tally sheets in the official count for each of six slices of the TREP — 5,580 tally sheets each. Together, the first five represent most of the 84 percent of tally sheets reported in the TREP before the interruption. The sixth (red) represents tally sheets that had not yet been reported.

**FIGURE 2**

**Six Different Timeslices of the TREP Scattered Throughout Cómputo**

Slice 6, representing the tally sheets purportedly “dosed” uniformly across the cómputo, was clearly neither suspiciously nor uniformly scattered through the official count.

Source: OEP (2019b) and author’s calculations.
**Conclusion**

Escobari and Hoover produce no credible evidence that the official results of the election should be doubted. Instead, their approach confirms our finding that the expanded MAS margin on the later tally sheets is explained by the mix of precincts considered before and after the interruption of the unofficial TREP count. Their alarm regarding “dosing” was based on a highly distorted presentation of the facts. Instead, careful examination simply reminds us that geography was even more critical to the order in which tally sheets were counted in the cómputo than to the order in the TREP.
References


