Late-arriving votes and electoral fraud: A natural experiment and regression discontinuity evidence from Bolivia*

Diego Escobari† Gary A. Hoover‡

August 17, 2023

Abstract

This paper uses a unique data set and a natural experiment to test if electoral fraud can exist in late-arriving votes. On the night of the 2019 Bolivian elections, the official vote counting system that was expected to publish the results real-time, suddenly stopped. When it resumed, the results had flipped. We estimate several difference-in-differences specifications using a 2016 referendum and the votes of other political parties. We find that the extent of the fraud is 2.51% of the valid votes, sufficient to change the outcome of the election. Our results are robust to polling-station-level shocks common across 2019 and 2016, as well as 2019 specific shocks. This controls for geography, socioeconomic characteristics, unobserved voting preferences, and endogeneity in the arrival of the polling station results. We report evidence of fraud that occurred before the shutdown and document a statistically significant discontinuous jump in the votes during the shutdown. We provide insights on how to apply our different identification strategies to test for fraud in other elections.

Keywords: Electoral Fraud; Natural Experiment; Regression Discontinuity

JEL Classifications: C21; D72; K42

*We thank comments by Eduardo Aleman, Binay Adhikari, Gustavo Canavire-Bacarreza, Francisco Cantú, Pablo Cuba-Borda, Jorge Dávalos, Jorge Derpic, Elaine Denny, Mariana Giusti Rodríguez, Yusaku Horiuchi, Irfan Nooruddin, Robert Nyenhuis, Jean-Baptiste Tondji, Greg Thorson, Walter Valdivia, and participants at the AsianPolmeth VIII & ASQPS IX, Bolivian Conference in Development Economics, SEA, LACEA, MPSA and WPSA meetings. We also thank Edgar Villegas for sharing data with us.

†Professor of Economics and PhD Program Director, Robert C. Vackar College of Business and Entrepreneurship, The University of Texas Rio Grande Valley, Edinburg, TX 78539, Phone: (956) 665-2104, Email: diego.escobari@utrgv.edu, URL: http://faculty.utrgv.edu/diego.escobari

‡Professor of Economics and Director of the Murphy Institute, Tulane University, New Orleans, LA 70118, Phone: (504) 865-5317, Email: ghoover@tulane.edu, URL: https://sites.google.com/site/garyhoovereconomics

1 Introduction

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). Furthermore, electoral fraud decreases accountability and helps autocratic leaders stay in power (Magaloni, 2009).

Testing for the existence of fraud is complex as neither the alleged victims nor the perpetrators have incentives to reveal the truth. The former will cry foul-play even when there is no record of irregularities (Magaloni, 2010), while latter will want to keep their activities hidden (Lehoucq, 2003). The existence of accusations can delegitimize the election process or can serve to justify a coup.[1] In Bolivia, the disputes around the 2019 presidential elections resulted in 37 deaths (Bjork-James, 2020), a president that resigned, claims that there was a coup (Lehoucq, 2020), a year of a transitional government, a substantially weakened democracy, and a deeply divided country.[2]

While using data from Bolivia, this article has a broader contribution by studying swings in late-arriving votes and by proposing different identification strategies to identify fraud that are new to the literature. It is common practice in contemporary democracies to publish the electoral results in real time. Delays between the closing of polls and the announcement of the final tally are known to generate doubt about the integrity of the election (Antenangeli and Cantú, 2019). For example, in the 2006 Presidential Election in Mexico, the official count showed that López Obrador held an advantage, but fraud allegations emerged when Calderón overtook the first place during the count of the last 2.5% of the polling stations.[3] A similar event occurred in Honduras, where after the incumbent's rival posted an early lead, vote-counting slowed to a crawl and with 89% of the votes counted, the results flipped.[4] More recently, delays in results reporting in the 2023 Presidential Election in Nigeria led to electoral fraud accusations.[5] Similar cases occurred during the 2017 Senate Primary elections in Argentina, the 2016 and 2021 Presidential Elections in Peru, and the 2017 Presidential Elections in Ecuador.[6] In the U.S., early counts in the 2020 Presidential Election and belief that they will persist contributed to enduring doubts about the integrity of the election (Eggers et al., 2021) with late-arriving mail votes contributing to the controversy (McGhee et al., 2021). Antenangeli and Cantú (2019) explain that concerns about undue delays in electoral results are present in about one in every five elections worldwide.

We use a regression discontinuity design and a natural experiment to test for the existence and estimate the magnitude of electoral fraud in the 2019 Bolivian presidential elections. At 7:40 pm on the night of the elections, the official preliminary vote counting system *Transmisión de Resultados Electorales Preliminares* (TREP) that was expected to publish the progress of the results real-time as the votes were tabulated by the electoral count, suddenly stopped.[7] At the moment of the shutdown, 84% of the valid votes had already been reported and a runoff between the incumbent *Movimiento al Socialismo* (MAS) and *Comunidad Ciudadana* (CC) appeared imminent.[8] However, when it resumed over 20 hours later, the vote count was giving an outright runoff-avoiding win to MAS. The shutdown created a natural experiment that separated polling stations between a control group that was less likely to be associated with fraud, and a treatment group of polling stations reported after the shutdown.[9]

We have a unique dataset that contains information that is superior to any of the other studies that looked at the 2019 Bolivian elections. We have timestamps with the minute and second of the arrival of the results of the 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.

To win the election a candidate either needed to obtain a majority (>50%) 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 MAS and if it decreased the votes of the main opposition party CC. We 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. As robustness check we relax the DiD parallel trends assumption to find that the gap between MAS and CC increased throughout the arrival of ballot boxes, consistent with fraud that took place before the shutdown.

The votes of other political parties help to provide various Difference-in-Difference-in-Differences (DDD) estimates. The results indicate that fraud increased the gap between MAS and CC by 2.51% of the valid votes, with 1.93% corresponding to pre-shutdown booths and 0.58% to shutdown polling stations. The evidence points out that fraud increased at the shutdown as MAS realized that earlier efforts were likely inadequate to get them to the 10% threshold to avoid the runoff. Our DDD 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, socioeconomic characteristics 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 the results of the polling stations to the Electoral Court.

An additional identification strategy is based on Regression Discontinuity (RD) and does not depend on the shutdown or on the availability of the 2016 constitutional referendum results. Our approach follows the RD-like graphs generated in Organization of American States (2019) and Nooruddin (2020b) that reveal a discontinuity in the data. To the best of our knowledge, our paper is the first to implement a full RD estimation and identify fraud. We find a statistically significant discontinuous jump when 0.95% of the polling stations results 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.

Previous work that aimed at capturing the size of the voting fraud includes Mebane Jr. and Kalinin (2010) who use second-digit mean tests. Beber and Scacco (2012) and Ichino and Schündeln (2012) focus on voters' registration. Moreover, Hyde (2007) and Fukumoto and Horiuchi (2011) assume away the existence of voters' heterogeneity across precincts, while Enikolopov et al. (2013) and Leeffers and Vicente (2019) study the role of observers. In addition, Callen and Long (2015) find that the announcement of a new monitoring technology reduced fraud, consistent with the monitoring of illegal activities. Fujiwara (2015) used a regression discontinuity design to show the importance of voting technology, and Kobak et al. (2016) use the frequency of reported round percentages to present a fraud indicator. Hausmann and Rigobon (2011) analyze diverse hypotheses of fraud in Venezuela. 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.[10] Our findings are also related to Foley (2013), who studied shifts as the vote progresses and Gallego et al. (2023), who studies vote-buying.[11]

While our main contribution is proposing two research designs to identify fraud, our results are important for Bolivia as well. 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. (2022) aim at supporting the hypothesis that there was no fraud due to the shutdown.

The Bolivian president Evo Morales stepped down on November 10, 2019, three weeks after the elections, following accusations of fraud, mass protests (Boulding et al., 2019), and an Organization of American States (2019) audit that found "serious irregularities".[12] 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 (Nooruddin, 2020b).[13] 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. Under the claim that the first term did not count because it was under a different constitution, 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—that qualified him saying that limits on the lengths of his tenure would violate his "human rights."[14] 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 favorable external conditions (Chumacero, 2019b).

The rest of the paper is structured as follows. Section 2 presents the data. The Difference-in-Differences results are reported and discussed in Section 3, along with evidence of widespread fraud (Section 3.1). Sections 4 and 5 present the results for the Difference-in-Difference-in-Differences and the Regression Discontinuity specifications. Section 6 reconciles our findings with related work, while Section 7 provides insights and examples of how our identification strategies can be used in other elections. Section 8 concludes.

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*. In English, TREP stands for "Transmission of Preliminary Electoral Results" and its role was to publish online and in real-time the preliminary results for every polling station as the information was received and recorded by the Electoral Court. 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 the results of the 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*, information that was openly shared in Nooruddin (2020a). 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, about here]

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.[15]

Note that shares are calculated only using valid votes, so having more blank or null increases 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 we expect to see a positive correlation, which is the case for all the parties except CC.

3 A Natural Experiment and Difference-in-Differences

The shutdown occurred at 7:40 pm on Sunday October 20, 2019 (the night of the elections) when TREP suddenly stopped posting the results. Prior to the shutdown it had already reported 84% of the valid votes.[16] The president of the Electoral Court (*Tribunal Supremo Electoral*), María Eugenia Choque had the authority to shutdown the TREP, and announced that they stopped transmitting the 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.[17] The representatives of the Organization of American States overseeing the election met with Electoral Court officials to stress the importance of keeping the TREP running.[18] Regardless, TREP remained inactive for an additional 23 hours. This disruption of the system resulted in a general public outcry as it was clear it was jeopardizing 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.

One important element in assessing the shutdown is the rural vote, as MAS is known to have 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.[19] In addition to geography, the gap between MAS and CC might have also been influenced, for example, by socioeconomic characteristics such as education or income.

[Figure 1, about here]

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 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, about here]

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:

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

where V is the share of the votes of CC and MAS. It makes sense to additionally use MAS$-$CC because the goal for MAS was to have a gap of 10% or more to avoid the runoff. Shut is the indicator variable that captures the shutdown, and Y19 is an indicator variable equal to one for the 2019 elections and zero for the 2016 referendum. β^{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 β^{DD} from the following regression equation:

$$\text{V}_{ijt} = \alpha_1 \text{Shut}_{ij} + \alpha_2 \text{Y19}_t + \beta^{\text{DD}} \cdot \text{Shut}_{ij} \times \text{Y19}_t + X_{ijt}\delta + \nu_{ij} + \varepsilon_{ijt}. \tag{2}$$

Note that in addition to polling station i in precinct 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 ε_{ijt}. The coefficient of interest (β^{DD}) captures the effect of the shutdown on votes during 2019, while using 2016 as a control. If Shut_{ij} is uncorrelated with the error term, following Enikolopov et al. (2013), β^{DD} can be interpreted as the size of the fraud.[20] 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.

Table 2 presents various sets of estimates of different versions of Equation 2. 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 2 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.[21] 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 13.28 (column 5) to negative 1.115 (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.

From the point estimates on Shutdown × Y2019 and Shutdown of column 5 (Table 2), we estimate that about 81.7% (13.28/(2.98+13.28)) of the increase in the MAS−CC gap after the shutdown can be explained by the votes in 2016. Note that this 81.7% 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 and potential endogeneity in the arrival of the polling station results. The remaining 18.3% is attributed to fraud.

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. Because the results arrive online, there is really no obvious reason why rural votes should generate a trend (i.e., increasing share of rural votes). The notion that rural areas should take longer to take a picture of the minutes and upload it makes no sense. It was during the 1980s and early '90s where Bolivians were accustomed to seeing the "rural vote effect" when the votes typically took longer to physically arrive from the rural areas. For example, in the 1985 presidential elections, the late-arriving rural votes reduced the difference between Victor Paz Estenssoro and Hugo Banzer.[22]

[Table 2, about here]

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.

3.1 Evidence of Widespread Fraud

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. Figure 3 presents the MAS votes in 2019 along with Yes votes in 2016 as a function of Arrival. 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.

[Figure 3, about here]

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, we cannot 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 3 reports the difference-in-differences estimates of an augmented version of Equation 2. 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.[23] 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 3, about here]

The statistically significant coefficient on Arrival × Y2019 in column 1 shows that the 2019 votes for MAS increased at a rate of 0.58% of the votes for every 10% of the ballot box minutes arriving at the Electoral Court. This corresponds to 2.88% (5.756%/2) of the total vote count. Note that this effect is different from the shutdown effect and cannot be explained by last name, socioeconomic status, 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. (2022) who use only the 2019 votes. They claim that this trend can be explained by less educated voting-booth jurors that take longer to report and that there are different education levels across voting booths within the same precinct. However, they do not have this within-precinct information.[24] Our results show that none of these polling-station specific factors can explain the trend captured by Arrival × Y2019 nor the within-precinct trend. Cantú (2014) explains that in Mexico voters are also assigned to polling stations according to their last names and argues that the only difference between voters at contiguous polling stations should be their last names. This argument is in favor of interpreting the 2.88% 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 × 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 3 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.[25]

4 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 it 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 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:

$$\begin{aligned}\beta^{\text{DDD}} &= \big(\mathbb{E}[\text{V}|\textsc{Shut}=1,\text{Y19}=1,\textsc{Treat}=1]-\mathbb{E}[\text{V}|\textsc{Shut}=0,\text{Y19}=1,\textsc{Treat}=1]\big)\\ &- \big(\mathbb{E}[\text{V}|\textsc{Shut}=1,\text{Y19}=0,\textsc{Treat}=1]-\mathbb{E}[\text{V}|\textsc{Shut}=0,\text{Y19}=0,\textsc{Treat}=1]\big)\\ &- \big(\mathbb{E}[\text{V}|\textsc{Shut}=1,\text{Y19}=1,\textsc{Treat}=0]-\mathbb{E}[\text{V}|\textsc{Shut}=0,\text{Y19}=1,\textsc{Treat}=0]\big).\end{aligned} \tag{3}$$

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 β^{DDD} from the following equation:

$$\begin{aligned}\text{V}_{ijt} &= \alpha_1\textsc{Shut}_{ij}+\alpha_2\text{Y19}_t+\alpha_3\textsc{Treat}_{ij}\\ &+\delta_1\textsc{Shut}_{ij}\times\text{Y19}_t+\delta_2\textsc{Shut}_{ij}\times\textsc{Treat}_{ij}+\delta_3\times\text{Y19}_t\times\textsc{Treat}_{ij}\\ &+\beta^{\text{DDD}}\cdot\textsc{Shut}_{ij}\times\text{Y19}_t\times\textsc{Treat}_{ij}+X_{ijt}\delta+\nu_{ij}+\varepsilon_{ijt}\end{aligned} \tag{4}$$

where the coefficient of interest is from the triple interaction $\textsc{Shut}_{ij}\times\text{Y19}_t\times\textsc{Treat}_{ij}$. 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.

[Table 4, about here]

Column 1 of Table 4 presents the estimates of Equation 4. All corresponding indicators as well as first-order interactions are included as controls, while the numbers in parentheses 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 results at the electoral court. The point estimates in column 2 indicate that fraud is estimated to be 2.51% of the votes, which increased the gap between MAS and CC. We find that 1.93% is due to fraud in pre-shutdown polling stations, while 0.58% is due to fraud in shutdown polling stations.[26]

When compared to the DiD in Equation 2, our specification in Equation 4 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.

5 Regression Discontinuity Evidence

An alternative fraud identification strategy arises 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 results 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.

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

$$\textsc{Fraud} = \mathbb{E}[(\mathrm{MAS{-}CC})_i^{\mathrm{Treatment}} - (\mathrm{MAS{-}CC})_i^{\mathrm{Control}} | \textsc{Arrival}_i = \overline{\mathrm{T}}]. \tag{5}$$

We can estimate Fraud using nonparametric kernel-based local polynomials on either side of the threshold following Hahn et al. (2001) and Porter (2003).

[Figure 4, about here]

[Table 5, about here]

Figure 4 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 Fraud in Equation 5. 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.[27]

Table 5 presents the sharp regression-discontinuity design estimates of fraud. The dependent variable is MAS−CC, while the running variable is 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. 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 4. 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 in 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 booth 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. This is consistent with the model in Little (2012), where there is an incentive to commit fraud independent of winning to try to get a convincing election result.

These RD design results are consistent with our DiD and DDD estimates. The rush of shutdown polling stations at the end of ARRIVAL, illustrated on the right-hand side of Figure 3, 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 3 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. 87-92) report that shows graphical evidence of an apparent discontinuity at 0.95. Idrobo et al. (2022) also test for discontinuities at 0.95; however, OAS works with TREP and *Computo* as the running variable while Idrobo et al. (2022) 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.

6 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 analysis presented in Nooruddin (2020b) and 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.[28] This report, released in the days following the election, also generated RD-like graphs that revealed some discontinuity in the data. In addition, 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. Curiel and Williams (2020) present a replication exercise that was commissioned by a think tank in Washington 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. (2022)'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 weaknesses. 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 3, 4, 5). 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 education, region, and rurality explain most of the time trend, but they have this information only at the municipality level. That is far from being a test to rule out fraud in vote shifts. We control for these factors in a much more flexible manner to find that they do 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. However, this assumes that the within-precinct trend is not fraud. They try to explain this trend by claiming that less educated voting-booth jurors take longer to report and that there are different education levels across voting booths within the same precinct. However, they do not have such information at the polling-station level. We control for these factors and they do not explain the trend, so their finding of a within-precinct trend is still consistent with fraud.[29]

Third, they contradict themselves. They argue that within-precinct variation, under their assumption that it can be explained by the level of education of voting-booth jurors and different education levels across voting booths within the same precinct, is important to rule-out fraud. However, when using the 2016 data (Table B.1), they match at the precinct level ignoring the importance of within-precinct variation. This makes their use of the 2016 votes 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.[30]

7 Insights for Other Elections

We provide various identification strategies that can be potentially employed to test for fraud in other elections. First, the RD approach in Section 5 can be used to test for fraud when votes can be ordered chronologically. This identification strategy is likely to work, for example, for Mexico in 2016, Argentina in 2016 and 2021, Honduras in 2017 and 2021, Ecuador in 2017, Peru in 2021, Brazil in 2022, or Nigeria, Guatemala and Paraguay in 2023. A shutdown is not necessary as votes can be separated between control and treatment with a cutoff driven by, for example, delays in the count. Antenangeli and Cantú (2019) point out that delays are present in one of every five elections worldwide. While our approach follows Organization of American States (2019) and Nooruddin (2020b), our results are the first to provide a full RD estimation that finds positive results. One advantage of this approach is that it does not depend on the availability of the votes from a comparable election or other political parties.

Second, our DiD approach of Section 3 illustrates that a previous election or the votes of other political parties can be employed as additional control groups to test or measure the size of election fraud. Availability is specific to the setting but the same methods can be employed when similar controls are also likely to exist. For example, events such as the repeated elections in Georgia in 2012, Ukraine in 2013, Israel in 2014, or Austria in 2016.[31] When more than one set of controls exist researchers or electoral observers might choose to employ DDD, similar to our Section 4. If suspected fraud is believed to be building over time, Section 3.1 illustrates how a test for a trend can be useful. This test does note rely on the existence of a shutdown or a discontinuity.

Third, when a chronological order is not available for all the votes or polling stations but votes can be separated into a fraud treatment versus control, a simple difference strategy might be enough. This can be combined with various levels of geography or precinct fixed effects. One benefit of this approach is that it would not need the results of a comparable election (e.g., the 2016 referendum). While we contemplated providing this difference estimates, for Bolivia this would need to assume that within-precinct heterogeneity is uncorrelated with the fraud treatment. We were not willing to make that assumption in this paper, but it is likely to hold in other settings where, for example, voters are allocated randomly across polling stations within the same precinct.

In addition to advancing the literature on methods to identify fraud that can be useful to researchers, election officials and observers, we provide some lessons learned. We argue that officials and observers should have information on time stamps along with the vote counts. This proved to be valuable to assess for irregularities and test for fraud, even without the existence of a shutdown. Moreover, a random allocation of voters across polling stations within the same precinct can also facilitate the process of testing for fraud. The Bolivian election highlighted the importance of real-time reporting of the results.

8 Conclusion

It is easy to argue that Evo Morales is Bolivia's most controversial figure of the last two decades. 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.[32] Since his exile in Mexico and now while living in Bolivia, he still declares himself as the winner of the 2019 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.[33] While a survey shows that 73% of Bolivians believe there was fraud in the 2019 elections, citizens around the world are deeply divided in their assessment of the 2019 elections.[34]

Using various identification strategies, in this paper we formally test for the existence and estimate the size of voting fraud in the 2019 Bolivian elections. Our first identification strategy relies on a natural experiment—the shutdown of the official vote counting system. We start my matching the 2019 polling stations with 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. 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, socioeconomic characteristics, and voting preferences across polling stations among a wide variety of unobserved shocks.

We use a RD design as an additional identification strategy to find results that are consistent with the DiD and DDD estimates. Overall, the results show strong evidence of fraud that is large enough to have changed the outcome of the election. Moreover, our results are relevant to other elections. We provide insights and details to help implement our different identification strategies to other contexts, the data requirements, and examples of election processes in other countries where our approach can be implemented.

Notes

[1] "Pool: Just a quarter of Republicans accept election outcome" (Domenico Montaro) *NPR*, December 9, 2020; "Myanmar coup: Does the army have evidence of voter fraud?" (Jack Goodman) *BBC*, February 5, 2021.

[2] "'Bolivia is shattered': Election crisis leaves deeply divided nation" (Monica Machicao, Daniel Ramos and Sergio Limchi) *Reuters*, November 10, 2019.

[3] "Mexican Court Rejects Election Fraud Challenges" *New York Times*, August 28, 2006.

[4] "Honduras's disputed election provokes a crisis" *The Economist*, November 13, 2017.

[5] "Nigeria's presidential election marked by long delays" *Al Jazeera*, February 25, 2023, and "Bola Tinubu wins Nigerian presidency amid fraud accusations" *Le Monde*, March 1, 2023.

[6] For Argentina, see, e.g., "Elecciones PASO 2017: Cristina Kirchner denunciará la 'trampa electoral' del Gobierno y apuntará a todos los votos peronistas" (Pablo de León) *Clarín*, August 14, 2017. For Peru, "El recuento apuntala la victoria por la mínima de Kuczynski en Perú" (Carlos E. Cué and Jacqueline Fowks) *El País*, June 6, 2016, and "Peru election race still too close to call" *Deutsche Welle*, June 8, 2021. For Ecuador, "Ecuador opposition leader alleges vote fraud" *Deutsche Welle*, April 4, 2017.

[7] "Bolivia elections: Concern as results transmission pauses" *BBC*, October 21, 2019.

[8] Most polls for a potential runoff showed MAS would lose (Chumacero, 2019a). Moreover, there was also the record of their recent defeat in the 2016 referendum. With this expectation, the need for a favorable 2019 election outcome for MAS was clear.

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

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

[11]Using survey data that includes Bolivia, Gallego et al. (2023) find that pre-electoral distribution of private goods has a limited electoral impact.

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

[13]Lehoucq (2020) has details of the outcome and aftermath of the elections.

[14]See, for example, "Evo Morales finds a way to run for re-election" *The Economist*, December 1, 2017, and "Evo Morales finally went too far for Bolivia" (Yascha Mounk) *The Atlantic*, November 11, 2019.

[15]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).

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

[17]"NEOTEC confirma que Choque llamó para cortar el TREP, sin argumentos válidos" *Página Siete*, October 31, 2019.

[18]Organization of American States (2019) states that it was an intentional and arbitrary freezing, with no technical basis.

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

[20]Our definition of fraud is consistent with Hausmann and Rigobon (2011), where it is defined as the difference between the elector's intent and the official vote tally.

[21]When the number of polling stations within the same precinct was not exactly the same, we used weights, while keeping the same alphabetical order.

[22]See, e.g., "Los votos rurales bolivianos rebajan la diferencia entre Bánzer y Paz Estenssoro" (The Bolivian rural votes reduce the difference between Bánzer and Paz Estenssoro) (Martin Prieto) *El País*, July 16, 1985.

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

[24]The 2020 version of their paper stated that this trend could be explained by last names, but they did not have information on last names either. We do have last names and we can replicate and extend their 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.

[25]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.

[26]There are 16% of shutdown ballot boxes. Hence, $(4.599\%/2)\times 0.84\% = 1.93\%$ and $((4.599\%/2)+1.335)\times 0.16 = 0.58\%$. Fraud due to Arrival is calculated with the base of 100% of the votes and the height of 4.599%.

[27]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.

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

[29]See Nooruddin (2020a) for his response to Curiel and Williams (2020) and Idrobo et al. (2022).

[30]It is not clear why they focus on citing the old 2019 version of our manuscript in their main text. This version did not have information on time stamps or the DiD, DDD, and RD results that were already included in the 2020 version. The 2020 version was available and they cite it in section D of their online appendix (see, e.g., Figure D.2).

[31]For Georgia, see, e.g., "Repeat elections in Georgia due to serious fraud" *Democracy & Freedom Watch*, October 14, 2012. For Ukraine, see, e.g., Kovalov (2014). For Israel, see, e.g., "Police to take extra precautions against fraud in repeat elections" (Nir Hasson) *Haaretz*, February 24, 2014. For Austria, see, e.g., "Austrian presidential election result overturned and must be held again" (Philip Oltermann) *The Guardian*, July 1, 2016.

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

[33]"Evo Morales agrees to new elections after irregularities found" (Dan Collyns) *The Guardian*, November 10, 2019.

[34]"Encuesta: 73% de la población considera que hubo fraude en las elecciones de 2019" *Erbol*, September 10, 2020.

References

Antenangeli, L. and Cantú, F. (2019). Right on time: An electoral audit for the publication of vote results. *Statistics, Politics and Policy*, 10(2):137–186.

Athey, S. and Imbens, G. W. (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. M. (2010). Political competition, policy and growth: Theory and evidence from the US. *The Review of Economic Studies*, 77(4):1329–1352.

Bjork-James, C. (2020). Mass protest and state repression in Bolivian political culture: Putting the gas war and the 2019 crisis in perspective. Human Rights Program, Harvard Law School, Research Working Paper Series HRP 20-003.

Boulding, C., Foxworth, R., Hummel, C., Nelson-Nunez, J., and Velasco-Guachalla, V. X. (2019). The real story behind Bolivia's protests. Washington Post Monkey Cage, November 20.

Callen, M. and Long, J. D. (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.

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

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

Curiel, J. and Williams, J. R. (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 Michael, 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.

Eggers, A. C., Garro, H., and Grimmer, J. (2021). No evidence for systematic voter fraud: A guide to statistical claims about the 2020 election. *Proceedings of the National Academy of Sciences*, 118(45):e2103619118.

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*, 110(2):448–452.

Foley, E. B. (2013). A big blue shift: Measuring an asymmetrically increasing margin of litigation. *Journal of Law and Politics*, 28:501–546.

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.

Gallego, J., Guardado, J., and Wantchekon, L. (2023). Do gifts buy votes? Evidence from sub-Saharan Africa and Latin America. *World Development*, 162:106125.

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.

Hausmann, R. and Rigobon, R. (2011). In search of the black swan: Analysis of the statistical evidence of electoral fraud in Venezuela. *Statistical Science*, 26(4):543–563.

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

Ichino, N. and Schündeln, M. (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. (2022). Do shifts in late-counted votes signal fraud? Evidence from Bolivia. *The Journal of Politics*, 84(4):1885–2311.

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

Kovalov, M. (2014). Electoral manipulations and fraud in parliamentary elections: The case of Ukraine. *East European Politics and Societies: and Cultures*, 28(4):781–807.

Leeffers, S. and Vicente, P. C. (2019). Does electoral observation influence electoral results? Experimental evidence for domestic and international observers in Mozambique. *World Development*, 114:42–58.

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

Lehoucq, F. (2020). Bolivia's citizen revolt. *Journal of Democracy*, 31(3):130–144.

Little, A. T. (2012). Elections, fraud, and election monitoring in the shadow of revolution. *Quarterly Journal of Political Science*, 7:249–283.

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

Magaloni, B. (2010). The game of electoral fraud and the ousting of authoritarian rule. *American Journal of Political Science*, 54(3):751–765.

McGhee, E., Paluch, J., and Romero, M. (2021). Vote-by-mail policy and the 2020 presidential election. Available at SSRN: https://dx.doi.org/10.2139/ssrn.3825939.

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

Mebane Jr., W. R. 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. C., 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.

Nooruddin, I. (2020a). Replication data for: OAS audit of Bolivia general election 2019. Harvard Dataverse, V3, https://doi.org/10.7910/DVN/SGOFSC.

Nooruddin, I. (2020b). Yes, Bolivia's 2019 election was problematic. Here's why. Washington Post Monkey Cage, March 10.

Organization of American States (2019). *Electoral Integrity Analysis, General Elections in the Plurinational State of Bolivia*. Final Report, Report to the General Secretariat, October 20.

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

Roe, B. E. and Just, D. R. (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. D. and Escobari, D. (2020). Bolivia's electoral fraud reckoning. Project Syndicate, March 17.

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.

Table 2: Difference-in-Differences Estimates

Dependent variable:	CC		MAS		MAS−CC		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables:							
SHUTDOWN × Y2019	-1.704***	-1.659***	1.095***	1.170***	2.981***	2.713***	2.854***
	(0.256)	(0.247)	(0.231)	(0.221)	(0.446)	(0.282)	(0.398)
SHUTDOWN	-6.582***		6.880***		13.28***	-1.115***	
	(0.372)		(0.364)		(0.718)	(0.210)	
Y2019	-13.80***	-13.49***	-1.866***	-2.359***	11.98***	11.15***	11.13***
	(0.0882)	(0.0918)	(0.0811)	(0.0789)	(0.154)	(0.102)	(0.141)
Constant	50.66***	49.44***	48.56***	49.92***	-2.154***	0.619***	0.437***
	(0.140)	(0.0443)	(0.139)	(0.0384)	(0.273)	(0.0726)	(0.0690)
Fixed Effects (F-statistic):							
Precinct	No	Yes	No	Yes	No	159	Yes
Polling Station	No	15.54	No	19.58	No	No	25
Observations	66,539	66,539	66,593	66,593	66,533	66,533	66,533
R-squared	0.103	0.950	0.016	0.956	0.035	0.934	0.966

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

Table 3: Difference-in-differences Estimates

Model:	SHUTDOWN	NOTREP
	(1)	(2)
Electoral Fraud Measures:		
SHUTDOWN × Y2019	1.060**	
	(0.440)	
NOTREP × Y2019		2.417***
		(0.811)
ARRIVAL × Y2019	5.756***	6.277***
	(0.545)	(0.494)
Fixed Effects (F-statistic):		
Last Name Initial	Yes	Yes
Precinct	Yes	Yes
Polling Station	25.48	25.53
Observations	65,810	65,810
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

Table 4: Difference-in-difference-in-differences Estimates

Model:	SHUTDOWN		NOTREP	
	(1)	(2)	(3)	(4)
Electoral Fraud Measures:				
SHUTDOWN × Y2019 × TREAT	2.897***	1.335***		
	(0.470)	(0.486)		
NOTREP × Y2019 × TREAT			3.024***	3.129***
			(0.704)	(0.705)
ARRIVAL × Y2019 × TREAT		4.599***		5.244***
		(0.628)		(0.626)
Fixed Effects (F-statistic):				
Last Name Initial	Yes	Yes	Yes	Yes
Precinct	Yes	Yes	Yes	Yes
Polling Station	3.260	3.237	3.263	3.233
Observations	136,316	134,793	136,316	134,793
R-squared	0.552	0.554	0.546	0.553

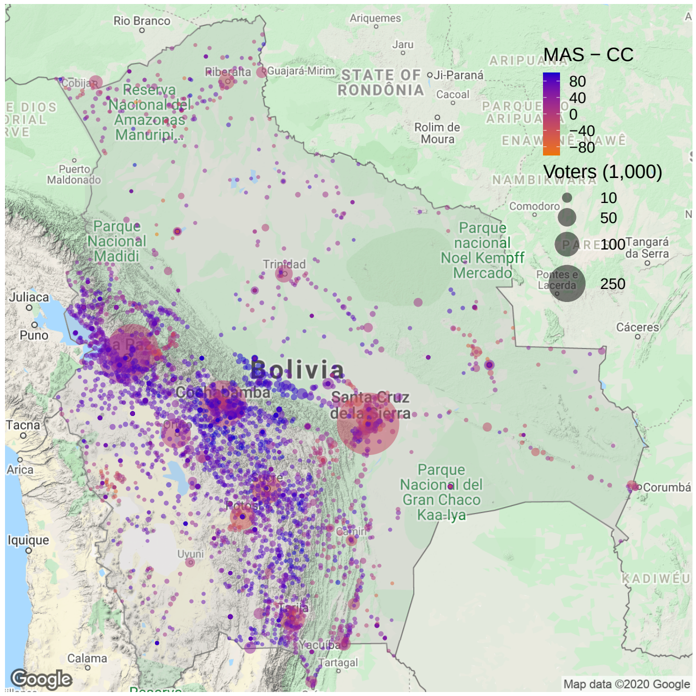
Notes: The dependent variable is MAS−CC. All specifications include the corresponding separate indicators (i.e., Y2019, TREAT, SHUTDOWN, and ARRIVAL) and first-order interactions (i.e., SHUTDOWN × Y2019, SHUTDOWN × TREAT, ARRIVAL × Y2019, ARRIVAL × TREAT, and Y2019 × TREAT) as controls. Last name initials 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

Table 5: 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

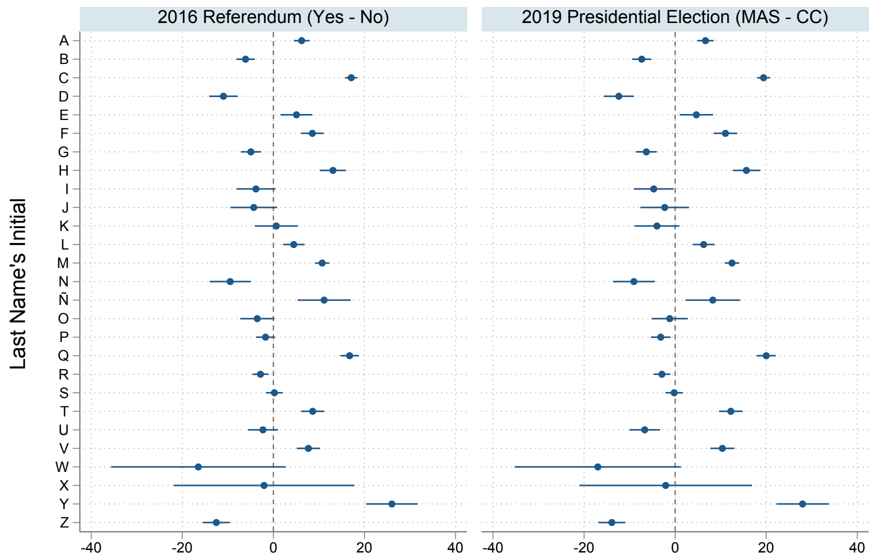
Notes: The dependent variable is MAS−CC. The running variable is the order polling stations were recorder as received by the Electoral Court (Arrival). The cutoff point it at $\overline{\mathrm{T}} = 0.95$. Robust bias-corrected 95% confidence intervals in brackets are based on Calonico et al. (2014). 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.

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



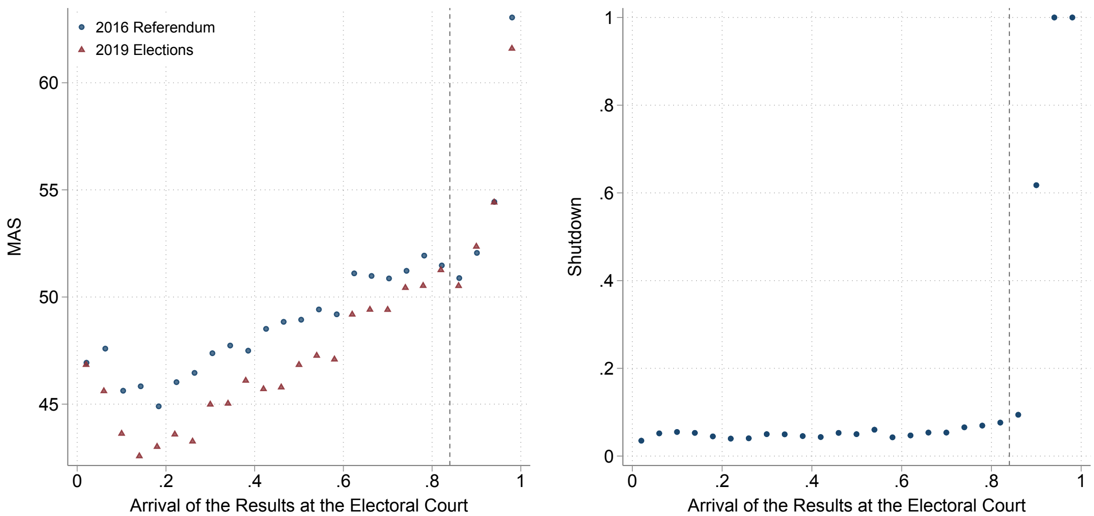
Notes: Votes at the locality level. Higher concentration of voters in urban areas are shown as larger bubbles. Larger metropolitan areas typically show relatively more support for CC. Smaller bubbles from rural areas and are more likely to support MAS in the highlands, and are closer to a balanced mix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

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



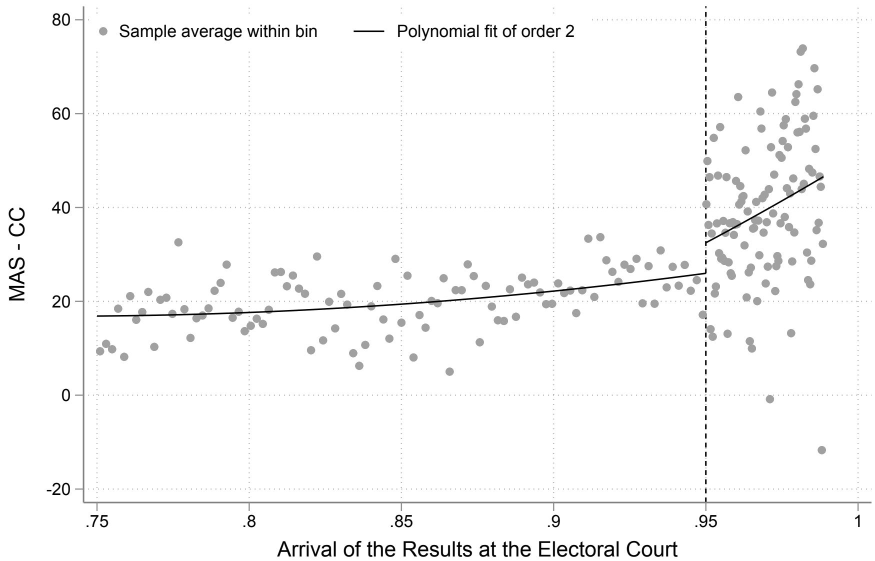
Notes: Last name's initial Fixed Effects point estimates with 95% confidence intervals. The dependent variable on the left-hand side of the figure is Yes−No from the 2016 constitutional referendum, while the dependent variable on the right-hand side is MAS−CC. All as shares of the valid votes.

Figure 3: 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 results at the Electoral Court, (Arrival). The right-hand side shows the share of shutdown polling stations as a function of Arrival.

Figure 4: Regression Discontinuity Plot.



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