



Analyzing Open Data to Support Democracy: a Study Case Inspecting Electoral Fraud in Bolivian General Elections

Sergio Peignier, Alexandre Foncelle

► To cite this version:

Sergio Peignier, Alexandre Foncelle. Analyzing Open Data to Support Democracy: a Study Case Inspecting Electoral Fraud in Bolivian General Elections. Digital Humanities BENELUX 2020, Jun 2020, Leiden, Netherlands. hal-02639204

HAL Id: hal-02639204

<https://hal.science/hal-02639204>

Submitted on 28 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyzing Open Data to Support Democracy: a Study Case Inspecting Electoral Fraud in Bolivian General Elections

Sergio Peignier¹ and Alexandre Foncelle²

¹Univ Lyon, INSA Lyon, INRA, BF2I, UMR0203, F-69621, Villeurbanne, France

²Inserm UMR-S 1028, CNRS UMR 5292, ImpAct, Centre de Recherche en Neurosciences de Lyon, Université Lyon 1, Bron, France

Abstract

The objective of this paper is to show how analyzing open data may help verify politicians' arguments. In this paper we considered the polemical 2019 Bolivian general elections as case study. We used open access electoral data and statistical tools, to assess two political arguments that aimed at explaining changes in the vote count trend, namely the arrival of rural votes, and electoral fraud. This study highlights and discusses the importance of open data access and the involvement of science in assessing political arguments.

1 Introduction

Karl Popper defines democracy, in opposition to tyranny, as it aims at allowing society to remove “bad rulers” without violence[1]. To meet this objective, democratic institutions are oriented towards transparency, which makes them more legitimate in the eyes of society[2]. However, in politics, instead of explaining concepts in a transparent rationality, argumentation is mainly used to persuade and seduce the public [3]. Political statements are often built upon beliefs and simplified causal reasoning [4]; and examples are often used as proofs, in a pseudo-scientific way. Hence, to support transparency, the objective of this paper is to show how analyzing open data may help verifying politicians' arguments. Here we use the 2019 Bolivian general elections as case study.

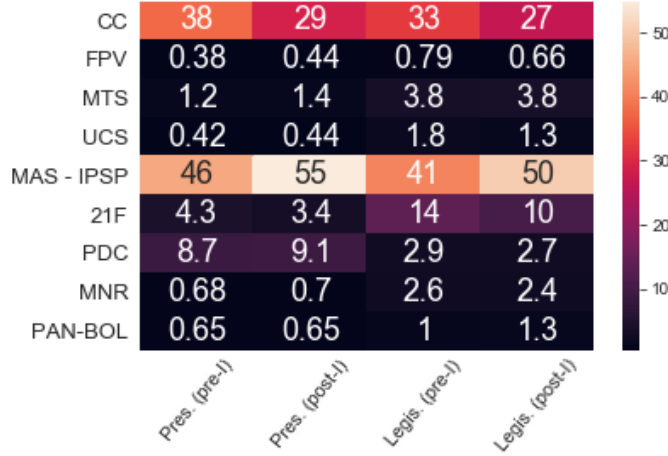


Figure 1: Parties vote percentages, for results published before the vote count interruption (*pre-I*) and new results reported after the interruption (*post-I*).

2 Context

Bolivia held polemical general elections the 20th of October 2019, when the then president, Evo Morales, ran for re-election, in spite of the results of a referendum rejecting any constitutional amendment allowing indefinite presidential re-elections[5]. Consequently, the opposition denounced that democracy was endangered, and an important part of the electorate was distrustful towards the election. To cope with this scenario, the Bolivian Electoral Institution undertook to publish the vote count online and live. According to the first results (accounting $\sim 87.5\%$ of the votes) the difference between the top two candidates, i.e. Evo Morales (MAS-IPSP party) and Carlos Mesa (CC party) was equal to 7.86%. Since this difference was lesser than 10%, according to Bolivian laws, a second tour was required[6]. Nevertheless, the vote count was interrupted unexpectedly for $\sim 24\text{h}$ [6], and after the delay, the voting percentages changed remarkably, as reported Figure 1. Indeed, the difference between MAS-IPSP and CC increased of 2.25%, moving from 7.86% to 10.11%, thus, Morales was re-elected without a second tour[7]. The opposition denounced an electoral fraud, while the government justified the changes claiming the arrival of rural votes endorsing Morales[7].

In this case-study, we used statistical methods to assess the **governmental rural votes** and the **opponents electoral fraud** hypothesis.

For the sake of clarity, we focus on the presidential election. Nevertheless figures also illustrate the legislative election results. Conclusions are analogous in both scenarios.

3 Data analysis

Data In this work we considered: i) vote counts published before the interruption (*pre-I*), ii) new counts published after the interruption (*post-I*). Both datasets report, for each pooling place, the number of votes per political party. Pooling places are characterized by their locations, hierarchically structured from departments to addresses. Here, we compared datasets votes distribution, at the “locality” level, since this territorial division, in between municipalities and pooling places, represents well-identified population nuclei.

Rural hypothesis According to the government, the voting distribution changes were due to the arrival of votes from rural localities. To assess this hypothesis, we identified 353 localities¹, that were only represented in *post-I*, and we evaluated their impact on the total count. These localities sum a total of 91,888 votes out of 5,888,004, which are indeed strongly oriented in favor of MAS-IPSP compared to CC (64,015 versus 11,873 votes).

However these votes only explain an increase of 0.886% in the difference between both parties, whereas 1.425% remains unexplained.

Intra-locality multinomial model Let us assume that vote counts, from pooling places within the same locality, follow similar distributions, regardless of being reported before/after the interruption. Given this founding hypothesis, we modeled intra-locality votes using a multinomial distribution, where each vote is an independent trial, leading to a success for one of $K = 9$ categories (parties). Let loc_i be the locality of pooling place i , and let n^i be its number of valid votes. Let $p_k^{loc_i}$ be probability of success of category k in loc_i . The expected number of votes for category k among n^i votes, is $p_k^{loc_i} \times n^i$. The bias between n_k^i , the observed number of votes for k in pooling place i , and the expected value is $\delta_k^i = n_k^i - p_k^{loc_i} \times n^i$, and the sum of biases for a set of pooling places is $\Delta_k = \sum_i \delta_k^i$. In practice, the model parameters were estimated by computing relative frequencies of votes in *pre-I*, and the sum of biases were computed for *post-I* pooling places.

According to our founding hypothesis, Δ_k should be distributed around zero. Nevertheless, as illustrated Figure 2, Δ_k registers a gain of 22,720 votes for MAS-IPSP, and a loss of 29,228 votes for CC. This represents a difference of 0.83% of presidential election valid votes.

Statistical significance To estimate the significance of observed biases, we estimated the probability p_w (p-value) of obtaining these biases by chance, using a paired Wilcoxon test, by comparing observed and expected *post-I* vote counts.

Then we used a bootstrap procedure to compute an empirical distribution of 100 p-values p_w^b by: 1) sampling randomly subsets of the *pre-I* (with same size

¹Accordingly to the governmental claim most of these new localities were rural, in fact, only 49 out of 353 belong to cities with more than 20 000 inhabitants.

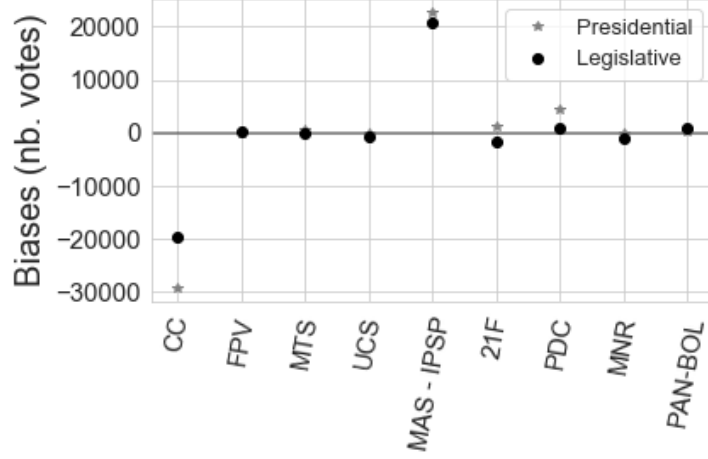


Figure 2: Sum of biases (in number of votes), per party, between *post-I* pooling place counts and expected values.

as *post-I*), 2) computing, for each subset, the biases with respect to an intra-locality multinomial model fitted on the remaining pooling places, 3) Computing p-values using Wilcoxon test.

The distribution of the log-odds $\log_2(p_w^b/p_w)$, between p_w^b and p_w , depicted Figure 3, suggests that it is very unlikely that MAS-IPSP and CC biases are due to chance, thus the fraud hypothesis is not excluded ².

This approach could be used in other cases where relatively fine-grained geolocalized voting results are available.

4 Conclusion

The rural hypothesis explained partially the increase of the difference between the top candidates. Nevertheless, intra-locality comparisons of pre- and post-interruption counts, exhibited significant biases favourable (resp. unfavourable) to MAS-IPSP (resp. CC) party, possibly due to electoral fraud.

To appease the socio-political crisis following the elections, the Organization of American States (OAS) conducted an in-depth audit of the elections, requesting the collaboration of society[9]. Hence, we sent a detailed version of this work and scripts to the OAS. Final OAS conclusions confirmed intentional manipulations of the election[10].

²It should be noted that different results could be obtained using different models or data sampling techniques however, as stated by George Box “all models are wrong, but some are useful” [8]

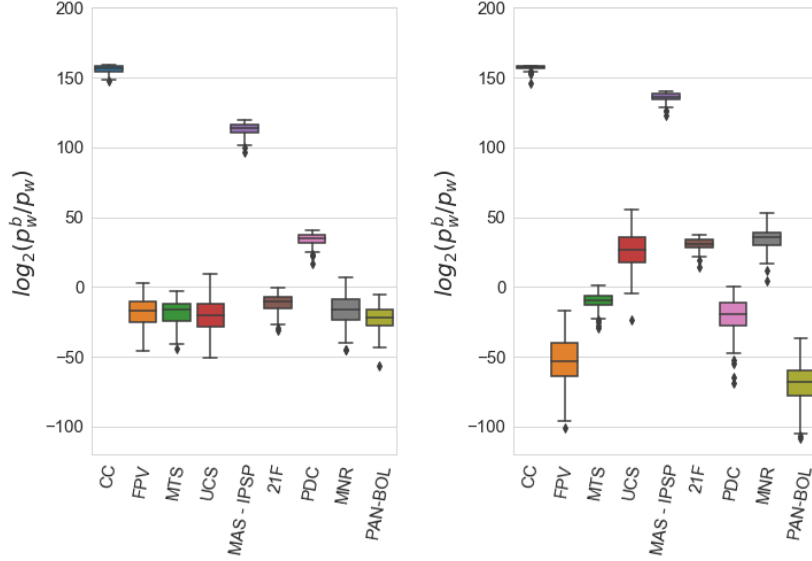


Figure 3: Distribution of log-odds between p_w and p_w^b , for presidential (right) and legislative (left) elections.

5 Discussion

The electoral data availability allowed society to participate in the electoral assessment (e.g., [11, 10, 12, 13, 14, 15, 16, 17, 18]), highlighting the importance of open data access and the involvement of science in assessing political arguments. This kind of approaches are increasingly important nowadays, given the growing spread of "fakes news" and "false arguments".

Nevertheless, there was no consensus over different works as depicted Figure 4. Some organizations were accused of leading biased investigations with political interests[19, 12]. Reports, published without undergoing peer-review, were presented, by some politicians and media, as depicting unquestionable truths. This phenomenon described as the "scientification of politics" and the "politicisation of science"[20], indicates that science is often used as authoritative argument, and scientists are indissociable from subjective judgments. Consequently, it seems crucial to integrate validation procedures such as independent analysis replication, and peer-review procedures.

References

- [1] The economist. From the archives: the open society and its enemies revisited, 2016.
- [2] James R. Hollyer, B. Peter Rosendorff, and James Raymond Vreeland.

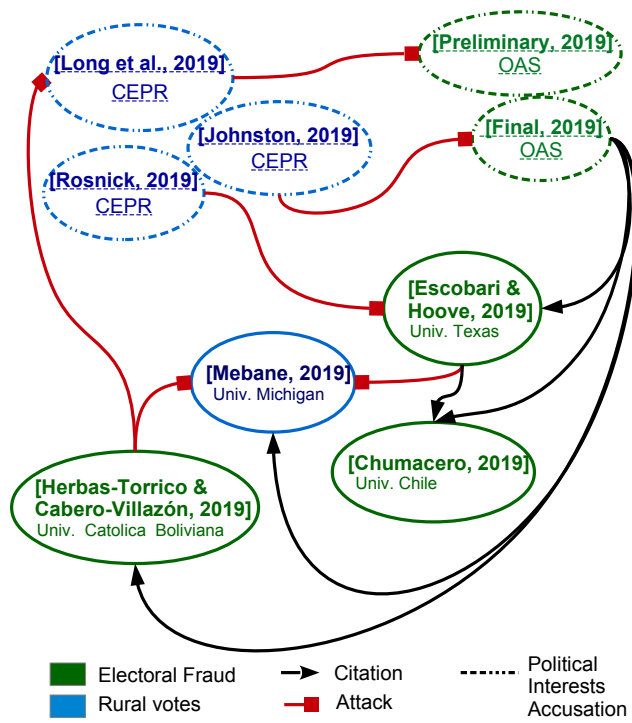


Figure 4: Related work citations network

- Democracy and transparency. *The Journal of Politics*, 73(4):1191–1205, 2020/02/18 2011.
- [3] Suhair Safwat. Speech acts in political speeches. *Journal of Modern Education Review*, 5:699–706, 07 2015.
 - [4] Patrick Charaudeau. Quand l’argumentation n’est que visée persuasive. l’exemple du discours politique. *Argumentation et communication dans les médias, Québec: Nota bene*, pages 29–49, 2005.
 - [5] BBC News Mundo. Evo morales: el tribunal electoral de bolivia lo habilita como candidato presidencial tras haber perdido el referéndum por la reelección, 2019.
 - [6] BBC News Mundo. Elecciones en bolivia: suspenden el recuento provisional de votos cuando todo apuntaba a una segunda vuelta entre evo morales y carlos mesa, 2019.
 - [7] BBC News Mundo Boris Miranda. Elecciones en bolivia: por qué hay cuestionamientos y denuncias de fraude sobre los resultados preliminares que sitúan a evo morales como ganador en primera vuelta, 2019.
 - [8] George EP Box. Robustness in the strategy of scientific model building. In *Robustness in statistics*, pages 201–236. Elsevier, 1979.
 - [9] Página Siete Digital. Equipo técnico de la oea pide que la “ciudadanía entregue material para análisis”, 2019.
 - [10] Organization of American States. Electoral integrity analysis, general elections in the plurinational state of bolivia, final report, 2019.
 - [11] Organization of American States. Electoral integrity analysis general elections in the plurinational state of bolivia october 20, 2019, preliminary findings report, 2019.
 - [12] Guillaume Long, David Rosnick, Cavan Kharrazian, and Kevin Cashman. ¿qué sucedió en el recuento de votos de las elecciones de bolivia de 2019? *CEPR*, 2019.
 - [13] Diego Escobari and Gary A Hoover. Evo morales and electoral fraud in bolivia: A natural experiment estimate. *Available at SSRN 3492928*, 2019.
 - [14] Rómulo Chumacero. El camaleón, el mutante y houdini: Resultados de las elecciones en bolivia. *Universidad de Chile*, 2019.
 - [15] David Rosnick. Unnatural claims in a ‘natural experiment’: Escobari and hoover on the 2019 bolivian elections. *CEPR*, 2019.
 - [16] Jake Johnston. Análisis preliminar de los hallazgos del informe final de la auditoría de la oea. *CEPR*, 2019.

- [17] Walter Mebane. Evidence against fraudulent votes being decisive in the bolivia 2019 election. *University of Michigan personal homepage*, 2019.
- [18] Boris C. Herbas-Torrico and Rene Cabero-Villazón. Análisis crítico de los estudios de mebane (2019) y long et al. (2019) sobre las elecciones de octubre de 2019 en bolivia. *Universidad Católica Boliviana “San Pablo”*, 2019.
- [19] Francisco Rodríguez et al. How not to defend the revolution: Mark weisbrot and the misinterpretation of venezuelan evidence. Technical report, Wesleyan University, Department of Economics, 2008.
- [20] Peter Weingart. Scientific expertise and political accountability: paradoxes of science in politics. *Science and public policy*, 26(3):151–161, 1999.