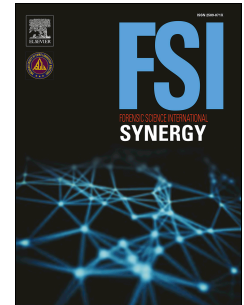# Journal Pre-proof

The forensics of fraud: Evidence from the 2018 Brazilian presidential election

Dalson Figueiredo Filho, Lucas Silva, Ernani Carvalho

Please cite this article as: D.F. Filho, L. Silva, E. Carvalho, The forensics of fraud: Evidence from the 2018 Brazilian presidential election, *Forensic Science International: Synergy* (2022), doi: https://doi.org/10.1016/j.fsisyn.2022.100286.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The forensics of fraud:
evidence from the 2018 Brazilian Presidential Election

Dalson Figueiredo Filho[1]; Lucas Silva[2]; Ernani Carvalho[3];

[1]Department of Political Science, Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil

[2]Department of Medicine, Universidade Estadual de Ciências da Saúde do Estado de Alagoas, Maceió, Alagoas, Brazil;

[3]Department of Political Science, Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil

Correspondence: Lucas Silva. Universidade Estadual de Ciências da Saúde do Estado de Alagoas; Rua Dr. Jorge de Lima, 113 - Trapiche da Barra, Maceió - AL, 57010-300; lucas.silva@academico.uncisal.edu.br; https://orcid.org/0000-0002-5013-6278

The forensics of fraud:
evidence from the 2018 Brazilian Presidential Election

ABSTRACT

Objective
This paper studies the integrity of the vote counting system in Brazil.

Method
We analyze data from the Superior Electoral Court (TSE) for the 2018 Brazilian presidential election to assess suspicious vote count patterns deploying five techniques commonly used to detect fraud: a) the second-digit Benford's law test; b) the last digit mean; c) frequency analysis of last digits 0 and 5; d) correlation between the percentage of votes and the turnout rate; and e) resampled Kernel density of the proportion of votes.

Results
The results show that the second-digit distributions for the three most voted candidates – Jair Messias Bolsonaro (PSL), Fernando Haddad (PT), and Ciro Gomes (PDT) – conform to Benford's law. We also find that last digit means and last digit frequency are within normal parameters, indicating no irregularities. Similarly, the fingerprint plot indicates a correlation coefficient that is consistent with the theoretical expectation of a fair election. The resampled Kernel density suggests that the vote count was performed without statistically significant distortions. These results are robust at different levels of data aggregation (polling station and municipality).

Conclusion
The joint application of digit-focused tests, regression-based techniques, and patterns in the distribution of vote-shares provide a more reliable method for detecting anomalous cases. Relying on this unified framework, we find no evidence of electoral fraud in the 2018 Brazilian presidential election. These results advance our current understanding of statistical forensics tools and may be easily replicated to examine electoral integrity in other countries.

Keywords: electoral fraud, forensic analysis, elections.

INTRODUCTION

Election fraud - clandestine, and illegal efforts to shape electoral outcomes[1] - distorts representation, undermines public accountability, and erodes government legitimacy[2]. Therefore, determining whether electoral results reflect voters' preferences is pivotal in representative democracies[3]. In this paper, we develop a new approach for diagnosing anomalies in electoral outcomes. Drawing on the work of Mebane[4], Myagkov, Ordeshook, and Shaikin[5], Beber and Scacco[6], and Rozenas[7], we propose a unified methodology that relies on the joint application of digit-focused methods (first, second, and last digits), regression-based techniques (correlation between turnout and votes), and the vote-shares distribution patterns (including invalid votes and coarse vote-shares). We demonstrate the advantages of our framework using data from the 2018 Brazilian presidential election.

A case study focused on Brazil represents a unique opportunity to learn from an outlier. According to Alvarez, Hall, and Hyde: "claims of electoral fraud are typically made by the losing candidate or party"[2]. Brazilian President Jair Messias Bolsonaro (affiliated with Social Liberal Party – PSL) was elected with 55,13% of valid ballots, at a 10-million votes lead from his opponent, Fernando Haddad (Workers' Party – PT). Yet, he has repeatedly raised suspicions about the integrity of the country's electronic vote-counting system, affirming for instance that "I believe, from the evidence I have in my hands, which I will reveal shortly, that I was elected in the first round. In my view, there was fraud"[8]; "I will pass on the presidential sash to anyone who cleanly beats me at the ballot box. If there is a fraud, I will not"[9]; "Without the printed vote [printed voting voucher], there won't be an election"[10]. Such statements have caused enormous discomfort and a negative impact on the credibility of the electronic ballot box. This attitude significantly increases the possibility of an institutional crisis since the judiciary is responsible for electoral governance and the president of the Superior Electoral Court (TSE) is always a Supreme Court (STF) Judge[11,12]. Given the role of the president in the presidential system, such an attitude against the Brazilian electoral system affects the public debate[13].

To evaluate Bolsonaro's claims of electoral fraud, we collected disaggregated data from the Superior Electoral Court (TSE) regarding the 2018 Brazilian presidential election. We use a combination of statistical techniques to spot anomalies: a) the second-digit test of Benford's law; b) the last digit mean; c) frequency analysis of last digits 0 and 5; d) correlation between the percentage of votes and the turnout rate; and e) resampled Kernel density (RKD) of the proportion of votes[14]. To increase transparency and reproducibility, we share the original data and computational scripts on an online open-access platform.

The remainder of the paper is divided as follows. The next section describes the materials and methods. The third part presents the statistical results. The last section concludes and discusses the study's limitations.

MATERIALS AND METHODS

All data used in this research are publicly available in the Electoral Data Repository of the Brazilian Superior Electoral Court[15]. Table 1 summarizes the main tests used to detect anomalies in the counting of votes in the 2018 Brazilian presidential elections.

Table 1.
Summary of diagnostic tests

Test	Definition	Theoretical expectation in a fair election	References
2BL	Distribution of the significant second digit in conformity with Benford's law	4.187	Mebane[4], Pericchi and Torres[16], and Hicken and Mebane[13]
UD	Last digit mean	4.5	Dlugosz and Müller-Funk[17], Beber and Scacco[6], and Skovoroda and Lankina[18]
	Frequency analysis of last digits 0 and 5	0.2	Beber and Scacco[6]
FPP	Correlation between the percentage of votes of the winning candidate and the turnout rate – fingerprint plot	r = 0 and cases grouped at the superior right corner of the scatterplot	Myagkov, Ordeshook, and Shaikin[5], Klimek et al.[19], and El-Dash[20]
RKD	Resampled Kernel density of the proportion of votes	0.07 (F value observed in Canadian elections in 2011)	Rozenas[7]

Source: Elaborated by the authors based on Hicken and Mebane[13] and Rozenas[7]

According to Mebane[4] and Hicken and Mebane[13], the second digit test (2BL) is one of the most widely used tools to detect anomalies in electoral data. The 2BL test compares the frequency of the second digit of the number of votes for each candidate and the theoretical distribution. Significant deviations, measured by the chi-square test, may indicate the presence of malicious activities[21]. Table 2 illustrates the theoretical expectation of the first and second digits distribution according to the Benford's Law.

3

Table 2.
Expected frequency of first and second digits according to the Benford's Law (%)

Digit	1st digit	2nd digit
0		11.97
1	30.10	11.39
2	17.61	10.88
3	12.49	10.43
4	9.69	10.03
5	7.92	9.67
6	6.69	9.34
7	5.80	9.04
8	5.12	8.76
9	4.58	8.50

Notes: Table shows the expected frequency of the first and second digits according to the Benford's Law distribution. The expected theoretical mean value of the second digit in a fair election is 4.187.

We also evaluate the distribution of the last digit of the vote count[6,17,18]. Mack and Stoetzer argue that "the last digit test is a notable method to detect election fraud. It is based on an assumption that a manipulator replaces the vote counts of an election result sheet with man-made numbers but will fail to make the numbers look random"[22]. Theoretically, manipulation-free data should exhibit an approximate mean of 4.5 and a uniform distribution of digits – each digit should appear 10% of the time.

Similarly, according to Beber and Scacco, "last digits will occur with equal frequency for a large class of theoretical distributions, and we argue that non-fraudulent electoral returns are likely to be drawn from such a distribution"[6]. In other words, assuming the inability of humans to produce random sequences, the expectation is that, in a fair election, the distribution of the last digit of the number of votes received by each candidate should be approximately uniform. On the other hand, the disproportionate occurrence of any number could indicate that the total number of votes was intentionally manipulated (fraud). This logic is at the base of the frequency analysis of last digits 0 and 5. In the absence of fraud, their average frequency should approach 0.2.

The fingerprint plot was suggested by Myagkov, Ordeshook, and Shaikin[5] and is widely used to detect signs of electoral fraud. According to El-Dash (our translation), "two types of fraud can be detected using this form of analysis: the addition of false votes for a candidate at the ballot box and the substitution of votes from one candidate to another"[20]. These count-tampering strategies produce specific statistical patterns, such as the correlation between valid votes and the turnout rate, and the clustering of cases in the upper corner of the scatterplot[19]. This is because the number of votes that can be transferred from candidate X to candidate Y is limited by the number of voters who turned out in a given polling station. Thus, the more votes

artificially added to the winning candidate, the more evident the pattern of association between the percentage of valid votes and the turnout.

Finally, we replicate the approach developed by Rozenas[7] to examine the proportion of votes based on the resampled Kernel density (RKD). The model is based on the "identification of irregular patterns in the distribution of party (or candidate) vote-shares across the polling stations. The method builds on an observation that in some allegedly fraudulent elections, there are many polling stations where the ruling party receives the percentage of votes that is evenly divisible by five – e.g., 50%, 65%, 75% (…) Intuition would dictate that in large electorates, it is exceedingly unlikely for a party to receive exactly such a coarse vote-share across many precincts without some form of vote falsification"[7]. This technique requires the use of the "Spikes" R package[23].

All statistical estimations were implemented using R Statistical, version 4.0.5. Following scientific best practices[24], the replication materials, including data and computational scripts, are available at the Open Science Framework (OSF).

RESULTS

a) Benford's law significant second digit

Table 3 presents the mean of the second digit distribution for the three most voted candidates in the first round of the 2018 Brazilian presidential elections.

Table 3.

Mean of the second digit in the first round of presidential elections using municipality data

Candidate (political party)	μ	95% C.I.*	p-value	N
Jair Bolsonaro (PSL)	4.208	4.133 – 4.285	0.584	5,696
Fernando Haddad (PT)	4.173	4.099 – 4.251	0.719	5,662
Ciro Gomes (PDT)	4.215	4.139 – 4.290	0.466	5,663

Notes: μ represents the observed mean of the second digit. N represents the number of Brazilian municipalities with available information; Theoretical expectation in a fair election (μ = 4.187); *Values calculated based on 1,000 iterations of non-parametric bootstrapping.

Using a 95% confidence level, we find no significant difference between the theoretical distribution of the second digit (2BL) and the frequency observed for the three most voted candidates during the first round of the 2018 presidential elections. Figure 1 displays the observed frequencies of each candidate with the theoretical distribution of the second digit in conformity with Benford's Law (2BL).
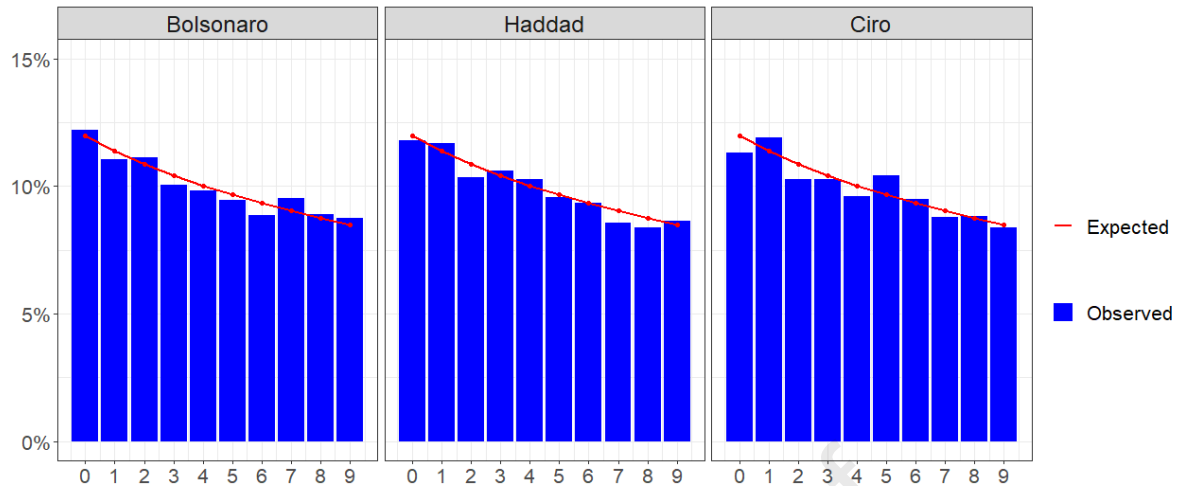
Figure 1. Second digit distribution and Benford's Law in the first round of presidential elections
Notes: The red line indicates the theoretical expectation in conformity with Benford's law for the second digit (2BL). The blue column represents the distribution observed from the results of the total vote count per municipality during the first round of the 2018 presidential elections.

The greater the proximity between the blue column and the red line, the lower the chi-square value. There was no evidence of anomalies, which is a sign of the absence of systematic fraud in the vote count. These results are consistent for the second round as explained in Table 4.

Table 4.
Mean of the second digit in the second round of presidential elections using municipality data

Candidate (political party)	μ	95% C.I.*	p-value	N
Jair Bolsonaro (PSL)	4.154	4.080 – 4.230	0.392	5,699
Fernando Haddad (PT)	4.197	4.120 – 4.270	0.783	5,682

Notes: μ represents the observed mean of the second digit using second round municipality data; N represents the number of Brazilian municipalities with available information; Theoretical expectation in a fair election ($\mu = 4.187$); *Values calculated based on 1.000 iterations of non-parametric bootstrapping.

The average votes for Bolsonaro (4.154) and Haddad (4.197) do not differ significantly from the parameter that assumes fair elections ($\mu = 4.187$). Also, we carried out the calculation of a 95% confidence interval based on 1,000 resamplings from the same dataset to ensure more robust results, and the evidence remained consistent. Figure 2 shows the frequencies of the second digit per candidate during the second round of the 2018 Brazilian presidential elections.
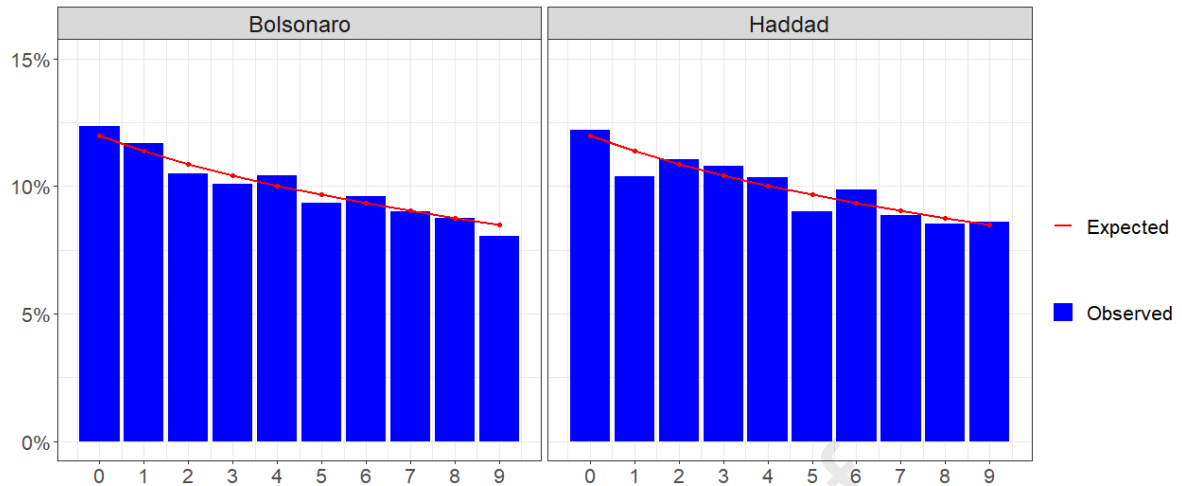
6

Figure 2. Second digit distribution and Benford's Law in the second round of presidential elections
Notes: The red line indicates the theoretical expectation in conformity with Benford's law for the second digit (2BL). The blue column represents the distribution observed from the results of the total vote count per municipality during the second round of the 2018 presidential elections.

b) Uniform distribution

b.1) Last Digit Mean

After examining the distribution of the second significant digit, the next step is to assess the extent to which the 2018 election data passes the last digit test, as described in Dlugosz and Müller-Funk[17] and Beber and Scacco[6]. Table 5 reports the last digit mean of the three most voted candidates during the first round of the 2018 presidential elections.

Table 5.
Last Digit Mean in the first round of presidential elections using municipality data

Candidate (political party)	μ	95% C.I.*	p-value	N
Jair Bolsonaro (PSL)	4.52	4.45 - 4.59	0.6	5,708
Fernando Haddad (PT)	4.49	4.42 - 4.56	0.8	5,704
Ciro Gomes (PDT)	4.49	4.41 - 4.56	0.7	5,705

Notes: μ represents the observed mean of the last digit mean; N represents the number of Brazilian municipalities with available information; Theoretical expectation in a fair election (μ = 4.5); *Values calculated based on 1,000 iterations of non-parametric bootstrapping.

The values obtained in the 2018 Brazilian presidential elections are close to the theoretically expected parameter (see the mean and the respective confidence intervals). For all candidates, the mean value is very close to 4.5. Comparatively, these results are different from those found by Hicken and Mebane[13] when analyzing the elections in Afghanistan (μ = 4.112) and South Africa (μ = 4.069).

When we analyze the results of the second round in Table 6, the results also remain close to 4.5.

Table 6.

Last Digit Mean in the second round of presidential elections using municipality data

Candidate (political party)	μ	95% C.I.*	p-value	N
Jair Bolsonaro (PSL)	4.43	4.36 - 4.50	0.07	5,708
Fernando Haddad (PT)	4.48	4.40 - 4.55	0.6	5,705

Notes: μ represents the observed mean of the last digit mean; N represents the number of Brazilian municipalities with available information; Theoretical expectation in a fair election ($\mu = 4.5$); *Values calculated based on 1,000 iterations of non-parametric bootstrapping.

To ensure more reliable evidence, we also observed the frequency of the last digit in candidate votes. Figure 3 presents the relative frequency of the last digit per candidate.
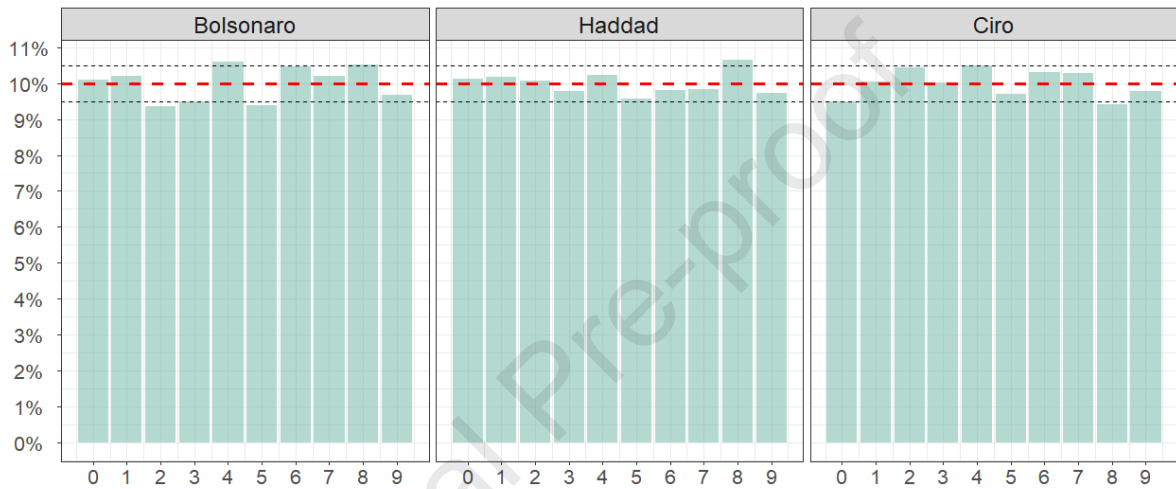
Figure 3. Frequencies of last digits in the first round of presidential elections by candidate using pooling station data

Notes: The red dotted line shows the uniform frequency that should be observed in manipulation-free distributions; black dotted lines indicate the value of .5% plus and minus the expected frequency

The red dotted line is the 10% parameter, considering that a fair distribution must be uniform[6,17,18]. Thus, considering the average of the last digit of the number of nominal votes received by each candidate during the second round of the 2018 presidential elections, the evidence reported here does not indicate any abnormality. Figure 4 shows the same pattern for the second round.
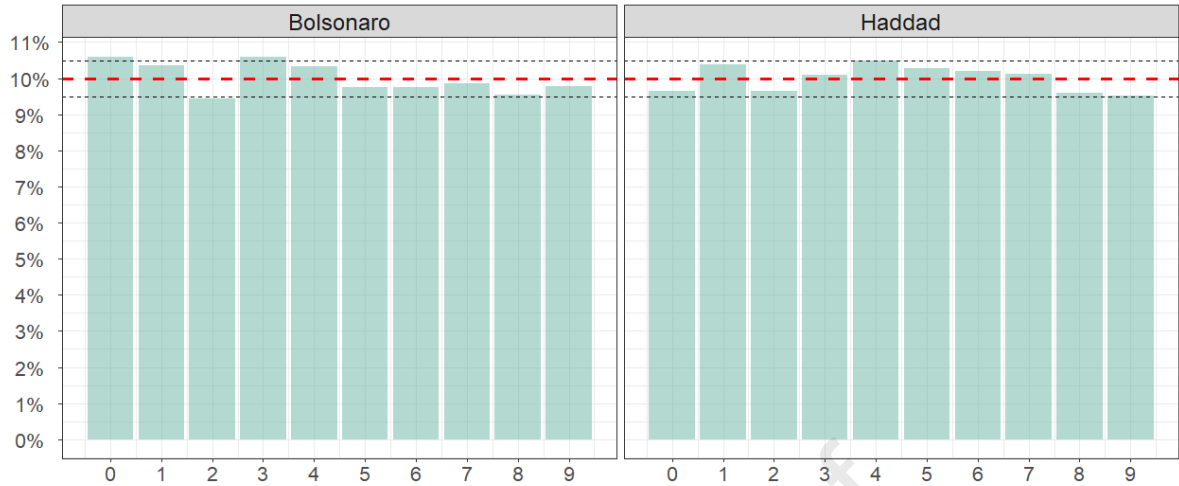
Figure 4. Frequencies of last digits in the second round of presidential elections by candidate using pooling station data
Notes: The red dotted line shows the uniform frequency that should be observed in manipulation-free distributions; black dotted lines indicate the value of .5% plus and minus the expected frequency

To confirm the validity of the result, we reproduced the analysis of the last digit to evaluate the distribution of four indicators that, in theory, are more difficult to manipulate and/or require a highly sophisticated statistical and computational effort: a) the number of eligible voters; b) the total number of voters that attended the polling stations; c) total number of nominal votes, and d) the count of total invalid votes. Figure 5 illustrates these distributions.
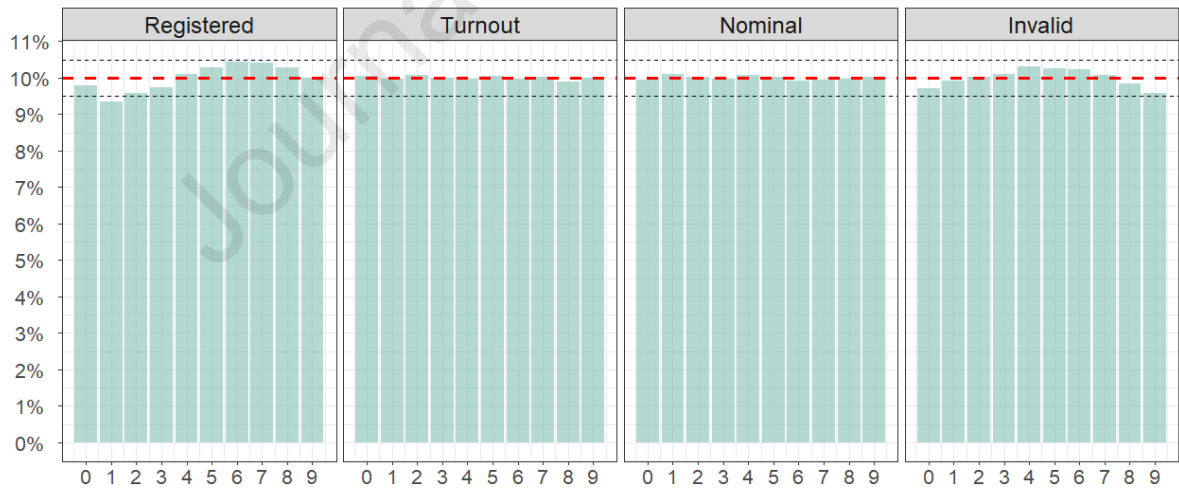

Figure 5. Frequencies of last digits in the first round of presidential elections by ballot type using pooling station data
Notes: The red dotted line shows the uniform frequency that should be observed in manipulation-free distributions; black dotted lines indicate the value of .5% plus and minus the expected frequency

For all indicators, the frequency of the last digit approaches a uniform distribution. These distributions are similar to the results found by Beber and Scacco[6] concerning the counting of votes in Sweden.

9

b.2) Digit frequency analysis of last digits 0 and 5

The next step is to evaluate the frequency of the last digits 0 and 5. Table 7 summarizes this information for the three most voted candidates during the first round of the 2018 Brazilian presidential elections.

Table 7.
Mean of last digits 0 and 5 in the first round of presidential elections using pooling station data

Candidate (political party)	μ	N
Jair Bolsonaro (PSL)	0.202	443,503
Fernando Haddad (PT)	0.200	454,414
Ciro Gomes (PDT)	0.198	452,851

Note: μ represents the observed mean of the last digit; N represents the number of Brazilian polling stations with available information; Theoretical expectation in a fair election ($\mu = 0.2$)

In a fair election, the theoretical expectation is that the mean of the relative frequency of digits 0 and 5 approaches 0.2. The counts of 443,503 polling stations resulted in a mean of 0.202 for Bolsonaro. For Haddad, the mean was 0.2 for 454,414 polling stations. Finally, a mean of 0.198 for 452,851 polling stations was found for Ciro Gomes. Following Beber and Scacco[6], the absence of digit count abnormalities should produce a distribution with a mean of 0.2. For the most preferred candidates, the observed mean of the last digit is statistically equal to the theoretical expectation in a fair election. Table 8 reproduces the analysis using data for the second round.

Table 8.
Mean of last digits 0 and 5 in the second round of presidential elections using pooling station data

Candidate (political party)	μ	N
Jair Bolsonaro (PSL)	0.199	454,119
Fernando Haddad (PT)	0.200	454,443

Notes: μ represents the observed mean of the last digit; N represents the number of Brazilian polling stations with available information; Theoretical expectation in a fair election ($\mu = 0.2$)

The results show no abnormality in the mean of last digits 0 and 5. The values are exactly equal to the expected theoretical parameter, assuming fair elections ($\mu = 0.2$).

c) Fingerprint plot

Figure 6 illustrates the correlation between the turnout rate and votes on the winning candidate for both rounds of the 2018 Brazilian presidential elections, on a polling station level.
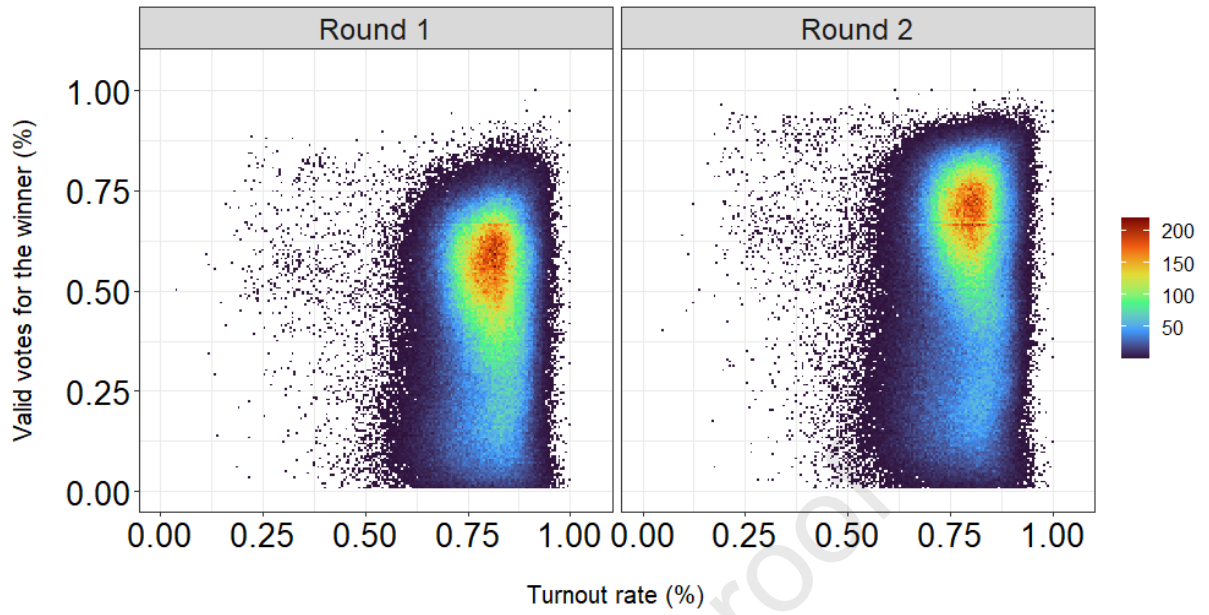
Figure 6. 2-d histograms of electoral districts per turnout rate and valid votes for the winning candidate

Notes: The election fingerprint plot shows a two-dimensional histogram associating the turnout (x- axis) with the percentage of nominal votes for the winning candidate (y-axis). The number of pooling station units is given by the color. In a fair election, the expected association between turnout and votes cast is zero. Conversely, above the average vote counts occurring with unusually high turnout are indicative of election fraud.

As observed in Figure 6, there is no pattern of association between the variables. In addition, there is no grouping of data in the upper right corner of the graph, which could be a possible sign of irregularity in the counting of votes. Figure 7 reproduces the association between the turnout rate and the proportion of valid votes obtained by the winning candidate from two random samples of 1% and 2%, respectively.
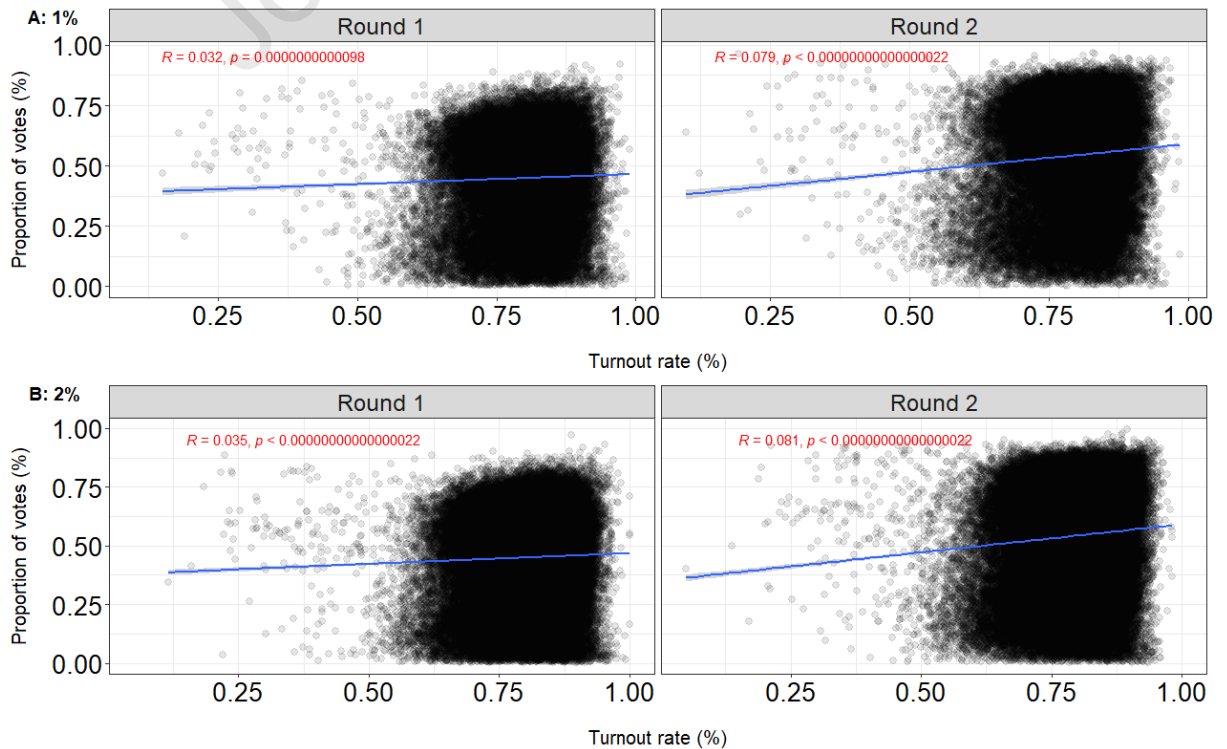
Figure 7. Association between turnout rate and valid votes of the winning candidate by electoral round (1% and 2% random samples, respectively)

Notes: The blue line represents the linear regression fit correlating the turnout (x-axis) with the percentage of nominal votes for the winning candidate (y-axis).

When considering a simple random sample of 1% of the polling stations (n = 90,789), the correlation between the turnout rate (%) and the proportion of valid votes (%) is 0.032 in the first round and 0.079 in the second round. Similarly, when considering disaggregated data from 181,578 polling stations, equivalent to a 2% sample of total observations, the association level between turnout and voting was 0.035 and 0.081 in the first and second rounds, respectively. The absence of correlation between the variables, demonstrated by the slope of the blue line, indicates the inexistence of suspicious vote count patterns.

d) Resampled Kernel density of the proportion of votes

Finally, we reproduce the model proposed by Rozenas[7], which relies on the proportion of votes received by each candidate as a function of the total number of eligible voters and the turnout rate. The results produce an F value that Rozenas called the estimated level of fraud with a credibility interval of 95%[7], as shown in Table 9.

Table 9.

RKD fraud estimates, with 95% credible intervals

	Estimate (F)	95% Credible Interval
United Russia, 2003	0.24	(0.19, 0.29)
United Russia, 2007	0.86	(0.74, 1.01)
United Russia, 2011	0.94	(0.87, 1.01)
Vladimir Putin, Russia, 2012	0.97	(0.89, 1.06)
Conservative Party, Canada, 2011	0.07	(0.02, 0.12)
Simulated elections		
0% contamination	0.01	(0.00, 0.05)
1% contamination	1.03	(0.95, 1.05)
2% contamination	2.03	(2.03, 2.09)
3% contamination	2.97	(2.94, 3.06)
4% contamination	4.01	(3.99, 4.06)
5% contamination	5.03	(5.03, 5.07)

Source: Rozenas[7]

Unlike previous tests, which set theoretical expectations regarding outcomes in fair elections, the method developed by Rozenas[7] reports the percentage of polling stations with suspicious votes. The lower this value, the greater the integrity of the vote counting system. In this paper, we adopted the parameter of 0.07, a level observed in Canadian elections that, to the best of our knowledge, were not considered suspicious. Figure 8 illustrates the results for the three most voted candidates during the first round of the 2018 Brazilian presidential elections.
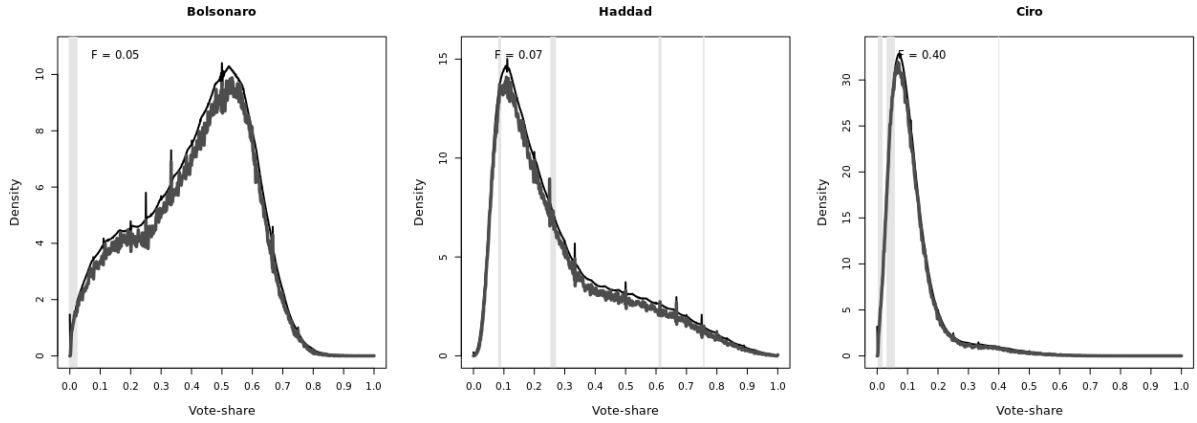
Figure 8. Kernel densities of observed vote-shares in the first round of the presidential election using pooling station data
Notes: Observed vote-shares are represented in bottom curves and their resampled upper envelopes in upper curves. Vertical grey bars indicate potentially falsified results.

According to Rozenas[7], the spikes of probability mass peak at coarse vote-shares and the vertical grey bars indicate potentially falsified results. For Bolsonaro, the percentage of suspicious polling stations is 0.05% (0.03 – 0.08, 95% CI) – lower than that identified by Rozenas[15] for the Canadian elections. For Haddad, the percentage was 0.07% (0.03 – 0.08, 95% CI). Finally, the F index for Ciro Gomes was 0.4% (0.04 – 0.11, 95% CI). In the worst-case scenario, the contamination percentage does not even come close to 0.5% of polling stations. Figure 9 repeats this procedure for the second round of the presidential election.
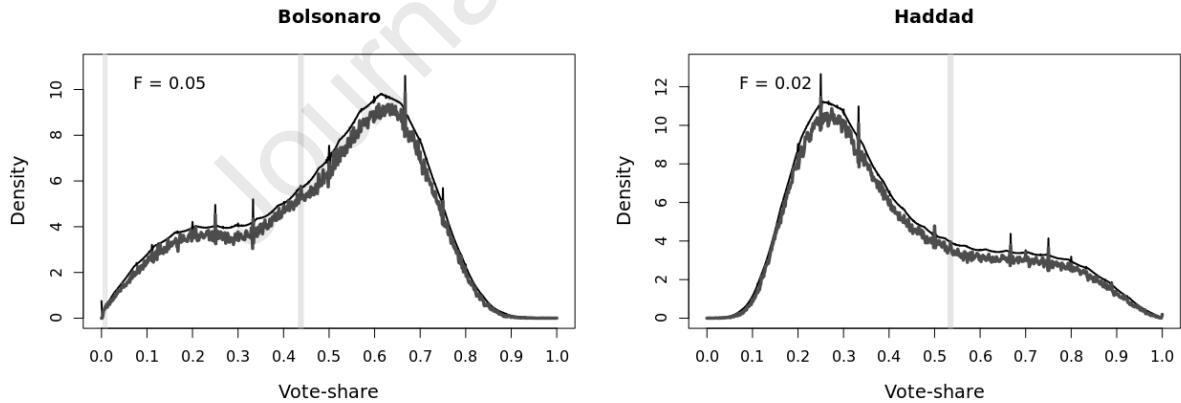

Figure 9. Kernel densities of observed vote-shares in the second round of the presidential election using pooling station data.
Notes: Observed vote-shares are represented in bottom curves and their resampled upper envelopes in upper curves. Vertical grey bars indicate potentially falsified results.

Considering the 454,119 polling stations that counted votes for Jair Bolsonaro (PSL) during the second round of the 2018 presidential elections, 0.05% (0.02 – 0.08, 95% CI) showed patterns considered suspicious according to the model proposed by Rozenas[7]. For Fernando Haddad (PT), this percentage was 0.02% (0.00 – 0.04, 95% CI).

13

DISCUSSION

If the ballot box is violated, universal suffrage no longer serves as a mechanism to ensure vertical accountability and representativeness within political power[13]. Therefore, illegal efforts to manipulate election results constitute a severe threat to democracy[2].

We find no evidence of irregularities in the 2018 Brazilian presidential elections. All parameters observed are within the theoretical expectation of fair vote counting. These inferences remain robust for the two rounds of the election and for multiple types of detection tools. Following the Beber and Scacco[6] approach of digit frequency, we were unable to detect any signs of abnormality. Considering the fingerprint analysis proposed by Klimek et al.[19], all estimated correlations were very close to zero, which is exactly the theoretical expectation in reliable vote counting processes. We did not identify any grouping of observations in the upper corner of the graph, which would be a sign of fraud. From the perspective of Rozenas[7], the estimate of fraud for Jair Bolsonaro and Fernando Haddad during the second round was lower than the values observed in the Canadian elections, which, to the best of our knowledge, were not placed under suspicion. Our findings are also supported by recent studies using longitudinal data from Brazil[26,27].

A limitation of our research design is the focus on only one election. Future research can investigate the extent to which vote counts in other periods and for different offices passes the unified framework we developed. Another promising perspective is to combine several electoral audit strategies and include political, social, and geographic variables. Given the complexity of the phenomenon, explanatory models should include both institutional and contextual factors on observed cases.

We are aware that the absence of anomalies does not constitute irrefutable proof that the accuracy of the vote counting was duly respected. As Mebane[4] has observed, nothing prevents the fraudster from distorting the data in conformity with the distribution of the second digit of Benford's Law (2BL), for example. We advance the literature of election forensics by developing a framework of joint application of multiple techniques. One thing is to manufacture data that conform to a single mathematical constraint; it is hugely more complex to forge a result that holds under different statistical assumptions. Ensuring that the farce remains undetected and capable of influencing the outcome of elections would require a level of organization so extreme that it could only be made possible by a widespread conspiracy.

Our findings advance the understanding of the integrity of the vote counting system in Brazil and can be helpful to inform the public debate about the main mechanism of accountability in democratic regimes: free and fair elections[28].

REFERENCES

1. Lehoucq, F. ELECTORAL FRAUD: Causes, Types, and Consequences. *Annual Review of Political Science* **6**, 233–256 (2003).

2. *Election Fraud: Detecting and Deterring Electoral Manipulation*. (Brookings Institution Press, 2008).

3. Deckert, J., Myagkov, M. & Ordeshook, P. C. Benford's Law and the Detection of Election Fraud. *Political Analysis* **19**, 245–268 (2011).

4. Mebane, W. Election forensics: the second-digit Benford's law test and recent American presidential elections. (2006).

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

6. Beber, B. & Scacco, A. What the Numbers Say: A Digit-Based Test for Election Fraud. *Political Analysis* **20**, 211–234 (2012).

7. Rozenas, A. Detecting Election Fraud from Irregularities in Vote-Share Distributions. *Political Analysis* **25**, 41–56 (2017).

8. VEJA. Bolsonaro afirma ter provas de que eleição foi fraudada. *VEJA* (2020).

9. Estado de Minas. Bolsonaro diz que só passa faixa presidencial com voto impresso. *Estado de Minas* (2021).

10. CNN Brasil. 'Se não tiver voto impresso, não terá eleição', diz Bolsonaro a ministro do TSE. *CNN Brasil* (2021).

11. Tarouco, G. da S. The Role of Political Parties in Electoral Governance: Delegation and the Quality of Elections in Latin America. *Election Law Journal: Rules, Politics, and Policy* **15**, 83–95 (2016).

12. Marchetti, V. Governança eleitoral: o modelo brasileiro de justiça eleitoral. *Dados* **51**, 865–893 (2008).

13. Hicken, A. & Mebane, W. A Guide to Election Forensics ∗. (2015).

14. Mebane, W. & Kalinin, K. Electoral Fraud in Russia: Vote Counts Analysis using Second-digit Mean Tests. (2010).

15. TSE. Repositório de Dados Eleitorais. (2022).

16. Pericchi, L. & Torres, D. Quick Anomaly Detection by the Newcomb—Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela. *Statistical Science* **26**, 502–516 (2011).

17. Dlugosz, S. & Müller-Funk, U. The value of the last digit: statistical fraud detection with digit analysis. *Adv Data Anal Classif* **3**, 281 (2009).

18. Skovoroda, R. & Lankina, T. Fabricating votes for Putin: new tests of fraud and electoral manipulations from Russia. *Post-Soviet Affairs* **33**, 100–123 (2017).

19. Klimek, P., Yegorov, Y., Hanel, R. & Thurner, S. Statistical Detection of Systematic Election Irregularities. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 16469–73 (2012).

20. El-Dash, N. A. Existe evidência de fraude nas eleições presidenciais brasileiras de 1994 a 2018? (2019).

21. Chandola, V., Banerjee, A. & Kumar, V. Outlier detection: A survey. *ACM Computing Surveys* **14**, 15 (2007).

22. Mack, V. & Stoetzer, L. F. Election fraud, digit tests and how humans fabricate vote counts - An experimental approach. *Electoral Studies* **58**, 31–47 (2019).

23. Rozenas, A. spikes: Detecting Election Fraud from Irregularities in Vote-Share Distributions. (2016).

24. Figueiredo Filho, D., Lins, R., Domingos, A., Janz, N. & Silva, L. Seven Reasons Why: A User's Guide to Transparency and Reproducibility. *Bras. Political Sci. Rev.* **13**, e0001 (2019).

25. Karp, J. A., Nai, A. & Norris, P. Dial 'F' for fraud: Explaining citizens suspicions about elections. *Electoral Studies* **53**, 11–19 (2018).

26. Eirado, C. R. da R., Silva, G. P. & Cordeiro, G. Análise da lei de benford de 2º dígito: Eleições Gerais para Presidente no Brasil e a Votação em Urnas Eletrônicas – 2002 a 2018. *R. Bras. Estat.* **78**, (2020).

27. Gamermann, D. & Antunes, F. L. Statistical analysis of Brazilian electoral campaigns via Benford's law. *Physica A: Statistical Mechanics and its Applications* **496**, 171–188 (2018).

28. Przeworski, A. *Why Bother with Elections?* (Polity Press, 2018).

Highlights

Brazilian President Jair Messias Bolsonaro has repeatedly raised suspicions about the integrity of the country's vote-counting system

These speeches cause enormous discomfort and a negative impact on the credibility of the electronic ballot box

Five electoral fraud detection tests are used to examine the latest presidential elections

The evidence presented here rules out the possibility of fraud in the 2018 presidential election in Brazil with a high confidence level

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: