10-27-2020

# Evo Morales and Electoral Fraud in Bolivia: A Natural Experiment and Discontinuity Evidence

Diego Escobari
*The University of Texas Rio Grande Valley*, diego.escobari@utrgv.edu

Gary A. Hoover

# Evo Morales and Electoral Fraud in Bolivia: A Natural Experiment and Discontinuity Evidence*

Diego Escobari† Gary A. Hoover‡

October 15, 2020

**Abstract**

This paper uses a unique data set and a natural experiment based on the shutdown in the official preliminary vote counting system to identify and estimate the size of electoral fraud in the 2019 Bolivian presidential elections. The 2016 Constitutional Referendum and the participation of other political parties serve as controls to estimate various difference-in-differences and difference-in-difference-in-differences specifications. The results show evidence of a statistically significant electoral case of fraud that increased the votes of the incumbent *Movimiento al Socialismo* and decreased the votes of the runner up *Comunidad Ciudadana*. We estimate that the extent of the fraud is 2.50% of valid votes, sufficient to change the outcome of the election. We report a break in trend and evidence of fraud beyond the shutdown. Our results are robust to polling-station-level shocks common across 2019 and 2016, as well as 2019 specific shocks. This controls for geography (e.g., rural vs. urban), unobserved voting preferences, voter's last names, and endogeneity in the arrival of the polling stations. We document a statistically significant discontinuous jump in the gap between the incumbent and the runner up during the shutdown.

*Keywords*: Electoral Fraud; Natural Experiment; Bolivia; Evo Morales; Difference-in-Differences; Regression Discontinuity

*JEL Classifications*: C21; D72; K42

*We thank comments by Binay Adhikari, Gustavo Canavire-Bacarreza, Francisco Cantú, Pablo Cuba-Borda, Jorge Dávalos, Jean-Baptiste Tondji, Walter Valdivia, and participants at the Bolivian Conference in Development Economics 2020. We also thank Edgar Villegas for sharing data with us.

†Department of Economics & Finance, The University of Texas RGV, Edinburg, TX 78539, Phone: (956) 665-2104, Email: diego.escobari@utrgv.edu, URL: http://faculty.utrgv.edu/diego.escobari

‡Department of Economics, University of Oklahoma, Norman, OK 73072, Phone: (405) 325-5857, Email: ghoover@ou.edu, URL: https://sites.google.com/site/garyhoovereconomics

# 1 Introduction

"If the fraud is proved, we go to the runoff" said Bolivian President Evo Morales on October 26, 2019 after being accused of electoral fraud in the elections that took place only six days earlier.[1] Due to its illegal nature, the identification of electoral fraud is difficult. Moreover, convincing identification of the magnitude of the fraud is methodologically challenging (Enikolopov et al., 2013). Capturing fraud is important as it slows leadership turnover, which is known to lower economic growth (Besley et al., 2010). It also decreases accountability and helps autocratic leaders stay in power (Magaloni, 2009).

In this article we use various estimators following a natural experiment to test for the existence and estimate the magnitude of electoral fraud in the Bolivian presidential elections of October 20, 2019. A shutdown of the official preliminary vote counting system *Transmisión de Resultados Electorales Preliminares* (TREP) gave us a rare opportunity that created a natural experiment to formally test for electoral fraud.[2] We take advantage of the shutdown to separate polling stations between a control group that is less likely to be associated with fraud, and a treatment group where fraud is more likely to have occurred. The Bolivian presidential elections were run under some unique rules. To win the election a candidate either needed to obtain the majority of the votes or get more than 40% with a difference of at least 10% over the runner up. We test whether the exogenous effect of the fraud treatment increased the votes of the incumbent *Movimiento al Socialismo* (MAS) and if it decreased the votes of the main opposition party *Comunidad Ciudadana* (CC) so that fraud would help to have a gap of at least 10%.

We have a dataset that contains information unavailable to other studies that looked at the 2019 Bolivian elections. Our time stamps of the arrival of polling stations at the Electoral Court, the names of the electoral judges associated with each polling station, and the booth identifiers that allow us to match the 2019 elections with the 2016 referendum at the polling station level. Using 2019 votes alone, we find preliminary evidence of fraud when we control for unobserved heterogeneity across precincts using fixed effects. We then provide various Difference-in-Differences (DiD) estimates using the 2016 constitutional referendum as control with flexible specifications that allow for heterogeneous voting preferences across ballot boxes. The results show how the shutdown had a statistically significant negative effect on CC and a positive effect on MAS. We also document a change in trend in the gap MAS−CC at the shutdown. As robustness check we relax the parallel trends assumption to find that the gap between MAS and CC continuously increased throughout the arrival of ballot boxes to the Electoral Court. This is

[1] Mitra Taj and Daniel Ramos "Bolivia's Morales vows second-round vote if fraud found in election, threatens siege of cities" Reuters, October 26, 2019.

[2] Roe and Just (2009) explain that natural experiments typically have a greater external validity than field experiments (see, e.g., Duflo et al., 2008).

consistent with a previously reported within-precinct trend. Moreover, we show it is consistent with fraud that took place even before the shutdown.

We additionally use the votes of other political parties in the election and provide Difference-in-Difference-in-Differences (DDD) estimates. The results indicate that fraud increased the gap between MAS and CC by 2.50% of the valid votes, with 1.92% corresponding to pre-shutdown booths and 0.58% to shutdown polling stations. These estimates are immune to polling station level shocks that are common across 2019 and 2016, and shocks that are specific to the 2019 elections (e.g., geography or voting preferences). Our DiD and DDD specifications include polling station fixed effects, which additionally allows us to control for any potential endogeneity in the arrival of polling stations to the Electoral Court.

Our last set of results present Regression Discontinuity (RD) evidence of fraud. To the best of our knowledge, this is the first paper to provide empirical evidence that RD design can be used to identify fraud. Our approach follows the intuition explained in Organization of American States (2019). We find a statistically significant discontinuous jump when 0.95% of the polling stations had arrived at the Electoral Court. This result is robust to the kernel type, bandwidth selectors, and the choice of order of the polynomial regressions. The set of booths that cause the increase coincide with a rush of fraud votes that were never reported on the TREP. Both, DDD and RD evidence show that the size of the fraud was big enough to change the outcome of the election.

Evo Morales eventually stepped down on November 10, 2019 following further accusations of election fraud, mass protests, and an Organization of American States (2019) audit that found "serious irregularities" in the elections.[3] In particular, the international audit discovered a manipulation of the reporting system, physically altered tally sheets, forged signatures, and was unable to confirm a first-round victory for Morales. Following the constitutional line of succession, Jeanine Añez became president, and all political parties in congress swiftly agreed to call for new elections.

This would have been Morales' fourth consecutive term in office. When he was first elected in 2005, the constitution did not allow for reelection, but he changed it in 2009 to allow for a single reelection. Based on a dubious interpretation of the law he was able to run for a third term in 2014. In 2019, Morales was on the ballot against the constitution and against a 2016 constitutional referendum that he lost. His participation followed a controversial ruling by the Supreme Court—completely under his control—qualified Morales saying that limits on the lengths of his tenure would violate Morales' "human rights".[4] It was apparent that he was

[3]Kay Guerrero and Dakin Andone "Bolivian President Evo Morales steps down following accusations of election fraud" CNN, November 10, 2019.

[4]See, for example, "Evo Morales finds a way to run for re-election" *The Economist*, December 1, 2017, and

willing to go great lengths to stay in power. His resignation ended close to 14 years of a once very popular president that had important achievements (see, e.g., Hicks et al., 2018) under extremely favorable external conditions (Chumacero, 2019b).

The 2019 Bolivian elections caught the attention of various researchers. Chumacero (2019a) presents a theoretical framework to understand fraud with an empirical section that finds different patterns in the polling stations with and without TREP. Moreover, Newman (2020) reports changes in the distributions due to the shutdown, Dávalos (2019) shows a positive relationship between the number of anomalies and MAS votes, and Mebane Jr. (2019) identifies fraud polling stations, but finds that those were not decisive. Finally, Curiel and Williams (2020) and Idrobo et al. (2020) aim at supporting the hypothesis that there was no fraud due to the shutdown. We discuss in Section 7 some important flaws in their analyses. To the best of our knowledge, ours is the first paper to formally test and capture the magnitude of the fraud.

Previous work aimed at capturing the size of the voting fraud includes Mebane Jr. and Kalinin (2010) who use second-digit mean tests, but need to make strong assumptions regarding the underlying voters' preferences. Beber and Scacco (2012) and Ichino and Schündeln (2012) focus on voters' registration rather that voting results. Moreover, Hyde (2007) and Fukumoto and Horiuchi (2011) assume away the existence of voters' heterogeneity across precincts, while Enikolopov et al. (2013) use the random assignment of independent observers in Russian parliamentary elections. In addition, Callen and Long (2015) study the role of political networks and weak institutions to find that the announcement of a new monitoring technology reduced fraud, consistent with the monitoring of illegal activities under corruption in Escobari (2012). Fujiwara (2015) used a regression discontinuity design to show the importance of voting technology. Kobak et al. (2016) use the frequency of reported round percentages to present a novel statistical fraud indicator. This large body of work on electoral fraud is part of the election forensics literature (see, e.g., Myagkov et al., 2009), where one of the goals is to diagnose the accuracy of reported election results.[5]

The rest of the paper is structured as follows. Section 2 presents the data, while we explain the natural experiment and present preliminary evidence on Section 3. The Difference-in-Differences results are reported and discussed in Section 4, along with a discussion on the trends (Section 4.1), and evidence of widespread fraud (Section 4.2). Sections 5 and 6 present the results for the difference-in-difference-in-differences specifications and the regression discontinuity. Section 7 reconciles our findings with related work, while Section 8 concludes.

Yascha Mounk "Evo Morales finally went too far for Bolivia" The Atlantic, November 11, 2019.

[5]For a review of the literature, see Lehoucq (2003).

## 2 Data

Most of the data used in this paper comes directly from the *Organo Electoral Plurinacional* (OEP), the official government body in charge of the elections. We obtained from the OEP two versions of the TREP and the final version of the *Computo*. As stated earlier, the stands for "Transmission of Preliminary Electoral Results" and its role was to publish online the preliminary results of the polling stations in real time. The first version of the TREP that we have corresponds to the shutdown of the system that occurred at 7:40 pm on the day of the elections. This version gives us the list of polling stations that were reported prior to the shutdown. The second version of the TREP that we have is the final version, which allows us to know the list of polling stations that were never reported on the TREP (4.4% of all the booths). All estimates in the paper come from using the final version of the *Computo*, the official final vote count as reported by the OEP. In addition, we have the list of names of the jurors at the polling station level, the exact time at which polling stations were received at the Electoral Court, exact geolocation of the localities, the final vote count of the 2016 referendum, and booth keys for 2016 and 2019 that allow matching at the polling station level.

We also have the sequence at which polling stations were reported in the final official vote count *Computo*. This information was openly shared by Irfan Nooruddin. Moreover, we have data shared by Edgar Villegas. This includes 104 different versions of the *Computo*, recorded approximately every hour between October 20, 2019 at 10:29 pm and October 25, 2019 at 7:20 am. We included an additional wave recorded on Friday, October 25, 2019 at 9:09 pm that contains 100% of the valid votes.

Table 1 presents a summary of the votes by political party. Column 2 shows how the difference between MAS and CC (10,57%=47.08%−36.51%) is above the 10% MAS needed to avoid the runoff. When breaking down this difference, we see from columns 3 and 4, that the margin in favor of Morales increased from 7.88% (45.74%−37.86%) in the polling stations reported before the shutdown to 24.65% (54.09%−29.44%) after the shutdown. A central element in this paper is to try to figure out if this sudden increase can be attributed to electoral fraud or has more benign explanations (e.g., geography, voter's preferences). Note that when comparing columns 3 and 4, only MAS and CC appear to have experienced relatively big changes at the shutdown as the other shares of political parties in the election appear relatively stable.[6]

Note that shares are calculated only using valid votes, so voting blank or null essentially

[6] In addition to Movimiento and Socialismo (MAS) and Comunidad Ciudadana (CC), the other political parties running were Partido Demócrata Cristiano (PDC), Bolivia dice No (21F), Movimiento Tercer Sistema (MTS), Movimiento Nacionalista Revolucionario (MNR), Partido de Acción Nacional Boliviano (PAN), Unidad Cívica Solidaridad (UCS), and Frente para la Victoria (FPV).

**Table 1:** Summary of the votes

| | Final | | Shutdown Shares[a] | | Correlations[b] | |
|---|---|---|---|---|---|---|
| | Votes | Share | Before | After | Blank | Null |
| Parties: | (1) | (2) | (3) | (4) | (5) | (6) |
| MAS | 2,889,359 | 47.08% | 45.74% | 54.09% | 0.203 | 0.112 |
| CC | 2,240,920 | 36.51% | 37.86% | 29.44% | -0.262 | -0.172 |
| PDC | 539,081 | 8.78% | 8.75% | 8.95% | 0.011 | 0.113 |
| 21F | 260,316 | 4.24% | 4.30% | 3.92% | 0.065 | 0.024 |
| MTS | 76,827 | 1.25% | 1.23% | 1.36% | 0.180 | 0.124 |
| MNR | 42,334 | 0.69% | 0.68% | 0.73% | 0.136 | 0.077 |
| PAN | 39,826 | 0.65% | 0.65% | 0.63% | 0.044 | 0.074 |
| UCS | 25,283 | 0.41% | 0.41% | 0.45% | 0.125 | 0.084 |
| FPV | 23,725 | 0.39% | 0.38% | 0.44% | 0.254 | 0.201 |
| Blank | 93,507 | 1.52% | 1.43% | 2.00% | 1.000 | 0.251 |
| Null | 229,337 | 3.74% | 3.64% | 4.22% | | 1.000 |
| Valid Votes[c] | 6,137,671 | | 5,155,958 | 981,713 | | |
| Polling Stations | 34,555 | | 28,975 | 5,580 | | |

*Notes:* [a] The numbers in columns 3 and 4 are calculated using the final *Computo* votes. [b] Correlations between blank, null (both divided by valid votes), and final shares. [c] Valid votes do not include blank and null votes.

helps increase the gap. Hence, in addition to the basic fraud strategies of simply increasing MAS or decreasing CC, shifting CC votes to blacks and nulls makes MAS−CC grow. This explains two additional elements in Table 1 that are consistent with fraud. First, blanks and nulls both increase with the shutdown. Second, there is a negative pair-wise correlation between final CC shares and both blanks and nulls. This is puzzling because this correlation is positive for all the rest of the parties.

# 3 The Shutdown as a Natural Experiment

On the day of the elections, Sunday October 20, 2019, TREP stopped posting the information online at 7:40 pm when it had already reported 84% of the valid votes.[7] The president of the Electoral Court (*Tribunal Supremo Electoral*), María Eugenia Choque announced that they stopped transmitting the TREP results to avoid "confusions" as they were planning to start transmitting the verified final results via the *Computo*. The firm in charge of the transmission later indicated that the shutdown was implemented under direct orders of the Electoral Court without any technical reasons. The representatives of the Organization of American States overseeing the election met with Electoral Court officials to stress the importance of keeping

[7]This 84% is calculated using the valid votes reported in Table 1, i.e., $\frac{5,155,958}{6,137,671} = 84\%$.

the TREP running. Regardless, TREP remained inactive for an additional 23 hours. This disruption of the system resulted in a general public outcry as it jeopardized the transparency of the process.

The shutdown helps our first identification strategy as it divides the polling stations in two. The first group, reported prior to the shutdown, serves as a control group that is less likely to be associated with fraud because the information on these polling stations was made public faster. The second group of polling stations, which was halted for over 23 hours, is our fraud treatment group. We later discuss the existence of fraud in the control ballot boxes.

For our initial test for differences between the control and the (fraud) treatment group, we estimate the following equation:

$$\mathrm{V}_{ij} = \beta^{\mathrm{D}} \cdot \textsc{Shut}_{ij} + \mu_j + \varepsilon_{ij}, \tag{1}$$

where $\mathrm{V}_{ij}$ denotes the share of the votes for $\mathrm{MAS}_{ij}$, votes for $\mathrm{CC}_{ij}$, or the difference, $\mathrm{MAS}_{ij}-\mathrm{CC}_{ij}$, in polling station $i$ which belongs to precinct $j$. $\textsc{Shut}_{ij}$ is the shutdown indicator variable equal to one if polling station $i$ in precinct $j$ was reported in the TREP after the shutdown, zero otherwise. $\mu_j$ is the precinct unobserved fixed effect, while $\varepsilon_{ij}$ is the remainder stochastic term.

We are interested in estimating $\beta^{\mathrm{D}}$, the effect of the shutdown on the votes. In addition to the votes of CC and MAS, it makes sense to use MAS−CC because the goal for MAS was to have a gap of 10% or more to avoid the runoff. If $\textsc{Shut}_{ij}$ is uncorrelated with the error term, $\beta^{\mathrm{D}}$ can be interpreted as the size of the fraud. We argue that the shutdown can be considered a natural experiment because there has been clear treatment exposure to a subset of polling stations. In addition, shutdown is exogenous to voting because voting had already finished by 7:40 pm. However, fraud was still likely to take place after voting was over. The important element in the estimation of Equation 1 is to control for potential differences in voting preferences before and after the shutdown.

### 3.1 Preliminary Evidence on the Shutdown

The first three columns of Table 2 present the pooled least squares regression results. The point estimate on the shutdown in column 1 shows that the polling stations after the shutdown saw a statistically significant 8.29% decrease in the votes for CC. Moreover, column 2 shows that the votes for MAS experienced a statistically significant 7.98% increase when they were not included in the TREP prior to the shutdown. From Table 1 we see that 981,713 of the votes remained to be reported after the shutdown, which implies that the shutdown increased the difference between MAS and CC by an estimated 159,636 votes (981,713×(7.98+8.29)/100) or about 2.60% of the valid votes. This magnitude is big enough to have switched the outcome of the election. Note that the magnitude of the decrease in CC is greater than the magnitude of the increase for MAS.[8] This is consistent with the goal of increasing the gap, where it is more efficient to decrease CC than to increase MAS.

**Table 2:** Difference Estimates

| Dependent variable: | CC | MAS | MAS−CC | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Variables:* | | | | | | |
| SHUTDOWN | -8.286*** | 7.975*** | 16.26*** | 7.243*** | 6.762*** | 0.365* |
| | (0.324) | (0.343) | (0.653) | (0.437) | (0.464) | (0.194) |
| Constant | 36.86*** | 46.69*** | 9.830*** | 11.28*** | 11.36*** | 12.39*** |
| | (0.136) | (0.134) | (0.266) | (0.162) | (0.151) | (0.0631) |
| *Fixed Effects*[a] *(F-statistic):* | | | | | | |
| Municipality | No | No | No | 129.6 | Yes | Yes |
| Locality | No | No | No | No | 23.49 | Yes |
| Precinct | No | No | No | No | No | 124.9 |
| Observations | 34,529 | 34,529 | 34,529 | 34,529 | 34,529 | 34,529 |
| R-squared | 0.017 | 0.016 | 0.017 | 0.640 | 0.740 | 0.958 |

*Notes:* The dependent variable is indicated in the column's heading. [a] Municipality, locality and precinct fixed effects. Robust standard errors in parentheses. The reported $F$-statistics are from the hypothesis that the corresponding fixed effects are jointly equal to zero. *** p<0.01, ** p<0.05, * p<0.1

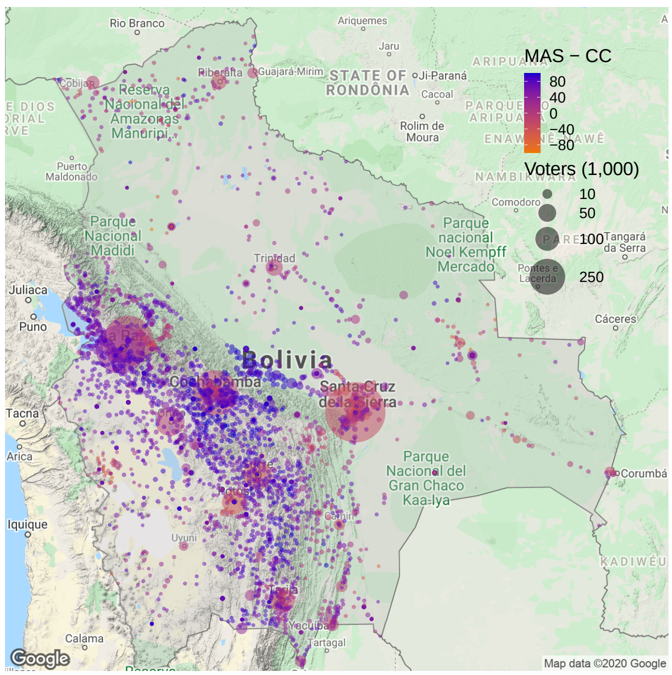
It is known that Morales has greater support in most of the rural areas. Figure 1 illustrates how larger metropolitan areas, presented as larger bubbles, typically have a relatively larger share of votes for CC, while the rural vote is more likely to support MAS in the highlands and is closer to a balanced mix between MAS and CC in the north and east part of the country.[9] In addition to geography, the gap between MAS and CC might have also been influenced, for example, by levels of education and income. Trying to use socioeconomic characteristics to control voting preferences is not a good option because the data is too aggregate.

Unobserved heterogeneity might be driving the results if SHUT in Equation 1 is correlated with one of the components of the two-way error term $\mu_j + \varepsilon_{ij}$, making it endogenous. As explained in Greene (2018), we can control for heterogeneous voters' preferences by simply including geographical region fixed effects $\mu_j$ that will control for any "rural vote" effect even if it has an unknown effect on voting and while allowing for differentiated effects of rural communities on the gap. We exploit the fact that we have 34,555 polling stations divided into 455 municipal-

[8]We run Zellner's seemingly unrelated regressions with the specifications in columns 1 and 2 to test for differences in the magnitude of the coefficient. A $\chi^2_1 = 6.65$ shows they are significantly different at a 1% level.

[9]The colors in the figure follow the official dark blue for MAS and orange for CC.

**Figure 1:** Role of Geography in the Gap MAS−CC




*Notes:* Votes at the locality level. Higher concentration of voters in urban areas are shown as larger bubbles. Larger metropolitan areas typically show relatively more support for CC. Smaller bubbles from rural areas and are more likely to support MAS in the highlands, and are closer to a balanced mix between MAS a CC in the north and east parts of the country.

ities, which are then further divided into 3,593 localities and 5,288 precincts. On average there are 75.95 polling stations per municipality, and only 6.53 per precinct.[10]

Following the two-way error structure in Equation 1, Table 2 allows for systematic variation in the differences in voting preferences across municipalities (column 4), localities (column 5), and across precincts (column 6). Comparing across these columns gives us two important takeaways. First, the coefficient on shutdown remains statistically significant even at the most disaggregated geographical fixed effects (column 6). Second, there is a substantial decrease in the magnitude of the coefficient. The identifying assumption in column 6 is that in the absence of fraud, within the same precinct, the vote margins between MAS and CC are the same before and after the shutdown.

One important drawback from the fixed effects is that as we move to smaller geographical units $j$, some of the variation in the dependent variable that could be attributed to be electoral fraud will be wiped out. For example, there are 1,475 precincts that have a single polling station. Those are located in rural areas that have little supervision and where fraud is relatively easier to implement. Column 6 eliminates those observations as they are spanned by the precinct fixed effects. Moreover, if electoral fraud affects all polling stations within the same precinct, we won't be able to detect it, as it is perfectly collinear with $\mu_j$. Identification is coming only from precincts that have polling stations before and after the shutdown. From the total of 5,288 precincts, 612 have polling stations only after the shutdown, and 3,177 have polling stations only before the shutdown. This means that column 6 relies only on 1,499 precincts (28.3% of the precincts) to identify $\beta^{\mathrm{D}}$ in Equation 1. This is consistent with the substantial increase in the R-squared as we move to include smaller geographical fixed effects. Note that from columns 3 through 6, the variation in the dependent variable explained by the model increased from 1.7% to 95.8% when including precinct fixed effects.

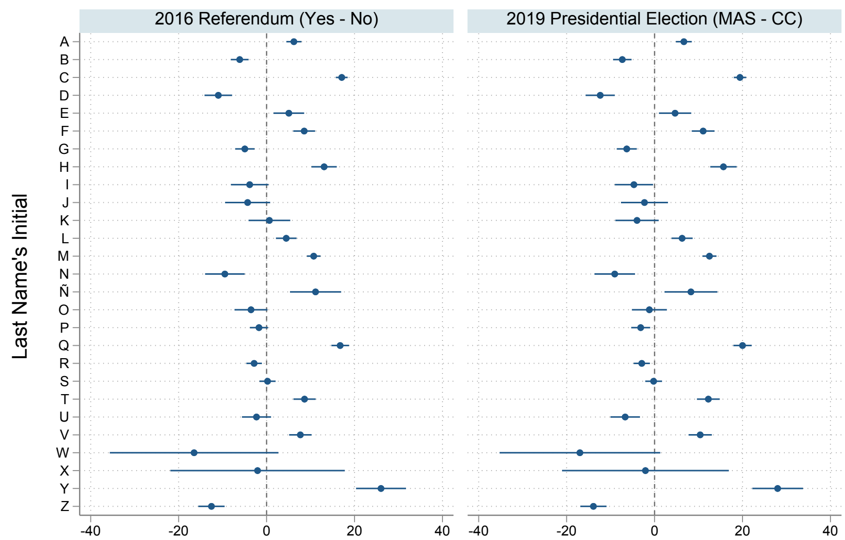
## 4 Difference-in-Differences: The Constitutional Referendum

Bolivia had a constitutional referendum on February 21, 2016, where the proposed constitutional amendment was to allow Evo Morales for an additional reelection. We use this referendum to serve as a control for the 2019 elections. Our argument is that the question that voters needed to answer in 2016 and in 2019 was essentially the same: Do you want Morales to stay as your president? In the 2016 referendum a positive answer was simply "Yes", while in 2019 it was a

[10]There are actually 340 municipalities in Bolivia, but we use the Electoral Court's definition of municipality when classifying geographical regions outside Bolivia, hence the total of 455.

vote for MAS. Hence, the "Yes" votes in 2016 can serve as a control for the MAS votes in 2019. In addition, knowing that the 10% between the first and the second was very likely to play a role in the 2019 final vote count, the opposition concentrated their votes on the runner up, CC. Hence, we can also use the difference YES−NO in 2016 to serve as a control for the MAS−CC gap in 2019.

**Figure 2:** Role of Last Name's Initial on Voting




*Notes:* Last name's initial Fixed Effects point estimates with 95% confidence intervals. The dependent variable on the left-hand side of the figure is YES−NO from the 2016 constitutional referendum, while the dependent variable on the right-hand side is MAS−CC. All as shares of the valid votes.

To illustrate the similarities between the referendum and the presidential election, Figure 2 presents the marginal effects of the voters' last name initials on YES−NO (left-hand side), and MAS−CC (right-hand side). The last names associated with each polling station are approximated using the last names of 204,989 jurors. Voters are assigned alphabetically, based on their last names, to polling stations within the same precinct. Jurors are selected randomly from the list of voters within the same booth; hence, with the names of the jurors we can approximate the last name initials associated with each booth. The point estimates in Figure 2 along with their 95% confidence bands are coming from OLS regressions of the differences as a function of last name initials dummies. The resemblance between 2016 and 2019 is remarkable. For example, the difference between YES and NO for voters whose last name starts with M (e.g., Morales) is about 10.3% in favor of Yes in 2016, and about 12.5% in favor of MAS in 2019.

Working with the 2016 data allows us to use the standard Difference-in-Differences (DiD) estimand:

$$\begin{aligned} \beta^{\text{DD}} &= \left(\mathbb{E}[\text{V}|\textsc{Shut}=1,\text{Y19}=1]-\mathbb{E}[\text{V}|\textsc{Shut}=0,\text{Y19}=1]\right) \\ &- \left(\mathbb{E}[\text{V}|\textsc{Shut}=1,\text{Y19}=0]-\mathbb{E}[\text{V}|\textsc{Shut}=0,\text{Y19}=0]\right) \end{aligned} \tag{2}$$

where V is the share of the votes, and $\textsc{Shut}_{ij}$ is the indicator variable that captures the shutdown. Moreover, Y19 is an indicator variable equal to one for the 2019 elections and zero for the 2016 referendum. $\beta^{\text{DD}}$ aims at capturing the population average difference before and after the shutdown, but after subtracting the population average difference before and after the shutdown for the control group (Y19 = 0) to remove biases associated with a common trend unrelated to the shutdown. We can estimate $\beta^{\text{DD}}$ from the following regression equation:

$$\text{V}_{ijt} = \alpha_1\textsc{Shut}_{ij} + \alpha_2\text{Y19}_t + \beta^{\text{DD}}\cdot\textsc{Shut}_{ij}\times\text{Y19}_t + X_{ijt}\delta + \nu_{ij} + \varepsilon_{ijt}. \tag{3}$$

Note that in addition to having polling station $i$ in geographical region $j$, we also have $t$, which keeps track of whether observations are from 2016 or 2019. $X_{ijt}$ is a matrix of control variables, while $\nu_{ij}+\varepsilon_{ijt}$ is the two-way error component. Including variables that change over $i$ (e.g., the fixed effects specifications), can not only help to control for confounding trends, but might also help reduce the variance of the error term $\varepsilon_{ijt}$. The coefficient of interest ($\beta^{\text{DD}}$) captures the effect of the shutdown on votes during 2019, while using 2016 as a control.

Table 3 presents various sets of estimates of different versions of Equation 3. The last three columns have MAS−CC as the dependent variable, with columns 6 and 7 including precinct and polling station fixed effects, respectively. Note that Equation 3 is very flexible when compared to simple difference-in-differences specifications where there is a single treatment and a single control group. The referendum and the elections are matched at the polling station $i$ level, which means that we have 28,975 controls and 5,580 treatments. $\text{Y2019}_t$ is included to control for permanent differences between 2016 and 2019, while $\textsc{Shut}_{ij}$ is included as our first approach to remove biases from comparisons that could be the results of trends unrelated to the treatment (Imbens and Wooldridge, 2009, p. 67). Identification in DiD comes from the parallel trends assumption. That is, these specifications assume that in the absence of the shutdown treatment, the path of the votes in 2019 would have followed the path observed in 2016, after controlling for systematic differences captured by Y2019, $X_{ijt}$, and the polling station fixed effects. We provide robustness checks to relax the parallel trends assumption.

The numbers in parentheses are robust standard errors, clustered by polling station. Bertrand et al. (2004) show the practical importance of controlling for clustering. The point estimates of the coefficients on Shutdown × Y2019 and Y2019 are fairly stable across specifications that have the same dependent variable. This is evidence that taking the difference between 2019 and 2016 appears to do a good job in controlling for heterogeneity across precincts and polling stations. Moreover, the estimate on Shutdown, which captures the difference in the gap before and after the shutdown for 2016, drops from positive 13.30 (column 5) to negative 1.090 (column 6). We interpret this as evidence that when using 2019 as a control, the gap between YES versus NO votes in the referendum for the shutdown polling stations is actually lower than for pre-shutdown polling stations.

By construction, the 16.26 point estimate on SHUTDOWN of column 3, Table 2, is the sum of the point estimates on SHUTDOWN $\times$ Y2019 and SHUTDOWN of column 5, Table 3 ($2.96+13.30$). Hence, it shows that about 81.8% ($13.30/16.26 \times 100$) of the increase in the MAS$-$CC gap after the shutdown can be explained by the votes in 2016. Note that this 81.8% already controls for heterogeneity across precincts and polling stations that did not change between 2016 and 2019. For example, systematic voting patters within each precinct and polling station, including differences between rural vs. urban voting preferences. The remaining 18.2% is attributed to fraud.

**Table 3:** Difference-in-Differences Estimates

| Dependent variable: | CC | | MAS | | MAS$-$CC | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Variables:* | | | | | | | |
| SHUTDOWN $\times$ Y2019 | -1.698*** | -1.657*** | 1.280*** | 1.160*** | 2.964*** | 2.699*** | 2.842*** |
| | (0.141) | (0.228) | (0.273) | (0.420) | (0.334) | (0.335) | (0.459) |
| SHUTDOWN | -6.588*** | | 6.695*** | | 13.30*** | -1.090*** | |
| | (0.292) | | (0.308) | | (0.610) | (0.212) | |
| Y2019 | -13.80*** | -13.49*** | -2.226*** | -2.357*** | 11.99*** | 11.15*** | 11.13*** |
| | (0.292) | (0.305) | (0.337) | (0.309) | (0.635) | (0.443) | (0.612) |
| Constant | 50.66*** | 49.43*** | 48.92*** | 49.92*** | -2.157 | 0.616** | 0.438 |
| | (1.693) | (0.162) | (1.694) | (0.130) | (3.380) | (0.298) | (0.290) |
| *Fixed Effects ($F$-statistic):* | | | | | | | |
| Precinct | No | Yes | No | Yes | No | 158.7 | Yes |
| Polling Station | No | 15.48 | No | 19.53 | No | No | 24.90 |
| Observations | 66,541 | 66,541 | 66,535 | 66,595 | 66,535 | 66,535 | 66,535 |
| R-squared | 0.103 | 0.949 | 0.017 | 0.955 | 0.035 | 0.934 | 0.965 |

*Notes:* The dependent variable is indicated in the column's heading. The numbers in parentheses are robust standard errors, clustered by polling station. The reported $F$-statistics are from the null hypothesis that the corresponding fixed effects are jointly equal to zero. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

When reducing the variance of the error term in the polling station fixed effect specification of column 7, SHUTDOWN is spanned by the fixed effects as this is a flexible specification that accounts for further nonlinearities in the path of the arrival of polling stations at the electoral court that are common to 2016 and 2019. The point estimates in columns 1 thorough 4 illustrate how the change in the gap was mostly due to a decrease in the votes for CC and to a lesser extent to an increase in MAS votes. This is consistent with the difference estimates reported in

Table 2.

## 4.1 Trends

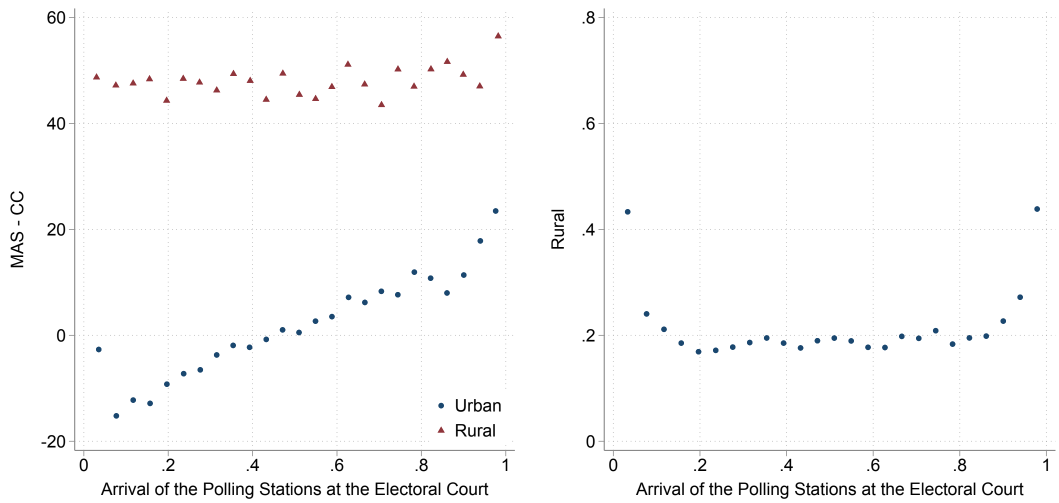
As voting in each polling station finishes, booth jurors count the votes in an open ceremony. Once this official count is over, jurors sign the booth minutes and then an Electoral Court employee takes a picture of the minutes and sends it *online* to the Electoral Court. We keep track of two important trends. First, the trend based on the arrival of the polling station minutes to the Electoral Court. The order of arrival of these polling stations is essentially the order in the TREP (i.e., the preliminary results published online). However, the TREP time stamps have one critical drawback. It happened that 4.4% of the polling stations were never reported on the TREP; hence, they are missing time stamps. This 4.4% batch is associated with fraud (Chumacero, 2019a), so we focus on the initial arrival of the minutes to the Electoral Court to be able to include this critical 4.4% batch. The second trend follows the *Computo*, and it is based on the timing at which the final official results are published online. The timing here largely follows the physical arrival of the booth minutes to the local Electoral Court. This is less critical for fraud as the picture of those minutes was already in possession of the Electoral Court.

The specifications in Table 3 already control for nonlinear trends on the arrival of the polling stations at the electoral court, the timing of the reporting on the TREP, and timing of the reporting on the *Computo*. This holds for any 2019 polling station level shock that is also captured by the same polling station in 2016 (e.g., rurality). Given the discussion in related work and in the media regarding the importance of the trends, it is worth estimating some specifications simply with precinct fixed effects to be able to identify the roles of different trends. The incumbent party justified the unusual increase in their votes after the shutdown by arguing that the polling stations from rural communities were yet to be considered and that those communities were more likely to support them. It is true that MAS has greater support in rural areas as we illustrate on the left-hand side of Figure 3. We can see that the gap between MAS and CC averages over 40% in rural areas, while it is significantly lower in urban areas. We define a rural polling station if it belongs to a locality that is at least 50 kilometers (31 miles) away from one of the largest 50 cities in the country, based on population.[11]

There are two additional takeaways worth noticing from Figure 3. First, the left-hand side suggest that there is an upward trend in the MAS−CC gap for urban areas, but not for rural

[11] We use the latitude and longitude of each of the localities to calculate the distances between cities and localities using Open Source Routing Machine (OSRM). We also used different distances and number of largest cities and they all convey the same basic message.

**Figure 3:** Arrival of the Polling Stations at the Electoral Court




*Notes:* The left-hand side shows the gap MAS−CC as a function of the arrival of the polling stations at the Electoral Court (Arrival), separated into rural and urban votes. The right-hand side shows the share of rural polling stations as they arrive to the Electoral Court. These are the arrivals of the booth minutes transmitted online.

areas. Second, from the right-hand side we observe that the rate at which urban and rural polling stations arrive at the Electoral Court appears to be relatively stable. The right-hand side of the figure illustrates the fraction of rural votes as a function of their arrival to the electoral court. For most of the process about 20% of the polling stations that arrive to the court at a given point in time originate from a rural area. Early in the process this fraction is about 40% probably due to some small precincts in rural areas that finish counting the votes relatively fast and are able to report relatively early. Moreover, there is a sudden increase at the end, probably driven by rural areas that do not have internet so that it takes longer for those to submit their results. This sudden increase is consistent with the hypothesis that rural votes might explain the increase in the gap. On the other hand, the left-hand side figure shows that the gap between MAS and CC also appears to have increased at the end for both urban and rural polling stations. A less publicized feature of the election is that the minutes of each of the polling stations arrive at the Electoral Court online. Hence, there is really no obvious reason why rural votes should generate a trend (i.e., increasing share of rural votes). It was during the 1980s and early '90s where Bolivians were accustomed to seeing the "rural vote effect" when the votes of the *Movimiento Nacionalista Revolucionario* typically took longer to physically arrive from the rural areas.

Let Arrival $\in [0, 1]$ be the variable that captures the order at which the picture of the polling station minutes arrives online at the Electoral Court. Moreover, we define the variable COMPUTO $\in [0, 1]$ as the order at which the Electoral Count publishes the official final results on the *Computo*. Idrobo et al. (2020) claim that a within-precinct "secular trend" can be explained by the last names of the voters, but provide no empirical evidence to support their claim. Hence, in addition to including ARRIVAL and COMPUTO, we want to have a specification with precinct fixed effects and the last name initials to test Idrobo et al. (2020)'s assertion. Columns 1, 3 and 5 of Table 4 estimate such a model, with CC, MAS and the gap as dependent variables.

From columns 1, 3, and 5 we see that COMPUTO is not statistically significant, suggesting that the order of reporting of the final results does not explain CC, MAS, or the gap. This result is quite interesting because COMPUTO is highly correlated with geography as the physical distance between each precinct and the local departmental courts clearly plays a role here. The non-statistically significant coefficients signal that the 2016 referendum works well to control for this confounding trend that affects both, 2016 and 2019. Consistent with Figure 2, the last name initials have an $F$-statistic of 102.4 (column 5), so they do explain the gap, but they do not prevent ARRIVAL from being positive and statistically significant. Consistent with the results in Table 3, the estimates associated with electoral fraud are highly statistically significant.

The discussion in the media also focused on whether there was a change in the trend. The results strongly support the hypothesis that there was a change in the levels, i.e., on average the gap was greater after the shutdown. A trend change would mean that the rate at which the gap opens increases after the shutdown. Columns 2, 4, and 6 aim at testing the hypothesis of a trend change by including the triple interaction ARRIVAL $\times$ SHUTDOWN $\times$ Y2019. Furthermore, these specifications include polling station fixed effects, which span the variables SHUTDOWN, ARRIVAL, COMPUTO, and ARRIVAL$\times$SHUTDOWN. Across all specifications there is a statistically significant and relatively large coefficient on the triple interaction, suggesting that there is a change in trend. The table also reports the marginal effects of the shutdown on the votes at the values of ARRIVAL $= 0.9$ and $0.95$. At both points the marginal effects are highly statistically significant.

## 4.2 Evidence of Widespread Fraud

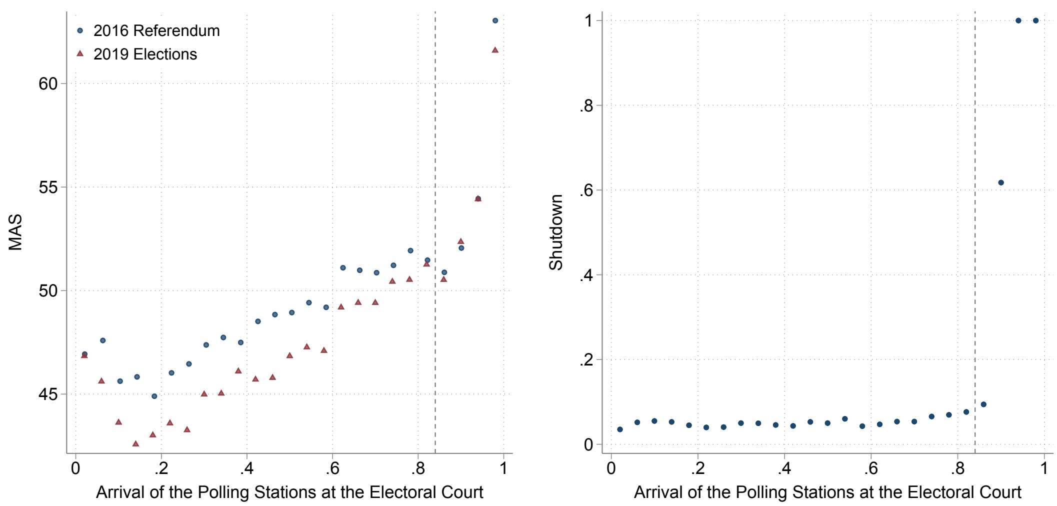
Figure 4 presents the MAS votes in 2019 along with YES votes in 2016 as a function of the arrival of the polling stations at the electoral court in 2019. We can observe that in the first 80% of the reported polling stations, average votes for YES in 2016 lie above average MAS votes in 2019. YES being above MAS is consistent with Morales' known loss of popularity between 2016 and 2019, including his defeat in the 2017 judiciary elections. However, at some point around the 90% mark, the difference flipped and it is MAS votes being above YES votes.

**Table 4:** Arrival and *Computo* trends, and break in arrival trend at shutdown

| Dependent variable: | CC | | MAS | | MAS−CC | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Variables:* | | | | | | |
| SHUTDOWN × Y2019 | -1.627*** | 0.418 | 1.315*** | -2.322*** | 2.954*** | -2.761*** |
| | (0.156) | (0.635) | (0.330) | (0.588) | (0.433) | (0.967) |
| SHUTDOWN | 0.838*** | | -0.733*** | | -1.577*** | |
| | (0.109) | | (0.190) | | (0.262) | |
| Y2019 | -13.49*** | -13.49*** | -2.345*** | -2.357*** | 11.15*** | 11.13*** |
| | (0.220) | (0.305) | (0.224) | (0.309) | (0.443) | (0.611) |
| ARRIVAL | -0.747*** | | 0.587*** | | 1.340*** | |
| | (0.124) | | (0.143) | | (0.242) | |
| COMPUTO | -0.159 | | -0.0512 | | 0.0890 | |
| | (0.149) | | (0.161) | | (0.256) | |
| ARRIVAL × SHUTDOWN × Y2019 | | -2.630*** | | 4.601*** | | 7.277*** |
| | | (0.853) | | (1.145) | | (1.588) |
| Constant | 50.17*** | 49.55*** | 49.38*** | 49.79*** | -0.831*** | 0.198 |
| | (0.168) | (0.152) | (0.171) | (0.129) | (0.311) | (0.280) |
| *Marginal Effect of* SHUTDOWN*:* | | | | | | |
| At ARRIVAL = 0.90 | | -1.949*** | | 1.819*** | | 3.788*** |
| $H_0$: $\beta^{DD} = 0$ | | [0] | | [0.000561] | | [2.52e-08] |
| At ARRIVAL = 0.95 | | -2.081*** | | 2.049*** | | 4.151*** |
| $H_0$: $\beta^{DD} = 0$ | | [0] | | [0.000412] | | [2.56e-08] |
| *Fixed Effects (F-statistic):* | | | | | | |
| Last Name Initial | 79.99 | Yes | 73.67 | Yes | 102.4 | Yes |
| Precinct | 92.04 | Yes | 107.8 | Yes | 130.9 | Yes |
| Polling Station | No | 11.89 | No | 15.62 | No | 19.43 |
| Observations | 65,817 | 65,817 | 65,866 | 65,866 | 65,811 | 65,811 |
| R-squared | 0.920 | 0.950 | 0.922 | 0.956 | 0.937 | 0.966 |

*Notes:* The dependent variable is indicated in the column's heading. The numbers is brackets are p-values. The numbers in parentheses are robust standard errors, clustered by polling station. The reported $F$-statistics are from the null hypothesis that the corresponding fixed effects are jointly equal to zero. *** p<0.01, ** p<0.05, * p<0.1

**Figure 4:** MAS vote share upon arrival at the Electoral Court




*Notes:* The left-hand side shows MAS (or Yes) votes as a function of the arrival of the polling stations at the Electoral Court, (Arrival). The right-hand side shows the share of shutdown polling stations as a function of Arrival.

This is a graphical representation of the DiD estimates that illustrates how 2019 shutdown ballot boxes behave differently. There are two subtle differences from similar figures in a typical DiD approach. First, we cannot rule out the existence of fraud in pre-shutdown polling stations. Second, while they are highly correlated, the order of arrival to the court is not exactly the same as the order at which polling stations were reported on the TREP. Hence, there is no running variable, borrowing the concept from regression discontinuity design, where we can locate a single point in a time-line that creates a before and after to separate the control and treatment groups. The right-hand side figure shows the fraction of polling stations in the shutdown group as a function of Arrival. On average, about 95% of the polling stations went directly from arrival to being reported on the TREP. However, about 5% of the ballot boxes were delayed after arrival and became part of the shutdown group. The vertical dotted lines at 84% serve as an approximate reference to signal that the shutdown occurred at the end, but about 38% of the shutdown polling stations were reported prior to the 84% mark.

In addition to showing how the gap turns to virtually zero when there is a rush of shutdown votes close to the end, the figure illustrates how the gap appears to be closing throughout. We now turn to assess the hypothesis that fraud existed in pre-shutdown ballot boxes. We do so by relaxing the parallel trends assumption and by disentangling the effect of timing of the arrival from the effect of the shutdown.

Table 5 reports the difference-in-differences estimates of an augmented version of Equation 3.

This specification follows Besley and Burgess (2004) and it is helpful to probe the robustness of the differences-in-differences identification. The first column shows that even after relaxing the common trends assumption, the shutdown has an statistically significant effect on the gap.[12] The second column works with the indicator variable NoTrep, which is equal to one for ballot boxes that were never reported on the TREP, zero otherwise. This is a group of 1,511 (4.37%) ballot boxes that are a subset of the shutdown booths. Its statistically significant coefficient is consistent with the estimate on shutdown. This NoTrep specification follows Chumacero (2019a), who focuses on this group to study irregularities. Prior to the election, a group of Bolivians foreseeing the possibility of fraud created a platform to oversee the election process (mivotobolivia.org and a mobile app). They collected thousands of pictures of the official ballot box minutes taken independently right after the local vote count at the polling stations. Chumacero (2019a) was able to access 1,004 of the pictures that correspond to ballot boxes never entered on the TREP. He reports that 45% had descriptions of observed problems, 99% contained discrepancies between the pictures and what was reported in the official results, 40% had mathematical mistakes, and 12% recorded more valid votes than the number of people registered to vote.

**Table 5:** Difference-in-difference Estimates

| Model: | Shutdown<br>(1) | NoTrep<br>(2) |
|---|---|---|
| *Electoral Fraud Measures:* | | |
| Shutdown $\times$ Y2019 | 1.064*** | |
| | (0.390) | |
| NoTrep $\times$ Y2019 | | 2.417*** |
| | | (0.576) |
| Arrival $\times$ Y2019 | 5.722*** | 6.245*** |
| | (1.067) | (1.141) |
| *Fixed Effects ($F$-statistic):* | | |
| Last Name Initial | Yes | Yes |
| Precinct | Yes | Yes |
| Polling Station | 25.37 | 25.41 |
| Observations | 65,811 | 65,811 |
| R-squared | 0.966 | 0.966 |

*Notes:* The dependent variable is MAS−CC. Both specifications include Y2019 and a constant. The reported $F$-statistics are from the null hypothesis that the corresponding fixed effects are jointly equal to zero. The numbers in parentheses are robust standard errors, clustered by polling station. *** p<0.01, ** p<0.05, * p<0.1

[12] After Besley and Burgess (2004) include state-specific time trends, analogous to our Shutdown $\times$ Y2019 interaction, their coefficient on labor regulation drops to zero.

The statistically significant coefficient on Arrival $\times$ Y2019 in column 1 shows that the 2019 votes for MAS increased at a rate of 0.57% of the votes for every 10% of the ballot box minutes arriving at the Electoral Court. This corresponds to 2.86% (5.722%/2) of the total vote count. Note that this effect is different from the shutdown effect and cannot be explained by last name, geography, nonlinearities on the arrival, or heterogeneity across polling stations that is captured by the 2016 votes. This result is consistent with the within-precinct trend reported by Idrobo et al. (2020) who use only the 2019 votes. They claim without any empirical evidence that this trend can be explained by last names; however, our results show that last names cannot explain the trend captured by Arrival $\times$ Y2019 nor the within-precincts trend.[13] Cantú (2014) explains that in Mexico voters are also assigned to polling stations according to their last names an argues that the only difference between voters at contiguous polling stations should be their last names. This argument is in favor of interpreting the 2.86% as electoral fraud.

Even most skeptics (e.g., Curiel and Williams, 2020) agree that the shutdown provided a clear motive to be worried about fraud. Studying the shutdown as a fraud mechanism makes sense because the shutdown buys time to implement it. For example, to rewrite the booth minutes and forge signatures. The interpretation behind the coefficient on Arrival $\times$ Y2019 follows the same logic: It takes time to implement fraud, so polling stations that took longer to arrive at the Electoral Court are more likely to be contaminated with fraud. Hence, Arrival captures a treatment that grows gradually, similar to gradual increases in minimum wages or a sequence of changes in employment regulations (see, e.g., Card, 1992, for a continuous treatment).

Note that the first bin on the far left of the left-hand side of Figure 4 shows no apparent difference between 2016 and 2019. These bins are largely coming from polling stations located in Argentina. The fact that they belong to a different time zone explains why they arrived earlier than the rest (all booths in Bolivia belong to the same time zone). The suspiciously high vote for MAS in Argentina was documented in the Organization of American States (2019) report. For example, 137 ballot boxes recorded over 90% of their votes for Morales. In addition, while participation rates increase on average 4.8% between 2016 and 2019 across all ballot boxes, for Argentina our data shows they increased by 154.6%. There is even a ballot box where Morales officially recorded 153% of the valid votes.[14]

[13] We can replicate and extend results. With 2019 MAS$-$CC as the dependent variable, we obtain a highly statistically significant coefficient of 1.5 on Arrival, after controlling for last name's initial, the shutdown and precinct fixed effects. However, that specification is not correct if the goal is to test for fraud as we explain at the end of Section 3.1.

[14] The official Electoral Court records show seven polling stations where MAS votes exceed the total number of valid votes, a behavior that does not occur for any of the other political parties.

# 5 Triple Differences: Other Political Parties

A more robust analysis than the DiD employed above can be obtained by using both the 2016 referendum and an additional control group within the 2019 election. The benefit of having two comparison groups is that we can remove any trends along these two dimensions of the data (see, e.g., Gruber, 1994; and Chetty et al., 2009). This relaxes the parallel trends assumption in our DiD approach. For the additional control group within the 2019 election, we use the votes obtained by two other political parties. The idea is that the votes of these parties serve as pseudo-outcomes that are known not to have been affected by the fraud treatment (Athey and Imbens, 2017). If there are still any unobserved factors within 2019, it is reasonable to argue that those factors should also affect the votes of these two other political parties.

We use MTS to match if with MAS, and 21F to match it with CC. The sample correlation in the pre-shutdown booths between MAS and MTS is 0.317, and between CC and 21F is 0.239. Let $\textsc{Treat}_{ij}$ be an indicator variable that is equal to one for MAS and CC, and zero for these within-2019 matched MTS and 21F. This will allow us to construct the following "triple difference" (DDD) estimand to capture the effect of the shutdown:

$$\begin{aligned}
\beta^{\text{DDD}} &= \big(\mathbb{E}[\mathrm{V}|\textsc{Shut}=1,\mathrm{Y19}=1,\textsc{Treat}=1]-\mathbb{E}[\mathrm{V}|\textsc{Shut}=0,\mathrm{Y19}=1,\textsc{Treat}=1]\big)\\
&- \big(\mathbb{E}[\mathrm{V}|\textsc{Shut}=1,\mathrm{Y19}=0,\textsc{Treat}=1]-\mathbb{E}[\mathrm{V}|\textsc{Shut}=0,\mathrm{Y19}=0,\textsc{Treat}=1]\big)\\
&- \big(\mathbb{E}[\mathrm{V}|\textsc{Shut}=1,\mathrm{Y19}=1,\textsc{Treat}=0]-\mathbb{E}[\mathrm{V}|\textsc{Shut}=0,\mathrm{Y19}=1,\textsc{Treat}=0]\big).
\end{aligned} \tag{4}$$

This DDD estimate starts by taking the difference before and after the shutdown in the 2019 election then nets out the shutdown change in means in the 2016 referendum and the shutdown change in means for the other political parties in the 2019 election. The goal is that we control for two kinds of potentially confounding trends: changes in voting preferences across polling stations that are common to 2019 and 2016, and changes in voting preferences across polling stations that are specific to the 2019 election.

We can estimate $\beta^{\text{DDD}}$ from the following equation:

$$\begin{aligned}
\mathrm{V}_{ijt} &= \alpha_1\textsc{Shut}_{ij}+\alpha_2\mathrm{Y19}_t+\alpha_3\textsc{Treat}_{ij}\\
&\quad+\delta_1\textsc{Shut}_{ij}\times\mathrm{Y19}_t+\delta_2\textsc{Shut}_{ij}\times\textsc{Treat}_{ij}+\delta_3\times\mathrm{Y19}_t\times\textsc{Treat}_{ij}\\
&\quad+\beta^{\text{DDD}}\cdot\textsc{Shut}_{ij}\times\mathrm{Y19}_t\times\textsc{Treat}_{ij}+X_{ijt}\delta+\nu_{ij}+\varepsilon_{ijt}
\end{aligned} \tag{5}$$

where the coefficient of interest is from the triple interaction $\textsc{Shut}_{ij}\times\mathrm{Y19}_t\times\textsc{Treat}_{ij}$. We include each indicator variable separately and all three first-order interactions to control for systematic differences across all combinations of these three groups. As before, $X_{ijt}$ is a set of controls and $\nu_{ij}+\varepsilon_{ijt}$ is the two-way error.

Column 1 of Table 6 presents the estimates of Equation 5. All corresponding indicators as well as first-order interactions are included as controls, while the numbers in parentheses

**Table 6:** Difference-in-difference-in-differences Estimates

| Model: | Shutdown | | NoTrep | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Electoral Fraud Measures:* | | | | |
| Shutdown × Y2019 × Treat | 2.881*** | 1.340*** | | |
| | (0.421) | (0.375) | | |
| NoTrep × Y2019 × Treat | | | 3.007*** | 3.129*** |
| | | | (0.771) | (0.573) |
| Arrival × Y2019 × Treat | | 4.563*** | | 5.211*** |
| | | (1.252) | | (1.321) |
| *Fixed Effects ($F$-statistic):* | | | | |
| Last Name Initial | Yes | Yes | Yes | Yes |
| Precinct | Yes | Yes | Yes | Yes |
| Polling Station | 3.259 | 3.236 | 3.262 | 3.232 |
| Observations | 136,318 | 134,794 | 136,318 | 134,794 |
| R-squared | 0.552 | 0.554 | 0.546 | 0.553 |

*Notes:* The dependent variable is MAS−CC. All specifications include the corresponding separate indicators (i.e., Y2019, Treat, Shutdown, and Arrival) and first-order interactions (i.e., Shutdown × Y2019, Shutdown × Treat, Arrival × Y2019, Arrival × Treat, and Y2019 × Treat) as controls. Last name initials and precinct FE are spanned by the polling station FE. The numbers in parentheses are cluster robust standard errors, clustered by polling station. The reported $F$-statistics are from the null hypothesis that the corresponding fixed effects are jointly equal to zero. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

are robust standard errors clustered by polling station, as suggested in Duflo et al. (2008). The positive and statistically significant point estimates on the triple interaction, reported in columns 1 and 3, are consistent with the previous DiD fraud estimates. Moreover, note that when we further include ARRIVAL × Y2019 × TREAT in columns 2 and 4, to find that across all specifications we have statistically significant estimates of fraud for both, due to the shutdown (and NOTREP), and throughout the arrival of the polling stations at the electoral court. The point estimates in column 2 indicate that fraud is estimated to be 2.50% of the votes, which increased the gap between MAS and CC. We say that 1.92% is due to fraud in pre-shutdown polling stations, while 0.58% is due to fraud in shutdown polling stations.[15]

When compared to the DiD in Equation 3, our specification in Equation 5 can account for unobserved trends in votes within the 2019 election that are not captured by the trends in the 2016 referendum. Hence, our DDD estimate is immune to two different types of shocks. It is immune to 2019 specific shocks and to shocks that are common to 2016 and 2019. The identification assumption for consistency of the DDD estimate is that there was no shock that differentially affected votes of only the treatment MAS−CC during 2019. Given that 2016 is clearly exogenous to 2019, and that voting took place before fraud, we believe this condition is likely to be satisfied. We include polling-station fixed effects simply to reduce the variance of the error term because, most likely, there is no remaining confounding trend that the polling-station fixed effects will control for.

# 6 Regression Discontinuity Evidence

An alternative fraud identification strategy can arise under Regression Discontinuity (RD) design if there is a sharp change in the gap MAS−CC. The intuition behind such an identification strategy is as follows. We want to estimate the fraud treatment effect where the observed "assignment" variable (or "running" variable) is the arrival of the ballot boxes to the Electoral Court (ARRIVAL). When ARRIVAL exceeds a known cutoff, e.g. 0.95 of the total vote count, then the remaining polling stations are more likely to be contaminated with fraud. We know that voting preferences or other factors behind voting can be different at different values of ARRIVAL. The idea behind this research design is that all non-fraud factors behind voting just below the cutoff are a good comparison to all non-fraud factors that affect voting just above the cutoff.

Formally, for each polling station $i$ in the data let the random variable $(\text{MAS}-\text{CC})_i$ denote

[15] There are 16% of shutdown ballot boxes. Hence, $(4.563\%/2)\times 0.84\% = 1.92\%$ and $((4.563\%/2)+1.34)\times 0.16 = 0.58\%$. Fraud due to ARRIVAL is calculated with the base of 100% of the votes and the height of 4.563%.

our outcome of interest.[16] The scalar regressor $\textsc{Arrival}_i$ is the running variable that determines the treatment assignment based on a known cutoff. Following the framework in Heckman and Vytlacil (2007) and Imbens and Wooldridge (2009), let $\{((\text{MAS}-\text{CC})_i^{\text{Control}}, (\text{MAS}-\text{CC})_i^{\text{Treatment}}, \textsc{Arrival}_i)' : i = 1, 2, \ldots, n\}$ be a random sample from $((\text{MAS}-\text{CC})^{\text{Control}}, (\text{MAS}-\text{CC})^{\text{Treatment}}, \textsc{Arrival})'$, with $(\text{MAS}-\text{CC})^{\text{Control}}$ and $(\text{MAS}-\text{CC})^{\text{Treatment}}$ being the outcomes without and with the electoral fraud treatment. $(\text{MAS}-\text{CC})_i$ is assigned to the electoral fraud treatment condition if $\textsc{Arrival}_i < \overline{\text{T}}$ and is assigned to the control (no fraud) condition if $\textsc{Arrival}_i \geq \overline{\text{T}}$ for a specific and known fixed value $\overline{\text{T}}$.

The observed outcome is

$$(\text{MAS}-\text{CC})_i = \begin{cases} (\text{MAS}-\text{CC})_i^{\text{Control}} & \text{if} \quad \textsc{Arrival}_i < \overline{\text{T}} \\ (\text{MAS}-\text{CC})_i^{\text{Treatment}} & \text{if} \quad \textsc{Arrival}_i \geq \overline{\text{T}}. \end{cases} \tag{6}$$

We identify $\textsc{Fraud}$ as the sharp average treatment effect at the threshold $\overline{\text{T}}$ and it is given by

$$\textsc{Fraud} = \mathbb{E}[(\text{MAS}-\text{CC})_i^{\text{Treatment}} - (\text{MAS}-\text{CC})_i^{\text{Control}} | \textsc{Arrival}_i = \overline{\text{T}}]. \tag{7}$$

We can estimate $\textsc{Fraud}$ nonparametrically following the regression-discontinuity design literature under mild continuity conditions. In particular,

$$\textsc{Fraud} = \lim_{\textsc{Arrival} \downarrow \overline{\text{T}}} \mathbb{E}[(\text{MAS}-\text{CC})_i | \textsc{Arrival}_i = \text{T}] - \lim_{\textsc{Arrival} \uparrow \overline{\text{T}}} \mathbb{E}[(\text{MAS}-\text{CC})_i | \textsc{Arrival}_i = \text{T}]. \tag{8}$$

Using kernel-based local polynomials on either side of the threshold we can estimate $\textsc{Fraud}$ following Hahn et al. (2001) and Porter (2003).

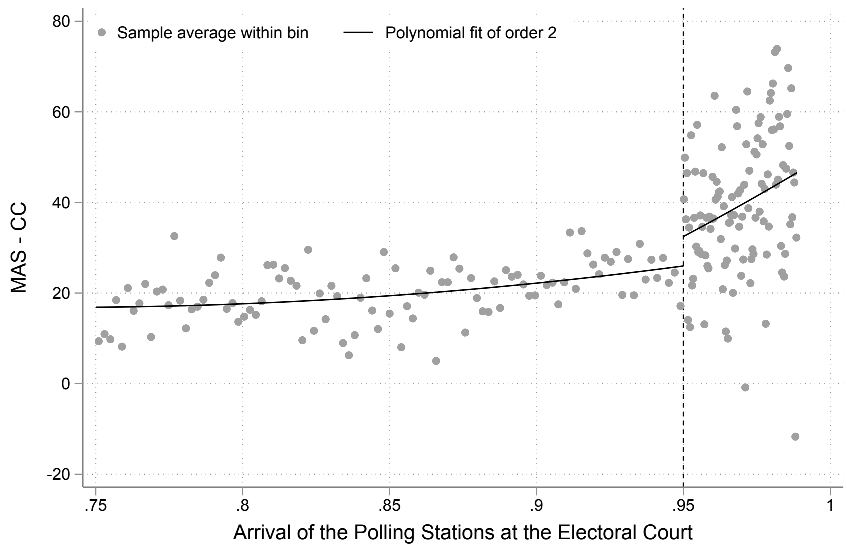
Figure 5 plots MAS−CC collapsed into bins along with second-order global polynomials estimated separately on each side of the 0.95 cutoff. The figure suggests that the gap increases significantly and discontinuously once it crosses the 0.95 threshold. The vertical distance at the discontinuity is analogous to the estimate of $\textsc{Fraud}$ in Equation 7. The last 1.13% of the polling stations on the right-hand side of the figure are the ones that do not have time stamps. We assume that they arrived at the end, and we do not use any of those MAS−CC values in the estimations.[17]

Table 7 presents the sharp regression-discontinuity design estimates of fraud as suggested by Equation 8. The dependent variable is MAS−CC, while the running variable is $\textsc{Arrival}$. We use the bias-corrected bandwidth selection approach proposed in Calonico et al. (2014). The robust 95% confidence intervals and the robust p-values reported in the table are based in this bias corrected RD estimator and the corresponding consistent standard error estimator.

[16] Defining precinct $j$ is unnecessary for identification in an RD design setting (Lee and Lemieux, 2010).

[17] When including them, the results are qualitatively the same as they are relatively far from the threshold and most optimal bandwidths were already excluding them.

**Figure 5:** Regression Discontinuity Plot.




*Notes:* MAS−CC collapsed into bins. The running variable is ARRIVAL, while the threshold is $\overline{\mathrm{T}} = 0.95$.

**Table 7:** Regression Discontinuity Estimates

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| FRAUD | 10.87*** | 12.24** | 14.38*** | 13.89*** |
| Robust 95% CI | [3.969 ; 22.735] | [3.23 ; 25.469] | [4.975 ; 25.523] | [6.225 ; 29.699] |
| Robust p-value | 0.00528 | 0.0114 | 0.00363 | 0.00271 |
| Kernel Type | Triangular | Uniform | Triangular | Triangular |
| BW Type | MSE | MSE | CER | MSE |
| BW Loc. Poly. ($h$) | 0.0343 | 0.0190 | 0.0203 | 0.0457 |
| Order Loc. Poly. ($p$) | 1 | 1 | 1 | 2 |
| Order Bias ($q$) | 2 | 2 | 2 | 3 |
| BW Bias ($b$) | 0.0458 | 0.0288 | 0.0458 | 0.0520 |
| Observations | 34,140 | 34,140 | 34,140 | 34,140 |

*Notes:* The dependent variable is MAS−CC. The running variable is the order polling stations were recorder as received by the Electoral Court (ARRIVAL). The cutoff point it at $\overline{\mathrm{T}} = 0.95$. Robust bias-corrected 95% confidence intervals in brackets are based on Calonico et al. (2014). The Mean Square Error (MSE) optimal bandwidth is obtained using MSE minimizing selection procedure, while the Coverage Error Rate (CER)-optimal bandwidth is obtained using CER minimizing selection procedure with the same bandwidths to the left and to the right of the threshold. *** $p<0.01$, ** $p<0.05$.

Different columns present robustness checks for different kernel types, bandwidth selectors, the choice of the weighted first or second order ($p$=1,2) polynomial regressions for both sides of the cutoff, and the order of the local polynomial bias estimator ($q$=2,3). The bandwidth ($h$) is measured as a fraction of the total number of polling stations and it is selected by either using the Mean Squared Error (MSE) or the Coverage Error Rate (CER) minimizing selection procedures with the same bandwidths on both sides of the threshold.

The point estimate in column 1 is consistent with Figure 5. It indicates that the gap after the 0.95 cutoff results in a statistically significant larger MAS−CC gap of 10.87%. We employ windows of various sizes around the cutoff when balancing the goal of focusing on observations close to the cutoff and using enough observations to obtain precise estimates. In the first column with a triangular kernel along with $p$=1 and $q$=2, MSE suggests a relatively large bandwidth of 0.034 (1,171 polling stations). The bandwidths suggested by a uniform kernel in column 2 and by CER in column 3 are more stringent. Moreover, column 4 experiments with different orders for the local polynomial ($p$=2) and bias ($q$=3), which results in a larger bandwidth.

Across all columns, the fraud estimate is statistically and economically significant, showing that it is robust to the kernel type selection, the bandwidth selection procedure, and the orders of the local polynomial and bias. For example, the point 13.89 estimate in column 4 is equivalent to 0.69% (13.89×0.05%) of the total vote count, just above the required margin to flip the results of the election. The point estimate column 1 is just below the margin. This illustrates some interesting behavioral elements. First, there is the possibility that fraud in the last 5% of the booths was not really needed. Second, people manipulating the last 5% of the minutes did not really know how many more votes they were looking to augment. Three, there is the chance that the goal was not just to get to 10% as it is reasonable to argue that the larger the margin over 10%, the easier it would be to claim there was no fraud.

These RD design results are consistent our the DiD and DDD estimates. The rush of shutdown polling stations at the end of ARRIVAL, illustrated on the right-hand side of Figure 4, coincides with the sharp increase in the gap captured by the RD estimates. Moreover, they coincide with the group of polling stations that never entered the TREP and that was reported in the DDD estimates to have a relatively high MAS−CC margin. We would also want to test for discontinuities at the shutdown, but as the right-hand side of Figure 4 illustrates, the shutdown polling stations are scattered throughout ARRIVAL and there is no well-defined cutoff. Our focus on the 0.95 cutoff follows the discussion of the statistical section of the (Organization of American States, 2019, p. 88) report that shows graphical evidence of an apparent discontinuity at 0.95. Idrobo et al. (2020) also test for discontinuities at 0.95; however, OAS works with *Computo* as the running variable while Idrobo et al. (2020) use a version of the TREP time stamps.

Testing for discontinuities with *Computo*, TREP, or ARRIVAL as the running variable could uncover different behaviors. It is more challenging to identify irregularities in the *Computo* because the minutes for those polling stations had already arrived at the Electoral Court. Moreover, the order at *Computo* is highly correlated with geography as the reporting depends on the booth minutes physically arriving to the local departmental courts. The TREP is missing 4.4% of the time stamps, so that is a significant drawback. Discontinuities on ARRIVAL will not capture centralized tampering of the results taking place on the Electoral Court. Discontinuities might simply be the result of fraud that takes place in different locations and that they coincide at the end, as pressure builds up to make sure the MAS−CC gap is big enough to avoid a runoff. Note that lack of discontinuities does not mean lack of fraud. If fraud intensifies continuously along the running variable, fraud will remain undetected. RD design isolates the treatment variation as a consequence of agents' inability to precisely control the assignment variable. In our case there is some control as fraudulent individuals can decide to delay the submission of forged minutes. However, they cannot go back in time. If a sufficiently large number of fraud polling stations build up close to the end, there is no point in delaying and a discontinuity might be unavoidable.

# 7 Reconciling with Related Work

Our results are consistent with Chumacero (2019a), who documents a number of irregularities including at booths that were never reported in the TREP, and evidence that digits do not follow Benford's law. Our findings are also in line with the various irregularities documented in the Organization of American States (2019) report, which includes falsification of signatures of poll officials, altered tally sheets and databases, and a broken chain of custody. OAS also documents that voting data transmission was redirected to two hidden and unauthorized servers.[18] Moreover, Newman (2020) reports changes in the distributions due to the shutdown, Dávalos (2019) shows a positive relationship between the number of anomalies and MAS votes, while Mebane Jr. (2019) uses an automated algorithm to find a small group of fraud polling stations.

There are two main studies that have results that are different from ours.[19] Curiel and Williams (2020) present a replication exercise that was commissioned by a think tank in Wash-

[18] The European Election Monitors also noted worrisome irregularities.

[19] Both of these studies, Curiel and Williams (2020) and Idrobo et al. (2020), coordinated their release with op-eds published in The Washington Post's Monkey Cage and The New York Times. They were widely disseminated across the globe, and Evo Morales and his supporters continuously cite these articles to argue that there was never any fraud.

ington DC. They claim their research found no reason to suspect fraud, but they do not present a formal test. Rather, they assume that the first 84% of the booths are fraud free and present a simulation exercise for the rest of the tally sheets. In a critical subset of small precincts (that amount to 2.2% of ballots) they assume that the comparison group is on average as polarized as the simulated group. However, fraud is easier to engineer in small precincts which are possibly rural and opposition is unlikely to be present (Cantú, 2019), so their method could be extrapolating wide margins artificially fabricated by fraud. Valdivia and Escobari (2020) point to various impressions in their analysis.

Idrobo et al. (2020)'s main result is to report *no* statistically significant coefficients when testing for discontinuous jumps at two points (at 7:40 pm and at 95%). They interpret this as evidence that late-counted votes do not signal fraud. Their analysis has fundamental flaws. First, they focus on the discontinuities on the TREP, but the TREP does not have 4.4% of the polling stations, a critical batch that has been documented to be associated with fraud (Chumacero, 2019a, and our Tables 5, 6, 7). Because they do not have time stamps for this 4.4%, they cannot test for discontinuities caused by this group. We do have time stamps for those booths, and the statistically significant discontinuous jump that we report coincides with this 4.4% batch.

Second, they argue that they can explain the pro-incumbent shift in vote share without invoking fraud. They show that regional and urban-rural divides account for two-thirds of the trend, but that is far from being a test to rule out fraud in vote shifts. We control for geography in a much more flexible manner to find that it does not explain the key shifts attributed to fraud. Moreover, they document a within-precinct trend (secular trend) and argue that it is helpful to rule out fraud explanations due to the shutdown, under the assumption that the within-precinct trend is not fraud. They try to explain this trend with the initials of the last names of the voters. However, they do not have such information. We do, and last names do not explain the within-precinct trend, so their finding of a within-precinct trend is still consistent with fraud.

Third, they contradict themselves. They argue that within-precinct variation, under their assumption that it can be explained by last names, is important to rule-out fraud. However, when trying to use the 2016 data they say "It is not possible to match voting booths across elections, because of how the booth identifiers changed." Hence, they match at the precinct level ignoring the importance of within-precinct variation. This makes their comparison between 2016 and 2019 incorrect because they are averaging out fraud and non-fraud polling stations that belong to the same precinct. We obtained the booth identifiers from the Electoral Court so can we match the 2016 and 2019 data at the polling station level. Doing so generates a very different pattern than the one they report.

## 8 Conclusion

It is easy to argue that Evo Morales is Bolivia's most controversial figure of the last decade. From his humble beginnings to being the Bolivian president that served the longest. He grew his popularity while organizing his fellow coca farmers against US-backed efforts to reduce cocaine production.[20] While in his exile in Mexico City and then Buenos Aires, he still declared himself as the winner of the elections, even though shortly after the OAS report was released, he said he would respect the findings of the audit and was calling for new elections.[21] While a recent survey shows that 73% of Bolivians believe there was fraud in the 2019 elections (Erbol, 2020), citizens around the world are deeply divided in their assessment of the 2019 elections.

In this paper we formally test for the existence and estimate the size of voting fraud in the 2019 Bolivian elections. The identification strategy that relies on a rare natural experiment—the shutdown of the official vote counting system. Our first set of estimates rely solely on 2019 data and uses pre-shutdown polling stations as the control group. Locality and precinct fixed effects estimates that control for unobservables show preliminary evidence of fraud. We then move to match the 2019 polling stations with a second control group, the 2016 referendum. Difference-in-Differences estimates show that the shutdown not only increased the gap between MAS and CC, but had a statistically significant effect on MAS and CC separately. We also document a change in trend at the shutdown.

When we allow trends in 2016 and 2019 to be different, we find evidence that fraud extends beyond the shutdown polling stations. Our results are consistent with a statistically significant within-precinct trend. We use the votes of other political parties as a third control group to obtain triple differences. Our difference-in-difference-in-difference estimates are immune to polling station level shocks that are common to 2016 and 2019, and shocks that are specific to the 2019 elections. Shutdown and the trend are both statistically significant, signaling widespread fraud that accelerated at the shutdown. The approach controls for last names, geography, and voting preferences across polling stations among a wide variety of unobserved shocks. In an alternative identification strategy, we use a Regression Discontinuity design to find results that are consistent with the DiD and DDD estimates. Overall, we provide strong evidence of fraud that is large enough to have changed the outcome of the election.

Evo Morales' party was not new to electoral fiasco. There were judiciary elections in 2017, but where all the candidates had already been preselected by a Parliament controlled by MAS. That was a clear violation of the independence between the branches of the government. Bolivians

[20] As president he expelled the US ambassador and the US Drug Enforcement Administration.

[21] https://www.theguardian.com/world/2019/nov/10/evo-morales-concedes-to-new-elections-after-serious-irregularities-found

had no power to protest, so 65% decided to vote null or blank. Yet, Morales still went ahead and appointed the preselected candidates.[22] A once very popular Morales was defeated in the 2016 and 2017 signaling that the 2019 elections would be challenging for MAS with most of the opposition vote concentrated in CC.

# References


Athey, S. and Imbens, G. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

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

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust Differences-in-Differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Besley, T. and Burgess, R. (2004). Can labor regulation hinder economic performance? Evidence from India. *The Quarterly Journal of Economics*, 119(1):91–134.

Besley, T., Persson, T., and Sturm, D. (2010). Political competition, policy and growth: Theory and evidence from the US. *The Review of Economic Studies*, 77(4):1329–1352.

Callen, M. and Long, J. (2015). Institutional corruption and election fraud: Evidence from a field experiment in Afghanistan. *The American Economic Review*, 105(1):354–381.

Calonico, S., Cattaneo, M. D., and Titunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

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

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

Card, D. (1992). The effect of unions on the distribution of wages: Redistribution or relabelling? NBER Working Paper No. 4195.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *The American Economic Review*, 99(4):1145–1177.


[22] https://latinamericanpost.com/18146-bolivia-judicial-election

Chumacero, R. (2019a). El camaleón, el mutante y Houdini: Resultados de las elecciones en Bolivia. University of Chile Working Paper.

Chumacero, R. (2019b). Skills versus luck: Bolivia and its recent bonanza. *Latin American Economic Review*, 29(7):1–27.

Curiel, J. and Williams, J. (2020). Bolivia dismissed its October elections as fraudulent. Our research found no reason to suspect fraud. Washington Post Monkey Cage, February 27.

Dávalos, J. (2019). Sistematicidad de la bi y trilocación de electores en las elecciones presidenciales Bolivia-2019. Universidad del Pacífico, Perú, Working Paper.

Duflo, E., Glennerster, R., and M., K. (2008). *Using Randomization in Development Economics Research: A Toolkit.* T. Schultz and John Strauss, eds., Handbook of Development Economics. Vol. 4. Amsterdam and New York: North Holland.

Enikolopov, R., Korovkin, V., Petrova, M., Sonin, K., and Zakharov, A. (2013). Field experiment estimate of electoral fraud in Russian parliamentary elections. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2):448–452.

Erbol (2020). Encuesta: 73% de la población considera que hubo fraude en las elecciones de 2019. Erbol, September 10.

Escobari, D. (2012). Imperfect detection of tax evasion in a corrupt tax administration. *Public Organization Review*, 12(4):317–330.

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

Fukumoto, K. and Horiuchi, Y. (2011). Making outsiders' votes count: Detecting electoral fraud through a natural experiment. *The American Political Science Review*, 105(3):586–603.

Greene, W. (2018). *Econometric Analysis.* Eight Edition, Pearson.

Gruber, J. (1994). The incidence of mandated maternity benefits. *The American Economic Review*, 84(3):622–641.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Heckman, J. J. and Vytlacil, E. J. (2007). *Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation,* ed. J. Heckman and E. Leamer, volume VI of *Handbook of Econometrics*, pages 4780–4874. Elsevier Science B.V.

Hicks, D., Maldonado, B., Piper, B., and Goytia Rios, A. (2018). Identity, patronage, and redistribution: Economic inequality in Bolivia under Evo Morales. *Journal of Economics, Race, and Policy*, 1(1):26–41.

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

Ichino, N. and Schündeln (2012). Deterring or displacing electoral irregularities? Spillover effects of observers in a randomized field experiment in ghana. *The Journal of Politics*, 74(1):292–307.

Idrobo, N., Kronick, D., and Rodríguez, F. (2020). Do shifts in late-counted votes signal fraud? evidence from bolivia. University of Pennsylvania, Working Paper.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

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

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Lehoucq, F. (2003). Electoral fraud: Causes, types, and consequences. *Annual Review of Political Science*, 6(1):233–256.

Magaloni, B. (2009). *Voting for Autocracy: Hegemonic Party Survival and its Demise in Mexico. Cambridge Studies in Comparative Politics*, Cambridge Univ Press, Cambridge.

Mebane Jr., W. (2019). Evidence against fraudulent votes being decisive in the Bolivia 2019 election. University of Michigan Working Paper.

Mebane Jr., W. and Kalinin, K. (2010). Electoral fraud in Russia: Vote counts analysis using second-digit mean tests. In *Annual Meeting of the Midwest Political Science Association*. Chicago, IL.

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

Newman, J. (2020). The OAS conclusions about the election integrity of the Bolivian election are correct. Working Paper.

Organization of American States (2019). *Electoral Integrity Analysis, General Elections in the Plurinational State of Bolivia*. Preliminary Findings, Report to the General Secretariat.

Porter, J. (2003). Estimation in the regression discontinuity model. Working Paper. Harvard University.

Roe, B. and Just, D. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics*, 91(5):1266–1271.

Valdivia, W. and Escobari, D. (2020). Bolivias electoral fraud reckoning. Project Syndicate, March 17.