Statistical Detection of Systematic Election Irregularities

Three Essays on Supervised and Unsupervised Machine Learning Approaches and the Attitudinal Consequences of Exposing Cheating

Lion Behrens

Mannheim 2023

Statistical Detection of Systematic Election Irregularities

Three Essays on Supervised and Unsupervised Machine Learning Approaches and the Attitudinal Consequences of Exposing Cheating

Inaugural dissertation
submitted in partial fulfillment of the requirements for the degree
Doctor of Social Sciences
in the Graduate School of Economic and Social Sciences
at the University of Mannheim

SUBMITTED BY
Lion Behrens

May 2023

Dean of the School of Social Sciences, University of Mannheim: Prof. Dr. Michael Diehl

First Supervisor: Prof. Thomas Gschwend, Ph.D. (University of Mannheim)
Second Supervisor: Prof. Dr. Richard Traunmüller (University of Mannheim)

First Examiner: Prof. Thomas Gschwend, Ph.D. (University of Mannheim)
Second Examiner: Prof. Dr. Richard Traunmüller (University of Mannheim)
Third Examiner: Prof. Dr. Thomas J. Bräuninger (University of Mannheim)

Date of the Oral Defense: August 28, 2023

To América and Aurora

Summary

This dissertation (i) develops statistical methodology for the detection of systematic irregularities in fine-graded election results (ii) and seeks to enhance our understanding of the attitudinal consequences of exposing individuals to information on electoral malpractice. While the range of methods for statistical fraud detection—a field often referred to as 'election forensics'—is constantly expanding, the same holds for the knowledge and strategies of micro- and macro-level agents of interference. Hence, this dissertation aims to contribute to the need of continuous innovation in the methodology of electoral anomaly detection. Without doubt, communicating negative findings about the integrity of electoral events is in itself likely to lead to a legitimacy loss of political institutions among the citizenry. This is why—next to making methodological contributions—this dissertation empirically investigates these decays in support, and hence also reflects on the role of statistical fraud detection in the tension between safeguarding democracy and producing democratic backlashes.

To locate my own contributions in the literature, I first motivate the need for the systematic study of numerical characteristics and statistical properties of (clean and fraudulent) voting returns from three exemplary recent national-level elections (*Chapter 1*). I line out the main challenges that the field of election forensic data analytics is faced with and demonstrate that approaches for statistical election fraud detection can be grouped into two distinct categories: Methods that exploit unique electoral circumstances which can be applied to very *particular* electoral events where these circumstances are given and methods that exploit distributional assumptions within voting returns that are present across heterogeneous electoral systems (but violated under manual interference) and hence can in principle be applied *globally*. I reflect on the main quality criteria that new statistical methods for fraud detection should possess. In discussing how these assets partially work against each other, I conclude that no single approach can fulfill all criteria and serve as a 'gold standard'. Rather, scholars are advised to consciously line out the advantages and shortcomings of developed approaches and communicate how new methods complement—rather than replace—existing methodology.

Building on this conceptual foundation, I introduce a novel approach for the statistical detection of irregularities that are indicative of election fraud (*Chapter 2*) exploiting the *specific* circumstance of concurrent electoral events. The method that I develop builds on the fact that in many countries, elections are not held as singular events. Rather, concurrent electoral contests which are administered side-by-side often take place simultaneously. I show that *undervoting irregularities*, which emerge if the same polling station documents different turnout levels across different electoral events, can be exploited for the detection of systematic irregularities if the extent of undervoting is related to the winner's vote share. I present a semi-parametric simulation model to estimate the share of polling stations with

undervoting irregularities at which vote shares were tainted. I apply this approach to a novel data set of recently contested Ecuadorian elections which report large extents of undervoting and simulated data for which the degree of fraud is known. I find that the proposed method reliably reverse-engineers true shares of fraud in synthetic data and that the empirical patterns which are inherent to Ecuadorian voting returns are well explained by systematic manipulation.

While the second chapter showcases a *novel empirical pattern* within fine-graded election results that has not received scholarly attention yet, the third chapter exploits supervised machine learning algorithms to present a *unified framework* for statistical election fraud detection based on existing forensic indicators (*Chapter 3*) that can be applied *globally*. I depart from the observation that the vast majority of global forensic tests that have been developed so far are respectively centered around one individual numerical characteristic within voting returns while being agnostic towards other features that have been successful in identifying fraud. As current indicators serve as standalone tests that don't inform each other, it is unclear how inconclusive results across different numerical attributes weigh into substantive conclusions. Subsequently, I present a data-generating protocol for simulating realistic micro-level training data which resemble data from empirical elections across a range of heterogeneous characteristics rather than one isolated pattern. I then train a supervised machine learning algorithm on synthetically generated data using a multivariate feature space, presenting a unified statistical framework for probabilistic election fraud detection that synthesizes multiple standalone tests with each other. Next to assessing laboratory performance in a simulation setting, I externally validate the proposed methodology on empirical data from Russia, Uganda and several Western European democracies.

Finally, I assess the attitudinal consequences of confronting individuals with credible information on systematic electoral manipulation (*Chapter 4*). In the past, scholars have shown that consciousness of election fraud lets individuals withdraw support from candidates, institutions and governments that are supposedly involved in manipulation. Together with my co-author Viktoriia Semenova, I argue that election fraud information will let individuals extrapolate legitimacy loss even to political institutions that are unrelated to electoral events and lead to decays of trust in the political system as a whole. Second, we argue that these spillovers are crucially shaped by the reactions of other political actors, as within-system corrections like court punishments of alleged fraud perpetrators can mitigate decays in diffuse support. We causally identify the main effect of fraud information and the moderating effect of political interventions from a pre-registered online survey experiment conducted in Colombia, Mexico and Russia. In addition, we present evidence from Bayesian matching estimators on cross-national survey data comprising over 48,000 individuals from 48 countries. We find that legitimacy loss of political institutions does indeed spill over to facets of the political system that are unrelated to electoral administration. Second, we show that political actors only have limited powers to mitigate citizens' alienation: Once information on fraud is shared, adequate punishment of fraud perpetrators may mitigate the negative effects of fraud information while this effect is far from being omnipresent across countries and institutions.

Taken together, this dissertation develops a novel statistical approach to detect systematic election irregularities in the presence of concurrent electoral events (*Chapter 2*), presents a unified framework for the joint evaluation of many statistical indicators using supervised machine learning (*Chapter 3*), and investigates the attitudinal consequences of exposing cheating (*Chapter 4*). While the methodological studies that I present are promising, the results from our survey experiment presented in *Chapter 4* cause this dissertation to close on a cautious note: Researchers, election observers and data scientists that are studying numerical anomalies in election results need to communicate statistical findings with care. As we learn from our experiment, the detrimental effect of credible election fraud information on political legitimacy loss is consistent, whereas the mitigating effect of political actors that step in and punish fraud perpetrators is not.

Replication Statement

Replication materials necessary to reproduce all results reported in this dissertation can be retrieved from the publicly available repositories on my Github account: `https://github.com/lion-be`.

Chapter 2 Behrens, Lion. "Detecting Unbalanced Election Fraud Approaches From Undervoting Irregularities".
Link: `https://github.com/Lion-Be/undervoting_irregularities`

Chapter 3 Behrens, Lion. "Quantifying Systematic Election Irregularities Using Supervised Machine Learning Algorithms".
Link: `https://github.com/Lion-Be/ml_detect`

Chapter 4 Behrens, Lion and Viktoriia Semenova. "Election Fraud Information, Punishment, and Political Trust: Evidence from a Survey Experiment in Colombia, Mexico, and Russia".
Link: `https://github.com/Lion-Be/fraud_spillover`

Acknowledgements

A dissertation project usually is the product of several—at times challenging—years throughout which I received a tremendous amount of support from a whole range of great people that have accompanied me through 2018-2023. Hereby, I would like to show my appreciation and gratitude.

First, I would like to thank my supervisor Thomas Gschwend for bringing me to Mannheim and providing me with continuous support throughout the years. No matter which idea I came with to you, you always encouraged me to follow through, challenged me on those aspects that were half-baked, and didn't stop demanding to finally "send it out". Even apart from my academic endeavors, you always had an open door and often pro-actively approached me to check on whether I am still doing well juggling my projects. Compiling a doctoral dissertation next to a demanding job as a research and teaching associate while in parallel founding a family with two beautiful children was not easy. I have gotten to know several peers along the way who where in similar situations and decided to quit their academic endeavor. You are the supervisor that I only could have hoped for and you are a major reason why I am able to present this dissertation now without having gotten lost along the way. I am grateful for your support and will be so in the future.

Second, I would like to thank the Collaborative Research Center 884 for generously funding much of my academic work and everyone associated to the Graduate School of Economic and Social Sciences for providing me with such an amazing environment for learning about political science, statistics and machine learning. I loved going to Grad School, and especially Rainer Gemulla, Paul Bauer, Richard Traunmüller, Sean Carey, Federico Nanni, Christoph Kern, Harald Schoen and Thomas Bräuninger have greatly contributed to this experience. Throughout the data collection for my dissertation projects, I was generously hosted at the Latin American Social Science Institute *FLACSO* by Santiago Basabe. You have provided me with access to your address book of valuable political figures in Ecuador's capital, which turned out to be key for the development of Chapter 2 of this dissertation.

I owe special thanks to my (former) colleagues at the Chair of Quantitative Methods, Marcel Neunhoeffer, Sebastian Sternberg, Benjamin Engst, Anna Adendorf, Oke Bahnsen, Sean Carey, Marie-Lou Sohnius, David Grundmanns, Oliver Rittmann, Domantas Undzėnas and Hannah Rajski. You have regularly inspired me and I routed for (and am routing) for you the same way I keep my fingers crossed for my own projects. You are doing great things and I hope you are having all the success that you wish for.

Several friends have made my almost five years at Mannheim University exceptionally special—yes, I am talking to you CDSS Cohort of 2018. Christoph, Viktoriia, Sascha, Hannah, Michael, Simon and David, you have proven as real friends and I so much enjoyed our Mensa lunches, coffee breaks, pub nights and weekend activities. I am sorry that as the years proceeded, my participation faded, but I guess

that's what happens with two little kids. A special shoutout, of course, goes to Viktoriia. Studying through geometric representations of OLS solutions, statistical regularization or different approaches to cross-validation were among the best Friday nights I've had in years. Also, you always lend me an ear when I needed to talk and have proved to be unstoppable in providing everyone with birthday and farewell presents that was in our orbit—something that we all appreciated. I am sure that life will reward you.

Lastly, my family have been those people that—up until today—motivate me to try to be the best version of myself, and also just let me be me. Mama and Papa, thank you for believing in me and supporting me for 28 years now. It's a great privilege that you let me walk my way and trust in the routes that I'm taking—this sets you apart from many others and I deeply acknowledge that. Hannah, thank you a lot for always standing by my side and for regularly pathing out ways that I can walk on. Things are so much easier with a bigger sister and I so much appreciate that you are there. Daniela, thank you for everything that you are doing for me, for you, and for our kids. I am so grateful to have you on my side and in many ways you truly are my "better half". When I wrote my B.A. thesis in 2016, I stated that "without you, I would have not taken the majority of those steps in life that I now feel proud of". Seven years later, this holds more true than ever. Thank you for 10 wonderful years and thank you for giving birth to our wonderful children. América Paola and Aurora Clara, I love you more than anything and you motivate me every day to continue. It is wonderful to see you grow and just as I have received so much help and support, I will support you. This thesis is dedicated to the two of you. And yes, América, I am now finally done with my dissertation, you can stop asking.

Contents

Summary **v**

Replication Statement **viii**

Acknowledgements **ix**

1 Introduction **1**
1.1 Motivating Examples 1
1.1.1 Bolivia 2019 1
1.1.2 United States 2020 5
1.1.3 Russia 2011 6
1.2 Research Questions 6
1.3 Electoral Integrity: Concepts and Definitions 8
1.3.1 Conceptualizations of Electoral Integrity 9
1.3.2 Electoral Integrity and the Statistical Detection of Election Fraud 13
1.4 Statistical Methods for Detecting Election fraud: A Primer 14
1.4.1 General Strategies 14
1.4.2 Examples of Statistical Approaches 16
1.4.3 Combination With Qualitative Investigation 21
1.4.4 Qualities of Individual Approaches 21
1.5 Plan of the Dissertation 22
1.6 Key Innovations and Contributions 24
1.6.1 Conceptual Contributions 24
1.6.2 Methodological Contributions 25
1.6.3 Empirical Contributions 26

2 Detecting Unbalanced Election Fraud Approaches 29
2.1 Introduction . . . 30
2.2 A Motivating Example . . . 31
2.2.1 Electoral History in Ecuador . . . 31
2.2.2 Undervoting Irregularities . . . 33
2.3 Election Fraud in the Presence of Concurrent Electoral Events . . . 36
2.3.1 Balanced and Unbalanced Election Fraud . . . 36
2.3.2 Undervoting Irregularities in the Absence of Unbalanced Fraud 38
2.4 A Simulation Model to Detect Unbalanced Election Fraud . . . 41
2.4.1 Stochastic Process of Concurrent Elections . . . 41
2.4.2 A Semi-Parametric Simulation Model . . . 43
2.5 Applications . . . 45
2.5.1 Alternative Explanations . . . 50
2.6 Conclusion . . . 53

3 Quantifying Systematic Election Irregularities Using Supervised Learning 55
3.1 Introduction . . . 56
3.2 Motivation for a Unified Statistical Framework . . . 58
3.2.1 Numerical Characteristics in Electoral Returns . . . 58
3.2.2 Motivation for Multivariate Learning from Synthetic Data . . . 63
3.3 Synthetic Data Generation . . . 65
3.3.1 Data Generating Methodology . . . 65
3.3.2 Comparison to Empirical Data . . . 70
3.4 Estimating the Percentage of Fraudulent Votes . . . 71
3.4.1 Synthetic Training Data for One Country Case . . . 71
3.4.2 Outcome Variable and Multivariate Feature Space . . . 72
3.4.3 Choice of Machine Learning Algorithm . . . 72
3.5 Applications . . . 73
3.5.1 Monte Carlo Simulation Study . . . 73
3.5.2 External Validation . . . 75
3.6 Conclusion . . . 78

4 Election Fraud Information, Punishment, and Political Trust 81
4.1 Introduction . . . 82
4.2 Election Fraud Information, Punishment, and Political Trust . . . 84
4.2.1 Election Fraud Information and Attitude Extrapolation . . . 85
4.2.2 Previous Literature . . . 87
4.2.3 Electoral Crimes and Punishment . . . 88
4.3 Matching Estimates from Cross-Sectional Survey Data . . . 91
4.4 Survey Experiment Design . . . 94
4.4.1 Case Selection . . . 94
4.4.2 Survey Design and Implementation . . . 94
4.4.3 Causal Identification . . . 95

4.5 Experimental Results 98
4.5.1 Statistical Modeling and Estimation 98
4.5.2 Empirical Results 99
4.6 Conclusion 103

5 Conclusion **105**
5.1 Research Questions and Answers 105
5.2 Implications and Avenues for Future Research 107
5.3 Concluding Thoughts 109

A Appendix for Chapter 2 **111**
A.1 The Distribution of Undervoting Irregularities 111
A.2 Prior Distributions for Parameter Estimation 112
A.3 R Code for Executing the Semi-Parametric Simulation Model 112
A.4 The Urban-Rural Divide in Undervoting Irregularities 114

B Appendix for Chapter 3 **117**

C Appendix for Chapter 4 **119**
C.1 Cross-National Evidence: Matching Analysis 119
C.1.1 Descriptive Statistics 119
C.2 Details on Case Selection 122
C.3 Survey Experiment 123
C.3.1 Descriptive Statistics 123
C.3.2 Data Quality 126
C.3.3 Raw Regression Results 130

Bibliography **149**

List of Figures

1.1 Discontinuity jump in cumulative MAS vote share, Bolivia 2019. . . . 3
1.2 Discontinuity jump in cumulative Haddad vote share for comparison, Brazil 2018. 4
1.3 Conceptualization of electoral integrity, funnel plot. 12
1.4 Digit distributions in raw vote totals against Newcomb-Beonford's law. 17
1.5 Standardized distribution of turnout shares across electoral units against a standard normal distribution. 18
1.6 Bivariate distribution between turnout shares and vote shares for the Spanish Socialist Workers' Party (PSOE) in Spain 2019 and for Vladimir Putin in Russia 2011. 19
1.7 Estimated density (Gaussian kernel) of the winning party's (or candidate's) vote share, Finland 2017 and Russia 2012 20

2.1 Irregularities on vote tallies, Ecuadorian Local Elections 2019 34
2.2 Undervoting irregularities across Ecuadorian elections 35
2.3 The relationship between undervoting irregularities and winner's vote shares (simulated) 40
2.4 The relationship between undervoting irregularities and winner's vote shares, General Elections 2017 47
2.5 The relationship between undervoting irregularities and winner's vote shares, Local Elections 2019 48
2.6 Undervoting irregularities and their assocation to Lenin Moreno's vote share, Presidential Election 2017. 50
2.7 The relationship between undervoting irregularities and winner's vote shares, General Elections 2017 52
2.8 The urban-rural divide in undervoting irregularities, General Elections 2017-A. 53

2.9 The urban-rural divide in undervoting irregularities, General Elections 2017-B. 54

3.1 Second significant digit distributions for winner's votes, cross-country comparison 60
3.2 Estimated density (Gaussian kernel) of the winning party's (or candidate's) vote share 61
3.3 Comparison of empirical and artificially manipulated data from Austria 2008, vote switching 63
3.4 Comparison of empirical and artificially manipulated data from Austria 2008, rounding fraud 64
3.5 Comparison of the distribution of winner's vote share values across all electoral units between empirical and synthetic data 68
3.6 Three-dimensional histograms of the number of electoral units with a given turnout and winner's vote share percentage, Western Democracies. 69
3.7 Three-dimensional histograms of the number of electoral units with a given turnout and winner's vote share percentage, Autocracies. 70
3.8 Estimated density (Gaussian kernel) of the winning party's (or candidate's) vote share 71
3.9 Monte Carlo simulation results on synthetic data with no data manipulation incorporated across different numbers of electoral units. . . . 74
3.10 Monte Carlo simulation results on synthetic data with different types and degrees of data manipulation incorporated across different numbers of electoral units. 75
3.11 Variable importance measures for the fifteen most important features in the Random Forest algorithm when predicting the percentage of votes that is subjected to manipulation. 77

4.1 Average treatment effects of election integrity perceptions on trust in political and non-political institutions from matching analyses on WVS Wave 7 (2017-2020) data 93
4.2 The effects of exposure to fraud information on confidence in political institutions 100
4.3 The effects of exposure to fraud information on confidence in non-political institutions 101
4.4 The effects of exposure to perpetrators' judicial punishment information on confidence in political institutions 102

A.1 The distribution of undervoting irregularities in Ecuadorian elections 111
A.2 The urban-rural divide in undervoting irregularities, General Elections 2017 114

A.3 The urban-rural divide in undervoting irregularities, General Elections 2017 . . . 115

C.1 Covariate balance before and after sample adjustment using direct exact matching . . . 121
C.2 Covariate balance before and after sample adjustment using coarsened exact matching . . . 121
C.3 The effects of exposure to fraud information on confidence in political institutions, only complete summaries of treatment texts . . . 137
C.4 The effect of exposure to fraud and punishment information on confidence in political institutions across countries . . . 138
C.5 The effects of exposure to fraud information on confidence in political institutions across countries, opponents and supporters . . . 139
C.6 The results of mediation analysis . . . 142

List of Tables

1.1 A sample of voting returns from Ecuador's 2017 presidential election. Documented are results for the two best performing candidates across ten voting tables of the canton Ibarra. Females and males place their ballots into separate ballot boxes. 15

2.1 Estimates of unbalanced election fraud, Ecuadorian Elections 2017 and 2019 49
2.2 The relation between undervoting irregularities and winner's vote shares, General Elections 2017 51

3.1 Parametrizations for synthetic data generation. 66
3.2 Application of supervised machine learning trained on synthetic data for six empirical elections. 76

4.1 Survey experiment, overview of experimental conditions 96

A.1 Descriptive statistics of undervoting irregularities for urban polling stations, General Elections 2017. 115
A.2 Descriptive statistics of undervoting irregularities for rural polling stations, General Elections 2017. 115

B.1 Descriptive statistics on electoral returns 117
B.2 Frequencies of digits as expected by Newcomb-Benford's law 117
B.3 List of feature variables and their abbreviations that are used to predict the percentage of votes that are subject to manipulation. 118

C.1 Summary statistics of key variables, World Values Survey Wave 7 (2017-2020) 119

C.2 Summary statistics of key variables, experimental survey data for Latin America . . . 124
C.3 Summary statistics of key variables, experimental survey data for Russia . . . 125
C.4 Dropout rates right after reading the treatment across experimental conditions, experimental survey data in Latin America . . . 126
C.5 Dropout rates right after reading the treatment across experimental conditions, experimental survey data in Russia . . . 127
C.6 Time required for completion in seconds (finished interviews), Experimental Survey Data for Latin America . . . 128
C.7 Time required for completion in seconds (finished interviews), Experimental Survey Data for Russia . . . 128
C.9 Data quality categories based on manual coding for the Russian sample 129
C.8 Data quality categories based on manual coding for the Latin American sample . . . 129
C.10 Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample . . . 130
C.11 Ordinal logistic regression estimates of treatment effects for Russian sample . . . 131
C.12 Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample . . . 132
C.13 Ordinal logistic regression estimates of treatment effects for the Russian sample . . . 133
C.14 Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample . . . 135
C.15 Ordinal logistic regression estimates of treatment effects for the Russian sample . . . 136
C.16 Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample . . . 144
C.17 Ordinal logistic regression estimates of treatment effects for the Russian sample . . . 146

List of Abbreviations

1BL First-Digit Benford's Law
2BL Second-Digit Benford's Law
AFP Agence France Presse
ANF Agencia de Noticias Fides
CNE Consejo Nacional Electoral *(engl. National Electoral Commission)*
EMB Electoral Management Body
OAS Organization of American States
TSE Tribunal Supremo Electoral *(engl. Supreme Electoral Tribunal)*

1
Introduction

"It's not the people who vote that count, it's the people who count that win the vote."

attributed to Joseph Stalin

1.1 Motivating Examples

1.1.1 Bolivia 2019

On October 20, 2019, Bolivians were called to the polls to participate in what turned out to be one of the most controversial elections the country had witnessed in its electoral history. In the presidential contest, the 2005-elected incumbent president Evo Morales faced eight challengers, and an opposition that was increasingly gaining strength against the Morales-led 'pink tide' which had grown the country to be a close ally of Venezuela's 'socialism of the 21st century'. The candidacy of Morales was itself disputed. Morales was eligible for re-election only after Bolivia's highest court had overruled (i) the country's 2009 constitution imposing a two-term limit on the presidency (ii) and the 2016 referendum in which 51% of voters rejected the abandonment of the country's presidential term limits. As the court argued, term limits in general violate the American Convention of Human Rights (Anria and Cyr 2019) and that as a result, the president cannot be prevented from running as a candidate.

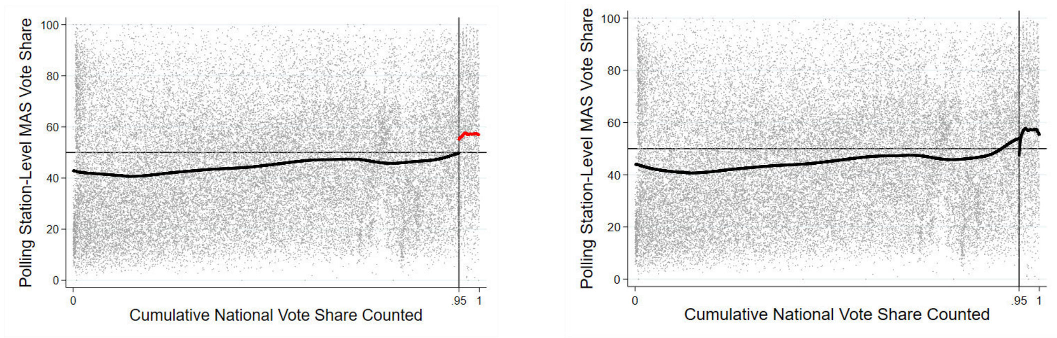
Morales led in the pre-election public opinion polls but needed a 10 percent margin over the second place candidate in order to avoid a run-off election which likely would have seriously threatened his subsequent presidency (ANF 2019). After the polls closed at 7:00 p.m. on election day, the *Tribunal Supremo Electoral* (TSE) started to post online preliminary results which were updated on a running basis. At about 7:45 p.m., the TSE announced preliminary results with 83.85 percent of votes processed. These had Morales in the lead with a margin of 7.87 percentage points—below the 10 point margin needed to secure a first round win (Johnston and Rosnick 2020; Idrobo, Kronick, and Rodríguez 2022). At that point, the online transmission of results stopped. The reason for this freeze remains disputed and Johnston and Rosnick (2020) asserted that the TSE provided no clear public explanation why the transmission of results had been put to an end unexpectedly. On the one hand, some political observers lamented a severe 'lack of expertise´ within the operating team of the TSE and attributed the freeze of the reporting system simply to an enormous technical mismanagement at the side of the TSE (Ferrufino and Cesar 2019). Contrary, critics saw the halting of results as a deliberate action to facilitate in-house tampering with the electoral material. The government itself later proclaimed that they had never intended to mirror 100% votes to the preliminary online system in the first place (Los Tiempos 2019). When online transmission continued the day after the election at 6:30 p.m. with 94.94 percent of votes processed, the incumbent president's margin had grown to 10.15 percentage points—right above the threshold to avoid a standoff, and was carried through up until the proclamation of final results.[1]

As a consequence of the transmission system shutdown and Morales's turn around, opposition leaders accused the TSE of having administered centralized fraud during the last stages of the counting process to avoid facing a united opposition in a runoff election (AFP 2019), election observation missions voiced 'deep concerns' (OAS 2019b), and Bolivia 'exploded in protest' (Kurmanaev, Anatoly and Cesar Del Castillo 2019). Under intense public pressure, on October 22, the Bolivian government requested the Organization of American States (OAS), one of the most reputable electoral observer missions globally, to officially evaluate the vote tabulation process. The OAS released their preliminary report in the morning of November 10. In an alarming and compromising report, the OAS alleged 'deliberate actions that sought to manipulate the results of the election' (OAS 2019a, p. 4). Next to the accusation of an "intentional and arbitrary freezing with no technical basis" of the transmission system, the OAS report alleged statistical patterns that the authors were unable to explain without evoking electoral fraud.[2]

First, the OAS outlined that when plotting the development of the cumulative national vote share of Morales across time during the counting stage, a sharp divergence from the previous trend can be observed around the 84% mark (OAS 2019a, p. 87), right at which the TSE announced the preliminary results and froze the transmission system. Most importantly, the OAS observed a striking discontinuity appearing at an "arbitrary point", the threshold of 95% counted votes (see Figure 1.1a), which they interpreted as a "massive and inexplicable increase in the number of votes for MAS in the final 5% of the votes counted". Overall, the OAS outlined that the last portion of the vote count is "sharply different than the trend just on the

[1] In the official results as announced on October 25, Morales's vote share of 47.08% of votes was tabulated 10.57 percentage points above the runner up vote share of 36.51%.

[2] The original report can still be accessed under `https://www.oas.org/fpdb/press/Audit-Report-EN-vFINAL.pdf`, last accessed April 26, 2023.



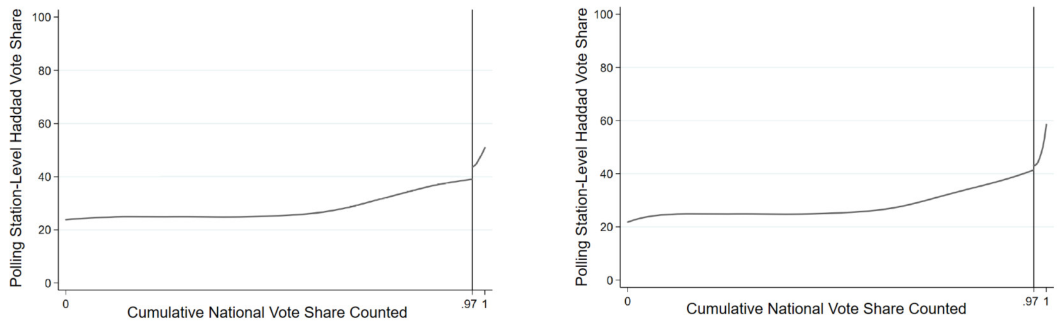
(a) **(b)**

Figure 1.1. Discontinuity jump in cumulative MAS vote share, Bolivia 2019. Alleged discontinuity jump in MAS vote share after 95% votes have been counted and its replication using data from $N = 33,038$ tally sheets from the preliminary results transmission system. Points represent the underlying raw results as reported by each polling station. *Panel A.* The original figure from OAS (2019a), p. 88. Local constant regression (polynomial of order zero) with handpicked bandwidths fit at each data point. *Panel B.* Local linear regression (polynomial of order one) implemented by Idrobo, Kronick, and Rodríguez (2022), all else left equal.

other side of the threshold" (p. 89) and recommended new elections under a newly elected TSE. In the evening of the same day, a conjunction of Bolivia's military chief and chief of the police pressured Morales to resign. During the night to November 11, within 24 hours after the OAS report had been released, Morales stepped down from office and fled to seek political asylum in Mexico.

Months later, the ad-hoc statistical analysis of the OAS had sparked the attention of several academic scholars and a whole range of articles had been developed that have put the OAS allegations under closer scrutiny (see Escobari and Hoover 2019; Johnston and Rosnick 2020; Newman 2020; Williams and Curiel 2020; Idrobo, Kronick, and Rodríguez 2022). As Idrobo, Kronick, and Rodríguez (2022) outline, little of the initial conclusions can be upheld after a deeper dive into the data.

- First, as micro-level election staff is selected at random from all voters registered at each polling booth and as higher educated staff can plausibly be expected to send in their results earlier than lower educated staff, localities with a higher share of educated voters are expected to send in their results earlier than localities with lower shares of these. As it is well known that in Bolivia, education is negatively correlated with support for Morales (Madrid 2012), votes from anti-Morales polling stations are hence systematically counted earlier than votes from Morales strongholds. Consistent with this mechanism being a function of simple demographic differences of election personnel rather than systematic fraud, the trend of late-counted votes after the 84% mark systematically over-favoring Morales flattens considerably even when applying rather crude controls for education, region and rurality.



(a) **(b)**

Figure 1.2. Discontinuity jump in cumulative Haddad vote share for comparison, Brazil 2018. Using local constant regression creates artifical discontinuity jumps around arbitrary values even in elections that the OAS monitored and explicitly endorsed. *Panel A.* Original figure from Idrobo, Kronick, and Rodríguez (2020), p. 46. Local constant regression (polynomial of order zero) with bandwidth from Fan and Gijbels (1996), p. 110-113. *Panel B.* Original figure from Idrobo, Kronick, and Rodríguez (2020), p. 46. Local linear regression (polynomial of order one) with bandwidth from Fan and Gijbels (1996), p. 110-113, all else left equal.

- Second, when re-examining the discontinuity claim presented in Figure 1.1a, Idrobo, Kronick, and Rodríguez (2022) show that, surprisingly, there simply "is no jump in vote share" at the cutoff reported by the OAS. In the original analysis, the OAS analysts used an estimator that is simply inappropriate for identifying regression discontinuities. The OAS created the pattern in Figure 1a by fitting two local constant regressions (using a polynomial of order zero) at each data point left and right to the 95% threshold and connecting the predicted values. However, this approach of local constant regression misrepresents the data at boundary points, in this case, at the arbitrary 95% threshold. As Cattaneo, Idrobo, and Titiunik (2020) outline, this constant fit has "undesirable properties at boundary points, where is precisely where regression discontuinty estimation must occur", an issue of local constant regression that is well known and profoundly discussed in the most popular textbooks on statistical learning such as *The Elements of Statistical Learning* (Hastie, Tibshirani, and Friedman 2009, Chapter 5). By accounting for this fallacy using local *linear* regression (see Figure 1.1b) and more sophisticated approaches (see Idrobo, Kronick, and Rodríguez 2022), the alleged discontinuity jump simply disappears.

- Third, in a previous version of their article, Idrobo, Kronick, and Rodríguez (2020) have analyzed shifts in late-counted votes of earlier elections that the OAS monitored and explicitly endorsed. When considering late-counted votes in Brazil's 2018 presidential election, the exact same appearance of spurious discontinuity jumps can be replicated using local constant regression if enough threshold values are tried out, and disappear using more appropriate higher degree polynomials (see Figures 1.2a and 1.2b). For the Brazilian election of 2018, the OAS did not interpret these statistical patterns as an indication of electoral tampering.

In total, the re-analyses of the OAS data dramatically weakened the statistical evidence that initially was presented as evidence for electoral tampering. While the lead-up to the election was legally disputed and while there do persist several reported issues regarding the technical set up and freeze of the online results transmission system, the main statistical irregularity that allegedly was in place—and that played an important role in overthrowing the, albeit controversial, incumbent government—turned out to be unwarranted after academic scholars had performed an in-depth scrutiny of the ad-hoc analyses.

1.1.2 United States 2020

In the United States presidential election of 2020, narratives of fraud were prevalent long before election day. Under the slogan "Stop the Steal!", the Republican candidate and incumbent president Donald J. Trump had managed to rally thousands of supporters well before election day to protest against what was perceived to be an allegedly fraudulent upcoming election. Similar to the patterns in Bolivia 2019, late-counted votes heavily shifted overall vote shares, but were well understood to stem from Democratic voters predominantly using postal votes (Curriel, III, and Williams 2021; Foley and III 2020; Li, Hyun, and Alvarez 2022), a pattern that had been predicted long before voting started.

After the election results came in, a different pattern sparked Republican observers' attention. A range of conservative news outlets observed that while first digits of Trump's county-level votes in Pennsylvania—historically one of the most competitive Swing states—follow a probability distribution called *Benford's Law*, the county-level votes of his competitor Joe Biden do not, and that this could be taken as evidence for the election being stolen[3]. A range of academic scholars stepped in, clarifying that given the peculiarities of the US electoral system, *first* digits of US county-level electoral results cannot be expected to follow the Benford's Law distribution in the first place, as these do not satisfy any of the conditions which have been laid out in the formal derivation of the law (Hill 1995; Hill 1996; Mebane 2020; Reuters 2020)[4]. In this particular example, scholars of election forensics clarified several misleading claims that spread across social media and conservative news channels.

[3]For instance, see `https://chance.amstat.org/2022/04/benfords-law-votes/`.

[4]See the discussion of the distribution of numbers in different digit positions of raw vote counts in Chapter 3.2.1 of this dissertation.

1.1.3 Russia 2011

In Russia, the parliamentary election of the new Russian state Duma held in December 2011 spiked mass protests with thousands of people gathering throughout Moscow (Schwirtz and Herszenhorn 2011) and St. Petersburg (The Guardian 2011) after a series of videos allegedly documenting malpractice by election officials and poll workers had started to spread on election day. Several analysts noted a compelling prevalence of exactly coarse turnout and vote shares that are multiples of five and especially appear in those polling stations where the governing party United Russia received overwhelmingly high support (0.6, 0.65, 0.7, 0.75, 0.8). In the immediate aftermath of the election, these spikes in the distribution around turnout and vote shares were deemed so improbable that they should rise suspicions that parts of the results were fabricated to meet certain "target values" (Mebane 2013; Gehlbach 2012).

After several years of academic scrutiny, the 2011 Russian election hat led to the development of two systematic approaches to systematically assess whether the number of coarse turnout and vote shares exceeds their statistically expected bounds (Kobak, Shpilkin, and Pshenichnikov 2016a; Rozenas 2017). While designing their methods slightly differently, both contributions concluded that the number of coarse vote shares drastically exceeded the frequency that can be expected giving the polling station design in the Russian electoral system, providing robust statistical evidence in favor of systematic data fabrication.

1.2 Research Questions

These three examples of contested elections show that in today's data-driven environment, public debates about the integrity of electoral events increasingly center around numerical features of published voting results that are perceived to be anomalous and discrepant to legitimate data. In general, statistical methods for anomaly detection are employed across a diverse range of domains.[5] Each of these domains is driven by the underlying assumption that natural (legitimate) data generating processes will manifest in an array of numerical and distributional characteristics (which take context-specific forms in each substantive domain) that are in place under clean procedures but violated when *systematic* illegitimate practices are present. In the greater scheme of developing statistical tools for numerical anomaly detection, the field of election forensics mirrors this baseline assumption to the particular use case of disaggregated electoral results which are available across a larger number of entities (precincts, polling stations).

[5]For an early overview, see Bolton and Hand (2002) spanning the identification of financial malpractice such as credit card fraud (Awoyemi, Adetunmbi, and Oluwadare 2017) or money laundering (Jullum et al. 2020; Sudjianto et al. 2010) from attributes of financial transaction data, telecommunications fraud obtaining services without having the necessary authority (Hilas and Sahalos 2005) or insurance fraud generating money from illegitimate medical claims (Li et al. 2008).

This dissertation departs from two key observations: The first observation is that there is a continuous need for methodological development in the field of electoral anomaly detection. First of all, this is because as our statistical tools develop, so do the strategies of micro- and macro-level agents of fraud yielding ever-new pictures and patterns of systematic interference. Second, this is because statistical analyses need to safeguard clean elections from unwarranted accusations of malpractice which are either strategically developed as an act to delegitimize the political opponent or voiced out of legitimate concern about seemingly spurious practices around numerical patterns in electoral results which turn out to be artefacts after closer investigations.

The second observation is that communicating negative findings about electoral integrity *in itself* is at risk of producing democratic backlashes. Distributing credible information about electoral malpractice—irrespective of these accusations being actually justified or not—will inevitably lead to attitudinal responses. Individuals who are exposed to this information are likely to detach from their political leaders and parties, the institutions that are related to electoral administration or the political system that electoral events are embedded in *as a whole*, while the scope of the loss in legitimacy and the factors that moderate this detachment are unclear.

In this dissertation, I speak to these two observations. I aim to both generate new insights on the methodological study of numerical anomalies as well as the attitudinal consequences of exposing cheating. In essence, the chapters of this dissertation link the two observations that I depart from to two central research questions that guide my endeavor and will be investigated (i) from different angles (ii) using diverse data structures and (iii) heterogeneous empirical approaches. The first research question relates to the fact that while strong numerical regularities can be found over and over again in fine-graded electoral data from substantially different electoral systems (see Pericchi and Torres 2011; Klimek et al. 2012), the idiosyncrasies that govern specific electoral designs are too specific to treat statistical regularities as 'universal laws' that are expected to be in place across all possible cases (Deckert, Myagkov, and Ordeshook 2011). Rather, the methodologies that are developed in this dissertation take a different path. The central underlying assumption is that context-specific characteristics of scrutinized elections crucially shape which statistical regularities can and cannot be observed under clean electoral processes and should drive the development of probabilistic tools to identify their violations under fraud.[6] Hence, the first research question that governs the

[6] For instance, compare the district-level data from Argentina 1931-1941 presented by Cantú and Saiegh (2011) and the county-level data from the US 2020 presented by Mebane (2020). In the first case, the frequency of numerals $d \in \{0, 1, \ldots, 9\}$ in the first digit of raw vote totals *can* (Argentina 1931-1941) be expected to decay logarithmically at a constant rate as stated by *Benford's law*. Due to the data aggregation that is taking place, first digits can be thought as the result of a statistical mixture distribution and span several orders of magnitude, data properties for which the law has been formally (and empirically) shown to hold (Hill (1995) and Hill (1996). In the latter case (US 2020), first digits of raw vote totals *cannot* be expected to decay logarithmically at a constant rate, as in a two-party system with relatively equal support rates across counties, numbers are artificially bounded, as also discussed in Chapter 3.2.1 of this dissertation.

first part of this dissertation can be formulated as:

> *Research Question 1: How can context-specific characteristics of electoral events be exploited for the statistical detection of systematic election irregularities?*

While with *Chapter 2* and *Chapter 3*, two building blocks of this dissertation are devoted to this question, the approaches with which context-specific characteristics are incorporated into the fraud detection prototypes that are developed are vastly different. In *Chapter 2*, an unsupervised semi-parametric simulation model is developed that is constructed around a numerical feature within electoral returns that can only emerge in the context of concurrent electoral contests which are administered side-by-side at the same localities—a contextual characteristic of *general elections* that is not universally applicable but comes with exceptional opportunities for the identification of systematic irregularities if concurrent elections are taking place. In *Chapter 3*, supervised machine learning methods are trained on numerical features of synthetic data that is simulated given the unique electoral characteristics of each empirical case at hand, incorporating the specificities of individual cases into what the supervised learning model believes to be clean or fraudulent numerical attributes rather than following 'universal laws' and testing against pre-defined null distributions.

The second research question speaks to the role of statistical fraud detection in the tension between *safeguarding democracy* through distributing credible fraud claims and potentially *producing democratic backlashes* by doing so. As several scholars have shown that credible information on systematic electoral malpractice updates citizens' attitudes on political matters (Williamson 2021; Robertson 2017; Reuter and Szakonyi 2021), the second research question asks:

> *Research Question 2: What are the attitudinal consequences of exposing cheating?*

In *Chapter 4*, I present a pre-registered survey experiment that is dedicated to several implications of this question. Together with my co-author Viktoriia Semenova, we line out how confronting individuals with credible fraud information leads to a decay in political trust even to political institutions that are unrelated to electoral administration which *cannot* be reliably restored even under within-system corrections from other political actors such as dismissal of electoral staff or court punishments of alleged perpetrators.

1.3 Electoral Integrity: Concepts and Definitions

Before I present the chapters of this dissertation that aim to contribute to the statistical detection of systematic election irregularities and our understanding of the attitudinal consequences of exposing cheating, it is crucial to clarify what I mean by electoral manipulation. Electoral malpractice is a multi-dimensional construct spanning various *periods* within the electoral cycle, different *actors* that perpetrate fraud, and manifold *objects* that manipulation can be targeted towards (Schedler 2002). Yet, while a range of scholars have provided different typologies and definitions of electoral malpractice, there is no consensus across the comparative literature about what (and what not) exactly constitutes good practice in the administration of elections and what counts as electoral wrongdoing (Vickery and Shein 2012; Schedler 2002). For instance, what about rigid voter registration rules which disproportionally affect different societal groups that are known to over-represent specific political preferences? How do we assess unequal campaign finances or specific incumbents' access to public and state-controlled media channels which generously report about the government platform in face of the ideal of free and fair campaigns (see Norris 2014)? This section will first map out different concepts of electoral integrity that have been elaborated by the comparative literature and will secondly provide a clear definition of electoral fraud that is used throughout this dissertation.

In short, the type of electoral tampering which the statistical methods that I outline here speak to corresponds to a minimalist definition of election fraud. Statistical tools for identifying systematic election irregularities relate to the intentional fabrication or alteration of vote totals on election day or during the counting stage by micro-level agents at the level of individual polling stations or election officials in vote tabulation centers.

1.3.1 Conceptualizations of Electoral Integrity

Political and electoral systems vary in great detail across the world, each coming with their own set of opportunities for political representation and challenges for holding electoral events. Since the 'electoral revolution' that surged since the mid-twentieth century led to a dramatic increase in the number of electoral events (Norris 2014), multiparty elections have become omnipresent across new democracies and electoral authoritarian regimes worldwide. Consequently, also electoral malpractice takes on a whole range of different forms that are too exhaustive to simply list here. However, it is reasonable to start from a general typology and subsequently identify the elements within this framework that the statistical detection of election fraud can speak to. The comparative literature has carved out at least three different approaches to systematically arrive at a definition of election integrity.

International norms and conventions
The first definitional approach defines electoral fraud as practices that violate internationally accepted norms and relates electoral integrity to internationally ratified standards and conventions. Norris (2014)) identifies at least two declarations of the United Nations that can serve as the foundation of how international conventions

guide international standards of electoral integrity. First of all, Article 21(3) of the Universal Declaration of Human Rights (1948) specifies

> *"the will of the people shall be the basis of the authority of government; this will shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures."*

Second, more specific agreements about the global norms that define the appropriate conduct of domestic elections can be found in Article 25 of the UN International Convent for Civil and Political Rights (1966), which defined the need for

- periodic elections in regular intervals
- universal suffrage that includes all sectors of society
- equal suffrage
- the right to stand for public office
- the right of all eligible electors to vote
- secrecy of ballots

(compare Norris 2014, p. 23-24.). The advantage of this approach is that it remains relatively broad, covering much of the aspects that are embedded into the electoral cycle and can form a normative framework for the execution of international electoral assistance, although international legal instruments are silent on many of the modern aspects of electoral conduct (Birch 2011, p. 12).

Additionally, reference to internationally set standards that are developed to hold globally often fail to consider culture- and country-specific habits and practices that are not easily defined as legitimate or not using international standards. For example, several countries selectively reserve a number of political positions to specific segments of society which distorts the "one person—one vote" principle. In Bolivia, where indigenous peoples are historically under-represented in elected bodies though constitute a relevant share of the society, the 2009 electoral law secures 7 out of 130 (5.4%) parliamentary seats for indigenous peoples of the lowlands (Barié 2022). In the United Kingdom, 26 out of the 757 seats within the House of Lords are reserved for Church of England bishops (Bown 1994). These practices provide the potential to openly create a systematic discrepancy between political representation as implied by factual voting results and final parliamentary seat shares. While in some countries, such practices are coined as illegitimate manipulation, in other countries, these regulations are deemed legitimate behavior to foster a balance of people between different societal groups. It will be hard, if not impossible, to define one international standard to grasp the range of different historical practices in ensuring political representation that are observed across the globe.

Sociological approach

A second approach lies in measuring election integrity through public mass surveys in order to grasp common *perceptions* towards what constitutes an actual violation of the electoral 'code of conduct'. This sociological—or "perceptual"—approach (see Birch 2011) carries some attractive properties as it (i) takes into account culture- and country-specific heritages and notions of which practices are within the realm of legitimate practices that an approach based on international norms is blind to (ii) and allows to focus on virtually any aspect of the conduct of elections that can be included in public opinion surveys that international agreements are silent about. Another intuitive advantage is that using surveys, levels of election integrity can easily be mapped across time within a country.

The obvious downside of such an approach is that here, election integrity easily becomes a fluctuating, almost arbitrary concept with little comparability across different political systems. Also, while blatant attempts to manipulate the voting process or outcomes might be jointly condemned, a lack of consensus might exist towards more intertwined topics around electoral conduct that are more nuanced.

Legalistic approach

Both conceptualizing election integrity via adherence to global norms in the form of international conventions or via cultural norms measured through mass surveys corresponds to a thick or 'inclusive' approach (Vickery and Shein 2012) to defining electoral integrity and identifying electoral malpractice. A thin—or 'restrictive'—definition of electoral malpractice defines all practices as illegitimate that explicitly violate existing domestic legal provisions. This law-based approach is attractive as it keeps the contextual flexibility of the sociological approach while at the same time being more exact than international conventions which often remain abstract and vague. For instance, some practices in private campaign finance might be legal in one country but warrant legal prosecution in a different country setting. With election management bodies (EMBs) embedded in countries' legal systems, the actors that define electoral wrongdoing and prosecute violations of the law are also clearly defined when using this legalistic notion of electoral integrity.

An obvious shortcoming of a law-based approach is that in emphasizing this "relative nature of election fraud" (Alvarez, Hall, and Hyde 2008, p. 9), it provides little resources for addressing manipulation in countries with weak legal provisions or where electoral laws are purposefully used to manipulate electoral outcomes. As Schedler (2002, p. 36) writes, "the modern history of representative elections is a tale of authoritarian manipulation as much as it is a saga of democratic triumphs". This "cultural, political or contextual relativism" (Vickery and Shein 2012. p. 6) is most visible in electoral autocracies where administrative bodes are not impartial. And even in those settings were electoral laws and bodies are establishing de-politicized, impartial rules, electoral malpractice—most certainly—exceeds the span of actions that are addressed by existing domestic laws.

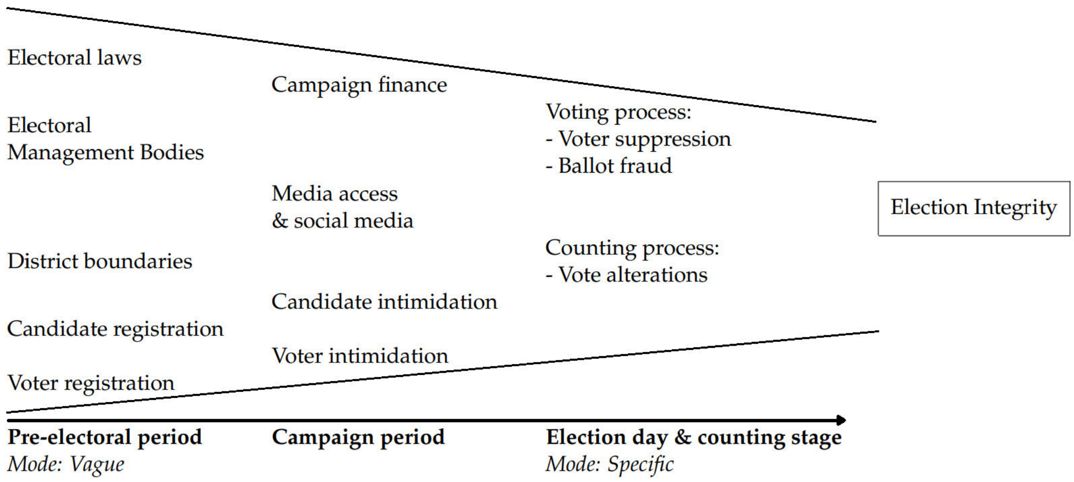


Figure 1.3. Conceptualization of electoral integrity, funnel plot. Documented are different arenas and fraud practices in which actors can intervene in order to manipulate election results. From left to right, arenas come closer to election day and interventions change from being vague to target specific voters or votes.

So far, I have sketched out different approaches to *define* what constitutes electoral wrongdoing and what separates legitimate from illegitimate actions. These different conceptualizations take different paths of *how they identify a discrepancy* between a certain practice of behavior and a normative baseline that is coined electoral malpractice. They agree upon that a discrepancy that has been identified constitutes an act of electoral malpractice. They disagree on how to construct normative baselines, either through consulting mass perceptions, international norms or domestic law.

Next to lining out the three different approaches of how to identify discrepancies, a straightforward follow-up task is to concretely define the different objects towards which electoral malpractice can be targeted in the first place. In line with Birch (2011), we can categorize different forms of electoral malpractice in terms of the objects they try to manipulate:

- electoral laws and institutions manipulating *the electoral framework*
- the formation of vote choice manipulating *the voter*
- the voting act manipulating *the outcome* itself

Figure 1.3 splits up these three objects of manipulation into a—schematic, but non-exhaustive—range of subcategories that adhere to either one of these classes. Manipulating the *electoral framework* is most commonly directed at drafting electoral laws that systematically favor selected parties or societal groups, for instance changing legal regulations of who can participate and strategically adopting or abolishing term limits to hinder or facilitate the participation of certain candidates. Other types of interfering with the electoral framework would be to strategically shape district boundaries (most prominently in majoritarian electoral systems) or strategically appointing leading positions in electoral management bodies to secure political influence on future administrative decisions. Manipulation of the *vote choice* most commonly relates to the campaign period in which vote choices commonly form, and spans issues such as unequal or illegal campaign finances, the misuse of public media outlets for partisan platforms by the current administration, algorithmically intruding social media networks, or violently intimidating candidates and voters at partisan rallies. Manipulation of the *voting act* relates to any practice that takes place on election day. This is what is most commonly referred to as 'electoral fraud' or 'electoral crimes', and historically has taken the form of violent voter suppression in front of polling stations within opposition strongholds, the organization of carousel voting in which voters are driven around to cast their vote multiple times, or tampering with the ballots that have been received.

1.3.2 Electoral Integrity and the Statistical Detection of Election Fraud

Now that I have mapped out the different dimensions of a concept as broad as electoral integrity, I proceed by defining which aspects of this concept the statistical detection of systematic election irregularities that this dissertation is concerned with speaks to and what I mean in the following with 'electoral manipulation'. The concept of manipulation that is relevant for the statistical methods that I outline is most closely related to a thin, legalistic definition of electoral fraud, but even adds some further restrictions.[7] What I treat as 'election fraud' here relates to *each practice on election day or in the counting stage which generates a discrepancy between the actual vote of the electorate and the reported results.* Essentially, this includes *adding* (or *removing*) votes to (from) the vote totals of one candidate (party), *shifting* votes from one candidate (party) to another and *invalidating* votes on election day or in the vote tabulation stage. Contrarily, it excludes every other type of behavior that is directed at manipulating the electoral laws, institutions, campaigns, or the vote.

That said, this does not mean that other forms of electoral manipulation cannot leave statistical traces in electoral data. For instance, if electoral violence is staged strategically in strongholds of certain political groups to hinder specific societal groups from turning out to vote, this might of course reflect in distributions of turnout and vote shares that are distorted in relation to previous elections in the same country where strategically placed violent acts were not present. However, tying election forensics tools to ballot fraud simply means that statistical procedures are not designed to flag such practices, but in their very inner workings are constructed to pick up different strategies of altering documented vote counts.

Now that I have clarified which aspects of the very wide-spanning and convoluted concept of election integrity I speak to in this dissertation, I proceed with a

[7] The scope of electoral manipulation that statistical tools from election forensics speak to is hence closely related to what Vickery and Shein (2012) coin 'election-related crimes'.

short overview of previously developed statistical tools for electoral anomaly detection to build some intuition for the reader. Afterwards, I will outline the main contributions that I myself add to the academic literature.

1.4 Statistical Methods for Detecting Election fraud: A Primer

The goal of this primer is to provide a short introduction to statistical methods for detecting election fraud. Statistical detection of numerical irregularities constitutes a prominent challenge across a variety of use cases. Shikano et al. (2019) have provided a review of statistical approaches to quantify anomalies in electoral returns. Sudjianto et al. (2010) have mapped out strategies for financial fraud detection such as credit (and debit) card fraud and money laundering in remarkable detail. Comprehensive reviews also exist for the fields of telecommunications fraud (Becker, Volinsky, and Wilks 2010), medical fraud (Ekin, Frigau, and Conversano 2021) or automobile ensurance fraud (Itri et al. 2019). Bolton and Hand (2002) have synthesized scholarly contributions into an excellent review of statistical fraud detection across this heterogenous range of fields. While many of the challenges and approaches share commonalities across the different subfields of forensic data analytics, every individual field presents its own idiosyncracies which make it unique.

The goal of this section is not to provide a comprehensive overview of all the different approaches to statistical election fraud detection that have been developed. Rather, before introducing my own methodological contributions, I aim to familiarize the reader with the general strategies, challenges and qualities that characterize this field of methods and sketch a couple of individual examples of fraud indicators.[8]

1.4.1 General Strategies

Statistical methods for election fraud detection depart from a dataset of fine-graded electoral results which are available across a large number of electoral units such as precincts, electoral districts or individual polling stations. A sample of such data from the Ecuadorian General Elections 2017, which will be analyzed in detail in Chapter 2, is presented in Table 1.1. Statistical approaches can be classified across at least two dimensions. First, all methods can be categorized to either exploit *unique characteristics* of individual electoral settings or to exploit numerical patterns that are expected to hold *globally*. Furthermore, methods can be sorted into supervised and unsupervised techniques. While supervised approaches rely on a pre-labeled dataset from which statistical methods learn to discriminate fraudulent from clean entities, unsupervised approaches typically quantify deviations from numerical or distributional assumptions that are supposed to be given (or explicitly modeled) if an election was clean.

[8] A more profound discussion of individual indicators that are relevant to this dissertation can be found in Chapter 3.2.1.

Canton	*District*	*Precinct*	*Table ID*	*Table Sex*	*Eligible*	*Turnout*	*Blank*	*Invalid*	*Cand. A*	*Cand. B*
Ibarra	Ambuquí	C. Borja	1	Male	328	253	21	15	60	118
Ibarra	Ambuquí	C. Borja	1	Female	330	281	12	18	44	144
Ibarra	Ambuquí	C. Borja	2	Male	330	263	13	25	56	122
Ibarra	Ambuquí	C. Borja	2	Female	330	273	18	25	33	134
Ibarra	Ambuquí	C. Borja	3	Male	330	251	15	23	34	122
Ibarra	Ambuquí	C. Borja	3	Female	330	276	9	28	45	143
Ibarra	Ambuquí	C. Borja	4	Male	330	245	13	29	46	131
Ibarra	Ambuquí	C. Borja	4	Female	329	262	15	23	37	148
Ibarra	Ambuquí	C. Borja	5	Male	329	194	18	14	50	126
Ibarra	Ambuquí	C. Borja	5	Female	303	189	6	21	46	132

Table 1.1. A sample of voting returns from Ecuador's 2017 presidential election. Documented are results for the two best performing candidates across ten voting tables of the canton Ibarra. Females and males place their ballots into separate ballot boxes.

However, in essence, virtually all statistical techniques for election fraud detection rely on comparing empirical quantities to expected values that are assumed to be in place when an election is clean. Large deviations from expectations are then taken as an indication of fraud. The main differences among the approaches lie in (i) how these expected values are computed (from theory, or—when empirically—in a supervised, or unsupervised manner) (ii) and whether expectations are assumed to hold globally for electoral events in general or whether these are inferred from specific characteristics of an individual election or country.

In the latter case, expectations usually are derived from circumstances that mimic a natural experiment, such as the random assignment of on-the-ground observers across polling stations (Hyde 2007; Enikolopov et al. 2013) or distributing voters to polling places in a way that is not correlated with political behavior (Cantú 2014).

Among those approaches that are designed to hold globally across elections, expected values under clean elections can be explicitly modeled from theoretical expectations (in *unsupervised* approaches) or learned in a *supervised* setting. It almost never becomes known whether individual observations indeed were fraudulent or not. Election forensics tools have therefore adopted the strategy of training supervised models on synthetic data sets that have been simulated by the researcher for which the degree of fraud is known (a strategy that also I undertake in Chapters 2 and 3). If these mimic the characteristics of clean elections (when no fraud is introduced into the data generating process) and supposedly fraudulent elections (under systematic vote alterations in the simulations), they provide useful prototypes for fine-tuning the development of new statistical methodology.

1.4.2 Examples of Statistical Approaches

Distribution of numerals in the digits of raw vote totals

Probably the most well-known approach to statistical anomaly detection in fine-graded electoral results such as the ones depicted in Table 1.1 exploits the frequency of different numbers (0,1,2,..,9) in the second or last significant digit[9] of raw vote totals for individual candidates or parties (located in the last two columns of Table 1.1). For a large class of data generating processes that include the composition of many electoral voting returns, well-grounded explanations exist that these distributions are far from random but can be described by a pre-defined pattern. Specifically, Newcomb-Benford's law (Newcomb 1881; Benford 1938) states that for suitable processes, the probability that the *first* significant digit is d ($d \in 1, 2, .., 9$) decays as an inverse-logarithmic function for early digits and approaches a uniform distribution for later digits. For *subsequent* digits, Hill (1995) and Hill (1996) provided a generalized version of the law postulating that the frequency of individual numbers d ($d \in 0, 1, 2, .., 9$) arising in the nth position ($n > 1$) can be defined as

$$P(d) = \sum_{k=10^{n-2}}^{10^{n-1}} log_{10}(1 + \frac{1}{10k + d}). \tag{1.1}$$

As Hill (1995) and Hill (1996) has formally derived, Equation (1.1) above holds asymptotically if observed numbers are generated as mathematical mixtures of different distributions without being naturally bounded towards a certain range of values.[10] That is, naturally observed vote totals that don't inherit manual manipulation are expected to follow Newcomb-Benford's law if these can be thought of as random samples that are not taken from one, but combined from many individual probability distributions. As Mebane (2006) argues, votes can be thought as stemming from hierarchical mixture population models in which at each electoral unit, at least two populations should be present: Those voters pertaining to partisan population strongly in favor of a candidate and the general population switching between candidates.

Figure 1.4 shows the expected frequencies of numbers 0,1,2,..,9 in the first, second and third digit and empirical frequencies within the second digit of raw vote totals for winning party (candidate) across six elections from five countries. As can be seen, empirical frequencies closely resemble those stated by Equation (1.1),

[9] The first significant digit of a number can be defined as its non-zero leftmost digit. Hence, the second significant digit of 60 is 0 and the second significant digit of 118 is 1.

[10] Essentially, this means that for vote totals that comprise three or more different digits across all electoral units such as 133, 221, and 148, the the individual numbers 0,1,2,..,9 should appear with unequal frequencies in the first and second digit, with low numbers being significantly over-represented. In the last digit position, individual numbers 0,1,2,..,9 should appear with (approximately) equal frequency. Interestingly, Beber and Scacco (2012) provide a different formulation and formal derivation of this uniform distributional property specifically for last digits, yet arriving at the same distributional assumptions as Hill (1995) and Hill (1996).

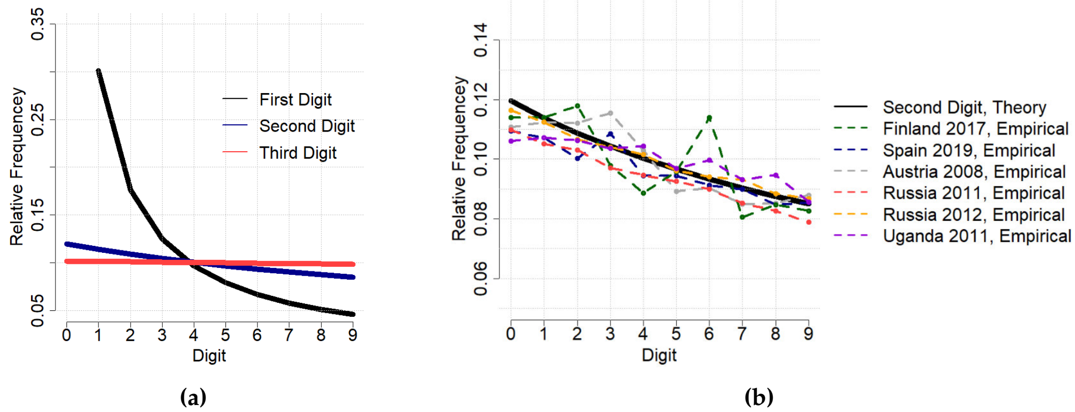


Figure 1.4. Digit distributions in raw vote totals against Newcomb-Beonford's law.*Panel A.* Expected frequencies of numbers 0,1,2,..,9 in the first, second and third digit. *Panel B.* Empirical frequencies within the second digit of raw vote totals for winning party (candidate) across six elections from five countries.

with the degree of fit being mostly a function of the number of observations (electoral units) that the data is respectively based on.[11] Deckert, Myagkov, and Ordeshook (2011), Mebane (2011), Beber and Scacco (2012), Mack and Stoetzer (2019) and Medzihorsky (2015) have extensively discussed the applicability and performance of *digit tests* for electoral anomaly detection.

In order to exploit the distribution of numerals within different digits, scholars usually test whether the empirically observed distribution differs significantly from its theoretical expectation stated in (1.1) using a χ^2-test ($df = 9$)

$$\chi^2_n = \sum_{i=0}^{9} = \frac{(d_i - d_i^*)^2}{d_i^*} \tag{1.2}$$

where d_i is the empirical frequency of a certain numeral in the nth digit and d_i^* is its theoretical expectation. The critical value against which the χ^2-statistic is evaluated for $df = 9$ is 16.92 at a significance level of 5%.[12]

Thus, *digit tests* in their traditional form are an example of an unsupervised statistical approach which—in principle—is designed to hold globally, where expected values under clean elections are theoretically derived. Digit tests are hence agnostic to context-specific characteristics, which largely determine their applicability or failure.

Skewness and kurtosis in the bivariate distribution of turnout shares and vote shares for the winning candidate

An approach that has gained increasing popularity in recent years is to look at the overall distribution of turnout or vote *shares* across all analyzed electoral units

[11] Details on the country-level data can be found in Appendix B.

[12] An alternative to the global χ^2 test for the emprirical fit to the overall distribution is testing deviations from particular empirical implications of (1.1) for significance, such as the mean of the last digit being 4.5 (see Hicken and Mebane 2017).

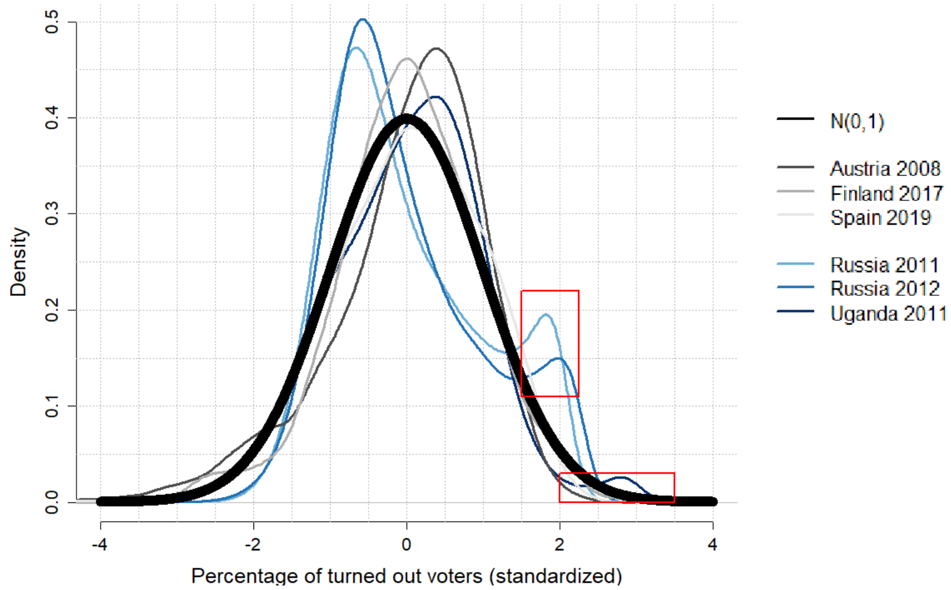


Figure 1.5. Standardized distribution of turnout shares across electoral units against a standard normal distribution. The black line depicts a standard normal distribution with a mean of 0 and standard deviation of 1. Colored lines represent standardized empirical distributions of turnout shares across electoral units from six elections in five countries.

that are available in the data. Methods that are constructed around the shape of these distributions base on recent work of applying concepts from statistical physics to quantitative social dynamics such as voting. In particular, studying a range of French elections since 1992, Borghesi and Bouchaud (2010) have shown that the distribution of turnout shares across towns is suprisingly stable over time, closely following a normal distribution whose parameters are (naturally) varying across elections, but whose *shape* across electoral units is remarkably stable.

When applying a simple transformation and representing turnout and vote share distributions as *logarithmic vote rates* (see Borghesi and Bouchaud 2010, p. 396) or standardized distributions, an even closer fit to Gaussian normality is reached *for clean elections*. Klimek et al. (2012) have shown that these distributional properties equally hold for Canadian elections and a range of Western democracies independently of the exact level of data aggregation that the data is based on (that is, whether data is available at the level of individual tables, polling stations, or low-level electoral districts).

For instance, Figure 1.5 displays standardized distributions for the percentage of turned out voters among all eligible voters for six elections out of the five countries of Austria, Finland, Spain, Russia, and Uganda. In Western democracies, standardized turnout distributions closely resemble a standard normal distributions. On the other hand, Klimek et al. (2012) as well as Myagkov, Ordeshook, and Shakin (2009) and Kobak, Shpilkin, and Pshenichnikov (2018) outlined that mechanisms such as systematic ballot box stuffing and deliberate wrong-counting in favor of one party (or candidate) leads to inflations of the distributions' right tail with extreme forms

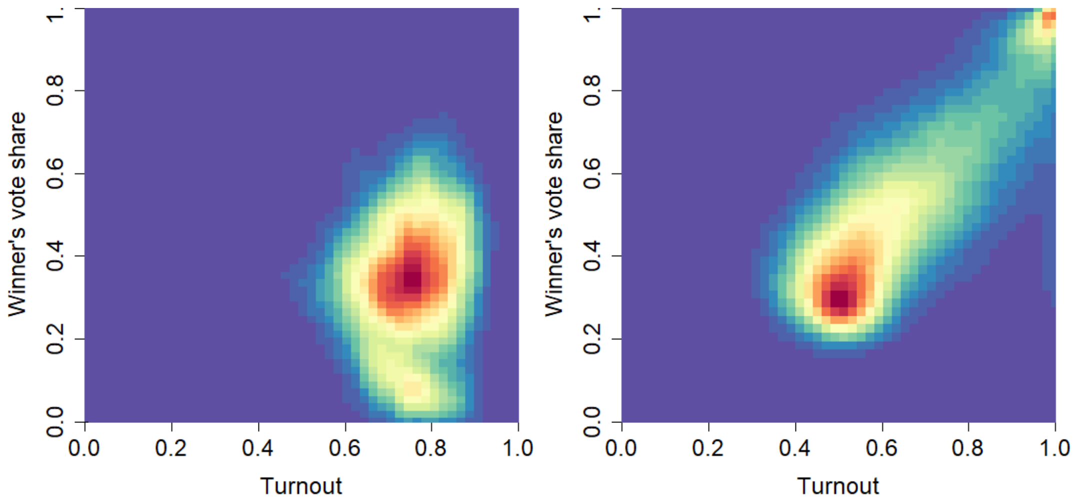


Figure 1.6. Bivariate distribution between turnout shares and vote shares for the Spanish Socialist Workers' Party (PSOE) in Spain 2019 and for United Russia in Russia 2011. Colors represent the number of electoral units with corresponding (x,y) coordinates.

of tampering producing clusters in the upper quintiles, as indicative from the standardized univariate turnout distributions from the Russian Duma elections 2011, Russian presidential elections 2012, and the Ugandan presidential elections 2011, which all three have bee subject to severe accusations of outright fraud.

To further showcase this phenomenon, Figure 1.6 compares the *bivariate* distribution between the percentage of turned out voters per municipality (Spain) or polling station (Russia) and the winning party for the two parliamentary elections from Spain 2019 and Russia 2011. As evident, while the Spanish data is reasonably well approximated by two orthogonal normal distributions, the bivariate distribution in Russia 2011 shows pronounced deviations from the statistical regularities that have been identified across heterogeneous country cases and sets of electoral systems among advanced industrialized democracies.[13]

To exploit these empirical patterns described by Borghesi and Bouchaud 2010 and their violations as documented by Myagkov, Ordeshook, and Shakin (2009), Klimek et al. (2012) have proposed a method for reverse-engineering levels of incremental and extreme fraud by modeling turnout and vote shares with two orthogonal Gaussian distributions and finding mechanisms of fraud that most closely resemble skewness, kurtosis and clusters between modeled and empirical distributions. The semi-parametric simulation model by Klimek et al. (2012) is an example of an unsupervised learning technique which, again, in principle can be applied globally but empirically *models* expected values rather than theoretically deriving these.

[13] Figures 3.6 and 3.7 in Chapter 3 show these distributions for additional datasets.

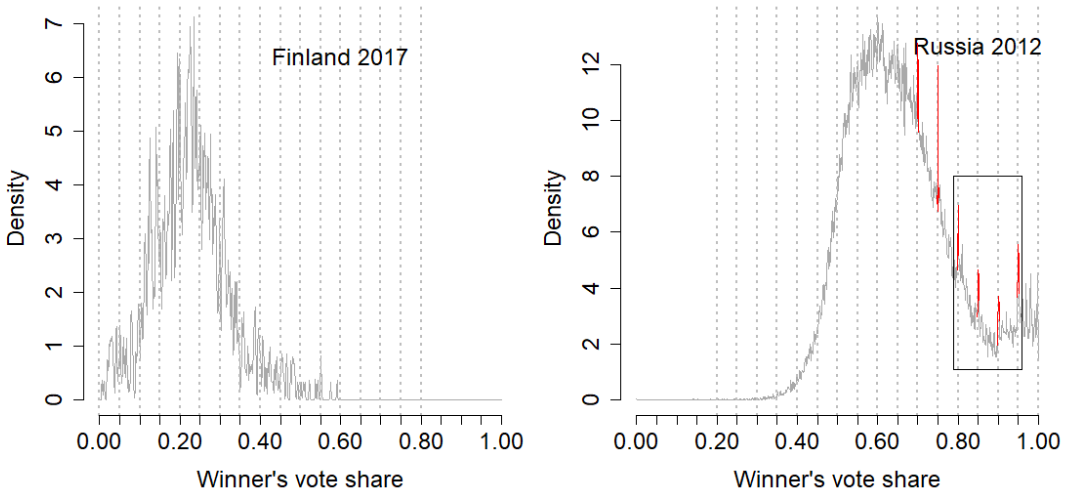


Figure 1.7. Estimated density (Gaussian kernel) of the winning party's (or candidate's) vote share. Finland 2017 ($n = 992$, bandwidth=0.001) and Russia 2012 ($n = 91,256$, bandwidth=0.0001). Spikes in the density mass around coarse shares (multiples of 5 and 10) colored red for Russian elections. Values at 100% not shown.

Inflated integer values that are multiples of '5' in turnout and vote shares

The most recent significant methodological contribution to the field of electoral anomaly detection stems from the observation that in election data that is supposedly tainted, the fraction of coarse integer percentages around turnout and votes share values is often considerably higher than what would be expected by pure chance, a phenomenon that appears if vote shares for the winner have been rounded up to meet certain target values. This feature has first been identified by Kobak, Shpilkin, and Pshenichnikov (2016a) in the history of Russian national-level elections in the period from 2004 onwards. While Rozenas (2017) shows that a sample of vote shares from a set of precincts is likely to exert noticeable spikes in the density mass at lower-order fractions even in the absence of any interference, both Kobak, Shpilkin, and Pshenichnikov (2016a) and Rozenas (2017) develop methods that closely follow the logic of the Bayesian posterior predictive check (Gelman et al. 2004), estimating the extent to which the observed share of exactly round vote shares exceed the values under a model of clean elections.

To showcase the empirical phenomenon of inflated coarse vote shares, Figure 1.7 presents estimated densities of the National Coalition Party in Finnish municipal elections 2017 and vote shares for Vladimir Putin in the 2012 Russian presidential elections. As can be seen, in the Finnish elections, no spikes around any integer percentages appear in the density mass. For the 2012 Russian presidential elections, spikes are noticeable and only appear around those vote percentages (>0.5) that favor Vladimir Putin.[14]

[14]These patterns are robust to slight adjustments in the bandwidth for density estimation.

1.4.3 Combination With Qualitative Investigation

It goes without saying that statistical election fraud detection needs to be performed hand-in-hand with qualitative assessments such as manual recounts of the ballots in selected polling places, screening vote tally sheets for inconsistencies in ink, handwriting or crossed-out numbers, observer missions that monitor proceedings on election day on the ground, or tracking electronic accesses to digital transmission systems used to aggregate the results.

There are several ways in which quantitative and qualitative inquiry can be intertwined. At the level of an election as a whole, usually qualitative anecdotal evidence suggests the presence of some level of systematic irregularities that has been documented in seemingly isolated instances. Statistical methods can then be used to estimate the prevalence of such fraudulent actions across the country. In this kind inquiry, anecdotal qualitative assessments forego the application of the election forensics toolkit. On the other hand, statistical approaches for election fraud can also be used much like their counterparts from other fields such as financial fraud detection. Especially using methodologies that allow to localize irregularities, suspicion scores can be constructed across regions, precincts or even individual polling stations. After ranking the individual localities by their likelihood of interference, qualitative assessments (such as manual recounts) can follow from statistical analyses.

Yet, the integrity of electoral events cannot be assessed by statistical analyses alone. What statistical methodologies can provide is setting 'red flags' that necessarily need to be followed by closer empirical scrutiny.

1.4.4 Qualities of Individual Approaches

Statistical methods for election fraud detection ideally hold the following qualities (Hicken and Mebane 2017):

- *Universality.* The use of a method should require as little information as possible. Ideally, all data that is needed can be found in the voting returns that have been published after a given election. Even if no information on context-specific third variables is present, this enables methods to be universally applicable across different countries, electoral events, and time points.

- *False negatives and false positives.* Methods should be sensitive enough to detect fraudulent interference when it was present and prevent producing results that are indicative of fraud if the underlying data stemmed from fraud-free processes. Failing this criterion will let elections that actually experienced vote fraud to be flagged as clean, providing misleading credibility or vice versa.

- *Degree of interference.* Not only should forensic approaches be able to signal *if* evidence for manual interventions is presented, but additionally allow for a *quantification* of the level of contamination that is inherent in voting returns. They should clearly distinguish elections that experienced small irregularities from voting figures that were vastly altered or even fabricated completely.

- *Probabilistic statements.* Methods for fraud detection should produce estimates of uncertainty. In the absence of on-the-ground monitoring and directly documented evidence, conclusions will by definition not be definite. Quantifying the confidence level of fraud claims is crucial in this regard.

- *Geographical localization.* Finally, ideal methods allow to geographically locate, or at least circumscribe which of the analyzed entities are the drivers of results that deviate from patterns expected under natural processes. For instance, if the extent of exactly coarse vote shares exceeds their expectation, the geographic localities of interest that warrant closer investigation are those with round vote shares. This distinguishes the appraoches developed by Kobak, Shpilkin, and Pshenichnikov (2016a) and Rozenas (2017) from digit tests, which merely present a test whether the overall distribution of digits across all units resembles its expectation or not.

No single method can fulfill all of these assets and there are a number of tradeoffs embedded in the development of each approach. For example, the most effective designs certainly stem from approaches that are tailored towards specific countries and exploit unique features of the electoral system that reliably indicate fraudulent activities yielding few false negative and false positive results (Cantú 2014). Yet they are hardly applicable in a cross-country setting. Vice versa, methods that rely on global assumptions that are assumed to be met in election data of all kinds are universally applicable, but typically produce a larger number of false positives and negatives. Developers of methodical approaches are not asked to design 'fit them all' methods that satisfy all criteria as this is seldomly possible, but to position their approaches among the qualities above and to correctly locate their strengths and weaknesses.

1.5 Plan of the Dissertation

In *Chapter 2*, I provide the first extension of the range of methods for probabilistic anomaly detection and develop an unsupervised method centered around simulated expected values under clean elections which focuses on a novel numerical characteristic that hasn't received scholarly attention yet. Other than existing approaches, the method that I outline does not exploit numerical patterns within data of one election but *across* simultaneous electoral events. I argue that in the presence of concurrent electoral contests on election day, ballot box stuffing and vote stealing can be detected from *undervoting irregularities* that emerge if protagonists of fraud fail to interfere into multiple races to equal extents. Conceptually, I introduce the distinction between balanced and unbalanced fraud approaches in the presence of several simultaneous electoral events. Methodologically, I develop a semi-parametric method of fraud detection building on the fact that if undervoting irregularities stem from administrative or human errors, discrepancies in turnout levels are unrelated to the winner's vote share. I illustrate the method on both (i) empirical data from recent Ecuadorian elections where undervoting irregularities are widespread (ii) and simulated data for which the degree of fraud is known.

In *Chapter 3*, I provide the second extension to the methodological state of the art by presenting a unified statistical framework using supervised machine learning to jointly evaluate different numerical indicators such as the ones discussed above that have been successful in identifying fraud. I speak to two challenges that stand out in the literature. First, the statistical tests that have been developed as standalone tests that are respectively centered around one numerical characteristic, don't inform each other, and easily lead to inconclusive results. Second, as a consequence of these missing links, many approaches are typically restricted to binary statements about electoral (mal-) practice and fail to provide estimates for the percentage of votes that is affected by manipulation. I speak to these shortcomings by fusing existing numerical indicators with supervised machine learning approaches. I define a protocol for simulating realistic micro-level electoral returns that resemble empirical data across a range of numerical characteristics rather than one isolated pattern. Subsequently, I train a machine learning algorithm on a multivariate feature space that takes into account characteristics of digits, turnout, and vote share distributions simultaneously and provides reliable estimates of the percentage of votes that have been tampered. I illustrate the approach on national-level elections that have been contested both publicly and in the academic literature from Russia 2011-2012, and Uganda 2011 as well as an array of Western European democracies and simulated data.

In *Chapter 4*, together with my co-author Viktoriia Semenova, I leave the scope of the methodological study of electoral integrity and investigate the attitudinal responses of individuals who have been confronted with credible information about electoral malpractice. Scholars have shown that consciousness of election fraud lets individuals withdraw support from candidates, institutions and governments that are supposedly involved in manipulation. We argue that election fraud information will let individuals extrapolate legitimacy loss even to political institutions that are unrelated to electoral events and lead to decays of trust in the political system as a whole. Second, we investigate how these spillovers are shaped by the reactions of other political actors such as court punishments of alleged fraud perpetrators. We present empirical evidence from a pre-registered online survey experiment in Colombia, Mexico and Russia. We find that legitimacy loss of political institutions does indeed spill over to facets of the political system that are unrelated to electoral administration. Second, we show that political actors only have limited powers to mitigate citizens' alienation: Once information on fraud is shared, adequate punishment of fraud perpetrators may mitigate the negative effects of fraud information while this effect is far from being omnipresent across countries and institutions.

1.6 Key Innovations and Contributions

This dissertation presents several contributions to the academic literature. In the following, I summarize the main innovations and contributions differentiating between (i) conceptual contributions, (ii) methodological contributions and (iii) empirical contributions.

1.6.1 Conceptual Contributions

Balanced and unbalanced fraud in concurrent elections

A large body of literature has investigated which numerical patterns in disaggregated voting results are likely to appear under natural data generating processes that resemble—largely—independent choices of a group of individuals and how these numerical patterns are distorted under systematic alterations of the results. In the development of many of these statistical tools, numerical distortions are directly linked to specific strategies of vote alterations that these are designed to flag. For instance, the semi-parametric simulation model by Klimek et al. (2012) is engineered to quantify the extent of two different forms of *ballot box stuffing*, namely *incrementally* adding votes for the winning party (candidate) across a large number of localities and setting the turnout and vote shares for the winner to *extreme* values that lie above 95% from the skewness and kurtosis of the bivariate turnout and vote share distribution. Rozenas's (2017) variation of the posterior predictive check for inflated coarse integer vote and turnout shares that are multiples of five (0.60, 0.65, 0.7) tests for the specific strategy of *rounding fraud* in which turnout or vote shares are rounded up at regional vote tabulation centers to meet certain 'target values'.

Next from the study of numerical patterns, it is hence of utmost importance to sharply define different *strategies* of fraud that can be executed in the first place in order to (i) understand how these will affect the numerical characteristics of electoral returns and to (ii) design statistical tools that pick up their specific distortions. In this dissertation, I contribute to the conceptual study of election fraud by defining two practices of electoral tampering *under the existence of concurrent electoral events* that have not been discussed in the literature yet, *balanced* and *unbalanced* fraud.

As I outline in Chapter 2, given that multiple electoral events are administered side-by-side at the same localities, agents of fraud are faced with a crucial choice. In a balanced fraud approach, all races are altered to exactly equal extents. That is, for every vote that is added to (or removed from) one electoral race, another vote is added to (removed from) all parallel contests, even those that are not of primary concern for the agents of fraud or their principals. In an unbalanced fraud approach, these numbers differ. Balanced fraud approaches keep the number of ballots that are observed for each electoral contest equal, but require extensive clandestine actions that are unrelated to the actual goal of electoral tampering.

Unbalanced fraud approaches efficiently meet fraud goals, but create a discrepancy between the number of ballots that are observed across concurrent elections, leading to undervoting irregularities.

1.6.2 Methodological Contributions

A model to detect unbalanced fraud approaches from undervoting irregularities Conceptually, one of the contributions of this dissertation is to define the two different strategies of systematic vote alteration of balanced and unbalanced fraud that can be in place under concurrent elections which are administered side-by-side (see above). The first methodological contribution that I present is to design, evaluate and apply an unsupervised statistical learning method to (i) detect (ii) and quantify the extent of unbalanced fraud approaches from systematic patterns in undervoting irregularities between concurrent elections. The semi-parametric simulation model that I present builds on the fact that if undervoting irregularities stem from administrative or human errors, discrepancies in turnout levels are unrelated to the winner's vote share. On the contrary, if undervoting irregularities are the result of systematically adding (or removing) ballots to (from) selected races in selected electoral contests, statistical associations with individual candidates' (parties') vote shares will appear.

As a general intuition, the semi-parametric simulation model detects unbalanced election fraud by (i) simulating a range of artificial concurrent elections from empirical input parameters that are either clean or manipulated to different degrees, (ii) quantifying the average numerical distance between the observed data and each set of simulated elections, (iii) and finding the set of artificial elections that–in expectation—minimizes the distance to the observed data. The fraud parameter that was used to construct this set of artificial elections serves as the estimate of fraud and iterating this procedure a large number of times is used to construct uncertainty intervals around the estimated fraud parameter.

A unified framework for existing approaches of probabilistic fraud detection The first methodological contribution of this dissertation is to develop a statistical method that is constructed around a *novel* numerical characteristic within voting returns that has not been systematically exploited to identify systematic irregularities. The second contribution presents a novel approach to probabilistic fraud detection but is based on the *existing* range of forensic indicators that have been developed in the academic literature. By training supervised machine learning methods on synthetically generated training data for which the degree of fraud is known, I present an approach to directly estimate the *percentage of votes* that are affected by systematic vote alterations.

First, I define a protocol for simulating realistic micro-level electoral returns that resemble empirical data across a range of numerical characteristics rather than one isolated pattern. Subsequently, I train a selection of machine learning algorithms on a multivariate feature space that takes into account characteristics of digits, turnout, and vote share distributions simultaneously and provides reliable estimates of the percentage of votes that have been tampered. This approach (i) merges different forensic indicators that have been developed as standalone tests which don't inform each other (ii) and directly allows to estimate the percentage of affected votes, circumventing the need of inferring rather broad (binary) statements about the integrity of an event in question.

A protocol to simulate artificial voting returns

The development of a supervised machine learning procedure to gauge the presence of systematic election irregularities comes with a more subtle contribution that warrants separate mention here. As I train supervised methods on synthetic data that take into account characteristics of raw digits, the skewness and kurtosis of turnout and vote share distribution, the association between these, and the share of exactly coarse vote shares, I outline a systematic protocol to simulate artificial voting returns that resemble empirical data across a whole range of numerical characteristics. This stands in contrast to the current practice of the field which construct statistical tests around one respective numerical pattern. The definition of this protocol means that the approach can easily be augmented to incorporate additional numerical characteristics that will spark scholars' attention in the future, and can serve as a baseline for the generation of synthetic training data that is representative across a range of substantively interesting indicators, countries and electoral systems.

1.6.3 Empirical Contributions

Last, this dissertation makes an *empirical* contribution to our understanding of how receiving credible information on electoral manipulation is tied to individuals' attitudes towards the political system. So far, scholars have shown that consciousness of election fraud lets individuals withdraw support from candidates, institutions and governments that are supposedly involved in manipulation. The study presented in *Chapter 4* goes beyond this state of the art in two ways. First, the study shows that election fraud information will let individuals extrapolate legitimacy loss even to political institutions that are unrelated to electoral events and lead to decays of trust in the political system *as a whole*—an attitudinal response to exposing cheating that has broader consequences than those that have so far been traced by the literature. Second, we acknowledge that fraud claims seldom rise in a vacuum but that the distribution of credible information often lead to correctives from within the political system. For instance, these correctives manifest themselves through the dismissal of electoral staff or court punishments of alleged fraud perpetrators. Once political actors step in, it is unclear whether such horizontal 'checks and balances' mitigate the detrimental attitudinal effect of fraud information. To our best knowledge, the study presented here is the first one to investigate how within-system corrections moderate decays in political trust.

The main empirical contributions to the scholarly understanding of the nexus between credible fraud information and political trust are that the consequences of administering election fraud for public support are even more detrimental than currently acknowledged by the literature and can't be easily prevented by adequate political interventions. On the one hand, this is because information on electoral misconduct even induces shifts in public support towards components of the political system that are no beneficiaries of manipulation and are not related to the administration of electoral administration in the first place. On the other hand, this is because once fraud information is disseminated, even credible punishments cannot completely account for the loss of trust in the political system.

2

Detecting Unbalanced Election Fraud Approaches From Undervoting Irregularities

Abstract: I argue that in the presence of concurrent electoral contests on election day, ballot box stuffing and vote stealing can be detected from undervoting irregularities that emerge if protagonists of fraud fail to interfere into multiple races to equal extents. Conceptually, I introduce the distinction between balanced and unbalanced fraud approaches in the presence of several simultaneous electoral contests. Methodologically, I develop a semi-parametric simulation model to detect and quantify systematic interference that stems from unbalanced fraud approaches. I validate the method on simulated data for which the degree of fraud is known. In addition, I apply it on empirical data from Ecuador's General Elections 2017 that showed large extents of undervoting irregularities and were marred by massive fraud accusations as well as from the Local Elections 2019 in which legitimacy was not largely contested. I demonstrate that while the developed model robustly reverse-engineers known levels of fraud in statistical simulations, the empirical patterns that are inherent to the 2017 voting returns are well explained by systematic manipulation while the 2019 contests are labeled as clean. This chapter highlights the relevance of contextual information for the practice of election forensics in general and improves our understanding of undervoting irregularities in particular.

Keywords: *Election forensics; Electoral integrity; Undervoting irregularities; Monte Carlo simulation, Ecuador.*

Author's note: A previous version of this chapter has been presented at the 12th Annual Conference of the European Political Science Association in Prague, June 23-25, 2022. I thank Sofía Garcia Durrer and Gloria Gerhardt for excellent research assistance as well as Oke Bahnsen and Viktoriia Semenova for helpful comments on prior versions of this manuscript. Substantial parts of this work were developed during a research stay at the *Latin American Social Science Institute FLACSO* in Quito, Ecuador. I thank Prof. Santiago Basabe Serrano for generously hosting me and organizing a series of expert interviews that greatly aided the development of this project.

2.1 Introduction

Since the 'electoral revolution' (Norris 2014) lead to a dramatic increase in elections during the second half of the twentieth century, direct elections at national scale have been almost unanimously adopted by countries across the globe. Doubts about electoral integrity are, however, by no means restricted to authoritarian regimes and new democracies. In a sample of 57 countries from the World Values Survey collected between 2017-2022, merely 15.8% of respondents asserted that votes are counted fairly and election officials are fair in their country. Substantial doubts are voiced even among developed democracies, as this share of respondents rises to merely 25.1% if only the twelve OECD member states that are part of the sample are considered.[1] Electoral events worldwide are regularly followed by intense scrutiny of their level of integrity.

The field of election forensics employs statistical methods to detect anomalies in voting returns that are indicative of systematic irregularities. Conventionally, this body of research is driven by the assumption that fine disaggregated election results which stem from fraud-free processes inherit a range of numerical characteristics that hold *globally* over different electoral systems but are violated under manual alteration of the data. For instance, existing research has focused on identifying anomalous patterns in the distribution of raw vote totals (Mebane 2008; Beber and Scacco 2012; Medzihorsky 2015), the share of polling stations that report exactly coarse integer percentages for turnout and the winner's vote share (Kobak, Shpilkin, and Pshenichnikov 2016a; Rozenas 2017), and systematic clusters, skewness and kurtosis within the bivariate distribution of turnout and support rates (Myagkov, Ordeshook, and Shakin 2009; Klimek et al. 2012).

A distinct approach to statistical election fraud detection lies in exploiting *specific* features of an electoral system that are inherent to selected country cases. Analyzing Mexico's 2010 gubernatorial elections, Cantú (2014) exploits the fact that within each electoral precinct, eligible voters are assigned to polling stations according to their childhood surname. As voting behavior is uncorrelated with voters' last name initials, Cantú identifies systematic interference from unexpected differences in turnout levels and vote shares across contiguous polling stations.

This chapter contributes to the literature on election forensics by exploiting the administration of simultaneous electoral events for the statistical detection of election fraud. The conduction of parallel events gives rise to the phenomenon of 'undervoting irregularities', which occur if *the same polling stations officially report diverging turnout levels across different electoral contests* and hence less (more) overall votes are observed for some races than for others. While at each individual polling station, the share of valid, invalid and spoiled ballots might differ across electoral contests, the total number of turned out voters necessarily needs to be identical across events. Given that no electoral laws are in place that formally restrict access

[1] Based on the Values Survey Wave 7 (Haerpfer et al. 2022).

to some electoral contests[2], these discrepancies are either the result of administrative errors or a consequence of fraudulent interference.

I argue that in the presence of concurrent electoral contests on election day, ballot box stuffing and vote stealing can be detected from undervoting irregularities that emerge if protagonists of fraud fail to interfere into multiple races to equal extents. Conceptually, I introduce the distinction between balanced and unbalanced fraud approaches in the presence of several simultaneous electoral events. Methodologically, I develop a semi-parametric method of fraud detection building on the fact that if undervoting irregularities stem from administrative or human errors, discrepancies in turnout levels are unrelated to the winner's vote share. I illustrate the method on both (i) empirical data from recent Ecuadorian elections where undervoting irregularities are widespread (ii) and simulated data for which the degree of fraud is known.

This chapter makes two contributions to research on electoral fraud. First, I coin the conceptual distinction between balanced and unbalanced fraud approaches that enhances our understanding of how agents of fraud behave given that concurrent electoral events are taking place. Second, I contribute to the growing literature on statistical tools to detect fraudulent interference from numerical characteristics in fine-graded voting returns.

The remainder of this chapter is organized as follows. The next section explains the phenomenon of undervoting irregularities and showcases them based on national- and local-level data from Ecuador. Section 2.3 introduces the conceptual distinction between balanced and unbalanced fraud approaches across multiple electoral events. The subsequent section outlines a semi-parametric simulation model estimating the degree of unbalanced fraud that is present. Lastly, I apply this method to two Ecuadorian elections in 2017 and 2019 and simulated elections for which the degree of fraud is known. I show that the empirical patterns that are inherent to Ecuadorian voting returns from the 2017 General Elections are well explained by systematic manipulation while the 2019 Local Elections are labeled as clean. Additionally, I trace empirical evidence for an alternative mechanism that would lead to similar empirical patterns in 2017. Taking advantage of Ecuador's urban-rural divide in educational attainment, little evidence is found for undervoting irregularities being the result of administrative incapacity of electoral staff in those regions in which left-wing support is traditionally highest.

2.2 A Motivating Example

2.2.1 Electoral History in Ecuador

Many countries hold concurrent electoral contests on voting day. In Latin America alone, 60 out of the last 95 elections at national scale were conducted alongside

[2] An example of restrictive electoral laws would be underage voters or non-citizen residents only being eligible to vote in local elections but not in national contests that are conducted side-by-side.

at least one parallel contest.[3] To motivate the idea behind the method, I examine election data from the country of Ecuador.

During most of the first part of the twentieth century, Ecuador's electoral history was deeply permeated by institutionalized manipulation administered by the Radical Liberal Party (PLRE, Torre 2015), whose rule included practices such as restricting voting rights of marginalized groups, intimidation of opposition supporters and the alteration of final vote counts on ballot day. After the liberal party's main competitor Velasco Ibarra's fifth non-consecutive presidency ended in 1972 with a military coup, the country experienced a—comparatively short— period of military rule with no national elections conducted before in 1979 power was handed over the constitutionally elected civilian Social Democrat Jaime Roldós Aguilera.

The 1978 electoral reform granted illiterates—which formed a large part of the country's indigenous population who *de facto* have been excluded from the right of suffrage— the right to vote. Ever since, elections in Ecuador did formally function as a legitimate process to select legislative representatives and public officials. After the 1979 handover of power, Ecuador's history of democratization was marked by a large series of presidential downfalls which were tightly coupled to a number of economic crises often triggered by fluctuations of world oil prices. Steady shifts between governments favoring liberal free-market economics and left-winged platforms fostering social equality and protectionist measures characterized Ecuadorian politics up until the beginning of the 21st century. These constant shifts lead to a total number of twelve different presidents in the period between 1979-2007, out of which only few could regularly end their presidential term. During these decades, the conduct of elections often fulfilled the mere purpose of officially restoring a delicate power balance that was continuously disrupted by repeated *coups d'état*.

The period of repeated party system collapse came to an end when in November 2006, Rafael Correa Delgado, an independent leftist with no partisan base, was elected president and went on to rule the country over three consecutive terms until 2017. Correa, a close ally of Venezuelan socialist leader Hugo Chavez, rapidly gained popular support implementing a platform that concentrated broad powers in the hands of the president, restored national control over the country's foreign owned oil industries and used flourishing oil revenues to implement extensive social spending alongside free secondary and post-secondary education. The second half of Correa's presidency was characterized by an increasingly authoritarian style, which—backed up by a series of constitutional referendums that received mass public support— witnessed a range of political reforms undermining the independence of the judicial system, growing control over media content by the government and the persecution of political opponents. As political institutions became more and more aligned with the president's *Correismo*, the Ecuadorian discourse about election integrity, while continuously revolving around voting day inconsistencies,

[3] Source: Own research by the author. Time frame covered: 2009-2020. Countries covered: Argentina, Belize, Bolivia, Brazil, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Venezuela, Chile, Urugay.

was augmented by a second dimension. Political observers increasingly criticized Correa's systematic politicization of electoral institutions such as the National Electoral Commission (CNE)—which was packed with Correa supporters—and his extensive presidential control over the country's main media outlets, which failed to provide a level playing field for political competition.

The discourse about election integrity escalated in the 2017 presidential election, when Correa-endorsed successor Lenín Moreno faced the Guayaquil-based liberal banker Guillermo Lasso in the runoff vote held on April 2. After several major exit polls during election night had predicted a win for Lasso who already declared victory and an end to 13 years of *Correismo*, the CNE declared Moreno to be the winner over a small margin of less than 2 percent of the votes. As a consequence, the country was shook by waves of protests that lasted several weeks, although several recounts, predominantly in the region of Guayas, reassured Moreno's victory.

Today, Ecuadorian national elections are characterized by repeated large scale protests that aim to question the legitimacy of election results, a large number of absentees (1 out of every 5 eligible voters) although the electoral system implements compulsory voting, and a large number of purposely spoiled ballots among those that do turn out (up to 17.9 percent in the country's most recent 2021 election) demonstrating the population's large distrust in the electoral process.

2.2.2 Undervoting Irregularities

Ecuadorian elections are regularly conducted as general elections in which multiple electoral races for different types of offices (*dignidades*) are conducted side-by-side. For instance, in the 2017 elections held on February 19, voters directly elected (i) the head of government in a presidential race, (ii) the members of the country's national assembly, (iii) parliamentary members of 24 regional assemblies, (iv) Ecuador's five national representatives for the Andean parliament—the deliberative body of the Andean community—and (v) cast votes in a nation-wide referendum prohibiting politicians and civil servants to hold bank accounts in countries with preferential tax regimes and low tax jurisdictions. On election day, voters are assigned to different polling stations according to their registered address. At the voting localities, voters get handed out different ballots corresponding to different electoral contests, which are inserted into separate ballot boxes. At the counting stage, votes for different electoral contests get tabulated in separate vote tally sheets (*actas*). The Ecuadorian electoral process has been described as highly complex by international

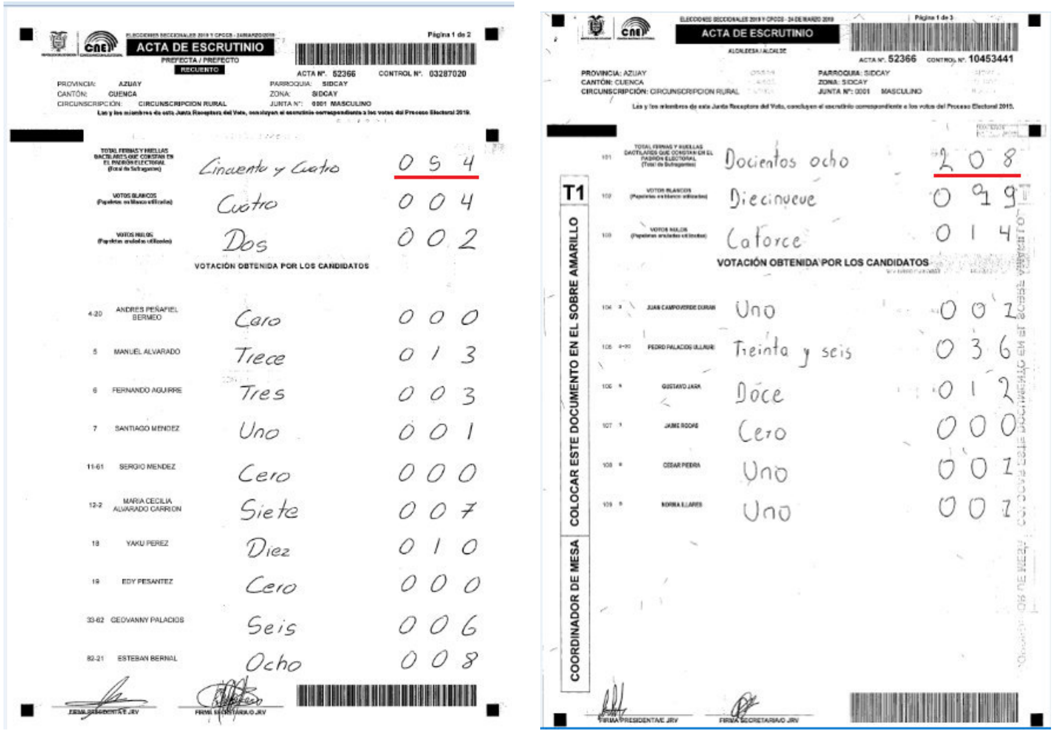


Figure 2.1. Irregularities on vote tallies. An empirical example of an undervoting irregularity in the Ecuadorian local elections on March 24, 2019 from the municipality Sidcay in Cuenca, Azuay. The left tally depicts the vote tabulation for provincial representatives (*prefectos*). The right tally tabulates votes for the city mayor (*alcalde*). The documented number of turned out voters is underlined in red and differs across both electoral races at the same polling station.

election observers monitoring democracy (EU 2009), who have voiced for a simplification of the procedures.

Perhaps unsurprisingly, Ecuadorian electoral returns are marred by empirical inconsistencies which are sometimes referred to as 'undervoting irregularities'. This peculiarity refers to the phenomenon of individual polling stations reporting diverging numbers of turned out voters across different electoral contests that are conducted side-by-side on voting day. Naturally, it is expected that some voters cast valid votes for some contests and submit empty or spoiled ballots for other electoral races depending on factors such as their saliency, perceived competitiveness or regional and political scope. However, it holds that across all electoral races, each polling station must report the same number of voters that turned out in the first place.

The left and right panel of Figure 2.1 show an empirical example of an undervoting irregularity for the Ecuadorian local elections conducted on March 24, 2019, whose legitimacy remained largely uncontested.[4] Depicted are two vote tallies (*actas*) from the same voting table (*junta receptora*) for two different electoral contests at a polling station in the municipality of Sidcay in Cuenca, Azuay. The left tally

[4]Material acquired digitally from the National Electoral Commission (CNE) on January 20, 2020 in Quito, Ecuador.

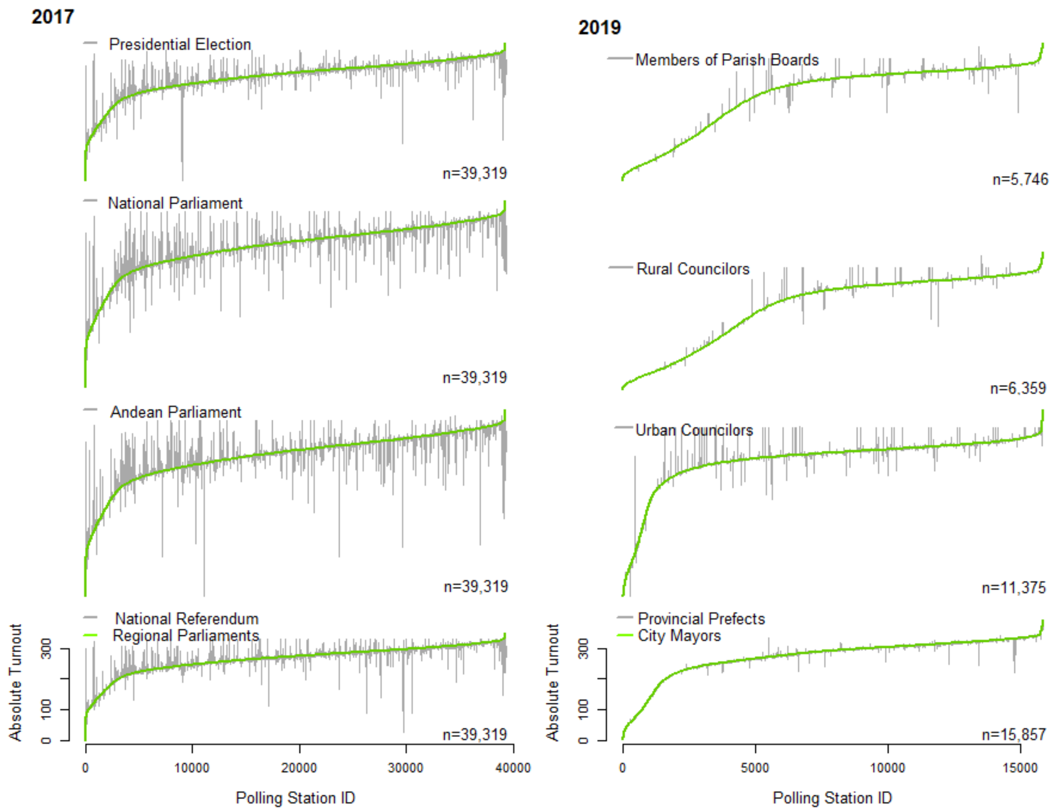


Figure 2.2. Undervoting irregularities across elections. Undervoting irregularities mapped across polling stations for five different electoral contests. Left panel: General Elections 2017. Right Panel: Local Elections 2019. Green lines indicate the absolute number of documented turned out voters for the election of state-level members of parliament (2017) and city mayors (2019). Grey lines indicate the absolute number of documented turned out voters in parallel elections at the same polling stations. Sample sizes differ for the 2019 elections as not all regional contests have been held in all provinces. Empirical patterns are insensitive to a change in baseline elections (green).

shows results for the election of provincial representatives (*prefectos*). Out of 350 eligible voters who are registered for this polling
station, a total of 54 ballots have been observed including four blank and two null votes. The right tally shows results for the election of the city mayor (*alcalde*), for which a total of 208 votes have been counted including nineteen blank and fourteen null votes. This documents a substantial irregularity as the total number or received ballots for each contest must necessarily be the same on both tallies. The magnitude of the inconsistencies is a multiple of the vote distances between the individual candidates.

Figure 2.2 maps the extent of undervoting irregularities across more than 39,000 Ecuadorian polling stations for two of the recent elections which publicly have been called into question. Discrepancies in the documented number of turned out voters emerge in up to 2,980 (1,528) out of a total of $n = 39,319$ ($n = 15,857$) polling stations in 2017 (2019) at which parallel contests were held. The magnitude of the discrepancies calls for the question whether these are due to administrative errors and the carelessness of low-level election officials or systematic manipulation on election day.

2.3 Election Fraud in the Presence of Concurrent Electoral Events

2.3.1 Balanced and Unbalanced Election Fraud

While multiple election day irregularities such as voter intimidation, electoral violence or the maladministration of polling stations have been observed in Ecuadorian elections (see for instance the Electoral Observer Mission reports OAS 2008, Carter Center 2007, EU 2009), this chapter focuses particularly on ballot fraud. Ballot fraud is of central interest for large accounts of the literature on election integrity. Perhaps the most direct documentation of manual alterations of vote counts is delivered by Cantú (2019b), who is providing a remarkable account of scanned vote tally-sheets from Mexico's 1988 presidential election which inherit crossed-out and altered numerals that are most prevalent in polling stations where members of the opposition were not present. Analyzing electoral returns from Russian presidential and Duma elections between 2000 and 2012, Kobak, Shpilkin, and Pshenichnikov (2016a) and Rozenas (2017) outline the specific mechanism of rounding fraud where vote shares for the favored candidate are simply rounded up at the vote tabulation stage, and both present statistical models that show that the number of exactly coarse integer vote-shares (e.g. 0.6, 0.65, 0.7) is significantly larger than expected by pure chance. Other methodological work on ballot fraud can be found in Beber and Scacco (2012), Myagkov, Ordeshook, and Shakin (2009), Klimek et al. (2012) and Callen and Long (2015). Notably, there are two distinct forms of ballot fraud that inflate (reduce) the number of turned out voters and thus are connected to undervoting irregularities: stuffing the ballot box with pre-prepared ballot papers or illegally withdrawing valid ballots from a particular race.

An explanation of the emergence of undervoting irregularities that can be put forward without the need of evoking systematic manipulation is that these might be manifestations of administrative challenges, human errors or the carelesness of electoral authorities during the voting or counting stage as polling places often fall short in personnel and material needed for faultless election conduct. Extensive election observation reports of the Organization of American States (OAS) and the European Union do a great job in describing the difficulties that Ecuadorian low-level election officials have faced, such as material arriving several hours after polling is supposed to start or shortage of electric light during the vote counting stage on election night (OAS 1998). Another explanation of undervoting irregularities is that some voters simply do not receive or hand back ballots for those races that they do not intend to vote for, either because election officials themselves are misinformed about valid procedures, voters are misinstructed at the voting booth or due to their own carelessness. This reasoning is especially plausible when taking into account statements by Electoral Observation Missions stating that a substantial share of low-level election officials did not receive sufficient training to carry out their election-day duties (Carter Center 2007; EU 2002; EU 2009; IRI 2003), even citing individual cases where volunteers waiting in queue were spontaneously mobilized to help out on election day (OAS 2006; OAS 2008).

In the case of Ecuador, the undervoting irregularities are indicative of a novel type of fraud that has not been described in the literature on election integrity. If low-level election officials or unauthorized individuals entering polling stations during voting or the counting stage selectively remove votes or add pre-prepared ballots to some of the ballot boxes, undervoting irregularities arise if ballot boxes concerning different electoral races are affected to unequal extents. An interview with a high-level representative of the Central Electoral Commission in Quito, Ecuador supports this view.

> *"What is happening? At certain tables, under the carelessness of the electoral authorities, they physically remove a number of votes, which constitutes an act of probabilistic fraud. If I know that 'Pedro' wins at Table A, I take one hundred votes from the ballot box. Most certainly, I will remove the majority of votes from 'Pedro'. I'll also remove votes from the rest of the candidates, but probably 'Pedro' is going to suffer the most from it. If I do this at ten tables, I achieve a relevant effect.*
>
> *It's simply groups of thieves [...] like the ones you see on the street. They tear the ballots out."*

This anecdote leads us to define two different strategies of fraud that can be present given that multiple electoral contests are conducted simultaneously: Balanced and unbalanced fraud. In a balanced fraud approach, all races are altered to exactly equal extents. That is, for every vote that is added to (or removed from) one electoral race, another vote is added to (removed from) all parallel contests, even those that are not of primary concern for the agents of fraud or their principals. In an unbalanced fraud approach, these numbers differ and undervoting irregularities emerge.

In the case of Ecuadorian elections, there are straightforward reasons substantiated in repeated reports of international Election Observation Missions to argue that undervoting irregularities are due to poor resources of administrative staff on election day combined with a complex voting system that conducts several contests side-by-side. Likewise, the sheer extent of turnout discrepancies and anecdotes such as the one above speak in favor of a different mechanism in which undervoting irregularities stem from unbalanced fraud approaches in which ballots are selectively added or removed. This chapter outlines a statistical approach to estimate the degree of undervoting irregularities that is due to unbalanced election fraud.

2.3.2 Undervoting Irregularities in the Absence of Unbalanced Fraud

In order to design a method that detects and quantifies unbalanced fraud approaches from undervoting irregularities and the winning candidate's (or party's) vote share, it is first important to understand that there is no expectation of a statistical relationship under human or administrative errors such as misinformed election officials, misinstructed voters, a loss of votes or miscounting. Before I describe a method for fraud detection, I hence first line out two properties of undervoting irregularities under random errors: In expectation, (i) observed vote shares for all candidates are equal to their true vote shares even in the presence of excessive errors and (ii) there is no statistical association between the extent of undervoting irregularities and the winning candidate's vote share across polling stations.

Let N_i denote the number of eligible voters across $i = 1, ..., n$ polling stations. $\mathcal{T}_i \in [0, N_i]$ denotes the absolute number of turned out voters for a particular electoral race of interest and the share of votes the winning candidate (party) received is denoted by $p_i \in [0, 1]$. Across all polling stations, observed turnout levels $\mathcal{T}_i$ and winner's vote shares p_i can be decomposed as

$$\mathcal{T}_i = \mathcal{T}_i^* + \mathcal{T}_i^\epsilon \qquad\qquad \mathcal{T}_i^\epsilon \sim \mathcal{N}(\mu, \sigma^2) \tag{2.1}$$

$$p_i = \frac{V_i}{T_i} = \frac{\mathcal{T}_i^*}{\mathcal{T}_i^* + \mathcal{T}_i^\epsilon} \underbrace{\frac{V_i^*}{\mathcal{T}_i^*}}_{p_i^*} + \frac{\mathcal{T}_i^\epsilon}{\mathcal{T}_i^* + \mathcal{T}_i^\epsilon} \underbrace{\frac{V_i^\epsilon}{\mathcal{T}_i^\epsilon}}_{\epsilon_i} \tag{2.2}$$

where $\mathcal{T}_i^*$ is the true number of turned out voters, $\mathcal{T}_i^\epsilon$ is the absolute number of discrepant votes from the true value either resulting from errors or fraud, V_i^* is the true absolute number of votes cast for the winner and V_i^ϵ is the number of votes for the winner among all lost or miscounted votes and V_i is the number of votes for the winner that is ultimately observed. Under the absence of fraud, the dispersion parameter σ is purely determined by structural factors such as the training of election officials and administrative or election day hurdles. Let us first consider the case in which undervoting irregularities exclusively emerged simply because less votes were cast for a particular electoral contest as turned out voters did not receive or hand back all relevant ballots and there is no miscounting of those ballots that have been received. In this case, the error terms $\epsilon_i (i \in \{1, ..., n\})$ in (2.2) simply reduce to zero and observed vote shares are equal to true vote shares as $V_i^\epsilon = 0$.

A second scenario is given by ballots not getting accounted for, lost or miscounted although these were cast by turned out voters. Using the loss of votes as a working example, it is intuitive that $\epsilon_i = \frac{V_i^\epsilon}{\mathcal{T}_i^\epsilon}$ consists out of a subset of the true votes, and hence it is straightforward that the erroneous votes themselves can be written as a function of the true votes. Let us imagine that for each individual polling station i, there are $j(j \in \{1, ..., J\})$ different hypothetical scenarios in which true votes can be lost, and

$$\epsilon_i^j = \frac{V_i^{\epsilon,j}}{\mathcal{T}_i^{\epsilon,j}} = \frac{V_i^*}{\mathcal{T}_i^*} + \xi_i^j \qquad\qquad \xi_i^j \sim \mathcal{N}(\mu_i, \sigma_i^2). \tag{2.3}$$

ϵ_i^j is denoted as the share of votes for the winner among all lost (or miscounted votes) at a particular polling station, which varies across J hypothetical realizations. In clean elections, not accounted, miscounted or lost votes constitute a truly random sample of the true votes and hence $\mu_i = 0$, $E[\xi_i] = 0$ and $E[\epsilon_i] = \frac{V_i^*}{\mathcal{T}_i^*}$. This means that the expected portion of votes that were cast for the winner among all lost or miscounted votes is equal to the portion of votes cast for the winner among all votes that were originally cast. Reformulating (2.2), we can now straightforwardly derive that at each polling station, the miscount or loss of votes—in expectation—affects all candidates proportionally to their electoral strength in clean elections as

$$\begin{aligned} E[p_i] &= \frac{\mathcal{T}_i^*}{\mathcal{T}_i^* + \mathcal{T}_i^\epsilon}\frac{V_i^*}{\mathcal{T}_i^*} + \frac{\mathcal{T}_i^\epsilon}{\mathcal{T}_i^* + \mathcal{T}_i^\epsilon} E\Big[\frac{V_i^\epsilon}{\mathcal{T}_i^\epsilon}\Big] \\ &= \frac{\mathcal{T}_i^*}{\mathcal{T}_i^* + \mathcal{T}_i^\epsilon}\frac{V_i^*}{\mathcal{T}_i^*} + \frac{\mathcal{T}_i^\epsilon}{\mathcal{T}_i^* + \mathcal{T}_i^\epsilon}\frac{V_i^*}{\mathcal{T}_i^*} \\ &= \frac{V_i^*}{\mathcal{T}_i^*}. \end{aligned} \tag{2.4}$$

Equation (2.4) shows that the vote share that is expected to be observed at a particular locality is equal to the true vote share even in the case of excessive amounts of administrative or human errors.

Furthermore, let u_i be the extent of undervoting observed at one particular polling station when comparing the election of interest to a baseline electoral race and let $u_i = \frac{|\mathcal{T}_i^\epsilon|}{\mathcal{T}_i}$ be the share of votes that are discrepant to a baseline race among the overall number of votes that have been observed in the main race. Across all polling stations, the covariance between the winner's vote share and the extent of undervoting is defined by

$$Cov(p, u) = \frac{\sum_{i=1}^n (p_i - \bar{p})(u_i - \bar{u})}{n} = \frac{\sum \big(\frac{V_i}{\mathcal{T}_i} - \frac{1}{n}\sum \frac{V_i}{\mathcal{T}_i}\big) \times \big(\frac{|\mathcal{T}_i^\epsilon|}{\mathcal{T}_i^* + \mathcal{T}_i^\epsilon} - \frac{1}{n}\sum \frac{|\mathcal{T}_i^\epsilon|}{\mathcal{T}_i^* + \mathcal{T}_i^\epsilon}\big)}{n}. \tag{2.5}$$

From Equation (2.5), it follows that even if large amounts of undervoting irregularities are present that are due to human errors or electoral maladministration, the extent of undervoting is unrelated to the winning candidate's (party's) vote share in expectation as

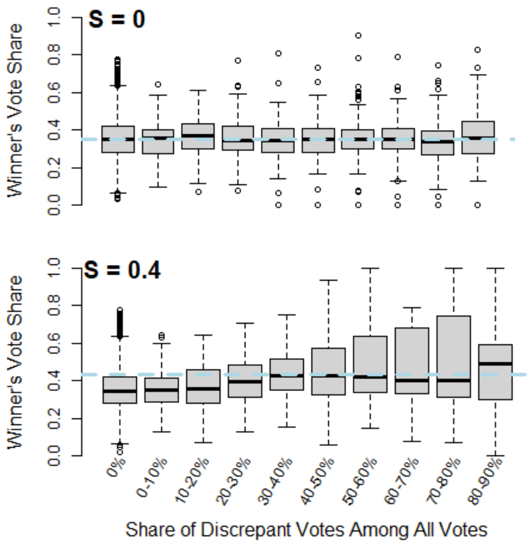


Figure 2.3. The relationship between undervoting irregularities and winner's vote shares (simulated). Boxplots summarize winner's vote shares in artificial data generated under the simulation model outlined in Section 2.4. Data was simulated for 40,000 polling stations out of which 3,000 were assigned undervoting irregularities. S describes the portion of polling stations with undervoting irregularities at which unbalanced fraud was executed. The share of discrepant votes among all votes is defined as $u_i = \frac{|T_i^{\epsilon}|}{T_i}$. The dashed blue line reports the average vote share of the winner (in favor of which results were altered) across all polling stations.

$$\lim_{T_i^{\epsilon} \to \infty} Cov(p, u) = 0. \tag{2.6}$$

Figure 2.3 illustrates this point using artificial election results for concurrent events simulated under the protocol which I outline in the subsequent section. The properties of undervoting irregularities in the absence and presence of systematic manipulation can be nicely illustrated by visualizing the conditional distribution of winner's vote shares separately for different groups of polling stations with different levels of undervoting. In clean elections ($S = 0$), the distribution of winner's vote shares is unrelated to the extent of undervoting as vote shares homogeneously vary around their mean value independently of the extent of irregularities that is observed. When unbalanced fraud is incorporated, distributions are inflated in their upper tail as polling stations document higher levels of irregularities ($S = 0.2$ and $S = 0.4$) and take on substantially above-average values as the amount of fraud that is introduced becomes extreme ($S = 0.8$).

From Equation (2.6), it is tempting to calculate a linear model and infer fraud from a hypothesis test on the relationship between undervoting discrepancies and winner's vote shares. The main shortcoming of a linear model is that the produced measure of association is hard to interpret for the purpose of statistical fraud detection and thus would reduce inferences to a simple significance test.[5] Therefore, I now outline a semi-parametric simulation method that is designed to estimate a substantive quantity of interest—*the share of polling stations with undervoting irregularities at which unbalanced fraud is conducted*, denoted by S. The goal is to verify the validity of the quoted anecdote and to estimate the prevalence of unbalanced election fraud across Ecuadorian elections.

2.4 A Simulation Model to Detect Unbalanced Election Fraud

The goal of the following model is to estimate the share of polling stations with undervoting irregularities at which unbalanced fraud is conducted. In order to make statistical inference on the presence of systematic irregularities across two electoral events for which data is observed, we first need to specify a main electoral race for which the quantity of interest S is estimated and a baseline electoral event to which discrepancies in turnout are quantified. As a general intuition, the semi-parametric simulation model detects unbalanced election fraud by (i) simulating a range of artificial elections (based on empirical input parameters from the observed data) which mimic the main race and are either clean or manipulated to different degrees, (ii) quantifying the average distance between the observed data and each set of simulated elections, (iii) and finding the set of artificial elections that–in expectation—minimizes the distance to the observed data. The fraud parameter that was used to construct this set of artificial elections serves as the estimate of fraud.

2.4.1 Stochastic Process of Concurrent Elections

I model the two concurrent electoral events as

$$T_i^* \sim \text{Binomial}(N_i, t_i^*), \tag{2.7}$$

$$T_i^\epsilon = \begin{cases} 0 & \text{if } D_i = 0 \\ \text{Norm}(0, \sigma) & \text{if } D_i = 1 \end{cases} \tag{2.8}$$

$$V_i^* \sim \text{Binomial}(T_i^* + T_i^\epsilon, v_i^*), \tag{2.9}$$

for each polling station $i = 1, ..., n$. The absolute number of turnout T_i^* is set as the turnout in the baseline race observed at each locality and is defined as a binomial draw where the population size is the number of eligible voters at polling station i (the number of people in the voter register that have been attributed to a particular locality) with the polling station-specific success probability t_i^*. D is an indicator variable documenting whether undervoting is observed at a particular polling station. Turnout discrepancies T_i^ϵ to the main electoral race are set to 0 if $D_i = 0$ and

[5] An example of a linear model predicting winner's vote shares from undervoting irregularities is presented in Table 2 in Section 2.5.

defined as a draw from a normal distribution with a mean of zero and standard deviation σ if $D_i = 1$ where values are rounded to integers. This means that most turnout discrepancies take on small values, while the probability for larger discrepancies is decreasing. The absolute turnout for the main race then—by implication—is $T_i = T_i^* + T_i^\epsilon$. If T_i^ϵ takes on a positive value, this means votes are added to the main electoral race. If T_i^ϵ is negative, less overall votes are observed in the main race than in the baseline event. The number of people who vote for the overall winning candidate (party) V_i^* in the main race is a binomial draw from the population size $T_i = T_i^* + T_i^\epsilon$ (the number of observed votes at a particular polling station) with the success probability v_i^*.

To arrive at a fully specified stochastic process of two concurrent elections, what is missing is to parameterize the functions producing $\{t_i^*, v_i^*\}$ which represent the unknown distributions of polling station-level turnout and winner's support rates. Since these success probabilities necessarily fall in the $[0, 1]$ interval, it is intuitive to model these as beta distributions

$$t_i^* \sim \text{Beta}(\alpha^t, \beta^t), \tag{2.10}$$

$$v_i^* \sim \text{Beta}(\alpha^v, \beta^v), \tag{2.11}$$

where $\{\alpha, \beta\}$ are scale and shape parameters estimated from empirically observed data.

The stochastic process of elections that underlies my simulation model generally relies on two assumptions. Evidently, it assumes that the data generating functions which produce observed turnout T_i^* and winner's votes V_i^* are described reasonably well by binomial draws with success probabilities parameterized by beta distributions and that undervoting discrepancies T_i^ϵ are well approximated by a normal distribution. This assumption is highly reasonable as data generated from these distributions resembles empirical data very closely and has been shown to hold equally well for other country contexts (see Rozenas 2017). Appendix Section A.1 compares data simulated from this stochastic process to empirical data from Ecuador and shows a close fit. A more subtle assumption is that empirically observed votes from the parallel contest that is set as the baseline race can be used to model the (latent) number of turned out voters and number of votes for the winner in the main electoral race T_i^*, V_i^* from Equations (2.1)-(2.2) *before* we introduce undervoting irregularities. It is important to note that this step does not assume that the baseline race itself is fraud-free. Rather, in the following, I present an approach to reverse-engineer the extent of *unbalanced* fraud between both races that is a consequence of intervening into both electoral races to unequal extents. This is not equal to the overall degree of fraud that might be present in the electoral data and does not assume that any of the used data was not manipulated at all. Rather, it exploits the fact that several simultaneous electoral events are taking place and infers the degree of unequal manipulation between the two.

2.4.2 A Semi-Parametric Simulation Model

The semi-parametric simulation model executes the following steps:

1. Set the overall number of polling stations n, the eligible voters per polling station N_i, and the number of polling stations with undervoting discrepancies $n^U = \sum D_i$ to their values in the observed data.

2. Estimate $\alpha^t, \beta^t, \alpha^v, \beta^v, \sigma$ from the observed data.

3. Sample values for T_i^*, T_i^ϵ from the distributions defined in (2.7)-(2.8). Sample values for V_i^* from Binomial(T_i^*, v_i^*) without incorporating undervoting yet.

4. Sample n^U polling stations at which undervoting irregularities are introduced. Set the share of polling stations with turnout discrepancies where unbalanced fraud is conducted to $S \in \{0, 0.02, 0.04, ..., 1\}$.

5. Iterate q times:

   (a) Add T_i^ϵ votes to those polling stations for which $D_i = 1$. Define the number of turned out voters in the main electoral race as $T_i = T_i^* + T_i^\epsilon$. For $n - n^U$ polling stations, the number of votes that is added (removed) from the winner is proportional to the winner's vote share V_i^*/T_i^* before undervoting discrepancies are introduced. For n^U polling stations, votes are added (removed) disproportionally. If $T_i^\epsilon > 0$, add a large share of votes to the winner and allocate the rest of the votes among the remaining candidates. If $T_i^\epsilon < 0$, a large share of votes is removed from the rest of the candidates and only few votes are removed from the winner.

   (b) For each polling station with undervoting discrepancies, construct u_i and p_i.

   (c) Compute the sum of the pointwise squared difference M_q between (p, u) from the main race in the simulated data and the observed data.

Step 1 assures that the actually observed number and sizes of electorates for which data is generated is represented by the model across polling stations. Step 2 assures that the empirically observed dispersion in turnout, winner's votes and turnout discrepancies is represented by the model. After having performed this protocol q times for each possible fraud parameter $S \in \{0, 0.02, 0.04, ..., 1\}$, the estimated portion of polling stations with undervoting irregularities that is supposed to be tainted is

$$\widehat{S} = \underset{S \in \{0,0.02,0.04,...,1\}}{\text{argmin}} \text{avg}(M). \tag{2.12}$$

Although the above protocol may seem complex, it has a very intuitive structure. It simply constructs synthetic data for two parallel electoral contests under the stochastic process defined in (2.7)-(2.9) using empirical input values and then constructs q different fraudulent elections for every fraud parameter in S and computes the distance between each simulated election and the observed data. The fraud parameter that—on average—leads to minimizing the distance between the observed and simulated data is the estimate of fraud $\widehat{S}$, that is, the share of polling stations with undervoting irregularities where unbalanced fraud is assumed to have taken place. The model is semi-parametric because—although resting on parametric distributional assumptions—when reverse-engineering the share of polling stations with undervoting discrepancies that are affected by fraud, it does not assume that the data generating process producing systematic alterations in the first place follows any particular functional form. Rather, the model flexibly finds the set of synthetic voting returns that are most similar to the observed data.

There are two types of uncertainty associated with this method, namely fundamental and estimation uncertainty. Estimation uncertainty simply follows from the fact that for particular values of $\{\alpha^t, \beta^t, \alpha^v, \beta^v, \sigma\}$ constructed in Step 2, statistical sampling is performed throughout Steps 3-5.[6] Uncertainty estimates that take into account estimation uncertainty simply estimate $\{\alpha^t, \beta^t, \alpha^v, \beta^v, \sigma\}$ from the data once using their maximum likelihood estimates, set q to any given value, and iterate over Steps 3-5 many times resulting in one estimate of the fraud parameter $\widehat{S}$ for each iteration. Uncertainty intervals can then straightforwardly be computed from the 92.5% and 97.5% quantiles of $\widehat{S}$. Fundamental uncertainty refers to the fact that the parameters estimated in Step 2 that define the distributions used for sampling themselves are random variables with their true values being unknown. Taking into account uncertainty in the parameters is straightforward in a Bayesian setting, in which the parameters are first parameterized by conjugate prior distributions that put equal weight on the full range of plausible values and Steps 3-5 are then repeated for different posterior draws.[7,8]

In its technical setup, my model shares similarities with both the approach developed by Rozenas (2017) to detect rounding fraud from spikes in the density distribution of winner's votes and turnout and with the simulation model by Klimek

[6]This just means that when executing the protocol several times, individual polling stations will be assigned different numbers of turned out voters, votes for the winner, and degrees of fraud as a result of statistical sampling.

[7]Details on prior distributions can be found in Appendix Section A.2.

[8]This also means that the frequentist approach (incorporating estimation uncertainty) and Bayesian approach (incorporating estimation and fundamental uncertainty) to this method do not differ considerably in their computational efficiency, as both rely on the same number iterations of the algorithm. In a frequentist setting, one simply estimates the parameters $\{\alpha^t, \beta^t, \alpha^v, \beta^v, \sigma\}$ in Step 2 once and executes the algorithm using these fixed values. In a Bayesian setting, one first updates priors for three univariate probability distributions and executes the algorithm using different values for $\{\alpha^t, \beta^t, \alpha^v, \beta^v, \sigma\}$ from different posterior draws in each iteration rather than iteratively using point estimates constructed from the data alone.

et al. (2012) to identify ballot box stuffing from skewness, kurtosis and clusters within the bivariate distribution of turnout and votes for a single race. As Rozenas (2017), I rely on a stochastic model of elections defined as a sequence of binomial draws for the absolute numbers of turnout and winner's votes which are parameterized by beta distributions. Other than Rozenas, I don't define my model as a version of the Bayesian posterior predictive check. Rather, following Klimek et al. (2012), I reverse-engineer the level of fraud by iterating over a sequence of fraud parameters and choosing the one that minimizes a pre-specified distance metric. Methodologically, the method that I outline here hence combines features from different forensic methods that have been proposed in the literature. What makes the approach unique is exploiting the execution of parallel events, the focus on undervoting irregularities, and the novel quantity of interest that is ultimately retained: the share of polling stations with undervoting irregularities in which unbalanced fraud was conducted.

For executing the algorithm, the user has to define the relevant variables for constructing T^*, T^ϵ, V^*, specify the types of uncertainty that should be incorporated when constructing the fraud estimate, set q to an arbitrary large number and define the number of times the algorithm iterates over Steps 3-5. In case parameter uncertainty is supposed to be incorporated in Step 2, the user needs to define the set up for MCMC sampling.[9] Appendix Section A.3 displays an exemplary execution of the function for the Ecuadorian General Elections of 2017.

For the application of the model, a natural question arising is which of the parallel electoral contests should be used as the baseline election. Ultimately, this choice needs to be informed from substantive reasons and there is no statistical fix. The electoral contest that was allegedly fraudulent is set as the main race for which the fraud parameter is estimated, with the parallel contest working as the baseline election. In the case of more than two concurrent elections taking place, the algorithm needs to be repeated for each election pair of interest. The practice of election forensics goes hand in hand with qualitative and on-the-ground observations from political observers and institutionalized electoral observation and is no panacea providing quick answers without the substantive engagement of the researcher.

2.5 Applications

To illustrate and validate the use of the proposed model, I apply it to a range of empirical elections and simulated electoral events for which the degree of fraud is known. The empirical data that I use are from the Ecuadorian General Elections of 2017 as well as the Ecuadorian Local Elections of 2019. All empirical data is available at the level of individual polling stations. In the 2017 General Elections, I

[9] In the laboratory setting reported in Table 1, the model performs well already for a small number of q and few posterior draws such as $q = 100$ and 500 posterior draws.

analyze results for $n = 39,322$ different localities. In the case of the 2019 Local Elections, the number of observations differs across electoral events, as not all regional contests have been held in all provinces. Prior to the analysis of the empirical data, I exclude the rare polling stations which either registered less than $n = 100$ eligible voters or who report more turned out than eligible voters. I exclude the smallest polling stations because extreme percentages for turnout and winner's vote shares easily become artefacts of small electorates.[10] Polling stations with more turned out than eligible voters are excluded in order to facilitate the estimation of latent turnout and support rates through Beta distributions, which force success probabilities to range in the $[0,1]$ interval.

I follow two different strategies to validate the results of the semi-parametric simulation model. For *internal* validation, I apply the method to a range of simulated data sets that mimic two simultaneous electoral events for which the degree of fraud is known. The simulated data follows the same stochastic process as outlined above. In order to ensure that the model works well for the 2017 and 2019 elections that are analyzed, parameters used for the data-generating process are set such that the extent of undervoting and distributions of turnout and winner's vote shares are comparable to those found in the empirical data. I apply the method to a fixed set of five artificial elections. In one of the elections, no fraud is introduced. In the remaining elections, the share of polling stations with undervoting irregularities at which unbalanced fraud is conducted varies between $\{0.2, 0.4, 0.6, 0.8\}$. The synthetic data that is constructed comprises $n = 10,000$ polling stations out of which $n^U = 1,000$ obtain discrepancies in turnout. All applications of the model that are presented incorporate both fundamental and estimation uncertainty and thus report Bayesian credible intervals as uncertainty estimates. Additionally, I *externally* validate the model by contrasting its performance on one set of elections that was accompanied by widespread accusations of fraud and followed up by massive public protests challenging the legitimacy of presented results (General Elections 2017) with another set of elections that were low-key, of significantly less political relevance, and remained largely uncontested (Local Elections 2019).

Figures 8 and 9 re-construct the plots on the relationship between undervoting irregularities and winner's vote shares for the General Elections 2017 and Local Elections 2019 that were introduced on simulated data in Figure 2.3. In the presidential contest, winner's vote shares refer to votes for the government and Correa-endorsed candidate (and ultimately elected president) Lenín Moreno. In the elections for the national and Andean assembly, winner's vote shares refer to the total share of votes that were cast for all candidates that ran for seats representing the government party MPAIS—*Movimiento Alianza País* (national assembly) or the electoral alliance between MPAIS and the Ecuadorian Socialist Party in the Andean elections. In the national referendum on civil cervants' and politicans' bank

[10]For an explicit test of voter rigging in small polling stations, see Jimenez, Hidalgo, and Klimek (2017).

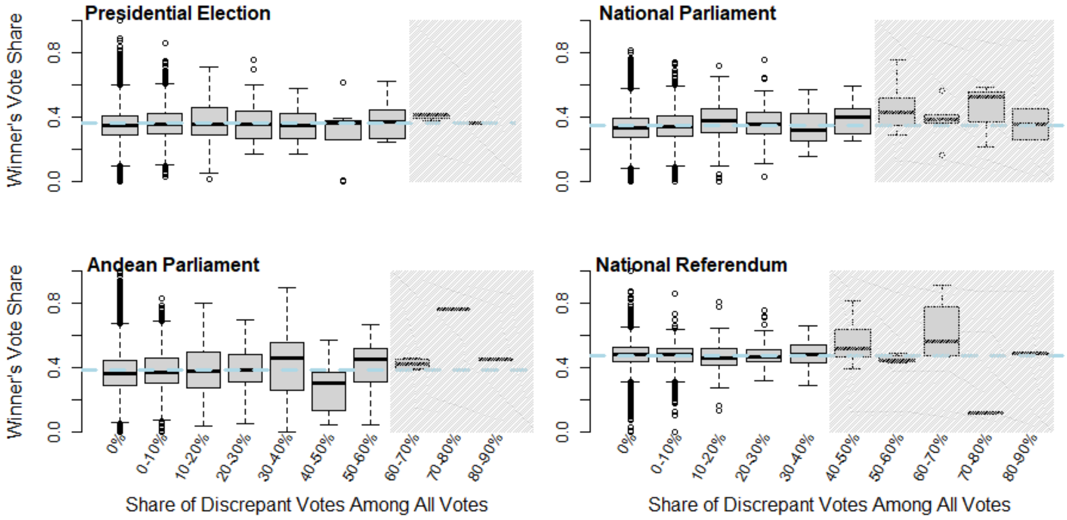


Figure 2.4. The relationship between undervoting irregularities and winner's vote shares (General Elections 2017). In the presidential contest, the winner is defined as Rafael Correa-endorsed candidate and election winner Lenín Moreno. In parliamentary elections, the winner's vote shares are defined as the total vote share of all candidates running for the government party Alianza País (national parliament) or the formed alliance with the Ecuadorian Socialist Party (Andean parliament). In the national referendum, the winner's vote share is the percentage of votes cast for the government-endorsed option.

accounts in international tax havens, winner's vote shares refer to the share of votes cast for the government-endorsed option of accepting the reform that was posted. In all four electoral contests, the government-endorsed options received the largest overall vote share. In the local elections, winners' vote shares are defined as the percentage of votes for the winning candidate in a specific locality.

As can be seen from the figures, across all electoral contests that are portrayed, the distribution of winner's vote shares varies homogeneously around the overall mean value for most extents of undervoting discrepancies. However, different empirical patterns emerge between the 2017 General Elections and the 2019 electoral contests for those localities that report large shares of undervoting. In ghe General Elections 2017, the vote shares of the election winner Lenín Moreno, the candidates of his associated party Alianza País, and the government-endorsed option in the national referendum administered on election day, however, are substantively skewed upwards in those polling stations that reported the most extreme values of undervoting. That is, vote shares for government-endorsed options are highest at those localities where irregularities are most extreme. This pattern does not emerge for the 2019 Local Elections which were not subject to major allegations of electoral manipulation.

The results from the semi-parametric simulation model for the empirically observed data and the batch of simulated elections are summarized in Table 1. Based on the model, undervoting irregularities in Ecuadorian voting returns of 2017 are

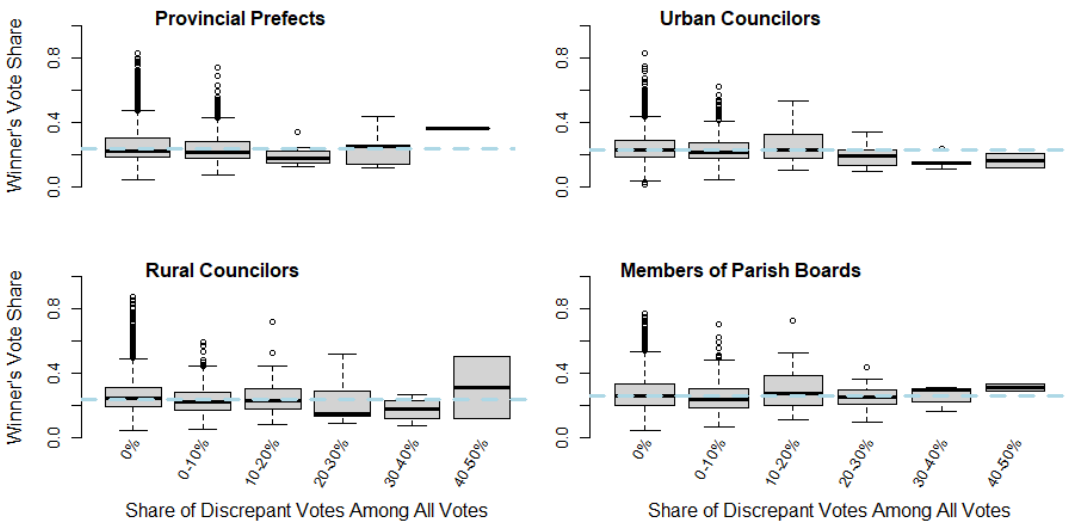


Figure 2.5. The relationship between undervoting irregularities and winner's vote shares (Local Elections 2019). In each electoral event, winner's vote shares are defined as the percentage of votes for the winning candidate in a specific locality.

well explained by unbalanced fraud approaches across the different electoral races. The share of polling stations with undervoting irregularities that is estimated to have witnessed unbalanced fraud ($\hat{S}$) ranges between 18% (elections for the Andean parliament) and 39% (presidential elections), which translates into an estimate of unbalanced fraudulent activity at 394 (or $1,162$) polling stations in the Andean (or presidential) elections. For the presidential elections, from the 95% credible intervals, we can say that with 95% probability this share lies between 24% and 52% of all polling stations with undervoting irregularities. For every of the four electoral events, the 95% credible interval around the estimated share of polling stations with undervoting irregularities that were subject to manipulation does not include the value 0. This means that the distortions in group-specific distributions of winner's vote shares in Figure 2.4 are indicative of unbalanced fraud having interfered with the voting process at a substantial share of localities across the country. When applying the semi-parametric simulation model to the individual electoral contests that formed the 2019 Local Elections, we can see that substantially smaller estimates of $\hat{S}$ are retained. Additionally, 95% credible intervals intersect with 0 for each of the four analyzed events, suggesting that the undervoting irregularities that were observed in the Local Elections of 2019 are not indicative of systematic manipulation.

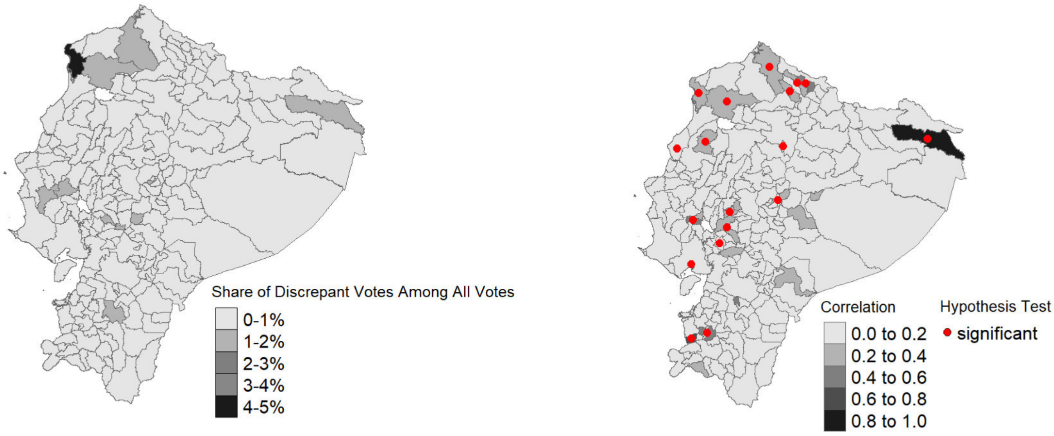
Across the five artificial elections that have been simulated using different degrees of unbalanced fraud, true values for S are reliably reverse-engineered by the model, yielding confidence in the estimates that are constructed for the empirical data.

	IDs	IDs with Undervoting	Estimate ($\widehat{S}$)	95% Credible Interval
Ecuador Local Elections 2019				
Baseline: City Mayors				
Members of Parish Boards	5,129	516	0.13	[0; 0.41]
Rural Councilors	5,530	513	0.11	[0; 0.39]
Urban Councilors	11,021	991	0.10	[0; 0.52]
Provincial Prefects	15,197	1,488	0.19	[0; 0.45]
Ecuador General Elections 2017				
Baseline: Regional Parliaments				
Presidential Election	39,319	2,980	0.38	[0.20, 0.56]
National Parliament	39,319	2,340	0.52	[0.26, 0.80]
Andean Parliament	39,319	2,192	0.54	[0.3, 0.79]
National Referendum	39,319	2,748	0.37	[0.20, 0.56]
Simulated Elections				
0% Fraud	10,000	1,000	0	[0, 0.18]
20% Fraud	10,000	1,000	0.25	[0.08, 0.37]
40% Fraud	10,000	1,000	0.42	[0.21, 0.60]
60% Fraud	10,000	1,000	0.54	[0.35, 0.79]
80% Fraud	10,000	1,000	0.82	[0.63, 0.99]

Table 2.1. Estimates of unbalanced election fraud. Semi-parametric simulation models incorporate fundamental and estimation uncertainty and rely on 100 posterior draws in Step 2 and $q = 50$ iterations of Step 5. Column 'IDs' refers to the overall number of polling stations at which both races were administered. Column 'IDs with Undervoting' refers to the number of polling stations in which undervoting discrepancies are observed in relation to the baseline race. The fraud estimate $\hat{S}$ refers to the portion of polling stations with undervoting discrepancies at which unbalanced election fraud is supposed to be conducted. The last column presents Bayesian credible intervals.

Figure 2.6 gives insight into the geographical distribution of undervoting irregularities in the General Elections 2017 and identifies the regional hotspots in which these are tied to unusually large vote shares for the winner in the presidential race—the most decisive electoral contest. Hence, Figure 2.6 showcases an exemplary follow-up analysis stemming from the results of the simulation model presented in Table 1 which can aid in identifying those localities that drive estimates of unbalanced election fraud and warrant most post-hoc attention by election observers and public electoral administration if the legitimacy of electoral results are contested.

Finally, to put the unbalanced fraud shares retained in Table 1 under further scrutiny, Table 2 reports a robustness test for the semi-parametric simulation model across all polling stations of the country and presents effect estimates of the extent of undervoting on winner's vote shares from linear multilevel regressions while simultaneously controlling for a range of control variables. As can be seen from models M2, M7 and M8, only for few elections unstandardized regression coefficients can be reliably distinguished from zero. This indicates that the semi-parametric



(a) Extent of undervoting irregularities.

(b) Correlations between extent of undervoting irregularities and Lenin Moreno's vote share.

Figure 2.6. Undervoting irregularities and their assocation to Lenin Moreno's vote share, Presidential Election 2017. (a) Map shows the average extent of undervoting irregularities across Ecuadorian cantons (excluding the Galápgos Islands) when comparing turnout in the presidential election to turnout for the baseline election of regional parliaments. **(b)** Map shows within-canton correlations between the extent of undervoting at a particular polling station and Lenin Moreno's vote share. The extent of undervoting at an individual polling station is defined as $u_i = \frac{T_i^{pres} - T_i^{reg}}{T_i^{pres}}$.

simulation model that I propose is more sensitive to detecting systematic irregularities than a parametric linear model. Figure 2.7 visualizes effects from Table 2.

2.5.1 Alternative Explanations

As outlined in Section 2.3 of this chapter, if discrepancies in turnout across multiple electoral races are produced at random due to the miscount or loss of votes, no statistical relationship is expected between the extent of undervoting and the winner's vote share. Through the application of the semi-parametric simulation model and additional statistical analyses, the former section outlined that the empirical patterns that are inherent to Ecuadorian voting returns of the 2017 General Elections are indicative of non-random processes producing undervoting irregularities and are well explained by the mechanism of unbalanced fraud.

However, not all mechanisms that are non-random equal fraudulent activity. Importantly, there is room for alternative explanations that do not evoke fraud which would lead to similar empirical patterns. Reconsidering Section 2.3.1, alternative explanations of high numbers of government vote shares emerging in those localities that report the above-average extents of undervoting can be derived. For instance, it is well documented that Latin America societies are described by an urban-rural divide in education (Queirolo 2013) with low-educated voters being heavily over-represented in rural, poor areas which simultaneously favored

	Dependent variable: Winner's vote share							
	Presidential Election		National Parliament		Andean Parliament		National Referendum	
	(M1)	(M2)	(M3)	(M4)	(M5)	(M6)	(M7)	(M8)
Extent of Undervoting	0.013	0.020*	0.005	0.001	−0.005	−0.002	0.045***	0.046***
	(0.010)	(0.009)						
Closeness of the Electoral Race		0.001***		0.001***		0.001***		0.0001***
		(0.00001)		(0.00001)		(0.00002)		(0.00001)
Number of Eligible Voters		−0.0002***		−0.0001***		−0.0002***		0.00000
		(0.00001)		(0.00001)		(0.00001)		(0.00001)
Percentage Turnout		0.056***		0.061***		0.067***		0.028***
		(0.004)		(0.004)		(0.005)		(0.004)
Percentage Null Votes		−0.120***		0.225***		0.249***		−0.186***
		(0.011)		(0.012)		(0.015)		(0.010)
Percentage Blank Votes		0.089***		0.414***		0.518***		−0.258***
		(0.016)		(0.018)		(0.022)		(0.015)
Constant	0.360***	0.333***	0.342***	0.266***	0.382***	0.300***	0.468***	0.465***
	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.004)	(0.005)
N Polling Stations (N Cantons)	39,319 (251)	39,319 (251)	39,319 (251)	39,319 (251)	39,319 (251)	39,319 (251)	39,319 (251)	39,319 (251)
ICC	0.68	0.61	0.68	0.58	0.59	0.56	0.49	0.50
R Squared	0.68	0.73	0.61	0.64	0.59	0.63	0.49	0.51

Table 2.2. The relation between undervoting irregularities and winner's vote shares, General Elections 2017. *Note: Table presents unstandardized coefficients from linear multilevel regression models with random intercepts across 251 cantons fitted with maximum likelihood estimation. Standard errors are reported in parentheses.* *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*

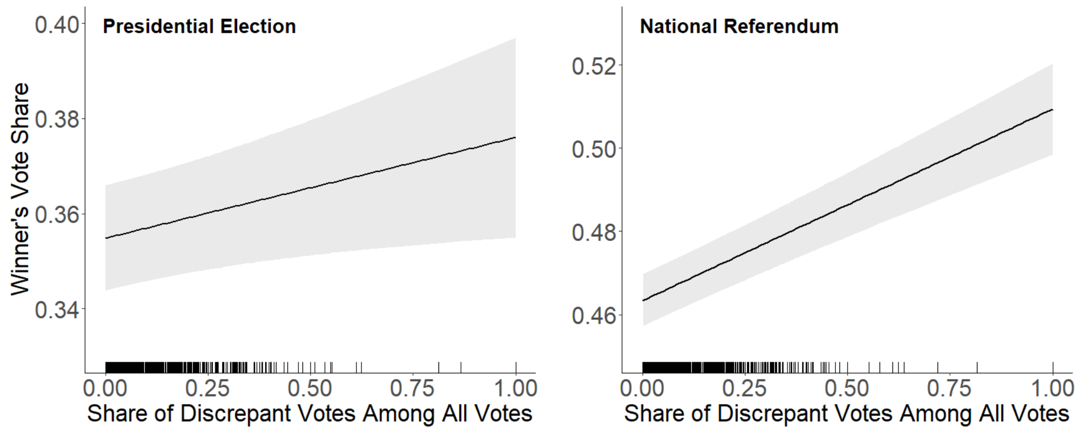


Figure 2.7. The relationship between undervoting irregularities and winner's vote shares, General Elections 2017. The figures plot expected values simulated under models M2 and M8 using 10,000 draws from a multivariate normal distribution defined by the vector of parameter estimates and their covariance matrix. All control variables reported in Table 2 are held constant at their mean values. Shaded regions visualize the 2.5% and 97.5% quantiles of the simulated expected values.

Rafael Correa's left-winged government in 2017. Low-level electoral stuff in rural, low-educated regions which form a sharp discrepancy to the country's urban metropolises might hence be systematically less-capacitated and reside in exactly those localities which predominantly favor government-endorsed options on the ballots. If empirical patterns such as those depicted by Figure 2.4 and the estimates of unbalanced election fraud retained by Table 1 are actually the result of administrative challenges and failures to administer elections with well-trained low-level electoral staff, then large shares of undervoting irregularities should predominantly be produced in rural localities.

Figures 8 and 9 trace empirical evidence for this alternative mechanism and visualize discrepancies between the presidential race as well as the election of members of the national parliament and the baseline election used in Table 1. If undervoting irregularities would predominantly be the result of an urban-rural divide in education and administrative capacity, turnout discrepancies would be over-represented in rural localities. As can be seen from both figures, there is no empirical evidence for an urban-rural divide in undervoting irregularities. While across both elections, rural localities seem to produce lower turnout levels in general, turnout *discrepancies* are not more or less prevalent among those entities that are located in rural regions (colored in red) in comparison to urban cities (colored in green). T-tests assessing differences in undervoting irregularities between urban and rural polling stations remain non-significant ($p>0.05$) across the two electoral contests presented here and the remaining contests presented in Appendix A.4 although these are based on excessive sample sizes. Appendix A.4 also reports further descriptive statistics on the patterns that are displayed in Figures 8 and 9. While there is a systematic relationship between the extent of undervoting and winner's vote shares, there is no geographical clustering of undervoting irregularities

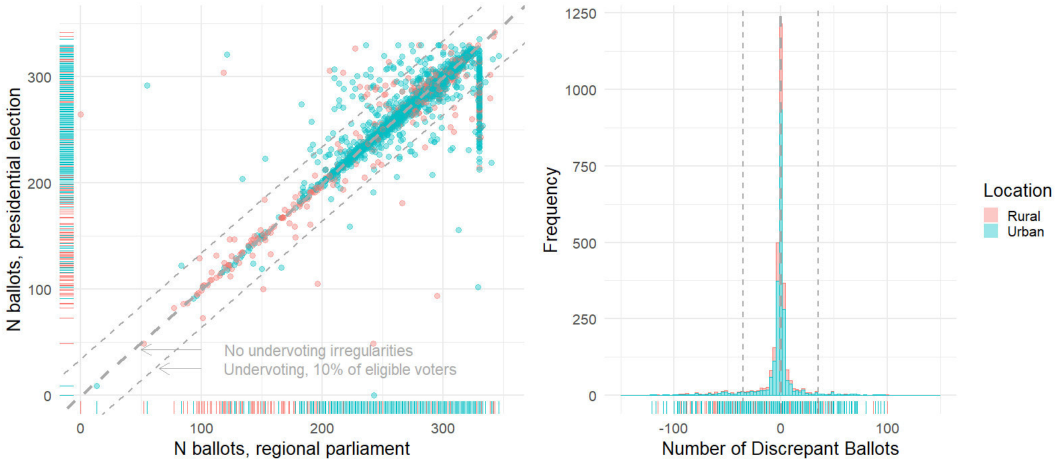


Figure 2.8. The urban-rural divide in undervoting irregularities, General Elections 2017. The left panel plots the absolute number of documented turned out voters for the presidential contest vs. the election of state-level members of parliament. The right panel depicts a histogram of the number of discrepant ballots between the two electoral events. Plots are generated separately for urban and rural localities. There are no significant differences in undervoting irregularities between urban ($n = 29,461$) and rural ($8,992$) localities, $t = 1.32$, $p = 0.188$.

across urban and rural localities.

2.6 Conclusion

This chapter exploited the execution of several concurrent electoral contests on election day for the statistical detection of election fraud. I presented the country case of Ecuador and showed that the execution of simultaneous electoral events can give rise to a phenomenon called 'undervoting irregularities', which occur if the same polling stations document different numbers of turned out voters for different electoral contests. A series of logically equivalent transformations showed that if undervoting irregularities are produced by a random process, all candidates are affected equally by these and there is no statistical expectation of a covariance between the extent of undervoting and the winner's vote shares across localities. Next to describing undervoting irregularities under random processes stemming from limited capacities of low-level election officials, I introduced the systematic mechanism of unbalanced fraud which occurs if protagonists of fraud fail to interfere into multiple electoral races to equal extents.

The chapter proposed a semi-parametric simulation method to detect unbalanced fraud approaches from undervoting irregularities and their relation to winner's vote shares. Using the method, practitioners of election forensics can estimate the share of polling stations with undervoting at which unbalanced fraud has been perpetrated and quantify the uncertainty of estimates under different statistical paradigms.

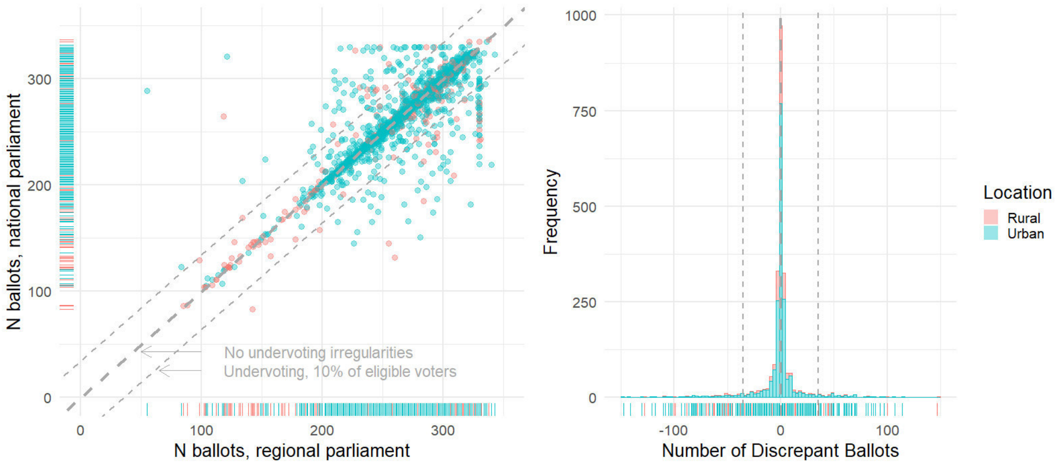


Figure 2.9. The urban-rural divide in undervoting irregularities, General Elections 2017. The left panel plots the absolute number of documented turned out voters for the election of national members of parliament vs. the election of state-level members of parliament. The right panel depicts a histogram of the number of discrepant ballots between the two electoral events. Plots are generated separately for urban and rural localities. There are no significant differences in undervoting irregularities between urban ($n = 29,461$) and rural ($8,992$) localities, $t = 0.84$, $p = 0.40$.

The method that I proposed only focuses on one very specific kind of fraud, namely the unequal manipulation between a main race of interest and a baseline election. This is not equal to estimating the overall degree of fraud that might be inherent to published electoral data as several different mechanisms of systematic manipulation that the model is not designed to pick up might be at place. Vice versa, the method does not assume that the baseline election itself is actually fraud-free. Rather, I present a statistical approach to reverse-engineer the degree of unequal intervention across multiple races.

Lastly, while the empirical patterns that are inherent to Ecuadorian General Elections of 2017 are well explained by unbalanced fraud approaches, it is important to note that there are alternative mechanisms that do not center around any kind of fraudulent activity which can produce similar empirical pictures. Practitioners of the method need to pay close attention to these alternative mechanisms. Ultimately, only careful data analyses providing robust evidence against the existence of alternative mechanisms that go along with the estimates from the semi-parametric simulation model speak in favor of systematic manipulation.

3

Quantifying Systematic Election Irregularities Using Supervised Machine Learning Algorithms

Abstract: The field of election forensics develops statistical methods that are designed to flag peculiarities in fine-graded voting returns which are indicative of election fraud. This study contributes to the application of supervised machine learning methods to the detection of systematic election irregularities. Other than prior contributions which have developed methodology to binary distinguish fraud-free from tainted elections or to quantify the number of precincts where fraud is expected, I present an approach to directly estimate the number of votes that are affected by systematic interference. In a Monte Carlo Simulation study, I confirm unbiased and robust laboratory performance on synthetic data. Additionally, I externally validate the approach and show that controversial elections in Russia and Uganda are robustly labeled as fraudulent estimating that between 4% (Uganda 2011), 8% (Russia 2012) and and 12% (Russia 2011) of votes have been altered. For the case of Russia 2011, this validates field-experimental evidence which arrived at similar figures. On the contrary, electoral results from three Western European democracies are labeled as clean. The results suggest that combining traditional election forensics techniques with modern machine learning approaches is considerably expanding the possibilities for the statistical detection of systematic election irregularities.

Keywords: *Supervised learning; Election Fraud; Electoral Integrity; Election Forensics; Monte Carlo Simulation.*

Author's note: A previous version of this chapter has been presented at the *Latin American Social Science Institute FLACSO* on February 6, 2020 in Quito, Ecuador. I thank all participants of the *Comparative Politics* workshop for fruitful discussions and stimulating ideas.

3.1 Introduction

Public controversies about the integrity of electoral events increasingly evolve around statistical patterns in election results that political observers find hard to explain without invoking fraud. Prominent examples include the discussions around large numbers of polling stations reporting coarse vote shares (Kobak, Shpilkin, and Pshenichnikov 2018) and unusual kurtoses in turnout and vote share distributions (Myagkov, Ordeshook, and Shakin 2009) in Russian *Duma* and presidential elections. Most recently in Bolivia, alleged discontinuity jumps in the incumbent vote share among late-counted votes led president Morales to resign from office under pressure from the military high command and flee into political exile after the country's 2019 presidential contest (Idrobo, Kronick, and Rodríguez 2020; Johnston and Rosnick 2020).

Scholars of political methodology frequently engage in the development of numerical approaches to separate anomalous patterns from fraud-free processes in fine-graded voting returns. Existing approaches in the statistical toolbox exploit unusual patterns in digit distributions of observed vote counts (Mebane 2008; Beber and Scacco 2012; Medzihorsky 2015), spikes in the density mass of turnout or vote share distributions (Kobak, Shpilkin, and Pshenichnikov 2016a; Rozenas 2017), and systematic clusters within the bivariate distribution of turnout and support rates (Myagkov, Ordeshook, and Shakin 2009; Klimek et al. 2012). Some scholars have combined traditional election forensics indicators with machine learning techniques. Cantú and Saiegh (2011) fed characteristics of the distribution of first digits into a naïve Bayes classifier to binary distinguish fraud-free from tainted elections. Levin, Pomares, and Alvarez (2016) and Zhang, Michael Alvarez, and Levin (2019) have used turnout and party-specific vote shares to identify polling stations that are at risk of different fraud mechanisms using tree-based ensemble methods.[1]

Three challenges stand out in the literature. First, the forensic tests that have been developed so far are respectively centered around one individual numerical characteristic of voting returns while being agnostic towards other features that have been successful in identifying fraud. As forensic indicators have been developed as standalone tests that don't inform each other, it is unclear how inconclusive results across distorted frequency distributions, spikes in the density mass of turnout levels, and skewness or kurtoses of turnout and vote share distributions

[1]There is a number of other contributions that employ machine learning technqiues for the purpose of election fraud detection. Cantú (2019b) and Warner et al. (2021) use original image databases of photographed vote tally sheets and statuary forms and employ Convolutional Neural Networks to visually identify irregularities such as inconsistencies in ink, color and handwriting, missing stamps or manually edited results based on human pre-labeled training sets. Montgomery et al. (2015) use Bayesian additive regression trees (Chipman, George, and McCulloch 2010) to predict expert-coded election integrity values for a large array of national elections. While these works do employ machine learning tools for the detection of election irregularities, they are only loosely related to the approach presented here as they don't fall into the category of statistical anomaly detection from numerical data characteristics, but automate qualitative coding decisions that could in principle also be performed by human subjects with sufficient time and effort.

should weigh into substantive conclusions. Second, as a consequence of these missing links, existing approaches are typically restricted towards rather broad statements about the presence and nature of fraud, such as diagnosing whether the election as a whole was free and fair or not (Mebane 2008; Beber and Scacco 2012; Medzihorsky 2015; Cantú and Saiegh 2011), whether evidence for the presence of a specific fraud mechanism can be collected or not (Levin, Pomares, and Alvarez 2016; Zhang, Michael Alvarez, and Levin 2019), and quantifying the number of data entities where fraud is expected to be present (Levin, Pomares, and Alvarez 2016; Zhang, Michael Alvarez, and Levin 2019; Klimek et al. 2012).[2] Directly quantifying the *percentage of votes* that are affected by manipulation is unfeasible for approaches that are screening one specific characteristic of voting returns for anomalous patterns. Third, as several of the methodological approaches are routed in statistical decision theory, scholars have questioned the frequencies of type-I (Deckert, Myagkov, and Ordeshook 2011) and type-II (Mack and Stoetzer 2019) errors of conventional forensic tests, which limits their applicability in real-life settings.

I speak to these three shortcomings by fusing existing election forensics indicators with a supervised machine learning approach. First, I define a protocol for simulating realistic micro-level electoral returns that resemble empirical data across a range of numerical characteristics rather than one isolated pattern. Subsequently, I train random forest regression trees, a flexible non-parametric learner, on a multivariate feature space that takes into account characteristics of digits, turnout, and vote share distributions simultaneously and provides reliable estimates of the *percentage of votes* that have been tampered. I first assess the performance of trained models in a Monte Carlo simulation study in which the degree of fraud within synthetic data is known. Finally, I illustrate and externally validate the approach on national-level elections that have been contested both publicly and in the academic literature from Russia 2011-2012 and Uganda 2011 as well as on an array of Western European democracies.

The main contribution of this study is to provide a unified statistical framework to evaluate electoral returns against different types of numerical anomalies that have been identified in fraudulent elections. This framework (i) allows to relate different forensic indicators to each other and quantifies their relative contribution in labeling an election as fair or foul (ii) and enables the direct estimation of the percentage of votes that are affected by fraudulent interference. Additionally, the proposed approach promises to control type-I and type-II error rates. On the one hand, this is because non-parametric models trained on simulated data are not bound to universally defined null distributions underlying conventional statistical tests which are agnostic to the characteristics of an empirical case. Rather, by learning from a large number of artificially simulated elections under the electoral

[2]This last proxy may be informative with high-resolution data down until the level of individual polling stations, but becomes increasingly vague and difficult to interpret as available data was aggregated to higher-level entities such as districts or precincts.

system of a specific country, models flexibly learn the case-specific level of data distortions that go along with certain degrees of fraud *given* the attributes of the empirical case under study. On the other hand, this is because voter behavior that is known to produce type-I errors for individual features such as strategic voting for digit tests (Hicken and Mebane 2017) or unequal voter mobilization for the kurtosis of turnout distributions (Klimek et al. 2012) are counter-weighted by alternative features in a unified statistical framework which merges a large number of numerical attributes. In the fraud detection prototype that is presented here, one anomalous pattern is not enough to label a collection of voting returns as tainted. Rather, the interplay and dependencies of different forensic indicators are taken into account as a whole.

The remainder of this chapter is structured as follows. I Section 3.2, I first provide an overview of three of the most prominent numerical approaches to identifying systematic irregularities in fine-graded voting results and motivate the need for multivariate learning from these. Section 3.3 outlines the protocol for synthetic data generation that resembles empirically observed data from a variety of countries across a range of dimensions and can incorporate different mechanisms of fraud. Section 3.4 describes the approach to combine supervised statistical learning with the different types of numerical indicators that were introduced. Finally, I validate and showcase the approach on synthetic data and a range of elections from Russia, Uganda, Austria, Finland and Spain.

3.2 Motivation for a Unified Statistical Framework

3.2.1 Numerical Characteristics in Electoral Returns

To provide an overview of different election forensics indicators and to motivate the approach presented here, I consider a range of electoral returns from national-level elections across the five country cases of Austria, Finland, and Spain as well as Russia and Uganda. Datasets cover a heterogeneous set of cases as they comprise parliamentary as well as presidential elections across proportional and majority voting rules and inherit between $n = 992$ (Finland 2017) and $n = 91,256$ (Russia 2012) electoral entities made available at different levels of aggregation by national election officials.[3] While in the cases of Austria, Finland and Spain, election integrity has widely been acknowledged by political observers, the validity of the national-level elections in Russia (2011, 2012) and Uganda 2011 has been put into

[3]The elections included are: Austria 2008 parliamentary election, Spain 2019 European parliament election, Finland 2018 municipal election, Russia 2011 parliamentary election, Russia 2012 presidential election, Uganda 2011 presidential election. Table B.1 in the Appendix provides an overview and descriptive statistics on the data that is used.

doubt both by public protests as well as in the academic literature.[4] I first outline three numerical regularities and how these are distorted under the presence of fraud.

The first characteristic refers to the distribution of numerals in the second and last significant digit of raw vote totals for different candidates (parties) across a large number of electoral units.[5,6] For a large class of data generating processes that include the composition of electoral voting returns, well-grounded explanations exist that these distributions are far from random but can be described by a predefined pattern. Specifically, Newcomb-Benford's law (Newcomb 1881; Benford 1938) states that for suitable processes, the probability that the *first* significant digit is d ($d \in 1, 2, .., 9$) decays as an inverse-logarithmic function. For *subsequent* digits, which are of interest here, Hill (1995) and Hill (1996) provided a generalized version of the law postulating that the frequency of numbers d ($d \in 0, 1, 2, .., 9$) arising in the nth position ($n > 1$) can be defined as[7]

$$P(d) = \sum_{k=10^{n-2}}^{10^{n-1}} log_{10}(1 + \frac{1}{10k + d}). \tag{3.1}$$

The law was initially known to apply to scale-invariant data composed of units that can be mapped across several orders of magnitude such as dollar amounts, distances, or weights (Pericchi and Torres 2011). Evidently, the raw vote totals for a specific candidate (party) that are observed across electoral units are not of this type, which has lead researchers to doubt whether the law is applicable to voting data at all (see, for instance, The Carter Center 2005). However, for unit-free data—such as the number of votes—alternative justifications exist and have been derived by Hill (1995, 1996) in his 'central limit theorem for significant digits' long before the rise of election forensics. As Hill (1995, 1996) has shown, Equation (3.1) above holds asymptotically if observed numbers are generated as mathematical mixtures of different distributions without being naturally biased towards a certain range of values. That is, naturally observed vote totals that don't inherit manual manipulation are expected to follow Newcomb-Benford's law if these can be thought of as random samples that are not taken from one, but combined from many individual probability distributions. One source of heterogeneity was suggested by Mebane (2006), who argues that votes can be thought as stemming from hierarchical mixture population models, in which at each electoral unit, at least two populations should

[4]For Russia 2011-2012 see Kobak, Shpilkin, and Pshenichnikov (2016a), Rozenas (2017) and Enikolopov et al. (2013). For Uganda 2011 see Klimek et al. (2012).

[5]An 'electoral unit' is defined as the officially reported unit of observation at the lowest level of aggregation in voting returns that is made public by election officials. Units might represent individual polling stations, districts or precincts.

[6]The first significant digit of a number (also described as the "leading digit") can be defined as its non-zero leftmost digit. Hence, the second significant digit of 350 is 5 and the second significant digit of 0.052 is 2.

[7]Table B.2 in the Appendix provides the expected frequencies for first, second, and third significant digits under these propositions.

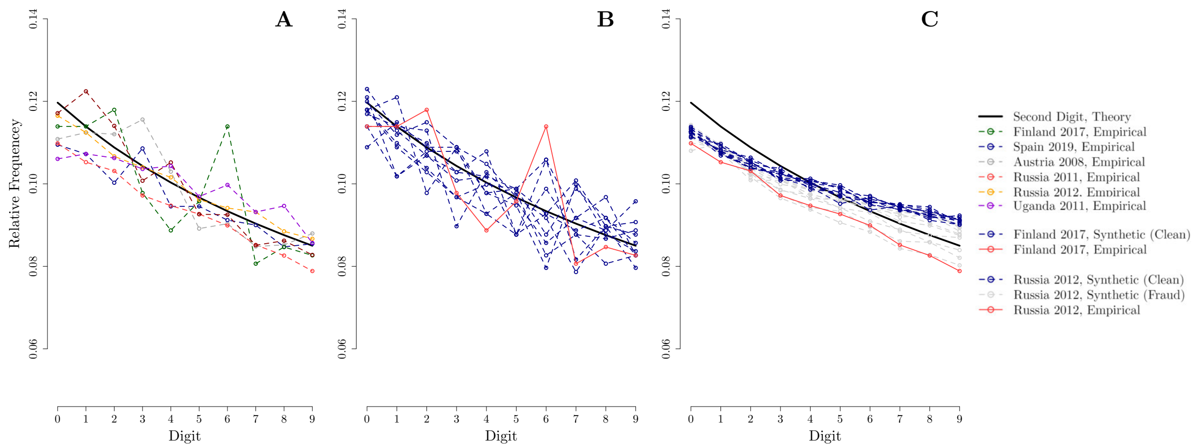


Figure 3.1. Second significant digit distributions for winner's votes, cross-country comparison. A: Generalization of Newcomb-Benford's Law for the second significant digit (black) plotted against empirical distributions from raw vote totals of the elections' winning party (candidate). B: Empirical data from Finland 2017 (red) plotted against synthetic data simulated from ten clean elections (darkblue). C: Empirical data from Russia 2012 (red) plotted against synthetic data generated without data manipulation (darkblue) and different levels of election fraud (grey). Empirical data from Finland 2017 is in line with clean processes. For Russia 2012, data generated with manipulation provides a considerably closer representation of the empirical distribution than simulated data without manipulation.

be present: Those voters strongly in favor of a candidate and the general population switching between candidates. Another source of heterogeneity stems from the process of data aggregation, in which election officials combine data from many individual tables, polling stations or low-level administrative units to produce publicized results at the level of individual districts. Certainly, the more heterogeneity is incorporated, the better will the data satisfy the formulations of the law. Furthermore, the closest approximation will be provided for distributions based on large sample sizes ($n \to \infty$) for which the mean is greater than the median and the data exhibits positive skewness (see Cantú and Saiegh 2011, p. 416).

These propositions come with at least two implications for their applicability to voting returns. First, numerical regularities for the *first* significant digit are easily violated for electoral units with relatively constant sizes without much data aggregation. For instance, if data from a two-party system is present on the level of individual polling stations each comprising a fixed number of voters (say 500), arising vote totals will fail to span the varying orders of magnitude necessary for first digits to follow a logarithmic decay[8], while subsequent digits are considerably less affected. This is why first digits may be useful for specific electoral system designs where the number of eligible voters per electoral unit spans several orders of magnitude (see Cantú and Saiegh (2011) for an application to historical voting returns from Argentina), but is unsuitable for generalized cross-country applications. Second, as the law applies to data that is the result of statistical mixtures,

[8]Naturally, first digits will almost exclusively range between 1,2,...,5.

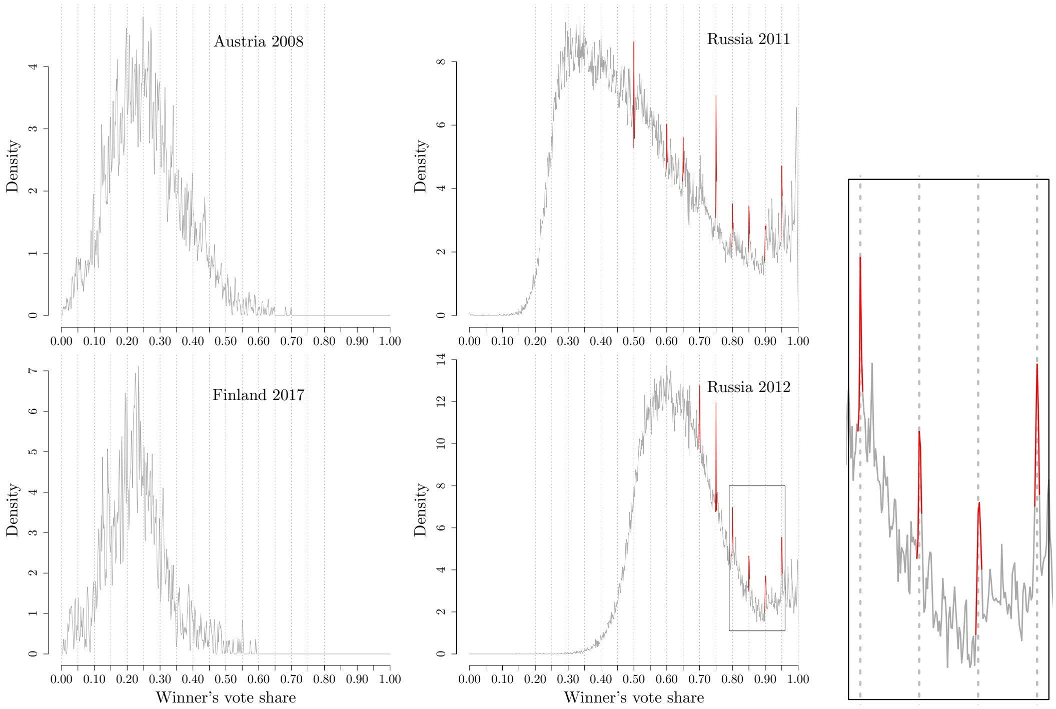


Figure 3.2. Estimated density (Gaussian kernel) of the winning party's (or candidate's) vote share. Austria 2008 ($n = 2,535$, bandwidth=0.001), Finland 2017 ($n = 992$, bandwidth=0.001), Russia 2011 ($n = 90,919$, bandwidth=0.0001) and Russia 2012 ($n = 91,256$, bandwidth=0.0001). Spikes in the density mass around coarse shares (multiples of 5 and 10) colored red for Russian elections. Values at 100% not shown.

data from the lowest level of data aggregation might be approximated worse than data that stems from aggregation. On the other hand, higher levels of data aggregation lead to a smaller number of data points that define the distribution, which naturally lets data deviate from the (generalizations) of Newcomb-Benford's law which is defined asymptotically. Choosing the right level of data aggregation is hence a trade-off between the degree of mixing and the resulting sample size.

Figure 3.1A (left) presents a graphical representation of the expected distribution of numbers in the second significant digit (black) against empirical distributions from Austria, Finland, Spain, Russia and Uganda. At first sight, the empirical distribution of numbers within the second significant digit strongly adheres to its expectation from Equation (3.1).

In order to exploit the distribution of numerals within different digits, scholars usually test whether the empirically observed distribution differs significantly from its theoretical expectation stated in (3.1) using an χ^2-test ($df = 9$)[9], or testing deviations from particular empirical implications of (3.1) for significance, such as the mean of the last digit being 4.5 (see Hicken and Mebane 2017).

[9]The test is formally defined as $\chi^2_n = \sum_{i=0}^{9} = \frac{(d_i - d_i^*)^2}{d_i^*}$ where d_i is the empirical frequency of a certain numeral in the nth digit and d_i^* is its theoretical expectation. The critical value against which the χ^2-statistic is evaluated for $df = 9$ is 16.92 at a significance level of 5%. This conventional test is electoral system- and context-agnostic.

The second characteristic that I focus on relates to skewness, kurtosis and clusters in the distribution of turnout and its bivariate distribution with the vote shares of the winning party (candidate). Considerations around turnout rates stem from the empirical observation that raw turnout and vote shares—although at times exerting positive levels of skewness and kurtosis— often closely resemble Gaussian distributions for elections that are clean.[10] As initially noted by Myagkov, Ordeshook, and Shakin (2009) and popularized by Kobak, Shpilkin, and Pshenichnikov (2018), several mechanisms of fraud such as ballot box stuffing and deliberate wrong-counting leads to inflations of the distributions' right tail, with extreme forms of tampering producing clusters in the upper quintiles. Klimek et al. (2012) have proposed a method for reverse-engineering levels of incremental and extreme fraud by modeling turnout and vote shares with two orthogonal Gaussian distributions and finding mechanisms of fraud that most closely resemble skewness, kurtosis and clusters between modeled and empirical distributions. Also, Levin, Pomares, and Alvarez (2016) and Zhang, Michael Alvarez, and Levin (2019) have used features of turnout and vote share distributions for the case of Argentina's 2015 general election.

Figure 3.6 (upper panel) plots data from three elections of Austria, Spain and Finland and Figure 3.7 (upper panel) replicates this plot for three elections from Russia and Uganda. As can be clearly seen, while the elections that are supposedly clean are well approximated by multivariate normal distributions, the latter distributions are inflated in their right tail and inhibit visible patterns of distortions that cluster around turnout and vote share levels above 90%.

The third characteristic that I focus on stems from the observation that in election data that is supposedly tainted, the fraction of coarse integer percentages around turnout and votes share values is often considerably higher than what would be expected by pure chance, a phenomenon that appears if vote shares for the winner have been rounded up to meet certain target values. This feature has first been identified by Kobak, Shpilkin, and Pshenichnikov (2016a) in the history of Russian national-level elections in the period from 2004 onwards. While Rozenas (2017) outlines that a sample of vote shares from a set of precincts is likely to exert noticeable spikes in the density mass at lower-order fractions even in the absence of any interference, both Kobak, Shpilkin, and Pshenichnikov (2016a) and Rozenas (2017) provide contributions to estimate whether their frequency exceeds the expected range of values. In Figure 3.2, I contrast estimated densities from Austria 2008, Finland 2017 and two Russian elections for the distribution of the winning party's (or candidate's) vote shares across all electoral entities. As is clearly noticeable, vote shares for United Russia (2011) and Vladimir Putin (2012) spike around exactly integer percentages that are multiples of '5' (55%, 60%, ..., 95%), while this pattern is not inherent to any of the two former elections.

[10] As Borghesi and Bouchaud (2010) have shown, rescaling raw distributions to represent logarithmic vote rates provides an even closer fit to normality.

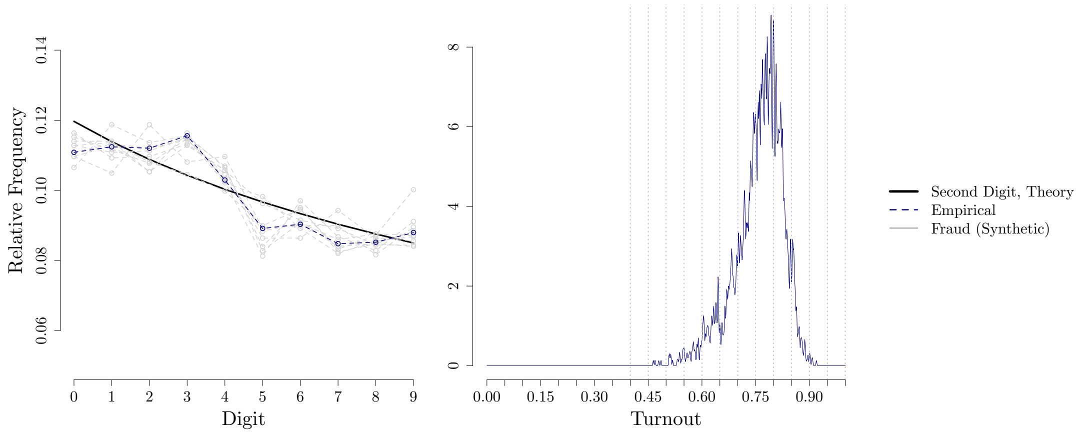


Figure 3.3. Comparison of empirical and artificially manipulated data from Austria 2008. Votes where switched from the *Österreichische Volkspartei (ÖVP)* towards the *Sozialdemokratische Partei Österreichs (SPÖ)*. Left figure shows distributions of second significant digits from empirically observed data of the SPÖ (blue) and electoral returns that have been frauded to different degrees (grey) against Newcomb-Benford's law (black). Right figure plots the distribution of turnout levels across all electoral units for empirical (blue) and frauded data (grey). Vote switching affects the distribution of digits, while turnout levels of empirical and frauded data are identical.

3.2.2 Motivation for Multivariate Learning from Synthetic Data

The main motivation for applying multivariate supervised machine learning techniques for election fraud detection is two-fold. For the first reason, reconsider Figure 3.1 which plots distributions of second significant digits for various country cases against their theoretically expected values. While the empirical distributions of second digits in Panel A approximately resemble their theoretical expectation at first sight, note that the degree of fit is *not* predominantly determined by political observers' and the academic literature's stances on the elections' fairness, but predominantly a function of the number of electoral units for which data is present in the first place, with Finland ($n = 992$) showing the largest deviation and Russia 2012 ($n = 91,256$) showing the closest approximation. This characteristic is routed in the statistical foundation of Newcomb-Benford's law which is defined asymptotically, and implies that larger sets of numbers will naturally provide closer fits than smaller sample sizes. Hence, without simulating the natural variability of digit distributions under clean and manipulated processes *given* the electoral system, the sizes and number of electoral units at hand, it is impossible to know which of the two distributions is a stronger indication for resemblance or violation of the law and whether a violation is to be detected at all.

Panel B contrasts the empirical distribution of Finland 2017 against distributions from ten simulated clean elections, which have been generated under the data generating protocol outlined in Section 3.3. Although the empirical deviation is largest in the cross-country comparison of Figure 3.1A, it is perfectly in line with the natural variation that we expect to see for the number and specific sizes of electoral units in Finland's electoral system. On the other hand, the comparatively

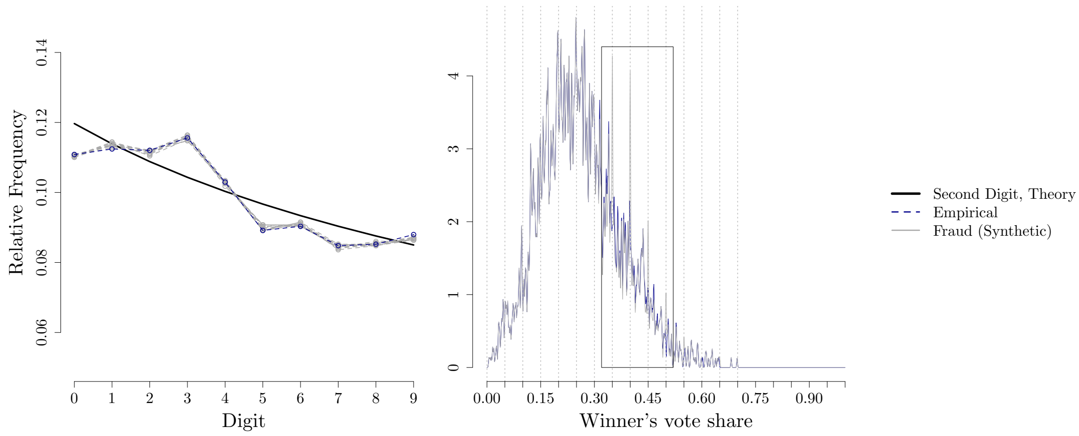


Figure 3.4. Comparison of empirical and artificially manipulated data from Austria 2008. Vote shares of the *Sozialdemokratische Partei Österreichs (SPÖ)* have been rounded up in 2% of locations. Left figure shows distributions of second significant digits from empirically observed data of the SPÖ (blue) and from ten artifically adapted electoral returns with 2% of contamination (grey) against Newcomb-Benford's law (black). Right figure plots the distribution of vote share values of the SPÖ across all electoral units for empirical (blue) and frauded data (grey). Rounding fraud sharply influences the distribution of vote shares, while digit distributions remain largely unaffected.

minor deviation that is observed for the distribution across $n = 91,256$ electoral units from Russia 2012 is clearly out of range of clean processes. When simulating data for Russia (see Section 3.3) and comparing the empirical distribution (red) to clean (blue) and tainted (grey) elections, the empirical distribution is considerably better approximated by simulated data incorporating varying degrees of data manipulation (Figure 3.1C).

The first motivation for the application of supervised learning algorithms trained on synthetic data is that models learning data distributions under a given electoral system (including the number and sizes of electoral units) flexibly take this natural variation into account. Rather than pre-defining expected values as in Equation (3.1), expected values under clean elections and manipulation are learned separately for each electoral system design at hand.

The second reason is presented in Figures 3.3 and 3.4. Essentially, it can be shown that different mechanisms of fraud affect the indicators that have been presented above in different ways, either by distorting some numerical regularities to a much larger extent than others or by not affecting some patterns that are screened for in the election forensics toolkit at all, wrongly indicating that the election was clean. To showcase this property, I used empirical data observed from the parliamentary election in Austria 2008 across $n = 2,535$ electoral units. For Figure 3.3, in between 1% and 30% of these units, votes have been taken away from the *Österreichische Volkspartei (ÖVP)*—the party with the second largest vote share in the election— and shifted towards the winning *Sozialdemokratische Partei Österreichs (SPÖ)* to different degrees (vote switching). As can be seen from the Figure, different magnitudes of fraud noticeably distort the distribution of second significant digits for the *SPÖ*. On the other hand, the turnout levels across affected units stay *exactly* the same, as no single ballot was added or removed, but only wrong-counted. Likewise, Figure 3.4 displays a scenario in which across all units, 2% of vote shares for the winning party have been rounded up to their closest integer multiple of '5', which displays rapidly in the density distribution of vote shares for the *SPÖ*, but leaves digits (almost) unaffected.

It can be argued that this is a positive feature since the detection of some distortions (but not others) will provide us with insights not only about the presence, but also the concrete nature of fraud. While this is true, the fact that individual indicators are constructed as standalone tests that don't take the simultaneous regularities or distortions of other numerical features into account, hinders our inference in a number of ways. First, the standalone application of different screening tools that are agnostic to each other easily yields an inconclusive picture. Second, simultaneously taking features of digits, coarse percentages, skewness, kurtosis, and clusters in the distribution of turnout and vote shares into account serves as a mechanism of avoiding false-positive statements as isolated anomalies in individual features are counter-weighted by alternative numerical characteristics. Third, relying on the interplay of heterogeneous characteristics within electoral returns enables the direct estimation of the number of votes that are affected by fraudulent interference, yielding a more nuanced approach to statistical fraud detection.

3.3 Synthetic Data Generation

The first step to training machine learning algorithms to directly estimate the number of manipulated votes from a whole range of forensic indicators is to create artificial voting returns that resemble the empirical characteristics of clean and frauded data across all relevant attributes. This goes beyond prior approaches in designing statistical methodology for election fraud detection which center synthetic data generation around the specific statistical pattern that they study and try to exploit (see Klimek et al. 2012; Kobak, Shpilkin, and Pshenichnikov 2016a; Rozenas 2017; Levin, Pomares, and Alvarez (2016); Zhang, Michael Alvarez, and Levin (2019)).

3.3.1 Data Generating Methodology

Clean Data

The synthetic data that are constructed under a given empirical case take on the form of $n(i = 1, ..., n)$ electoral units on the micro-level (polling stations, for instance) with N_i number of eligible voters per unit. I consider the problem of choosing between two candidates c (or parties) $(c \in \{A, B\})$ where fraud can happen in

Country	Year	Election	f_i	f_e	f_s	c
Austria	2008	Parliamentary election	0	0	0	NA
Finland	2017	Municipal election	0	0	0	NA
Spain	2019	European Parliament election	0	0	0	NA
Russia	2011	Parliamentary election	0.32	0.1	0.01	1.5
Russia	2012	Presidential election	0.35	0.09	0.01	1.5
Uganda	2011	Presidential election	0.49	0.02	0	1.5

Table 3.1. Parametrizations for synthetic data generation. Values have been chosen such that simulated elections closely resemble empirical patterns.

favor of both camps.[11] I first line out how synthetic data can be tailored to mimic clean elections, before fraud is introduced. To simulate data for n electoral units resembling one empirical election, the following protocol is applied.

1. Set n and N_i to their empirical values.

2. Across the n entities, both turnout t and the winner's vote share s are respectively defined as $\mathcal{N}(\mu_t, \sigma_t)$ and $\mathcal{N}(\mu_s, \sigma_s)$ where the parameters are estimated from empirical data using

$$\mu_t = t_{(n+1)/2}, \quad \sigma_t = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - \bar{t})^2_{t_i<q}} \quad , \tag{3.2}$$

$$\mu_s = s_{(n+1)/2}, \quad \sigma_s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(s_i - \bar{s})^2_{t_i<q}} \quad . \tag{3.3}$$

q is the turnout share level that defines the 75% quantile of the empirical turnout distribution t. Note that the means of the synthetic distributions are specified via the *median* values of empirical distributions and the distributions' dispersion parameters are estimated using only those entities that belong to the first three quartiles of the empirical turnout distribution. When simulating data for empirical cases that are supposedly clean, this comes with little sacrifice as median values closely resemble empirical means and constrained standard deviations closely resemble empirical standard deviations. However, when simulating clean data under electoral systems for which supposedly only tainted empirical data is available, medians and constrained dispersion parameters are substantially less affected by ballot box stuffing or vote stealing which smears out the right tail of turnout and vote share distributions. To generate data points, n values are sampled from these distributions and

[11] In elections with more than two eligible candidates (parties), it is feasible to define the election's winner and runner-up as $\{A, B\}$.

only values between 0 and 1 are accepted in order to truncate each distribution towards the acceptable range of values.

3. Keeping N_i and the sampled values for t_i fixed, construct m random permutations of winner's vote share values s and find the permutation that minimizes

$$\min_m \sum_{d=2}^{D} KL(P_d||Q_d) = \min_m \sum_{d=2}^{D} P_d(x)\ log \frac{P_d(x)}{Q_d(x)}, \tag{3.4}$$

where $\sum_{d=2}^{D} KL(P_d||Q_d)$ is the sum of the Kullback-Leibler distances between the frequency distributions of synthetic digits P_d and theoretically expected digits Q_d as specified by Equation (3.1) for second ($d = 2$) and last significant digits. The number of permutations to evaluate can be set freely. In practice, already values of $m = 100$ work reasonably well across many sensible choices of n.

Each of the n electoral units of size N_i hence observe an overall turnout level of t_i, of which a share of s_i observations vote for the winning candidate or party. The first step of the protocol ensures that the actual number of electoral units n and electorate sizes N_i represent empirical values. Explicitly incorporating the electorate sizes N_i for each unit additionally ensures that the fraction of exactly coarse vote shares across the n units is representative. The second step ensures that the distribution of turnout and vote shares are in line with the empirical patterns that are widely observed for elections that are supposedly clean. Finding the optimal permutation of sampled winner's vote share values in the third step assures that synthetically generated raw vote totals closely follow theoretical expectations as specified by the generalizations of Benford's law in their second and last digit given the distribution of electoral unit sizes N_i and overall turnout levels t_i.

Introducing Fraud

Introducing manipulation into data that has been created under fair processes can be performed in manifold ways. For instance, one of the best documented historical instances of widespread fraudulent interference into elections has been compiled for the province of Buenos Aires during Argentina's 'infamous decade' 1931-1941 (Ciria 1974), in which the ruling Conservative coalition extensively used voter coercion and intimidation, tampered voter registers, restricted access to polling stations, and stuffed ballot boxes in order to maintain power from Socialist contenders (Cantú and Saiegh 2011). The historical evidence that has been compiled is suggestive of widespread acts of ballot fraud that has been conducted across almost every polling station across Buenos Aires in the form of 'vote switching', that is moving registered votes from the opposition towards the Conservative party (Cantú

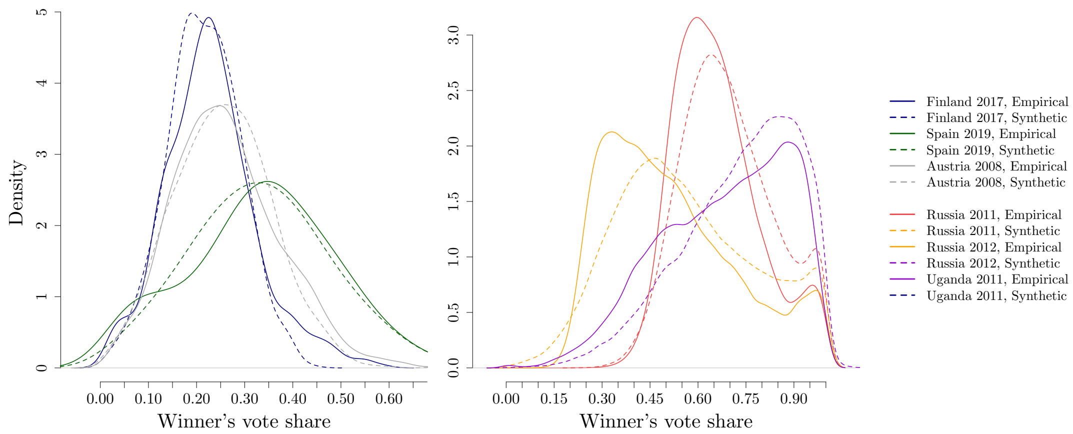


Figure 3.5. Comparison of the distribution of winner's vote share values across all electoral units between empirical and synthetic data. Figures plot estimated densities (Gaussian kernels) with bandwidths set according to the number of electoral units for each country. Left panel generated from synthetic data with no data manipulation. Right panel generated from synthetic data with pre-specified levels of fraud. Empirical and synthetic distributions are almost identical.

and Saiegh 2011, p. 216). In recent Russian electoral history, evidence is suggestive of mechanisms described as 'ballot box stuffing', where urns are getting filled with ballot sheets of voters that never showed up at the polling station in the first place, inflating levels of turnout (Klimek et al. 2012). As Klimek et al. (2012) argue, observed skewness and clusters are indicative of one mechanism of 'incremental fraud', in which turnout levels are adjusted across a wide range of polling stations to a small degree, and 'extreme fraud', in which almost all votes are counted in favor of the winning party at few selected places. Statistical patterns that are suggestive of these mechanism have also been documented in electoral returns from Uganda's publicly contested 2011 presidential election (Klimek et al. 2012).

A generalized fraud detection methodology that holds value across heterogeneous contexts needs to be adaptive to each of these scenarios. Departing from the protocol outlined above, data manipulation is introduced as a three-step process.

4. Randomly select a fraction of f_i electoral units at which incremental fraud takes place, a fraction of f_e electoral units at which extreme fraud takes place, and a fraction of f_s electoral units at which vote shares are rounded up for either of the two candidates (parties) $\{A, B\}$.

5. Across $n \times f_i$ electoral units, incremental fraud is defined as $\mathcal{N}(f_i^c, \sigma_i)$ and across $n \times f_e$ electoral units, extreme fraud is defined as $1 - \mathcal{N}(f_e^c, \sigma_e)$ where c is an exponent and

$$\sigma_i = \sigma_e = \sqrt[4]{\frac{1}{n}\sum_{i=1}^{n}(s_i - \bar{s})^2_{s_i > q}} \;, \tag{3.5}$$

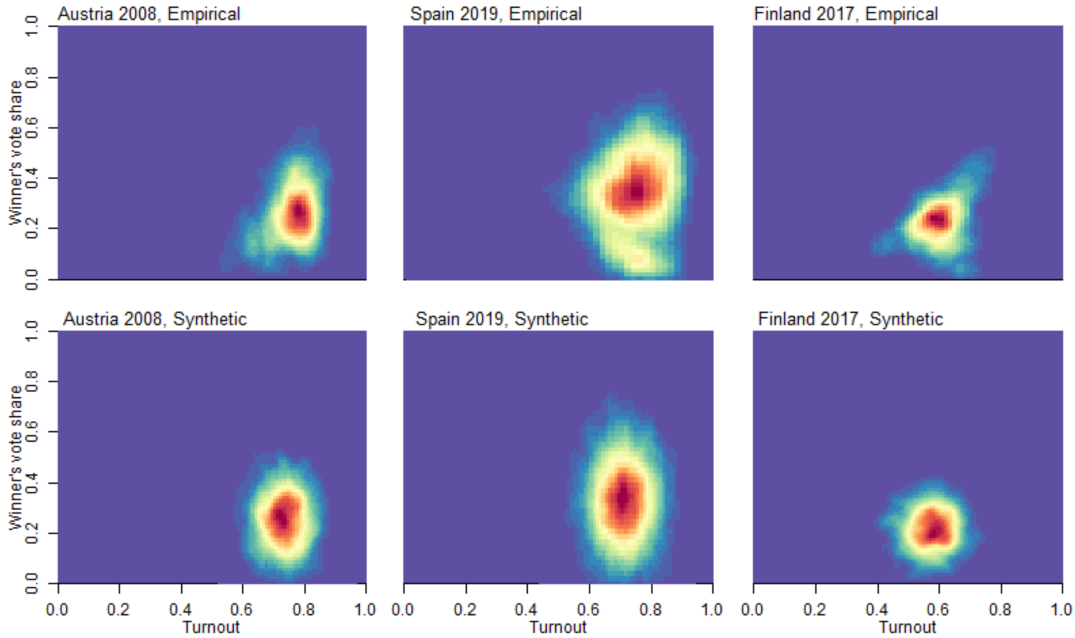


Figure 3.6. Three-dimensional histograms of the number of electoral units with a given turnout and winner's vote share percentage, Western Democracies. Colors represent the number of electoral units with corresponding (x,y) coordinates. Upper panel generated from empirical data. Lower panel generated from synthetic data with no data manipulation.

where q is the winner's vote share level that defines the 50% quartile of the empirical distribution s in the case of σ_i and the 75% quartile of the empirical distribution s in the case of σ_e. Sampling $n \times f_i$ and $n \times f_e$ values constructs the share of votes at the particular electoral unit that is affected by fraud.

6. In $n \times f_s$ electoral units, vote shares for the selected candidate (party) are rounded up to their closest integer percentage that is a multiple of '5' (0.1, 0.15, ..., 1).

There are two ways that assure that the data generating protocol can reconstruct heterogeneous types of fraud that might take place in different empirical contexts. First, the exponent c defines the relation between fraud *occurrence* f_i, f_e and fraud *intensity* f_i^c, f_e^c. For incremental fraud, values larger than 1 carry the assumption that the share of moved votes is smaller than the share of entities where fraud occurs overall. Likewise, for extreme fraud, values larger than 1 imply that that the share of moved votes is bigger than the share of entities at which votes are tampered. Since values of c can be varied freely, different relationships between fraud occurrence and intensity can be represented under this protocol and diverse fraud mechanisms can be reconstructed. Second, each type of fraud can be implemented in three distinct ways. *Ballot box stuffing* implies taking the sampled fraction of votes that should be affected from all non-voters and adding it to the vote count of the

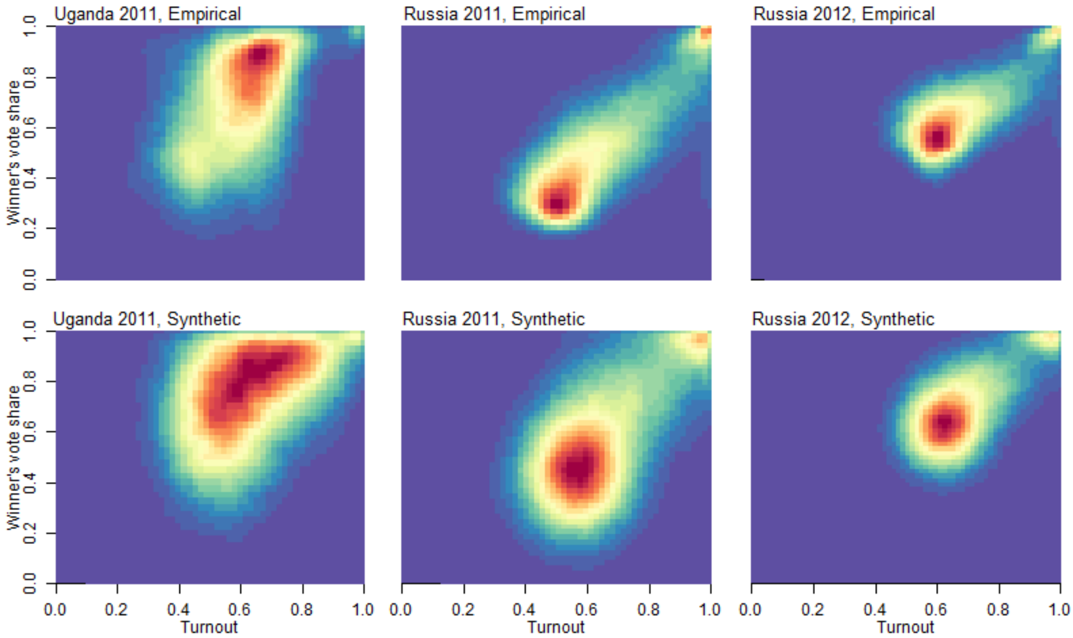


Figure 3.7. Three-dimensional histograms of the number of electoral units with a given turnout and winner's vote share percentage, Autocracies. Colors represent the number of electoral units with corresponding (x,y) coordinates. Upper panel generated from empirical data. Lower panel generated from synthetic data with pre-specified levels of manipulation.

winning camp. *Vote stealing* is defined as removing the sampled fraction of affected votes from the loosing camp as if these votes have not been observed in the first place. *Vote switching* implies taking the sampled fraction of affected votes from the loosing camp and adding it towards the winning party (candidate). All three types of fraud can be combined witch each other, intensifying the level of data intrusion. If vote tampering leads to turnout or vote shares exceeding the 100% level, they are respectively set to 100%.

3.3.2 Comparison to Empirical Data

In order to apply supervised machine learning to election fraud detection, it is required that the synthetic data which learners are trained on resemble numerical characteristics of empirical data. At first, I applied the protocol outlined above to the cases of Austria, Finland and Spain without introducing any kind of data manipulation. Second, I repeated the procedure for the elections from Russia and Uganda choosing fraud parameters such that the distribution of digits, turnout and vote shares are represented well. Table 3.1 summarizes the parameters that have been used for synthetic data generation for each of the empirical cases.

The figures that I subsequently discuss are respectively based on generating one set of electoral returns under these parametrizations. Figure 3.1 considers distributions of second significant digits for the two exemplary cases of Finland 2017 and Russia 2012. Figure 3.5 plots empirical distributions of winner's vote share

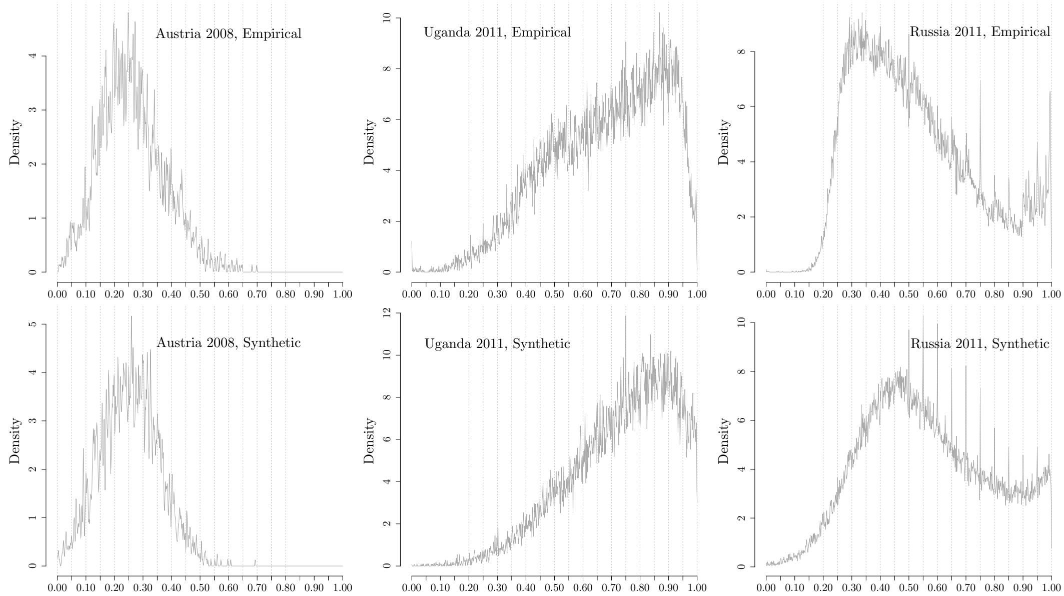


Figure 3.8. Estimated density (Gaussian kernel) of the winning party's (or candidate's) vote share. Austria 2008 ($n = 2,535$, bandwidth=0.001), Finland 2017 ($n = 992$, bandwidth=0.001) and Russia 2011 ($n = 90,919$, bandwidth=0.0001). Upper panel generated from empirical data. Lower panel generated from synthetic data with pre-specified levels of manipulation. Values at 100% not shown.

values against the synthetic counterparts. Figures 3.6 and 3.7 consider the bivariate distribution between vote shares and turnout levels across all electoral units. Figure 3.8 considers spikes around integer vote shares of the winning party (candidate). As can be seen, the synthetic data that has been simulated under the specific constraints of respective electoral systems and the fraud parametrizations that have been used provide close fits between synthetic and empirical data across all countries under study, as digit and vote share distributions, skewness, kurtoses and clusters in bivariate vote share-turnout distributions, and spikes in the density mass around exactly coarse vote share values are well represented. As the outlined data generating methodology can reconstruct the patterns of heterogenous empirical cases, it is useful in training machine learning methods on synthetic data in order to identify fraud in future real-world settings.

3.4 Estimating the Percentage of Fraudulent Votes

3.4.1 Synthetic Training Data for One Country Case

In order to study one empirical election and estimate the number of manipulated votes, I first make use of the protocol outlined above to simulate a large number of artificial clean and fraudulent elections under the given number and sizes of electoral entities of the particular empirical case at hand. First, $7,500$ clean elections are simulated. Afterwards, in order to produce synthetic fraudulent elections with different types and degrees of interference, I combine the following values for the fraud parameters $f_i = \{0.01, 0.02, ..., 0.5\}$, $f_e = \{0.01, 0.02, ..., 0.1\}$ and $f_s = \{0.01, 0.02, ..., 0.05\}$ with three types of fraud (ballot box stuffing, stealing, switching) in a full factorial design yielding $50 * 10 * 5 * 3 = 7,500$ synthetic elections for a given empirical case that were frauded with different mechanisms and to different degrees. This results in a total of $n = 15,000$[12] artificial elections simulated under one particular country case under study which have been either constructed without manipulation, or tainted to different degrees and using different mechanisms of fraud. The complete set of synthetic data is then split into a training set of 12,000 (80%) synthetic elections and a hold-out validation set of 3,000 (20%) synthetic elections that is used to estimate final test error rates. Using the 12,000 training observations, I then perform 5-fold cross-validation in order to calibrate a statistical learner.

3.4.2 Outcome Variable and Multivariate Feature Space

As an outcome variable, I calculate the percentage of votes that has been affected by fraud during the simulation of each synthetic election. The predicted outcome is hence a continuous variable ranging between [0,1], which takes on the value 0 if a synthetic election is clean and positive values for tainted elections.

In order to predict the percentage of affected votes, I employ a range of fifteen explanatory features that pick up on heterogeneous numerical characteristics of electoral returns such as the distribution of digits, skewness, kurtosis and clusters in turnout and vote shares, and the share of exactly integer vote shares that are multiples of '5'. Table B.3 in the Appendix provides an overview of the numerical characteristics that are used for training.

3.4.3 Choice of Machine Learning Algorithm

The fundamental problem of statistical inference concerns approximating an unknown target function $g(y)$ that generates a set of labeled examples y by using a k-dimensional vector of inputs $x = [x_1, ..., x_k]$ to construct $f(X)$ such that

$$y = f(X) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2). \tag{3.6}$$

A variety of tools have been proposed to flexibly approximate the unknown traget function $f(X)$ from pre-labeled outputs. First, although this is not a problem of causal estimation, the interpretability of the used model is key. Ultimately, it is not enough to receive a prediction about a specific mechanism or magnitude of fraud, but we also are in need of a straightforward interpretation to *why* elections are classified as such and which (combination of) features did contribute the most to an election being flagged as fraudulent. At the same time, it is important to note that this is not a task where the data-generating process of of the percentage of

[12]Ultimately, the number of simulated synthetic elections depends on computational resources and constraints.

affected votes follows a simple linear-additive pattern that is imposed by parametric regression-based approaches. As the models that I apply should provide state-of-the art predictive performance, I make use of Random Forests as these provide a suitable trade-off between both the interpretability and flexibility.

After the Random Forest model has been calibrated using synthetic training data, the final prediction function is applied to the empirical data from the country case under study to predict the percentage of fraudulent votes giving the specific country and election and hand.

3.5 Applications

To validate the methodology that is outlined in this chapter and to showcase its application on a range of empirical cases, I now first present results from a Monte Carlo Simulation study which investigate the behavior of the used Random Forest algorithm on simulated cases for which the degree of fraud is known. This serves as an internal validation confirming that the proposed methodology can reliably estimate the percentage of tainted votes in a laboratory setting. Afterwards, I apply the method described above to the six empirical elections from Austria 2008, Finland 2017, Spain 2019 as well as Russia 2011, Russia 2012 and Uganda 2011. This second step serves as an external validation, confirming that the approach presented here labels elections from Western European democracies that have not been put into doubt as clean and highly controversial elections for which a range of evidence in favor of systematic manipulation exists as foul. Additionally, the application to empirical country cases provides insights into the types of inferences that can be performed using tree-based ensemble learners trained on synthetic data of electoral returns.

3.5.1 Monte Carlo Simulation Study

To study the behavior of the outlined methodology when applied to artificial cases for which the degree of fraud is known, I first of all generate data for single artificial elections simulated under the data generating protocol outlined above. Afterwards, I treat these simulated datasets as if they were stemming from actual empirical elections and test whether the outlined approach can correctly estimate the percentage of votes that was actually manipulated in the data generating protocol. In all artificial elections, individual electoral units are respectively composed of $1,000$ eligible voters. The parameters that I vary during the simulation study are

- number of electoral units n: 500, 600, 700, 700, 900, 1000, 2000
- share or polling stations with incremental fraud f_i: 0.02, 0.03, 0.04, 0.05
- share of polling stations with extreme fraud f_e: 0.01, 0.015, 0.02

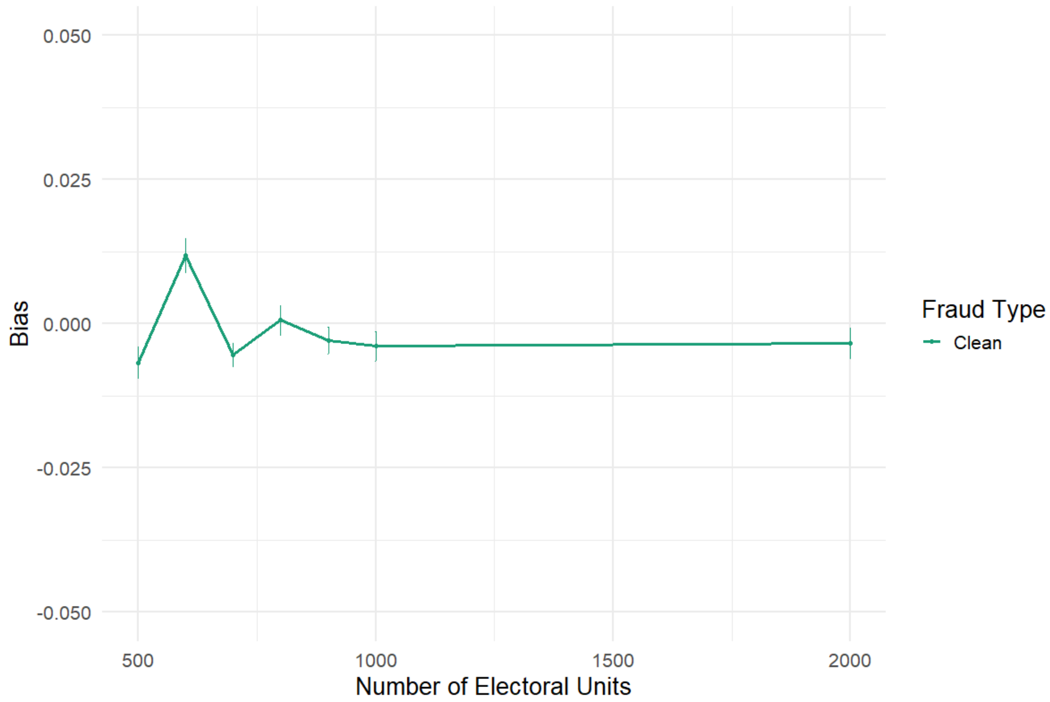


Figure 3.9. Monte Carlo simulation results for synthetic data with no data manipulation incorporated across different numbers of electoral units. Each electoral unit comprises an electorate of 1,000 potential voters. Numbers of electoral units that are considered are 500, 600, 700, 800, 900, 1000, 2000.

- type of fraud: ballot box stuffing (adding votes), vote stealing (removing votes), vote switching

Furthermore, in those elections in which vote manipulation is implemented, vote shares for the winning party are rounded up to their closest integer percentage that is a multiple of '5' in two percent of polling stations. These parameters are then combined in a full-factorial design, yielding a total of 189 of artificial fraudulent elections, which are complemented by seven clean elections each comprising a different number of electoral units. This yields a total of 196 artificial elections for which the degree of fraud is known. For each of these 196 artificial elections, all steps described in Section 3.4 is applied. For each artificial election, I compute the difference between the true percentage of manipulated votes that is stemming from the data generating protocol and the predicted percentage of manipulated votes stemming from the Random Forest algorithm.

Figures 3.9 and 3.10 summarize the results. As can be seen from Figure 3.9, the predicted number of manipulated votes fluctuates around the true number of zero votes for clean elections given that the number of electoral units that data was simulated for is small. However, already for 800 electoral units, predicted values are close to the true number of zero tainted votes, which equally holds if the number of electoral units is increased to 1,000 or 2,000. This provides us with confidence that the share of manipulated votes is reliably predicted to be zero when elections

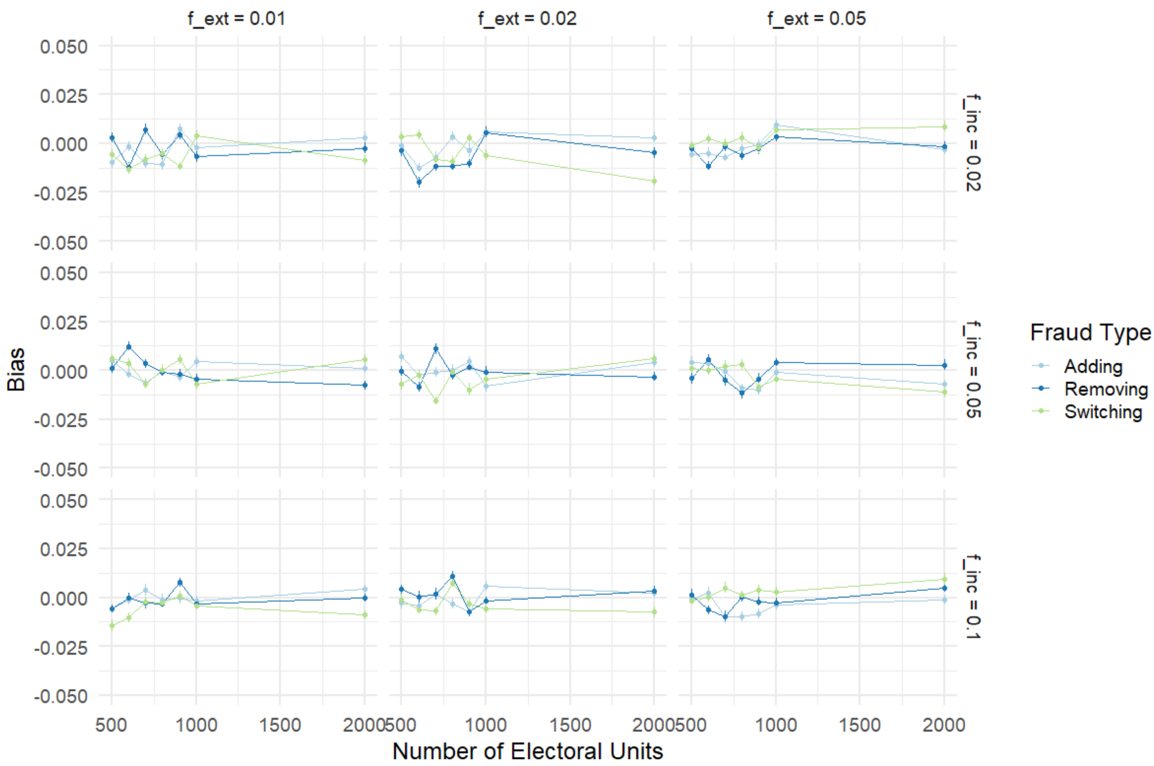


Figure 3.10. Monte Carlo simulation results for synthetic data with different types and degrees of data manipulation incorporated across different numbers of electoral units. Each electoral unit comprises an electorate of 1,000 potential voters. Numbers of electoral units that are considered are 500, 600, 700, 800, 900, 1000, 2000.

are clean already for moderate numbers of electoral units for which voting results are available.

Furthermore, these empirical patterns equally hold when we consider those artificial elections that were manipulated in Figure 3.10. Independently of the extent and type of vote alterations, the tree-based ensemble learner that I employ here yields estimates that are very close to the true percentage of tainted votes, with the predictive performance improving for larger numbers of electoral units that data is available for.

3.5.2 External Validation

One potential criticism of the Monte Carlo Simulation exercise above is that—by definition—the 196 artificial datasets that are used as a ground truth are simulated from the same data generating protocol which the proposed methodology uses for training the Random Forest algorithm to predict the affected percentage of votes. While this internal validation is a necessary step to evaluate the models' performance, sufficient confidence in the proposed approach is only generated if it generates valid estimates for cases of actual elections for which the data generating process is unknown, but academic experts and political observers largely agree on them being either fair or subject to significant manipulation.

Country	Year	Election	Electoral Units	Type of Electoral Unit	Estimate
Western Democracies					
Finland	2017	Municipal	992	Municipalities (*kunnat*)	0.02 [0; 0.08]
Austria	2008	Parliamentary	2,535	Communities (*Gemeinden*)	0.04 [0; 0.06]
Spain	2019	European Parliament	6,622	Municipalities (*municipios*)	0.03[0; 0. 06]
Electoral Autocratic Regimes					
Russia	2011	Parliamentary	90,919	Polling stations	0.12 [0.11; 0.13]
Russia	2012	Presidential	91,256	Polling stations	0.08 [0.08; 0.09]
Uganda	2011	Presidential	23,754	Polling stations	0.04 [0.03; 0.05]

Table 3.2. Application of supervised machine learning trained on synthetic data for six empirical elections. Documented is the predicted percentage of votes that was subject to manipulation as well as 95% uncertainty intervals. Random Forest regression has been trained on $12,000$ artificial elections simulated for each country. 95% uncertainty intervals are calculated from the prediction model's performance on the $3,000$ artifical elections used as test data and defined as $1.96 * sd(\hat{y} - y)$.

The remainder of this chapter hence applies all steps outlined in Section 3.4 to the six elections of Austria 2008, Finland 2017, Spain 2019, Russia 2011, Russia 2012 and Uganda 2011. The advantage of using elections that lie a couple of years in the past is that the public and academic debate had sufficient time to converge to a joint judgement, which makes these historic election suitable for externally validating the presented approach. The three elections from Western European democracies have—up until today—not been contested. For detailed scrutiny of the used elections in Russia and Uganda, see Kobak, Shpilkin, and Pshenichnikov (2016a), Rozenas (2017) and Enikolopov et al. (2013), Klimek et al. (2012) as well as Schwirtz and Herszenhorn (2011) and The Guardian (2011). As outlined in Section 3.4, for each of the empirical datasets, a total of $3,000$ artificial elections are simulated out of which $1,500$ are clean and the other half is tainted to different degrees whereas a total of $12,000$ (80%) are used for training and $3,000$ (20%) for testing model performance.

Table 3.2 summarizes the results across the six elections under scrutiny. Presented is the predicted percentage of manipulated votes as well as their 95% uncertainty interval calculated based on the $3,000$ artificial elections that were held out for model evaluation and were not used for model training. As can be seen, while small percentages of manipulated votes are predicted for the three elections from Western European democracies, uncertainty intervals reliably include 0 which labels these elections as clean. On the contrary, percentages of manipulated votes between 4% ([3%; 5%]) and 12% ([11%; 13%]) are predicted for the three elections from electoral autocratic regimes.

Interestingly, the estimate for the Russian parliamentary elections of 2011 is close to the result of Enikolopov et al. (2013), who sent independent observers to 156 out of 3,164 polling stations in the city of Moscow and estimated the actual share of votes for the incumbent United Russia party to be about 11% lower than the official count, as fully observed polling stations in the treatment group on average reported a vote share of 36% for United Russia, whereas polling stations at

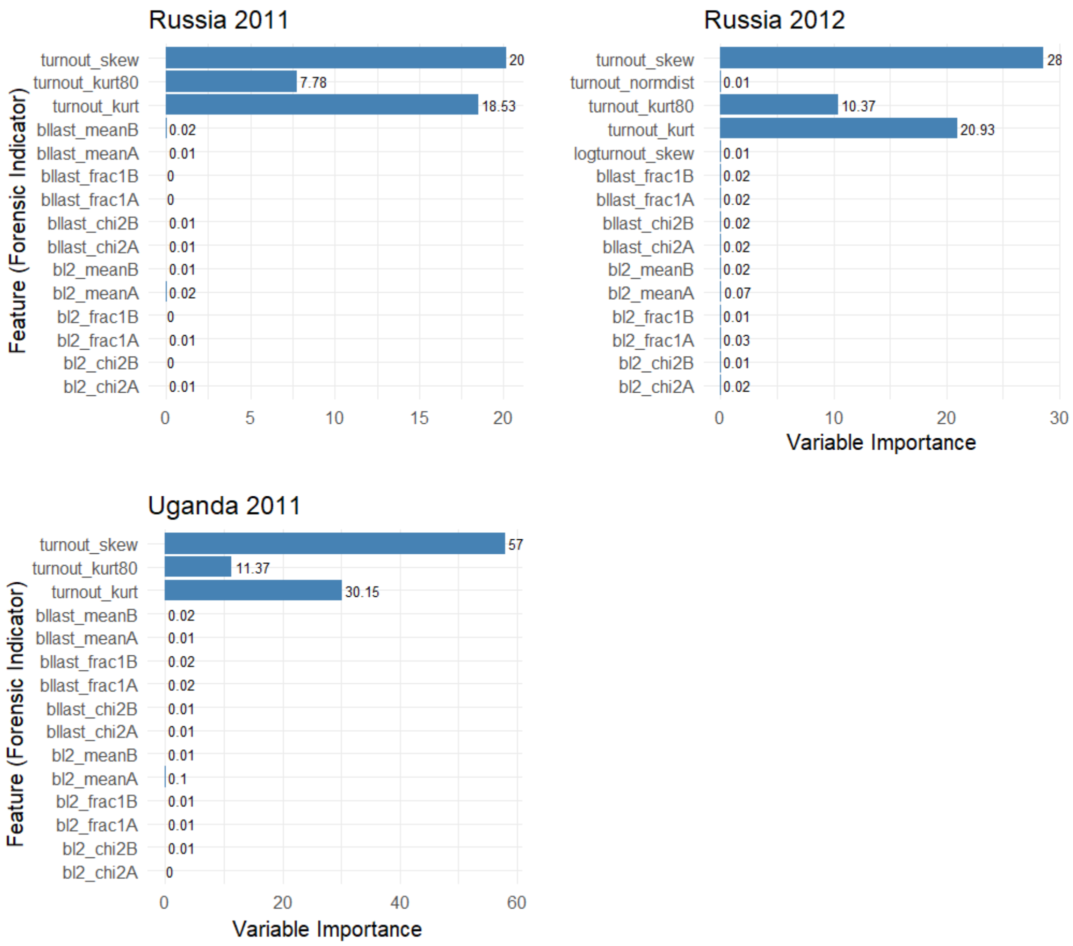


Figure 3.11. Variable importance measures for the fifteen most important features in the Random Forest algorithm when predicting the percentage of votes that is subjected to manipulation. Variable importance is defined as the average decrease of Gini impurity when a variable is chosen to split a node. Descriptions of all feature variables that are reported on vertical axes can be found in Appendix Table B.3.

which randomly assigned observers were not present reported an average of 47% for the incumbent party. Apart from the these congruent results being a further external validation check of the methodology that is developed here, this also further deepens academic evidence that the 2011 Russian *Duma* elections were deeply flawed.

Finally, Figure 3.11 provides insight into the variables that yielded most predictive power in predicting known tainted vote percentages across artifical elections during model training. As can be seen, for all three elections for which substantial shares of manipulated votes are predicted, the main drivers behind these predictions seem to be the skewness and kurtosis in *turnout* distributions, that the used prediction model predominately picks up on when predicting the percentage of manipulated votes. Distributions of numerals in the second and last significant digits as well as the number of coarse vote shares are relevant to a far lesser degree, indicating that the results presented in Table 3.2 can be interpreted as mostly indicative of ballot box stuffing.

3.6 Conclusion

This chapter departed from the observation that public controversies around the integrity of electoral events increasingly evolve around statistical patterns in published voting returns that are hard to explain without invoking fraud. After showcasing several prominent indicators that have been constructed by the field of *election forensics*, it outlined an approach to apply supervised machine learning methods in order to combine the predictive power of standalone indicators that in themselves are agnostic to each other and only allow vague statements about the nature of manipulation. The methodology presented performed well in a Monte Carlo simulation study in which true manipulation rates could reliably be predicted by a Random Forest regression model even for datasets that were merely composed of a moderate size of electoral units, while clean simulated elections were reliably labeled as such. Additionally, externally validating the method on six actual empirical electoral events, elections from Western European democracies were labeled as clean, whereas considerable percentages of manipulated votes where predicted for the highly contested Russian parliamentary election 2011, the Russian presidential election 2012 and the Ugandan presidential election of 2011.

The main advantages of the outlined approach are three-fold. On the one hand, this framework allows to relate different forensic indicators to each other and quantifies their relative contribution in labeling an election as fair or foul. Second, training supervised models on any particular target variable of choice enables the direct estimation of the percentage of votes that are affected by fraudulent interference. Third, as models flexibly learn from a large number of artificially simulated elections under the specific electoral system design of choice rather than adhering to pre-specified theoretical null-distributions that are hypothesized to arise under clean data generating processes, the proposed approach promises to control type-I and type-II error rates.

The central challenge of this chapter is to design a data generating process that simulates synthetic data fulfilling two main criteria: First of all, they need to replicate distributional characteristics of observed election data across a large range of dimensions rather than just one particular variable that a standalone forensic indicator is constructed around. Second, the type and degree of manipulation that is inherent to synthetically generated data needs to be known.

A range of possibilities comes to mind in order to satisfy the first criterion. Most notably, *generative* machine learning models have taken the world by storm in recent years, producing fake images and video sequences (Harshvardhan et al. 2020) and tabular data (Neunhoeffer, Wu, and Dwork, 2021) from real-world observed data that they are trained on. Of course, it would be straightforward to take an observed dataset of fine-graded voting returns as input and to synthetically replicate it any given number of times. This approach would create artificial data which differs from the observed values but keeps initially observed distributional properties fixed—even those that have not been explicitly specified by the researcher. While applying generative models intuitively seems promising, this approach comes with the disadvantage that after we have created synthetic data, it comes with little use for actually identifying manipulation. No follow-up machine learning model can be trained to detect manipulation in such synthetic data as we did not know the type and extent of manipulation that was underlying the observed data that was synthetically replicated in the first place. Against the rise of sophisticated generative models for tabular data, manually defining a data generating process as outlined in Section 3.3 of this chapter comes as a more primitive approach. Yet, manually incorporating manipulation is necessary as for follow-up supervised learning, the target variable that models are trained on (the percentage of votes that are subject to manipulation) needs to be known for every synthetic dataset at hand.

Naturally, the material presented in this chapter does not come without shortcomings. First of all, while incorporating manipulation manually in the data generating process of synthetic data comes with the advantage that target values are known for every synthetically generated dataset, it also comes with the assumption that the way that fraud is manually incorporated actually resembles real-life strategies. This may or may not be the case, and most importantly cannot be validated (as we would not need statistical models for anomaly detection in the first place if an external validation would exist). Second, an implicit assumption of the approach presented in this chapter is that the manipulation that is present in the data is administered on the same level of observation that the synthetic dataset was generated on. However, if fraud is executed on the micro-level (say polling stations) but only aggregate data actually is available (say precincts), it is unclear how this will affect the performance of trained supervised models. Third, an additional implicit assumption of the data generating protocol presented here is that fraud is administered in favor of exactly one party. While especially in electoral autocratic settings, this assumption will typically hold, it is an open question how simultaneous manipulation in favor of different camps will affect model results.

Summing up, this chapter presents a unified statistical framework which combines the advantages of many of the forensic indicators that have been developed so far and can be easily extended to new and unforeseen real-life scenarios when the data generating process outlined in Section 3.3 is adjusted. The framework can also straightforwardly incorporate new statistical indicators that will be developed in the future, given that the numerical characteristics that they are formulated on can be assumed to hold globally and do not pick up idiosyncrasies of single country cases that do not generalize. As this chapter has shown, training supervised machine learning models on synthetically generated data is a promising extension of current methodologies for the statistical detection of systematic election irregularities.

4

Election Fraud Information, Punishment, and Political Trust: Evidence from a Survey Experiment in Colombia, Mexico, and Russia

Abstract: In developing democracies and even electoral autocracies, credible allegations of electoral manipulation regularly lead to political interventions such as the dismissal of electoral staff or court punishments of alleged perpetrators. So far, scholars have shown that consciousness of election fraud lets individuals withdraw support from candidates, institutions and governments that are supposedly involved in manipulation. In this study, we first argue that election fraud information lets individuals extrapolate legitimacy loss even to political institutions that are unrelated to electoral events. Second, we investigate whether political interventions mitigate decays in diffuse support. Using a pre-registered online survey experiment in Colombia, Mexico, and Russia ($n = 2,057$), we (i) present empirical evidence for the spillover effect of election fraud information to political institutions unrelated to the fraud stimulus (ii) and show how spillovers largely persist even after political interventions. Our findings hold important implications for the study of developing democracies and electoral autocracies.

Keywords: *Election Fraud; Diffuse Support; Political Trust; Electoral Courts; Survey Experiment; Causal estimation.*

Author's note: This chapter is co-authored with Viktoriia Semenova.

This chapter draws on empirical evidence from an online survey experiment that has been pre-registered prior to data collection. The pre-registration plan covers the substantive hypotheses, experimental design, and exact measurements and can be accessed via the Open Science Framework under `https://osf.io/jyc2n/`.

A previous version of this manuscript has been presented at the 11th Annual Conference of the European Political Science Association (EPSA). We are grateful to Harald Schoen, Thomas Bräuninger, Thomas Gschwend as well as all panel and workshop participants of the University of Mannheim's CDSS workshop 'Political Science' for helpful comments on previous versions of this manuscript.

4.1 Introduction

Which consequences do information about electoral fraud have for citizens' relationship towards their political system? Since the 'electoral revolution' that surged since the mid-twentieth century led to a dramatic increase in the number of electoral events (Norris 2014), multiparty elections have become omnipresent across new democracies and electoral authoritarian regimes worldwide. The conduct of these, however, is frequently accompanied by publicly voiced doubts about their integrity. For instance, nearly 80% of all federal elections in non-established democracies nowadays are monitored 'on the ground' by international observers (Kelley 2012b; Hyde 2011) and around half of all observed elections had led international missions to declare problems of moderate or high magnitude (Kelley 2012b; Kelley 2012a).

Among the citizens themselves, credible information about electoral crimes can hold several behavioral and attitudinal consequences. Becoming conscious of electoral malpractice has been shown to lead to participation in popular protests and violent uprisings (Daxecker 2012). The literature has furthermore amassed a wealth of knowledge on the effects of election fraud perceptions on individuals' attitudes towards their political authorities. First, scholars have examined how information about electoral misconduct shape individuals' evaluations of the electoral process itself. For instance, Robertson (2017) shows that providing citizens with critical reports of election observation missions considerably reduces their perceived levels of electoral integrity. Second, citizens that are conscious about misbehavior withdraw support from those candidates that are allegedly involved in malpractice (Reuter and Szakonyi 2021) and express lower levels of legitimacy for the political regime that surged out of an electoral process that is perceived to be fraudulent (Williamson 2021).

This study holds two main contributions. First, we draw on theories of information processing and outline a mechanism of *attitudinal spillover* which states that individuals extrapolate specific fraud allegations to their confidence in the political system itself. The theory argues that even when information about fraud is attributed to unique political actors, citizens tend to relate these to political institutions that are unconnected to electoral administration. In contrast to the prior literature, the presence of such attitudinal spillovers predicts that consciousness about electoral misconduct will not let individuals merely detach from political authorities that can be directly linked to misbehavior and the regime that surged out of an illegitimate election process, but rather holds implications that are considerably more detrimental. Following the mechanism that we outline, consciousness about electoral malpractice can lead citizens to withdraw approval from the political system as a whole.

Second, we investigate whether the spillover effect induced by fraud information is endogenous to the reactions of other actors of the political system. Credible allegations of electoral manipulation regularly induce political interventions in developing democracies and even electoral autocracies. In the year 2015 in Colombia, a controversial congressman was stopped by local authorities two days before the country's regional elections while transporting a sum of over 200 million Colombian *pesos*. Under allegations of vote buying, the congressman was dismissed from office (El Espectador 2015). After the Bolivian presidential elections of 2019, the Organization of American States has voiced criticisms around statistical patterns among late-counted votes (OAS 2019a), shortly after which ex-president Evo Morales was asked to resign by the country's military chief and head of the police (Idrobo, Kronick, and Rodríguez 2020). In Russia, rumors and allegations of vote irregularities were widespread after the country's 2016 legislative elections. After allegations culminated, Russian election chiefs sacked seven officials from polling stations in a region were contestations were most severe from their posts, proclaiming that observed irregularities were 'isolated incidents' that do not put the results of the election into doubt (Reuters 2016).

Scholars of electoral integrity have put little effort into understanding how such political interventions to fraud allegations interplay with political attitudes. Notably, attitude shifts induced by information on electoral malpractice might be *exacerbated* (amplification effect) or *mitigated* (suppression effect) by how other political actors respond. For instance, alleged perpetrators being removed from their posts in the electoral commission or public court rulings on electoral crimes may send out signals of the political system's professionalism, autonomy, and commitment to a fair electoral process (Kerr and Wahman 2021). Successful convictions of alleged perpetrators might therefore function as signals of at least some level of horizontal 'checks and balances', mitigating individuals' depressed levels of diffuse support. On the other hand, especially for those individuals that hold positive views of the government or political regime in place, interventions like these might provide legitimacy to claims that otherwise are dismissed as political rhetoric. Under this mechanism, political interventions might *induce* rather than prevent spillover effects. Grasping such dynamics is crucial for understanding the real-life impact of fraud information, as citizens are not only exposed to disseminated information about electoral malpractice, but also perceive how actors of the political system *respond*.

We present evidence from a pre-registered online survey experiment conducted in Colombia, Mexico and Russia ($n = 2,057$) assessing (i) the presence of attitudinal spillovers of election fraud information to political institutions that are unrelated to electoral events and (ii) how punishment of alleged perpetrators exacerbate or mitigate decays in political trust. Since much of the previous literature has focused on the analysis of large cross-national survey data, our empirical analysis first showcases that these are unable to answer questions such as those that we pose here using 48,953 respondents across 48 countries from Wave 7 (2017-2020) of the World

Values Survey (WVS). Even after applying a range of state-of-the-art matching algorithms for causal estimation combined with various robustness checks and a Bayesian estimation approach, we cannot distinguish spillover effects on political institutions (that are dictated by theory) from spillovers on non-political institutions (that should not be present in theory). We then present evidence from our experiment adding two original findings to the current literature. First, exposing individuals to information about electoral misconduct induces negative spillovers to trust in components of the political system that are not tied to elections among government opponents and supporters. Second, across a range of subsample analyses, we find no evidence for an amplification effect of electoral staff dismissal or court punishments among government supporters and opponents. However, for most group comparisons, spillover effects persist after disseminating information on punishments of alleged perpetrators. Finally, we show that if multiple credible punishments are in place, negative shifts in attitudes can be mitigated.

The main conclusion of our study is two-fold. On a general level, the consequences of administering election fraud for public support are even more detrimental than currently acknowledged by the literature. This is because information on electoral misconduct even induces shifts in public support towards components of the political system that are no beneficiaries of manipulation and are not related to electoral administration. Second, we cast light on the under-acknowledged role that political interventions which punish alleged perpetrators of fraud play. A combination of several credible punishments *can* mitigate (or even remove) negative attitude shifts of individuals exposed to fraud information, although this effect is far from consistent across countries and institutions. This study hence closes on a cautious note: While the spillover effect of fraud information is consistent, the mitigating effect of effective punishments is not.

4.2 Election Fraud Information, Punishment, and Political Trust

In this section, we outline a theory of how the acquisition of new information about the integrity of domestic elections will affect the amount of trust that citizens place in the institutions of their broader political system. Essentially, this comes down to defining an argument of why individuals will extrapolate information about electoral misconduct to political institutions that are unrelated to electoral administration. Afterwards, we discuss how the interventions of other political actors can *amplify* or *mitigate* such spillover effects.

4.2.1 Election Fraud Information and Attitude Extrapolation

Scholarly contributions that examine the attitudinal nexus between citizens and the state commonly refer to the work of David Easton (1965) and Easton (1975) on 'system support' as a joint conceptual heritage. The theoretical distinction that is most relevant for our argument is the classical discrimination between diffuse and specific levels of support. Specific support refers to the relationship between members of a system and the specific actions and decisions of political authorities that reside within its institutions. As such, specific support relates to the evaluations of the day-to-day actions of political leaders, and are highest if perceived outputs match citizens' articulated demands (Easton 1975, p. 438). In contrast, diffuse support describes individuals' generalized attachment to the political system. According to Easton, *"[diffuse support] refers to evaluations of what an object is or represents [..] not of what it does. [..] Whereas specific support is extended only to the incumbent authorities, diffuse support is directed towards offices themselves as well as their individual occupants. More than that, diffuse support is support that underlies the regime as a whole and the political community."* (Easton 1975, pp. 444-445). Hence, diffuse support is *a priori* expected to be more durable than citizens' performance evaluation of specific political authorities. While positive evaluations of actors' performance is volatile and comes with consistent rise and fall, diffuse political support for the entity of the political system is in general thought to be long-lasting.

Early work on the concept of political support did almost exclusively focus on the relation between citizens and the state in the context of the United States and other advanced industrialized democracies (Easton 1965; Easton 1975; Citrin 1974). Importantly, already in their seminal work on popular support for authoritarian regimes, Geddes and Zaller (1989) have argued that political reasoning in democracies and autocracies can be expected to operate in similar ways and a range of studies have evaluated concepts derived from the distinction of specific and diffuse support in autocratic settings as well (Reuter and Szakonyi 2021; Frye and Borisova 2019). In addition, it has been shown that measurement equivalence of the most prominent operationalizations of diffuse support holds across a variety of regime types (Schneider 2017).

In the first place, we can expect that credible fraud information evolving around electoral contests will lower citizens' confidence in such. For instance, both Robertson (2017) as well as Bush and Prather (2018) show that confronting voters with criticisms from election observer groups reduces their evaluations of electoral quality and the legitimacy of the electoral process. In the literature evolving around system support, it has long been argued that attitudes about the performance of individual objects that are commonly associated with specific support can spill over to more generalized attachments towards the political system (Bowler and Karp 2004). It is important to note that this goes contrary to an assumption by which citizens' evaluations of *political actors* are unrelated to their evaluation of their *political* *institutions*.

Empirically, spillover-like effects are a well-established phenomenon in various branches of attitudinal research. These can be understood as specific manifestations of a more general psychological principle commonly referred to as the 'halo effect' by which individuals ascribe characteristics to a person or an object based on their evaluation of other empirically observable object-related characteristics even if the individual traits are unrelated to each other (Thorndike 1920; Palmer and Peterson 2016). Such spurious inferences may result from individuals' inability to differentiate between different characteristics and may even occur if there is sufficient information to allow for independent assessments in the first place (Nisbett and DeCamp Wilson 1977). Regarding citizens' evaluations of actors and institutions, it has been shown that trust in national institutions transcends to trust that is placed in the international arena, extrapolating federal-level experiences to European institutions (Torcal and Christmann 2019) and international organizations (Dellmuth and Tallberg 2015). Studying attitudinal spillovers between national institutions, Bowler and Karp (2004) show that political scandals of individual politicians have the power to erode confidence in executive institutions and the government in general. Notably, such spillover effects may either be the result of evaluating a series of repeated outputs over a long time series that can change even fundamental beliefs, or chief, salient, and decisive short-term experiences that transform into fundamental attitudes more rapidly.

We hypothesize that information about electoral fraud provide the kinds of short-term information that dis-attaches from the volatile performance of political actors and transforms into generalized evaluations even of other components of the political system. Essentially, this is based on a two-step argumentation line. First, as elections lie at the core of democratic accountability and are the one crucial element common to all and even minimalist definitions of democracy (Przeworski, Stokes, and Manin 1999), systematic misbehavior that evolves around the decisive process of elections is likely to be taken as informative not only of what a specific political object does, but even towards the system that it represents. Hence, the central place of well-conducted elections in the constitution of a democratic political system lets evaluations of the electoral process fundamentally differ in their nature from perceived output that is generated through the short-term and volatile performance of individual office holders. This provides election-related information with the general *possibility* for producing spillovers. Second, it has been shown by a variety of authors that citizens tend to fail in *distinguishing* their attitudes towards individual components of the multidimensional political system. This is most evident as the political sphere is usually described to be too complex to understand even for highly informed individuals (Zaller 1992) and as citizens need lower-complexity informational cues to maneuver their perceptions of political affairs. Empirically, scholars have found that support levels for different political institutions or entities are highly correlated with each other and are often hard to disentangle within individual respondents (Hooghe and Marien 2012; Mishler and Rose 2001). It is these two observations that build the premises from which motivate the first central claim of this chapter. The centrality of election-related information for citizens' evaluations of the political system which provides the possibility for spillover fused with the general tendency of individuals to fail distinguishing support for different institutions leads us to formulating the first main hypothesis:

Hypothesis 1: When exposed to information about electoral fraud, individuals show less confidence in institutions of the political system that are unrelated to electoral administration.

4.2.2 Previous Literature

While examining the empirical interrelations between operationalizations of specific and diffuse support is a decade-old endeavor, the attempt to link system support with election fraud information is rather new. Our specific research strategy tabs into a broader field of previous studies that have examined related phenomena which are relevant for our hypothesis. A branch of studies focused on the relation between 'objective' measures of electoral manipulation and average levels of diffuse support. Mauk (2019) globally assembles expert-coded judgements of federal-level electoral integrity from the Varieties of Democracy dataset and relates these to national levels of political trust, finding little evidence that objectively coded factual levels of electoral integrity are related to country-specific average values of political trust. Exploiting largely exogenous variation in a survey conducted in Moscow around the 2011 Russian *Duma* elections, Frye and Borisova (2019) reach similar conclusions and find that simply the mere event of an allegedly fraudulent election does not significantly reduce levels of diffuse support when comparing those individuals that have been surveyed after the election with the respondents whose data has been collected beforehand. These studies carry the obvious shortcoming that they calculate effects of fraud indicators that have been collected *a posteriori* on all individuals that might potentially have become aware of such information. However, as Mauk (2019) outlines, actual electoral malpractice does not necessarily need to be related closely to citizens' individual perceptions of electoral integrity (see also Ham 2015), since these crucially depend on factors such as a sufficiently free media environment to report about electoral inferences and one's individual political interest to become informed through media channels.

A different group of authors directly exposes individuals to information about electoral malpractice and investigates how becoming aware of misbehavior affects citizens' beliefs about the electoral process (Robertson 2017; Bush and Prather 2018) and their support for candidates that are allegedly involved in malpractice

(Reuter and Szakonyi 2021). These studies shed great light into individuals opinion-formation dynamics as a response to sensitive information, but restrict their analyses to attitudes that are directly linked to the electoral process or to specific evaluations of office holders rather than examining underlying attachments towards the political system.

Using data from the World Values Survey, Norris (2014) and Norris (2019) exploits a cross-sectional design and shows that even when controlling for a range of attitudinal and socio-demographic factors, expert evaluations and perceptions of electoral integrity are still correlated to an array of items as wide as confidence in elected institutions such as parliaments and governments, overall satisfaction with performance of democracy and respect for human rights. Obviously, using such cross-sectional strategies, it as hard to disentangle whether perceptions of electoral integrity and institutional confidence are simply observed jointly, or if one determines the other, falling short in testing a spillover theory as outlined here. Even if a directional effect exists, the causality chain might well go into the opposite direction. It is not less reasonable to assume that stable underlying beliefs such as confidence in political authorities pre-structure individuals' evaluations of specific political events such as electoral contests. In the piece that is most relevant for our research, Williamson (2021) shows how confronting citizens with condemnations of international election monitors can reduce expressed legitimacy in the political regime that surged out of an allegedly fraudulent process. Using correlational analysis from eight Arab countries and a survey experiment conducted in authoritarian Egypt and Morocco, he shows how perceptions of electoral misconduct hinder both attitudinal and behavioral compliance with a regime's rule. This investigation of individual conformity with the direct beneficiary of misbehavior is considerably different from our spillover perspective which investigates effects even towards components of the multidimensional political system that are unrelated to fraud information as coined by the Easton's concept of diffuse support.

4.2.3 Electoral Crimes and Punishment

In this second part of our theoretical scrutiny, we calibrate our theoretical expectation and outline how spillover effects might be moderated if third-party system actors become active as a response to fraud allegations. While accounting for the reactions of political actors has—to the best of our knowledge—not been incorporated into any study on election fraud information so far, it is at the same time crucial for understanding attitudinal dynamics stemming from exposing cheating, as fraud allegations are never observed in isolation but are accompanied by political developments that are either permissive or marked by intervention. How does the spillover effect of election fraud information behave against interventions from within the political system?

The intervention that is most relevant to our argumentation line here is punishment. After information on malicious behavior has been exposed, it is not unusual that functionaries in the electoral commission that are responsible for electoral administration need to step down from office or forcefully lose their posts (El Espectador 2015; Idrobo, Kronick, and Rodríguez 2020; Reuters 2016). Additionally, in recent decades, the judiciary has played an increasingly important role in electoral politics. Courts have emerged as an important actor that settles electoral disputes and frequently intervenes in pre- and post-electoral stages when the electoral conduct is in doubt (Eisenstadt 2002; Kerr and Wahman 2021). The topics that are covered by electoral tribunals range from issues revolving around constructing valid and comprehensive voter registers, the confirmation of candidate or party lists and the regulation of campaign resources up to sensitive issues such as election day fraud and vote manipulation. Court rulings on electoral crimes are highly salient for the electorate as they provide citizens with key non-partisan political information which regularly makes headlines in federal newspapers. Therefore, punishments are likely to directly affect the dynamics of attitude extrapolation. We focus on two specific arguments: the *amplifying spillover* argument and the *spillover suppression* argument.

The Spillover Suppression Effect

The line of reasoning emphasizing the suppression potential of punishments builds on the idea that functioning punishment mechanisms within the electoral commission or interventions of the judiciary into the electoral process signals information about the quality and independence of the underlying political system. Electoral commission punishment or court rulings may be interpreted as a sign of autonomy and professionalism which goes in counter to information about electoral fraud signaling system deficiencies. Punishments may lead to individual perceptions that the system of checks and balances in the country works reasonably well and that the political system does indeed have the capacity for self-correction if elections fail to meet shared standards. In this line of argumentation, successful punishments show that it's not the political system *as a whole* that is foul, but that state institutions do have the capacity to offer counterweights to malpractice. As a consequence, interventions by electoral commissions or courts may reduce the spillover to decays in diffuse support.

The Spillover Amplification Effect

On the other hand, there is reason to believe that information on election fraud can lead to an amplification of spillovers. This argumentation line is rooted in the empirical observation that electoral quality is so routinely disputed in new democracies and authoritarian regimes that, opposition parties' or the international community's protests may simply be perceived as a conventional part of the game (Kerr and Wahman 2021). From this perspective, defeated candidates are incentivized to publicly condemn the electoral process in order to avoid seeming weak in front of their voter base and to discredit the authority of the political opponent (Lindberg 2006). The potential spillover effects from acquiring information about electoral misconduct may hence be depressed by doubts whether the allegation itself is credible. When alleged perpetrators of electoral crimes are subject to punishment, the presence of real convictions in turn provide an official recognition that the election process was not free and fair and send credible signals about the trustworthiness of fraud claims. Under this logic, punishments provide individuals with detailed information about the nature and scope of electoral malpractice and may serve as a heuristic device for them to reliably evaluate electoral fairness based on the statement of third-party actors. As such, punishments can be expected to lead to a *stronger* spillover effect, as they confirm the deficiencies in the political system as suggested by information on the presence of electoral manipulation.

Heterogeneous Effects for Government Supporters and Opponents

While in principle, both argumentation lines can be put forward, we can expect the interplay between information on election fraud and punishment effects to vary across *supporters* and *opponents* of the government or political regime. We argue that heterogeneous punishment effects across supporters and opponents might be routed in the logic of 'Bayesian belief updating' (Bullock 2009; Hill 2017). As showcased by a wide branch of research on the winner-loser gap in political support (Nadeau, Daoust, and Dassonneville 2021; Cantú and García-Ponce 2015) and literature on the opposition in authoritarian regimes (Reuter and Szakonyi 2021), government opponents do *a priori* take on considerably more negative attitudes towards the political process than supporters. Regime supporters have *ex ante* beliefs that are considerably more in line with a well-functioning political system than regime opponents. The sources of this imbalance can be manifold.

For one, they can be a manifestation of regime supporters and opponents selectively exposing themselves to different kinds of news. In authoritarian states and developing democracies, it's safe to assume that regime supporters are considerably more exposed to pro-regime propaganda or state-owned media outlets that particularly present the government in a favourable light. These arguments relate to differences in information *acquisition* that supporters and opponents self-select into.

Additionally, the way that both groups *process* the same kind of information might lead to differences in *ex ante* beliefs about the political system. Even if regime supporters have been exposed to fraud information in the past, it is likely that these are simply discounted as anti-government agitation. Reuter and Szakonyi (2021) show that when revealing information about systematic interference, especially *regime supporters* withdraw support from regime candidates that allegedly engaged in fraud as it these respondents for which the information actually makes a difference. Opponents, on the other hand, already hold *ex ante* beliefs that elections are tainted and have already incorporated expectations about election fraud into their pre-existing belief *before* being exposed to new information about electoral manipulation. As a consequence, government opponents likely don't need to perceive official interventions to be convinced that a certain fraud allegation is credible. Rather, interventions might challenge their pre-existing belief of a foul system and mitigate—or even remove—attitudinal spillovers as a consequence of fraud.

When tracing the impact of punishment, the following hypotheses hence guide our empirical scrutiny:

> *Hypothesis 2a: The attitudinal spillover effect of election fraud information is stronger for regime supporters when they are exposed to information about within-system interventions.*
>
> *Hypothesis 2b: The attitudinal spillover effect of election fraud information is weaker for regime opponents when they are exposed to information about within-system interventions.*

4.3 Matching Estimates from Cross-Sectional Survey Data

Before we describe our survey experiment, we first present a placebo test using various algorithms for statistical matching and cross-sectional survey data to motivate our experimental design. As we outline in Section 4.2.2, much of the previous literature on the interrelation between perceptions of electoral integrity and institutional trust has exploited (cross-national) survey data to identify the dimensions of political trust for which perceptions of electoral integrity matter, relating to our *Hypothesis 1* (for instance Norris 2014; Norris 2019). We first showcase that even when applying a range of state-of-the-art matching algorithms that balance covariates across compared groups, we cannot distinguish spillover effects on political institutions (that are dictated by theory) from spillovers on non-political institutions (that should not be present in theory). Afterwards, we turn to our survey experiment.

To showcase how observational survey data is unsuitable to study the type of questions that we pose here and to motivate our experiment, we turn to data from the World Values Survey (Haerpfer et al. 2022). Because perceptions of electoral fraud are not randomly assigned among respondents, individuals may differ from each other in ways that are related to their fraud perception as well as their diffuse support for political institutions. The World Values Survey includes a rich set of covariates that make it possible to condition on possible differences between individuals and construct balanced (sub-)samples. While survey items relating to Easton's (1965, 1975) concept of diffuse support are part of the core questionnaire and asked consistently throughout all waves, a comprehensive battery of questions assessing respondents' perceptions of their country's electoral integrity has not been introduced before Wave 6 (2010-2014, c.f. Norris 2014). In particular, Wave 7 (2017-2020) of the cumulative data file provides questions on issues that are relevant as confounders of the relationship between fraud perceptions and diffuse support. We hence exploit data from Wave 7 of the World Value Survey covering 48 different countries across democratic and electoral authoritarian regimes. Table C.1 in the Appendix provides an overview of our main dependent and independent variables as well as all covariates used for sample adjustment.

Regarding the measurement of the core variables, we use the classic measure of political trust grasping confidence in different institutions on a four-point scale as our dependent variables. The fraud information treatment is asking whether the respondent believes ballots in the country to be counted free and fairly and measured on a four-point scale, which we dichotomize for the purposes of matching, with zero being more negative perceptions of election integrity. Due to the nature of dependent variable, we present results from ordered logit models.

We do not provide a detailed discussion of the different matching algorithms that underly our analysis. The interested reader is referred to Morgan and Winship, 2007 for a comprehensive overview of the statistical underpinnings. We employ three algorithms: direct exact matching; a less restrictive coarsened exact matching (King and Nielsen 2019); and the widely-used propensity score-based nearest neighbor matching (without replacement). Exact and coarsened exact matching balance observations across our multivariate vector of covariates and drastically reduce sample size to balanced datasets of of $n = 580$ (exact matching) and $n = 2,475$ (coarsened exact matching). Propensity-score models match on a pre-determined balance score and keep our sample size at $n = 42,246$.[12] For exact and coarsened exact matching, we first create balanced datasets, and then calculate ATEs using Bayesian ordered logit models using non-informative priors and $1,000$ posterior samples. Following Alvarez and Levin (2021), for our propensity-score based models, we employ Bayesian estimation in our first-stage model in order to account for uncertainty in the balance score arriving at a posterior distribution of $1,000$ propensity scores. Results presented from propensity-score models are subsequently based on $1,000$ separate matching procedures for these $1,000$ different propensity scores per individual, again yielding a posterior distribution for the ATE.

First, we estimate the spillover effect of election integrity perceptions on political institutions that are unrelated to electoral administration. Second, we perform

[1]Figures C.1 and C.2 in the Appendix present measures of covariate balance between those individuals falling into the 'fraud' and 'no fraud' perception condition before and after our sample adjustment. After matching, differences in covariates between 'treatment' and 'control groups' are almost completely removed. We obtain a very high degree of balance on all covariates.

[2]One issue with our procedure is that the attitudinal covariates are not strictly preceding the main variables of interest and might thus be affected by fraud perceptions and confidence in institutions rather than predicting 'treatment' status. We report estimates including the attitudinal measures, but note that the results are robust to model specifications which only use socio-demographic information for covariate adjustment.

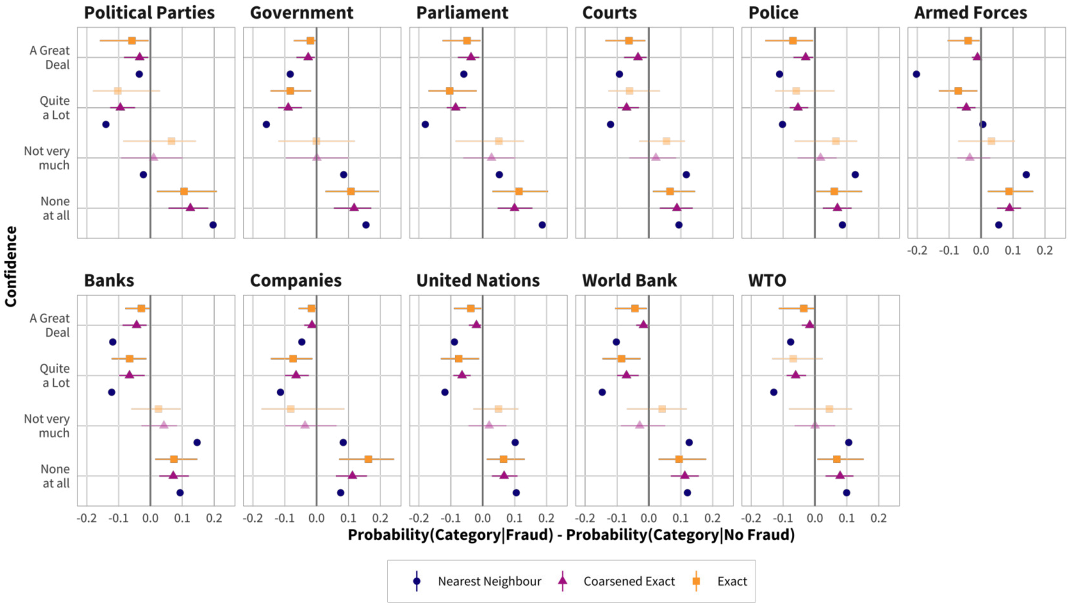


Figure 4.1. Average treatment effects of election integrity perceptions on trust in political (upper panel) and non-political (lower-panel) institutions from matching analyses on WVS Wave 7 (2017-2020) data. Individual figures report the difference in the predicted probability to tick each of the four categories of the dependent variable between respondents who perceive ballot counting to be foul vs. fair. Respondents with negative perceptions of electoral fairness show less confidence in political and non-political institutions. Point estimates are means, with 95% credibility intervals depicted with point-ranges.

a placebo test by estimating the same effects on dimensions of trust that cannot be expected to be subject to attitudinal spillover. If the spillover effect of election integrity perceptions is well-identified, we expect to observe effects on *political institutions* while treatment and control units necessarily are similar in their attitudes towards *non-political institutions* that do not form part of the domestic political system.

Figure 4.1 visualizes the average treatment effect of the variable of interest, fraud perceptions, on political trust from the ordered logit regressions across the three matched datasets for political and non-political institutions. Perceived prevalence of election day ballot fraud in the federal elections of one's country shows to robustly decrease confidence in political institutions that are unrelated to electoral administration. Yet, we also observe similar effects of fraud information for non-political institutions such as the World Bank or the WTO, which are not dictated by theory and suggest that the effect of fraud perceptions is not causally identified. These results demonstrate the unsuitability of cross-sectional survey data for the analysis of the type of questions that we pose and motivate our experimental design that we present in the following.

4.4 Survey Experiment Design

We now test our arguments on (i) the attitudinal spillover of election fraud information and (ii) the moderating role of punishments using data from a pre-registered[3] online survey experiment that was conducted in Colombia, Mexico, and Russia in April-June of 2021. We first discuss the case selection for the experimental part of our study, describe our sample and provide an overview of the questionnaire. Lastly, we elaborate on the the causal identification strategy.

4.4.1 Case Selection

We focus on Colombia, Mexico, and Russia as these countries, on the one hand, share a variety of features relevant for our experimental design. In particular, we study middle-income countries with party systems that have shown a sizable degree of stability throughout past decades. These countries also share a large history of public controversies around electoral fraud. On the other hand, these countries are sufficiently different from one another to observe if the mechanisms that we outline travel to diverse political and cultural contexts. Additionally, the three countries provide a temporally comparable political environment for the study of electoral misconduct, as Mexico (July 2021) and Russia (September 2021) held federal elections later in 2021 and Colombia in May 2022. More details on how the three countries compare in terms of their electoral history of fair and foul elections are discussed in the Appendix.

4.4.2 Survey Design and Implementation

The respondents from all three countries were recruited using the crowd-sourcing platform *Yandex.Toloka*.[4] Evidently, not all societal groups are (equally) represented on crowd-sourcing platforms (e.g., Bartneck et al. 2015; Berinsky, Huber, and Lenz 2012). As a consequence, our survey was predominantly conducted among a population of urban internet users who are somewhat younger than the general population and have obtained some level of higher education[5]. Attitudes towards incumbents and political authorities are divided enough among this group (Robertson 2017) for us to be able to test both hypotheses. Most importantly, the socio-demographic profile of our survey respondents specifically targets those population groups that are particularly important for political dynamics such as gathering and sharing sensitive regime information and boosting their publicity by carrying them

[3]The experimental design and all hypotheses have been pre-registered using the Open Science Framework. The pre-registration plan can be accessed via `https://osf.io/jyc2n/`.

[4]As stated in the pre-analysis plan, we initially intended to use *Amazon Mechanical Turk* for recruiting participants from the two Latin American countries. As the number of available *MTurk* workers from Mexico and Colombia turned out to be far from sufficient, we deviated from the pre-analysis plan and collected all participants from *Yandex.Toloka*. All data that has been genereted through *Amazon Mechanical Turk* is not reported nor included in the analysis.

[5]Tables C.3 and C.2 in the Appendix report summary statistics on the socio-demographic profile of survey respondents.

to the streets in protests. This makes the adopted sampling strategy advantageous over nationally representative surveys for this particular study.

In addition, while the specific attitudes of these surveyed groups might not be representative of the population as a whole, there is a wealth of evidence amassing that the factors which shape these attitudes are. Research has shown that treatment effects within attitudinal research based on data collected using crowd-sourcing platforms is similar to those found in representative surveys (Clifford, Jewell, and Waggoner 2015; Coppock 2019).[6] Hence, while we expect that our sampled group differs from the general population in terms of their socio-demographic profile descriptively, it is fair to assume that the general patterns around their reactions to the experimental stimuli hold in the general population.

The survey was designed to take about ten minutes to complete and participants were presented with the survey in either Russian or Spanish (Colombia: $n = 517$, Mexico: $n = 481$, Russia: $n = 1,334$). When constructing our questionnaire, we mimicked all question formulations from the World Values Survey Wave 7 core questionnaire as closely as possible. For all original questions, questionnaire text was validated through cognitive interviews with native speakers.

4.4.3 Causal Identification

After answering a range of introductory questions on political attitudes, political knowledge, party affiliation, and general trust, respondents were randomly assigned to one of four experimental conditions, i.e. they were required to read a paragraph of text related to a hypothetical election (see Table 4.1). The text of the first group is neutral and states some basic facts about a hypothetically upcoming election to elect the country's national legislative body, and the respondents are presented with a status quo outcome. The second text contains general information about the election scenario identical to that of the first group, but additionally explicitly exposes respondents to information about malpractices that were allegedly performed on election day.[7] The information presented to the third and fourth group additionally exposes respondents to information about punishments from within the political system directed at those individuals who are allegedly responsible for electoral crimes. For the third group, the punishment is limited to actions being undertaken by the superior electoral commission, i.e. the alleged perpetrators lose their positions at the commission. For the fourth group, the punishment includes legal actions in which alleged perpetrators were legally convicted for the performance of electoral crimes as well as personal consequences within electoral commissions for which the alleged perpetrators have worked.

[6] Similar conclusions have been reached independently from each other across a variety of disciplines (Bartneck et al. 2015; Yang et al. 2015).

[7] We follow Reuter and Szakonyi (2021) and omit the information source to avoid convoluting the effects of *information* and *source credibility*.

(1) Control group: Neutral information
On Sunday, *[6 June 2021/19 September 2021/13 March 2022]*, legislative elections are scheduled to be held in *[Mexico/Russia/Colombia]*. More than [2,000/2,000/1,000] registered candidates will compete for the *[500/450/280]* parliamentary seats of the *[Chamber of Deputies/State Duma/Congress of Colombia]*. The results will be determined by nearly *[90/110/36]* million *[Mexicans/Russians/Colombians]*. Imagine that the elections have already passed and suppose that as in the current convocation, *eight/four/twenty* parties retained seats in the assembly.

(2) Treatment group: Fraud information
[Neutral information]
In this hypothetical scenario, suppose that after election day, it becomes known that ballot-box stuffing and alterations of vote tallies in favor of the incumbent party perpetrated by individuals working for electoral commissions were widespread. Suppose that these electoral misconducts and manipulation practices took place across several regions of the country.

(3) Treatment group: Fraud information with electoral commission punishment
[Neutral information]
[Fraud information]
Furthermore, suppose that as a consequence, individuals allegedly responsible for fraud lost their position in the electoral commissions that they served in.

(4) Treatment group: Fraud information with court and electoral commission punishment
[Neutral information]
[Fraud information]
Furthermore, suppose that as a consequence, individuals allegedly responsible for fraud lost their position in the electoral commissions that they served in. Also, legal action was brought against these individuals, who were convicted for electoral crimes by responsible courts.

Table 4.1. Survey experiment, overview of experimental conditions.

In order to ensure that exactly one fourth of all respondents per country are placed in each of the four experimental groups and to avoid sparse data problems that might arise from extreme scenarios under complete randomization, we apply a randomized block design separately for each of the countries. Additionally, to be able to compare the results across regime supporters and opponents, we randomize treatments within these two groups.

After being confronted with the treatment conditions, respondents are presented a battery of questions measuring their levels of political trust in individual institutions of their political system. Specifically, we ask each participant the following question: *"Upon receiving this information, how much confidence would you have in the following organizations or institutions?"*. The battery of items includes *"(1) the* *armed forces, (2) the police, (3) the justice system/courts, (4) the central electoral commission, (5) the government, (6) the parliament and (7) political parties*" in randomized order for each participant. Responses are collected on a four-point scale ranging from "none at all" to "a great deal". Similarly, we include the same list of randomly-ordered non-political institutions to validate the identification of spillover effects as a placebo test as in Section 4.3.

A range of important characteristics of the experimental design can be noted in relation to the spillover theory outlined above. First, the text fragments that are presented to members of the treatment groups explicitly discuss the *mechanisms* of alleged election-day fraud. This should allow us to avoid variations in the interpretations of our fraud treatment, for instance associations of vote buying in Mexico (Cantú 2019a) and ballot box stuffing in Russia (Klimek et al. 2012, and hence prevent heterogeneous effects related to vagueness in the treatment. Second, referring to 'individuals working for electoral commissions', we explicitly state the alleged *perpetrator of fraud*. Mentioning the perpetrator is crucial as it allows us to directly study if trust is extrapolated to different bodies that are unrelated to the administration of elections.

Importantly, not all of the institutions included in our battery on political trust allow us to unambiguously identify spillover effects. For example, it is hard to disentangle whether fraud in new democracies or electoral authoritarian regimes constitutes the actions of micro-level agents or whether these practices are instructed from party and/or state representatives. For citizens who expect the latter, information about election day fraud may actually function as informational cues implying partisan involvement from political authorities. Among these participants, changes in diffuse support for the institutions of the government, parliament, political parties and central electoral commission may not necessarily be the result of an attitudinal spillover following from trust extrapolation as we theorize.Instead, it can be a straightforward withdrawal of political support from the perceived perpetrator of electoral misconduct.

To counter this possibility, we ask for respondents' levels of diffuse support towards a number of institutions that are clearly exogenous to our fraud treatment. Without spillover, trust in institutions such as the armed forces, the police, or the justice system are unaffected by information on election day manipulation, as members of these institutions are unrelated to fraudulent interference practiced by individuals working for electoral commissions on voting day. In contrast, decays in diffuse support towards these institutions as a response to the fraud treatments present genuine spillover effects.

4.5 Experimental Results

4.5.1 Statistical Modeling and Estimation

To test Hypothesis 1 (the effect of exposure to the information about electoral fraud on confidence in institutions), we estimate the mean difference between the control group (1) and treatment group (2) in Table 4.1. To evaluate Hypotheses 2a and 2b (heterogeneous effects of punishment across supporters and opponents), we estimate the differences between group (2), which only received fraud information, and either group (3) or (4), who received punishment information on top of fraud information. To estimate varying effects across government supporters and opponents, we construct a binary moderator based on our pre-treatment measure of party affiliation, distinguishing those respondents that consider themselves supporters of a party that was in government or the opposition at the time of the survey.

Since our dependent variables comprise four ordered categories j, with $(j = 1, ..., 4)$, we again estimate a set of ordered logistic regressions with the following specifications:

$$ln\left(\frac{\Pr(y_i \leq j)}{\Pr(y_i > j)}\right) = \underbrace{\underbrace{\alpha_j - (\beta_1 \text{ Control}_i + \beta_2 \text{ Punishment}_i + \beta_3 \text{ Judicial Punishment}_i)}_{\text{Main Specification (H1)}} \times \beta_4 \text{ Opponent}_i}_{\text{Heterogeneous Effects Specification (H2)}}, \tag{4.1}$$

where y_i is the level of diffuse support of an individual i with $(i = 1, ..., n)$ for an institution, and Control, Punishment, and Judicial Punishment are binary indicators for membership in the experimental groups (1), (3), and (4) [8]. As outlined in the preregistration plan, we pool the available data across countries and use all available observations that fulfill basic data-quality criteria for the main analysis.[9][10]

For parameter estimation, we employ a fully Bayesian framework for statistical inference as implemented in Stan (Stan Development Team 2020). Specifically, we rely on Hamiltonian Monte Carlo sampling in which priors are defined to follow Student-t distributions centered around zero and take on a sufficiently large variance to ensure that prior distributions are uninformative and do not favor any of the substantial hypotheses. We run a set of four Markov chains out of which we discard the first 10,000 as warm-up and use the following 10,000 samples to describe posterior distributions, which results in a total of 40,000 post-warmup samples.

[8] Individuals who only received fraud information serve as the reference category in our analysis.

[9] Respondents are excluded if they completed the survey in less than three minutes.

[10] As our theory—in which attitudes towards elections *spill over* to attitudes of other political institutions—implies that it is the changes in trust in elections that are responsible for the differences in trust for other institutions of the political system, one might trace empirical effects using mediation models. In our questionnaire, we have included a question that accounts for trust in the electoral system. Figure C.6 in the Appendix re-presents our main results explicitly incorporating trust in elections as a mediator, adjusting the model formulation in Equation (1). As detailed in the Appendix, all substantive conclusions are unchanged and mediation effects behave as hypothesized.

We check for model convergence using the Gelman-Rubin diagnostic and consider models converged if the discrepancy measure stays below 1.1 (Gelman et al. 2004).

4.5.2 Empirical Results

We start off my describing evidence for the spillover effect of fraud information as stated by Hypothesis 1 (Figure 4.2), validate it against a placebo test on non-political institutions (Figure 4.3), and trace spillover dynamics across countries (Figure C.4). Secondly, we evaluate Hypotheses 2a and 2b (Figure 4.4). For each figure, we first calculate individuals' probabilities of selecting each of the four answer categories, from *A great deal* to *None at all*. In the figures themselves, we present the differences in probabilities for each answer category between two respective experimental conditions.

Spillover Effect of Fraud Information

Figure 4.2 summarizes evidence for the spillover effect of fraud information as stated by Hypothesis 1, comparing trust in political institutions between respondents in experimental groups (2) and (1). As we can see, respondents who received fraud information are substantially less likely to voice higher degrees of confidence in political institutions than respondents of the control group. Naturally, this effect is most evident for trust in the Central Electoral Commission, whose affiliates have performed vote rigging as in our experimental treatment. However, these decreasing levels of political trust hold equally for institutions that are directly related to legislative elections and hence and might be perceived as endogenous to the fraud treatment (political parties, parliament) as well as political institutions that are not related to the electoral event and can be treated as exogenous. With the exception of the police, information about election day electoral produces decays in political trust that spill over to political institutions which are unrelated to electoral events, providing robust empirical evidence for Hypothesis 1. Regarding the non-existing effect on the institution of the police, we would expect an average respondent to have little direct contact with most political institutions we study, yet one could argue that the police, is, potentially, the institution with which respondents are more likely to interact in their daily life. For that reason, trust in police may be naturally stronger and more insensitive than trust in other institutions, making the spillover for the police less pronounced.

Figure 4.3 repeats the placebo test that we reported on in Section 4.3 and provides further evidence that the spillover effects of election fraud information are well identified—respondents of the two different treatment conditions do not differ in their trust in non-political institutions that are unrelated to our outlined theory, going against the notion that the reported effects might be a methodological artefact of the survey design.

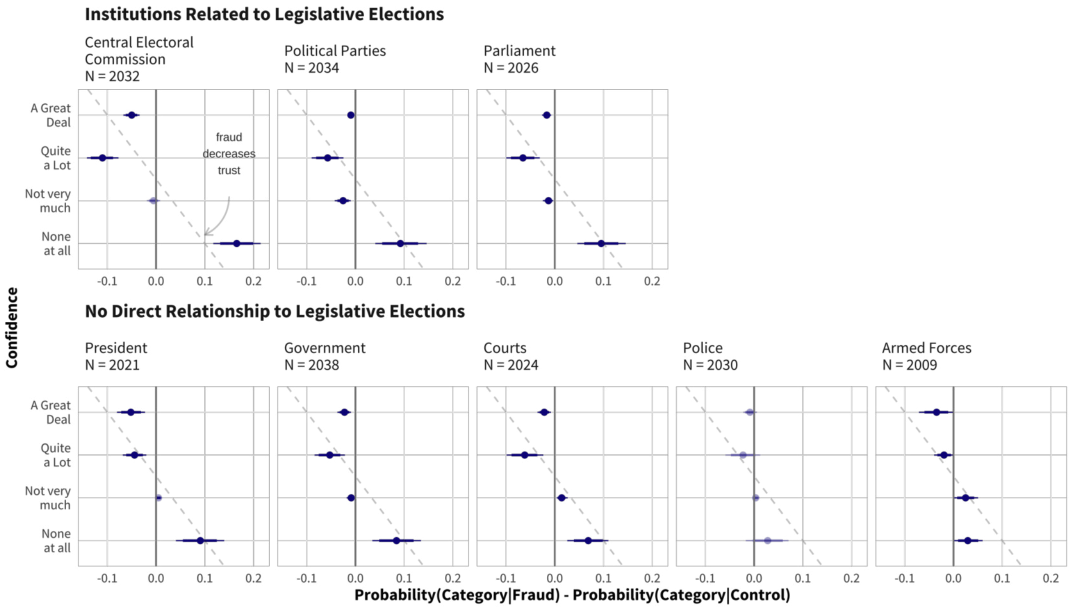


Figure 4.2. The effects of exposure to fraud information on confidence in political institutions. Plots depict medians and 83% (bold) and 95% (thin) highest-density intervals of differences in probabilities for choosing respective categories based on draws from the posterior predictive distributions. Transparent point ranges include zero in the 95% HDCI. Probabilities are calculated based ordered logit model estimates as lined out in Equation 1. The dashed line schematically depicts the hypothesized relationship between categories.

After having described main effects, we also investigate whether experimental results differ across countries. As Figure C.4 shows, while *some* hypothesized effects do exist for Colombia and Russia (for the institutions of political parties, the parliament and the president), main effects in the pooled sample are mostly driven by the Mexican subsample, which shows strong effects across seven out of the eight analyzed institutions. First, we do note that there is evidence suggesting that Mexican respondents simply did a better job in grasping and revising the treatment text. If we restrict the analysis to those respondents that correctly summarized their treatment text in the follow-up attention check[11], spillover effects are present for a range of institutions in all three countries (see Figure C.3.) More substantively, these country-level differences could be attributed to the straightforward mechanism of Bayesian belief updating which is restricted if prior expectations are already in line with the treatment. On the one hand, if baseline trust in political institutions (as captured in the control group in our experiment) could simply be very low, allowing very little room for any belief updating as a consequence to fraud. In other words, if there exists little trust in political institutions in the first place, information on fraudulent elections can hardly show any negative effects as little belief updating is taking place. This is likely to explain the fewer number of

[11] We manually coded respondents' summaries into correctly or incorrectly the scenario presented to them.

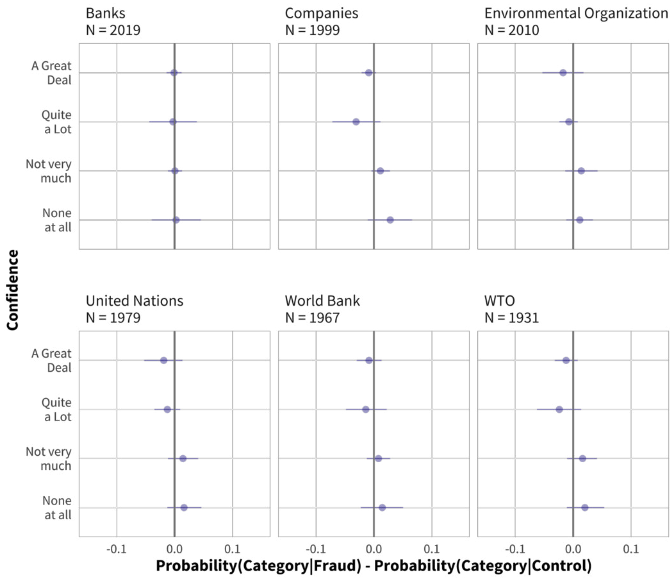


Figure 4.3. The effects of exposure to fraud information on confidence in non-political institutions. Plots depict medians and 83% (bold) and 95% (thin) highest-density intervals of differences in probabilities for choosing respective categories based on draws from the posterior predictive distributions. Transparent point ranges include zero in the 95% HDCI. Probabilities are calculated based ordered logit model estimates as lined out in Equation 1. The dashed line schematically depicts the hypothesized relationship between categories.

spillover effects in Colombia and Russia. Colombia shows the lowest levels of trust across all eights institutions *across all respondents* and *within the control group*, with government opponents making up 82% of our sample (compared to 65% in Mexico and 59% in Russia). In Russia, it is primarily regime supporters that show to be sensitive to treatment information (see Figure C.5) while few effects are found for opponents who are already sceptical of the regime, likewise speaking in favor of a mechanism of Bayesian belief updating that acquires higher baseline levels of trust in general.[12]

Needless to say, across all examined scenarios, it is clear that information on electoral malpractice robustly leads to decays in political trust for more than one political institution that is not tied to electoral administration.

System-response Effect

Regarding the moderating role of political interventions as a response to the circulation of fraud information, Hypotheses 2a and 2b expected the amplification

[12]Specifically for the case of authoritarian Russia and past cases of severely restricting opposition candidates from running in the election, a second explanation lies in the possibility of the control condition which re-states the status-quo not being interpreted as 'neutral' by regime opponents. Rather, an election in which the status quo persists might re-inforce their beliefs of an authoritarian political system, again not allowing for much updating after being confronted with additional information on electoral malpractice.

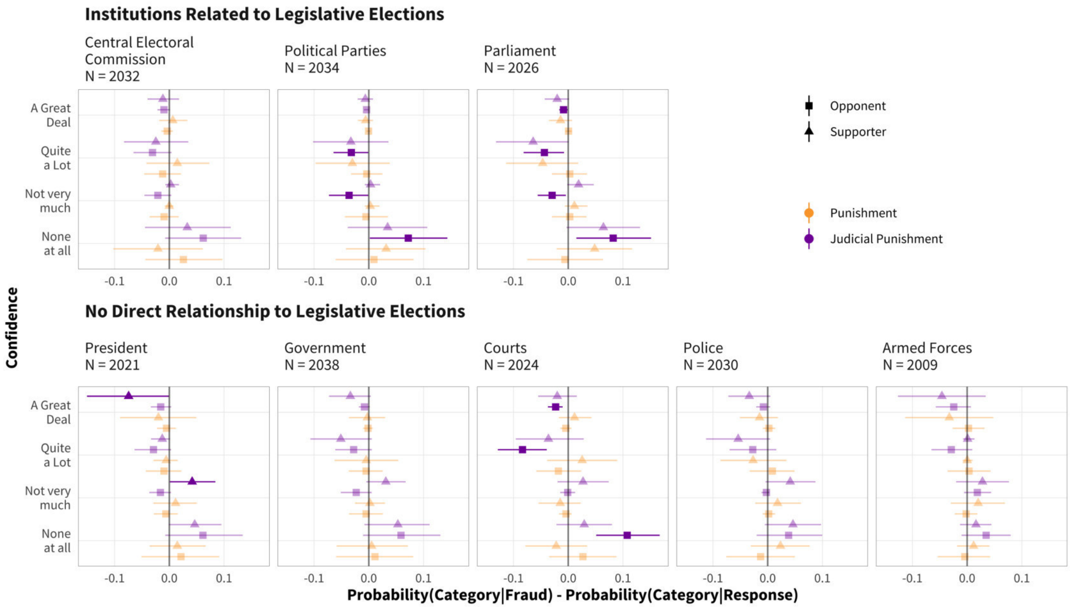


Figure 4.4. The effects of exposure to perpetrators' judicial punishment information on confidence in political institutions. Plots depict medians and 83% (bold) and 95% (thin) highest-density intervals of differences in probabilities for choosing respective categories based on draws from the posterior predictive distributions. Transparent point ranges include zero in the 95% HDCI. Probabilities are calculated based ordered logit model estimates as lined out in Equation 1. The dashed line schematically depicts the hypothesized relationship between categories.

of spillovers in comparison to the fraud condition (2) for government supporters and a suppression of spillovers in comparison to the fraud condition (2) for government opponents. Figure 4.4 plots differences in predicted probabilities to tick any of the four trust categories between those that only received fraud information (experimental group 2) and those that additionally received information on punishments (experimental groups 2 and 3). There are several key take aways. First, punishment information does predominantly not change spillover dynamics as for most institutions, punishment information does—on average—not alter levels of political trust in comparison to those who only received information on fraud. Second, if any, we observe evidence for spillover suppression across opponents and supporters given that experimental texts mention electoral commission *and* judicial punishment. Naturally, this effect is most evident for trust in courts, as respondents that received information on electoral commission and judicial punishment exert higher confidence in courts than those who merely perceived information on election fraud. Yet, for some unique comparisons, spillovers to other institutions are mitigated as well. These patterns persist once we allow for heterogeneous effects across countries, too, and when we restrict the sample to those individuals who clearly finished reading the treatment text.

To sum up, we find sufficient support for the first hypothesis, indicating the damaging effect of information about fraud on institutional trust across both samples. At the same time, the effect of within-system responses remains ambiguous. While adequate punishment of fraud perpetrators may decrease the negative effects of fraud information, this effect is far from universal and omnipresent across institutions.

4.6 Conclusion

While nation-wide multiparty electoral events are nowadays omnipresent across democracies and (electoral) autocracies alike, these are frequently accompanied by doubts about their integrity, outright accusations of fraud, and large-scale public protests. While literature on the citizen-system nexus has already amassed a wealth of knowledge on the behavioral and attitudinal consequences that becoming aware of electoral malpractice has for the citizenry, scholars have predominantly studied individuals' evaluations of the electoral process itself (Robertson 2017), the candidates that are supposedly involved in manipulation (Reuter and Szakonyi 2021) and the government that stemmed out of an allegedly fraudulent election (Williamson 2021). On the other hand, those contributions that have explicitly focused on how perceptions of election fraud influence individuals' attitudes towards components of the political system that are not tied to electoral administration have relied on cross-sectional survey data (Norris 2014; Norris 2019).

At the same time, widespread accusations of electoral fraud regularly induce political interventions by exactly those political elites that allegedly are the beneficiaries of manipulation such as the dismissal or judicial punishment of electoral staff (compare Reuters 2016). To our best knowledge, no contribution has so far focused on how credible within-system responses shape individuals' attitudinal responses to allegations of fraud.

In this chapter, we have first empirically showcased the limitations that cross-sectional survey data holds for the study of the attitudinal nexus between citizens and the political system. Even when applying a range of matching algorithms combined with a Bayesian estimation procedure, spillover effects on political institutions cannot be distinguished from placebo effects that should not be in place in theory, which casts serious doubts on the causal identification of attitudinal spillovers as a whole. Second, overcoming these limitations, we have presented results from a pre-registered online survey experiment conducted in Colombia, Mexico and Russia. Our experimental results document that disseminating credible information on election fraud does indeed induce decays in diffuse support for political institutions that are unrelated to electoral administration. Third, we showed that these attitudinal spillovers largely persist even if information on electoral manipulation is accompanied by information on credible punishments of micro-level agents of fraud.

Our study holds two main implications for developing democracies as well as contemporary authoritarianism. The first is that the negative effect of election fraud information is substantially more detrimental than currently acknowledged by the academic literature. This is because information on electoral misconduct even induces shifts in public support towards components of the political system that are no beneficiaries of manipulation and are not related to electoral administration. Second, our findings hold implications for the practice of election monitoring itself. Our chapter well aligns with a set of studies that have highlighted the cost among civil society when election observation missions expose cheating (Daxecker 2012). Our preliminary findings suggest that when large-scale observation missions that are perceived and framed as credible players in the field claim election malpractice to be at place, such exposure may have detrimental effects that may hinder, rather than foster, the consolidation of a democratic society. This is especially relevant against the backdrop of widespread criticism that has been voiced against recent election observation missions proclaiming early conclusions about electoral malpractice that later do not uphold more intensive scrutiny (Idrobo, Kronick, and Rodríguez 2020). While the spillover effect of fraud information is consistent, the mitigating effect of effective punishments is not. Third-party actors who monitor and evaluate the legitimacy of electoral events are hence advised to cautiously reflect negative assessments of electoral integrity before disseminating information.

5 Conclusion

5.1 Research Questions and Answers

This dissertation departed from two central observations which guided the research endeavors that have been presented in this work. The first observation was that as manipulation strategies of micro- and macro-agents of fraud are constantly under change and as new ways to taint electoral events are constantly developing, there—likewise—is a continuous need for developing statistical methodology to identify these strategies. The second observation was that any method that is developed in the field of statistical electoral anomaly detection is operating in a tension between defending electoral integrity and producing democratic backlashes itself. This is because credible information about electoral malpractice will—if unsubstantiated or not—likely produce attitudinal responses among the recipients of such information, with the scope of attitudinal decays being far from known.

The aim of this dissertation then was to develop statistical methodology for the detection of systematic irregularities in fine-graded election results and to enhance our understanding of the attitudinal consequences of exposing individuals to information on electoral malpractice. In developing statistical methodology, the goal was to go beyond the approach of testing universal 'statistical laws' such as the generalization of *Benford's law* presented in Equation (1.1) and its test in (1.2), but to explicitly incorporate *context-specific characteristics* into the fraud detection prototypes that were presented. This led to the first research question:

> *Research Question 1: How can context-specific characteristics of electoral events be exploited for the statistical detection of systematic election irregularities?*

The commonalities of *Chapters 2* and *3* are that both present approaches to context-specific anomaly detection that (i) explicitly make use of the ideosyncracies of different political (and electoral) systems (ii) and incorporate these using statistical simulation. The difference between both presented methodologies is that they incorporated context-specific information in vastly distinct ways.

In *Chapter 2*, I made use of the fact that elections often are not administered as standalone events but take place side-by-side with other electoral events that are conducted at the same localities. Motivating the example of the Ecuadorian General Elections 2017, I noted that once voters arrived at the polling booth, they directly elected (i) the head of government in a presidential race, (ii) the members of the country's national assembly, (iii) parliamentary members of 24 regional assemblies, (iv) Ecuador's five national representatives for the Andean parliament—the deliberative body of the Andean community—and (v) cast votes in a nation-wide referendum prohibiting politicians and civil servants to hold bank accounts in countries with preferential tax regimes and low tax jurisdictions. I lined out that if multiple elections are administered side-by-side, the phenomenon of *undervoting irregularities* can occur, namely different numbers of ballots (turned out voters) being reported across different electoral events for the same polling station. After showcasing that turnout discrepancies can either be the result of administrative errors or unbalanced fraud approaches in which ballots are added to (removed from) different races to unequal extents, I presented a statistical model which produces point estimates and uncertainty intervals for the share of polling stations with undervoting irregularities at which unbalanced fraud is expected to have affected the results. In *Chapter 2*, context-specific information is incorporated in the statistical detection of systematic election irregularities by explicitly including information from parallel electoral contests into the developed methodology.

In *Chapter 3*, the approach to include context-specific information in its essence was to iteratively simulate synthetic voting returns that resemble actual electoral results across a range of numerical characteristics *given the number of electoral units, type of electoral units and number of eligible voters per unit for each particular country*. Subsequently, these synthetic voting returns were used to directly implement different types and degrees of fraud during the data generating process. After training random forest regression trees on numerical features of the synthetic data to quantify the percentage of votes that were affected by manipulation, the developed approach was applied to datasets of different levels of aggregation of three Western democracies and three contested elections from electoral autocracies labeling the former as clean and the latter as tainted to different degrees. In this chapter, context-specific information is incorporated in the statistical detection of systematic election irregularities by explicitly simulating expected values of a *combination* of forensic indicators under the circumstances of a particular country case—in contrast to *a priori* defining these as a statistical law that is expected to hold globally across time and space.

After having presented these methodological contributions, this dissertation turned to the second key observation that was guiding my research endeavor and investigated the following question:

Research Question 2: What are the attitudinal consequences of exposing cheating?

In *Chapter 4*, together with my co-author Viktoriia Semenova, we first reviewed the past literature that was devoted to this research question and outlined that consciousness of election fraud was shown to let individuals withdraw support from candidates, institutions and governments that are supposedly involved in manipulation (Williamson 2021; Reuter and Szakonyi 2021). Presenting evidence from a pre-registered survey experiment that we conducted in Colombia, Mexico and Russia, we provided two contributions to the literature. First of all, we showed that election fraud information will let individuals extrapolate legitimacy loss even to political institutions that are unrelated to electoral events and lead to decays of trust in the political system as a whole. Second, we showed that while within-system corrections such as court punishments of alleged fraud perpetrators can mitigate decays in diffuse support, this mitigating effect is far from consistent across countries and institutions. The attitudinal consequences of exposing cheating hence go beyond what is currently acknowledged in the literature, as they are (i) more far reaching than just affecting those institutions that are directly involved in cheating (ii) and are hard to mitigate even when actors step in to provide 'checks and balances'.

5.2 Implications and Avenues for Future Research

The findings of my dissertation come with several implications that open up avenues for future research. In particular, I would like to highlight three aspects that I consider to be important implications of the evidence that has been presented in the last chapters.

Alternative Explanations for Systematic Undervoting

In *Chapter 2*, I paid substantial formal detail to showcasing that if turnout (or ballot) discrepancies between individual races result from random errors, then—independently of their extent—these will be unrelated to variables such as the winner's vote share. Secondly, I showcased that if undervoting irregularities are systematic—for example when stemming from unbalanced fraud approaches in favor of one party (candidate)—, then the statistical relationships between undervoting and other variables can be exploited to reverse-engineer the extent of unbalanced fraud in a simulation-based approach.

However, it is very clear that unbalanced election fraud is not the only systematic factor that can produce undervoting irregularities in a non-random manner. A whole range of other dynamics could be present that lead to similar empirical patterns as those that are observed under unbalanced election fraud. In *Chapter 2*, I discussed systematic administrative failures as a function of voters' education levels as one potential factor, yet provided empirical evidence against this factor driving the systematic relationship between undervoting and winner's vote shares that was observed in the Ecuadorian General Elections 2017. Future research could continue to assemble systematic factors that explain the extent of undervoting, either as a means to improve political representation, or as a way to carve out rival explanations to unbalanced fraud approaches for forensic data analytics.

Generative Models for Synthetic Data Construction

In *Chapter 3*, I developed a data generating protocol to construct synthetic voting returns that resemble factual election results across a range of dimensions. In doing so, I departed from first explicitly defining the numerical attributes that should be re-constructed in the synthetic data, and developed a simulation strategy that resembles these features well. The advantage of this approach, naturally, is that it allows for efficient synthetic data generation that includes only those properties that are known (and implemented) by the researcher. *Generative models* have taken the world of statistical modeling and machine learning by storm in the recent decade (Harshvardhan et al. 2020), and are applied to a range of domains as wide as developing fake images and videos and differentially private data for two-dimensional observational data (Neunhoeffer, Wu, and Dwork 2021).

It is intuitive that models for synthetic data creation such as adversarial neural networks (see Harshvardhan et al. 2020) might fruitfully be applied in the field of forensic data analytics as well. The application of such methodology yet comes with the drastic shortcoming that these approaches use factual datasets to construct synthetic counterparts. When dealing with empirically observed electoral returns, however, we are never really sure which data generating process actually led to a factually observed dataset, that is, we usually do not know whether some dataset is clean or was manipulated.[1] This *uncertainty in class membership* poses a central challenge. If we don't know how clean election data across, for instance, more than 90,000 Russian polling stations look like, which merit do we get out of synthetically replicating factually observed data that we cannot explicitly label as clean or tainted?

Scholarly work on generative models for constructing synthetic electoral returns could certainly tackle these issues. One strategy might lie in reproducing synthetic data from datasets that are supposedly clean (such as voting returns from advanced industrialized democracies from uncontested elections) and afterwards implementing fraud manually. Certainly, this would increase the 'representativeness' of synthetic data even across numerical attributes that have not been explicitly specified by the researcher. Yet, this comes against a computational cost and uncertainties around which data generating process the empirical data that has been synthetically replicated entails.

[1] If we would know, there would be no need to perform forensic data analytics in the first place.

Conditions for Attitudinal Spillovers of Fraud Information

In *Chapter 4*, my co-author Viktoriia Semenova and me showed that while exposing individuals to credible information on election manipulation does robustly decrease the trust that they place in political institutions even if these are no beneficiaries of malpractice and unrelated to electoral administration in the first place, the mitigating effect of horizontal within-system interventions is far from consistent. When respondents additionally received information on electoral commission and court punishments of alleged perpetrators, they only marginally differed from those respondents that just received fraud information and did so only for a small subset of political institutions, such as the courts that were involved in the punishment themselves.

As substantial shares of electoral events in developing democracies and authoritarian regimes are accompanied by accusations of election fraud, a whole range of questions that relate to the findings that we present in *Chapter 4* are naturally arising. For instance: What are the contextual and individual factors that determine whether within-system interventions can mitigate decays in diffuse support? What is the role of source credibility when distribution information on electoral manipulation? Do spillovers only arise from credible sources? Which sources lead to spillovers and which do not? Questions like these are only to be answered throughout a series of research studies, and there is no indication that these will lose their empirical relevance in subsequent years.

5.3 Concluding Thoughts

Nowadays, around half of all national-level electoral events result in some form of allegation about manipulation and fraud (Rozenas 2017). The public discourse about the integrity of electoral events increasingly focuses around numerical patterns in published election results that are perceived to be fraudulent. After the Bolivian elections of 2019, an audit team of the Organization of American States examined the development of Evo Morales's vote shares across the counting stage, and diagnosed a striking discontinuity between the percentage of votes that was counted in favor the incumbent president before and after an 'arbitrary threshold' of 95% counted votes (OAS 2019a). The main criticism that was voiced by the OAS stemmed from the results of two local constant regression models and inferences about their behavior at boundary points, an inference which is known to lead to erroneous conclusions (Hastie, Tibshirani, and Friedman 2009; Cattaneo, Idrobo, and Titiunik 2020). After the US 2020 presidential election, a range of news outlets stated that while Donald Trump's absolute vote totals followed a probability distribution called *Benford's law* in county-level election data, the vote totals of his main contestor did not, and took this as an empirical indication for a stolen election. After the 2011 parliamentary election in Russia, a range of political observers noted that in the polling station-level electoral results, the prevalence of exactly coarse turnout and vote shares that are multiples of five in those polling stations in which the government party United Russia received overwhelmingly high support (0.6, 0.65, 0.7, 0.75, 0.8) was striking and—intuitively—odd.

Against this background of a public discourse and political observers in particular that are increasingly aware that—as in many fields and use cases—also electoral returns inherit numerical characteristics that are given under clean elections and violated under fraud, the field of statistical detection of systematic election irregularities finds itself in a two-fold role. One the one hand, statistical methods for anomaly detection can be useful tools to provide election observer missions and the citizenry with the necessary tool set to quantify anecdotal claims of electoral malpractice and placing eye witnesses' anecdotal reports on a solid methodological footing using statistical methods. On the other hand, especially in an area of democratic backlashes in advanced industrialized democracies (Norris and Inglehart 2019), the same tools need to be increasingly used to safeguard clean and legitimate elections from populist incumbent narratives of stolen elections that are increasingly developed already long before election day.

This puts developers of statistical approaches such as the ones that I have presented in this dissertation in a delicate position. On the one hand, we need to fine-tune methodologies in such a way that these will robustly filter out anomalies in electoral events that actually were manipulated to not place wrong legitimacy in authoritarian elections whose outcomes do not represent the will of the people. On the other hand, models need to be safeguarded against producing too many type-I errors, diagnosing systematic irregularities where they can easily be explained by alternative circumstances.

After all, statistical detection of systematic election irregularities is by no means a panacea that erases the need for systematic qualitative investigations, selective recounting of votes and rigorous election observation especially in those events in which political tensions are high. Only in conjunction with these alternative strategies, it can realize its full potential.

A

Appendix for Chapter 2

A.1 The Distribution of Undervoting Irregularities

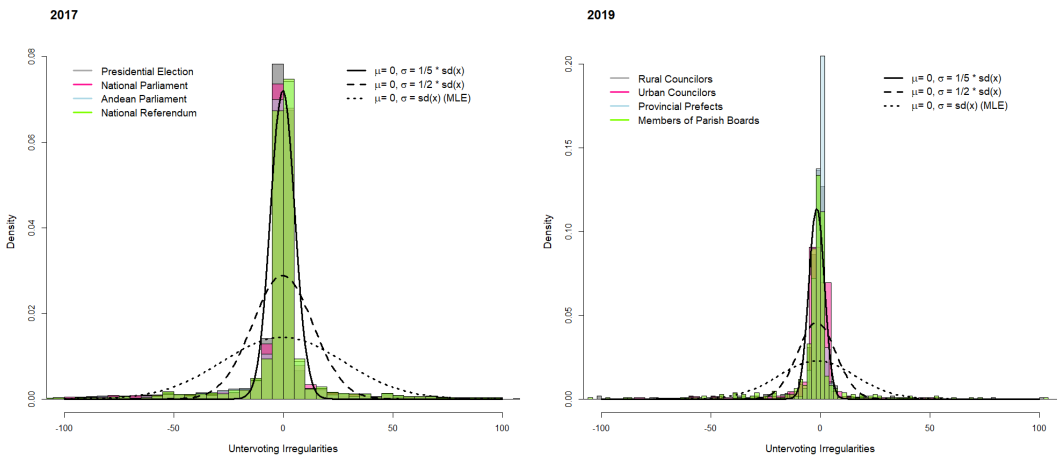


Figure A.1. The distribution of undervoting irregularities in Ecuadorian elections. Left panel: General Elections 2017. Right panel: Local Elections 2019. Histograms visualize the discrepancies in absolute numbers of documented turned out voters between the election of state-level members of parliament (2017) and city mayors (2019) and four concurrent elections. Black lines depict normal distributions scaled by different dispersion parameters that are estimated from the data.

Observed turnout levels $\mathcal{T}_i$ can be decomposed as

$$\mathcal{T}_i = \mathcal{T}_i^* + \mathcal{T}_i^{\epsilon}{}_i, \tag{A.1}$$

where $\mathcal{T}_i^* \in [0, N_i]$ is the true number of total votes cast and $\mathcal{T}_i^{\epsilon} \in [0, N_i]$ is the absolute number of votes that has been added (removed) by error or fraud. Across all polling stations $i \in \{1, ..., n\}$, turnout discrepancies $\mathcal{T}_i^{\epsilon}$ are distributed as

$$\mathcal{T}_i^{\epsilon} \sim \mathcal{N}(\mu, \sigma^2). \tag{A.2}$$

Figure A.1 shows that the normality assumption fits the data well. Undervoting irregularities are dispersed around a mean of $\mu = 0$. The maximum likelihood estimate of the dispersion parameter is given by $\sigma = \sqrt{\frac{\sum_{i=1}^{n} \mathcal{T}_i^{\epsilon} - \mu}{n}}$ and provides a poor fit to the kurtosis of the distributions. Estimating the dispersion parameter as

$$\sigma = \frac{1}{5} * \sqrt{\frac{\sum_{i=1}^{n} \mathcal{T}_i^{\epsilon} - \mu}{n}} \tag{A.3}$$

provides a close fit to undervoting irregularities (i) across different elections (ii) and years.

A.2 Prior Distributions for Parameter Estimation

In order to sample from the distributions outlined in Equations (7)-(9), parameters $\{\alpha^t, \beta^t, \alpha^v, \beta^v, \sigma\}$ can either be estimated using maximum likelihood estimation from the empirical data that is being used. Alternatively, to incorporate fundamental uncertainty of the parameters and iterate the algorithm across a range of different parameter values, parameters can be parameterized by prior distributions. The execution of the semi-parametric simulation model is then repeated once for each posterior sample. The following prior distributions underlie the execution of the model:

$$\begin{aligned}
\sigma &\sim \text{InvGamma}(0.001, 0.001) \\
\alpha^t &\sim \text{N}(0, 1000) \\
\beta^t &\sim \text{N}(0, 1000) \\
\alpha^v &\sim \text{N}(0, 1000) \\
\beta^v &\sim \text{N}(0, 1000)
\end{aligned}$$

A.3 R Code for Executing the Semi-Parametric Simulation Model

The code snippet below documents an exemplary execution of the semi-parametric simulation model by the user for data from the Ecuadorian General Elections 2017. The election for members of regional parliaments (*asambleístas provinciales*) is set as the baseline electoral contest. Estimated is the share of polling stations with undervoting at which unbalanced fraud between the election of interest and the baseline electoral contest is perpetrated. Function calls incorporate fundamental and estimation uncertainty and output estimated shares together with 95% credible intervals.

```
library(EnvStats)
library(fields)
library(foreign)
library(rstan)

# load data at polling station level, General Elections 2017
load("actas17.Rdata")

# delete polling stations with <100 eligible voters
actas17 <- actas17[-which(actas17$ELECTORES_REGISTRO_pres<100),]

# run model, presidential election
actas17 <- actas17[-which(actas17$turnout_pres>1),] # exclude if turnout > 1
ecu17_pres <-
  est_fraud(eligible = actas17$ELECTORES_REGISTRO_pres,
            turnout_main = actas17$SUFRAGANTES_pres,
            turnout_baseline = actas17$SUFRAGANTES_asam_prov,
            winner_main = actas17$MORENO_pres,
            uncertainty = c("fundamental", "estimation"),
            n_iter = 100,
            n_postdraws = 500,
            n_burnin = 400,
            seed = 12345
            )

# run model, national parliament election
actas17 <- actas17[-which(actas17$turnout_nac>1),] # exclude if turnout > 1
ecu17_nac <-
  est_fraud(eligible = actas17$ELECTORES_REGISTRO_asam_nac,
            turnout_main = actas17$SUFRAGANTES_asam_nac,
            turnout_baseline = actas17$SUFRAGANTES_asam_prov,
            winnershare_main = actas17$winnershare_asam_nac,
            uncertainty = c("fundamental", "estimation"),
            n_iter = 100,
            n_postdraws = 500,
            n_burnin = 400,
            seed = 12345
            )

# run model, Andean parliament election
actas17 <- actas17[-which(actas17$turnout_andean>1),] # exclude if turnout > 1
ecu17_andean <-
  est_fraud(eligible = actas17$ELECTORES_REGISTRO_andino,
            turnout_main = actas17$SUFRAGANTES_andino,
            turnout_baseline = actas17$SUFRAGANTES_asam_prov,
            winnershare_main = actas17$winnershare_andino,
            uncertainty = c("fundamental", "estimation"),
            n_iter = 100,
            n_postdraws = 500,
            n_burnin = 400,
            seed = 12345
            )

# run model, national referendum
actas17 <- actas17[-which(actas17$turnout_referend>1),] # exclude if turnout > 1
ecu17_referendum <-
  est_fraud(eligible = actas17$ELECTORES_REGISTRO_consulta,
            turnout_main = actas17$SUFRAGANTES_consulta,
```

```
            turnout_baseline = actas17$SUFRAGANTES_asam_prov,
            winner_main = actas17$Si_consulta,
            uncertainty = c("fundamental", "estimation"),
            n_iter = 100,
            n_postdraws = 500,
            n_burnin = 400,
            seed = 12345
            )
```

A.4 The Urban-Rural Divide in Undervoting Irregularities

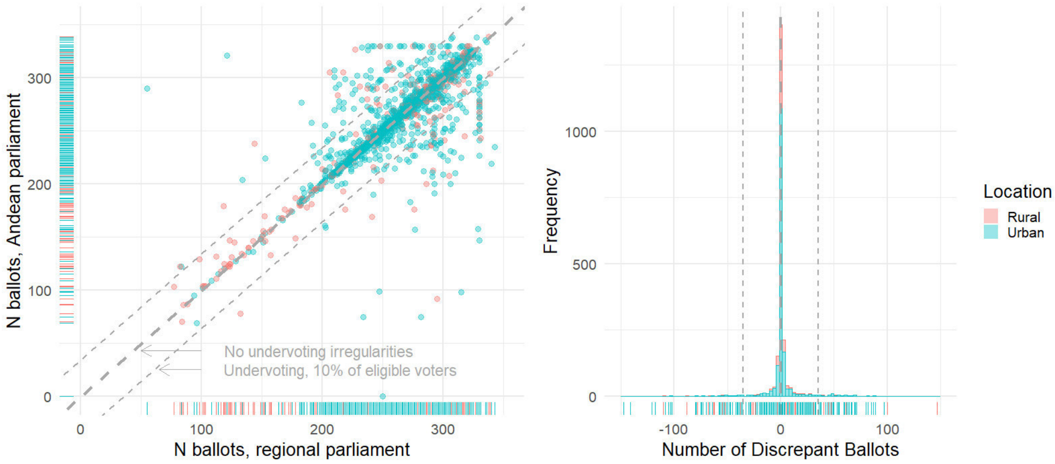


Figure A.2. The urban-rural divide in undervoting irregularities, General Elections 2017. The left panel plots the absolute number of documented turned out voters for the election of members of the Andean parliament vs. the election of state-level members of parliament. The right panel depicts a histogram of the number of discrepant ballots between the two electoral events. Plots are generated separately for urban and rural localities. There are no significant differences in undervoting irregularities between urban ($n = 29,461$) and rural $(8,992)$ localities, $t = 0.11, p = 0.91$.

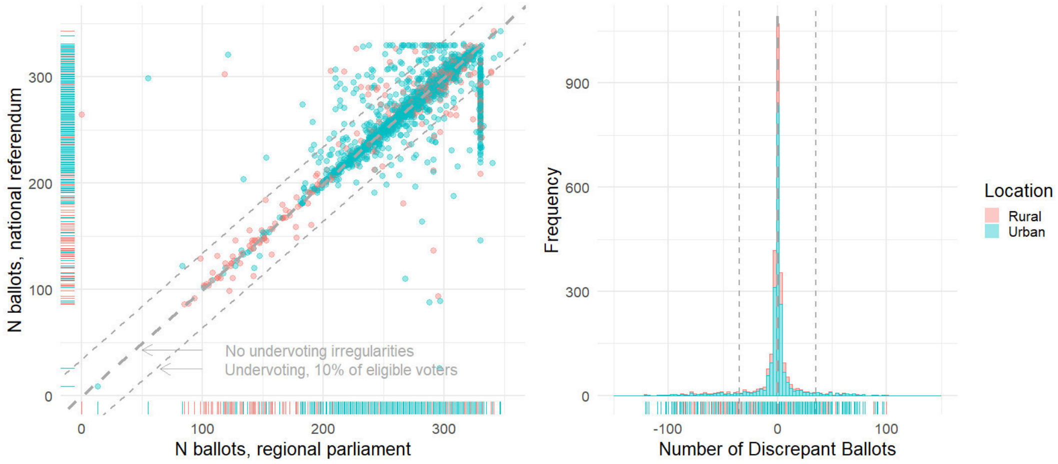


Figure A.3. The urban-rural divide in undervoting irregularities, General Elections 2017. The left panel plots the absolute number of documented turned out voters for the national referendum vs. the election of state-level members of parliament. The right panel depicts a histogram of the number of discrepant ballots between the two electoral events. Plots are generated separately for urban and rural localities. There are no significant differences in undervoting irregularities between urban ($n = 29,461$) and rural ($8,992$) localities, $t = 1.34, p = 0.18$.

Statistic	N	Mean	St. Dev.	Min	Max
Undervoting, presidential vs. regional parliament	29,515	−0.179	5.698	−243	237
Undervoting, national vs. regional parliament	29,514	−0.049	5.397	−147	234
Undervoting, Andean vs. regional parliament	29,514	−0.053	5.921	−250	235
Undervoting, Referendum vs. regional parliament	29,515	−0.121	6.261	−270	244

Table A.1. Descriptive statistics of undervoting irregularities for urban polling stations, General Elections 2017. Documented are discrepancies in the raw numbers of turned out voters.

Statistic	N	Mean	St. Dev.	Min	Max
Undervoting, presidential vs. regional parliament	9,000	−0.288	7.232	−201	265
Undervoting, national vs. regional parliament	9,000	−0.098	4.720	−128	147
Undervoting, Andean vs. regional parliament	9,000	−0.060	5.145	−203	100
Undervoting, Referendum vs. regional parliament	9,000	−0.233	7.092	−201	265

Table A.2. Descriptive statistics of undervoting irregularities for rural polling stations, General Elections 2017. Documented are discrepancies in the raw numbers of turned out voters.

B

Appendix for Chapter 3

Country	Year	Election	n	N_{min}	N_{mean}	N_{max}
Austria	2008	Parliamentary election	2,535	103	7,495	266,391
Finland	2017	Municipal election	992	19	6,421	778,028
Spain	2019	European Parliament election	6,622	2	3403	1,623,091
Russia	2011	Parliamentary election	90,919	100	1,196	22,671
Russia	2012	Presidential election	91,256	100	1,204	79,711
Uganda	2011	Presidential election	23,754	100	583	1,766

Table B.1. Descriptive statistics on electoral returns. n refers to the number of electoral units that are available for each country-election. N_{min}, N_{mean} and N_{max} refer to the minimun, mean and maximum electoral unit size N_i across all units.

	0	1	2	3	4	5	6	7	8	9
First Significant Digit	-	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046
Second Significant Digit	0.120	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.088	0.085
Third Significant Digit	0.102	0.101	0.101	0.101	0.100	0.100	0.099	0.099	0.099	0.099

Table B.2. Frequencies of digits as expected by Newcomb-Benford's law. For the first significant digit, frequencies are generated by $P(d) = log_{10}(1 + 1/d)$ for digit d ($d \in 1, 2, .., 9$). For subsequent digits, frequencies are generated from $P(d) = \sum_{k=10^{n-2}}^{10^{n-1}} log_{10}(1 + \frac{1}{10k+d})$ for digit d ($d \in 0, 1, 2, .., 9$) arising in the nth position ($n > 1$). The value '0.301' in Row 1, Column 2 means that across the set of analyzed numerical entries, the number '1' should appear 30.1% of the time as the first digit.

Abbreviation	Variable Description	Variants
Digit characteristics		
Second digit		
bl2_frac1	Fraction of number ‘1’ among all second digits across electoral units	Separately for Party A and B
bl2_mean	Mean of second digit across all electoral units	Separately for Party A and B
bl2_chi2	χ^2-statistic between observed and expected frequency of numbers in the second digit	Separately for Party A and B
Last digit		
bllast_frac1	Fraction of number ‘1’ among all last digits across electoral units	Separately for Party A and B
bllast_mean	Mean of last digit across all electoral units	Separately for Party A and B
bllast_chi2	Chi^2 statistic between observed and expected frequency of numbers in the last digit	Separately for Party A and B
Characteristics of turnout distribution		
turnout_skew	Skewness of turnout distribution	Once
turnout_kurt	Kurtosis of turnout distribution	Once
turnout_kurt80	Kurtosis of turnout distribution only taking into account values above 80% turnout	Once
turnout_normdist	Kullback Leibler divergence between observed turnout distribution and a normal distribution estimated with mean and variance from observed data	Once
Characteristics of vote share distributions		
share_skew	Skewness of vote share distribution	Separately for Party A and B
share_kurt	Kurtosis of vote share distribution	Separately for Party A and B
share_kurt80	Kurtosis of vote share distribution only taking into account values above 80% vote share	Separately for Party A and B
share_normdist	Kullback Leibler divergence between observed vote share distribution and a normal distribution estimated with mean and variance from observed data	Separately for Party A and B
Frequency of coarse vote shares		
coarse_frac	Share of electoral units with vote shares for the winning party that are a multiple of ’5’	Once

Table B.3. List of feature variables and their abbreviations that are used to predict the percentage of votes that are subject to manipulation. Features are derived from existing indicators developed in the election forensics literature.

C

Appendix for Chapter 4

C.1 Cross-National Evidence: Matching Analysis

C.1.1 Descriptive Statistics

Table C.1 contains the summary statistics for complete cases, World Values Survey Wave 7 (2017-2020).

	N	Mean	St. Dev.	Min	25%	75%	Max
Confidence in Armed Forces	42,586	2.865	0.945	1	2	4	4
Confidence in the Police	42,586	2.629	0.943	1	2	3	4
Confidence in Courts	42,586	2.568	0.941	1	2	3	4
Confidence in Parliament	42,586	2.187	0.926	1	1	3	4
Confidence in Government	42,586	2.384	0.984	1	2	3	4
Confidence in Parties	42,586	2.034	0.874	1	1	3	4
Confidence in Companies	42,586	2.398	0.851	1	2	3	4
Confidence in UN	42,586	2.445	0.943	1	2	3	4
Confidence in Banks	42,586	2.563	0.930	1	2	3	4
Confidence in WTO	42,586	2.439	0.911	1	2	3	4
Confidence in World Bank	42,586	2.419	0.948	1	2	3	4
Fraud Perception	42,246	2.841	1.028	1	2	4	4
Fraud Perception (binary)	42,586	0.651	0.477	0	0	1	1
Political Interest	42,586	2.390	0.955	1	2	3	4
Generalized Trust	42,246	0.202	0.401	0	0	0	1
Female	42,246	0.495	0.500	0	0	1	1
Age	42,586	41.634	15.832	16	28	54	103
Family Savings	42,586	2.049	0.902	1	1	2	4
Rural	42,246	0.339	0.473	0	0	1	1
Perceived Political Corruption	42,586	7.736	2.423	1	6	10	10

Table C.1. Summary statistics of key variables, World Values Survey Wave 7 (2017-2020).

- Confidence in political institutions is measured on a 1 to 4 scale, with 1 meaning *None at all* and 4 depicting *A great deal*.

- Political interest is measured on a 1 to 4 scale, with 1 meaning *Not at all interested* and 4 depicting *Very interested*.
- Generalized trust is a binary indicator, with 0 meaning *Need to be very careful* and 1 depicting *Most people can be trusted*.
- Fraud perception is measured in WVS on a 1 to 4 scale based on the agreement with the following statement: "How often in country's elections: Votes are counted fairly", with 1 meaning *Not at all often* and 4 depicting *Very often*.
- Fraud perception (binary) is the variable used in our matching analysis and is a binary indicator, with 0 being comprised of *Not at all often* and *Not often* and 1 of *Fairly often* and *Very often* answers to the respective WVS question.
- Family savings during past year is measured on a 1 to 4 scale, with 1 indicating Save money, 2—*Just get by*, 3—*Spent some savings and borrowed money*, and 4—*Spent savings and borrowed money*.
- Perceived Political Corruption is measured on a 1 to 10 scale, with 1 meaning *There is no corruption in my country* and 10 depicting *There is abundant corruption in my country*.

b. Covariate Balance

Figures present measures of covariate balance between those individuals falling into the 'fraud' and 'no fraud' perception condition before and after our sample adjustment. For each covariate that we match on, we report the standardized bias as measured by the difference in means between both groups of individuals scaled by the pooled standard deviation. Dots to the right (left) of the dashed vertical line are indicative of a higher incidence of respective characteristics among those individuals that perceive the elections of their country as unfair (fair). As indicated by the grey circles, fraud perceptions are most prevalent among those individuals that report to never turn out to vote, obtain low levels of interpersonal trust, and perceive political corruption at large scale in their country.

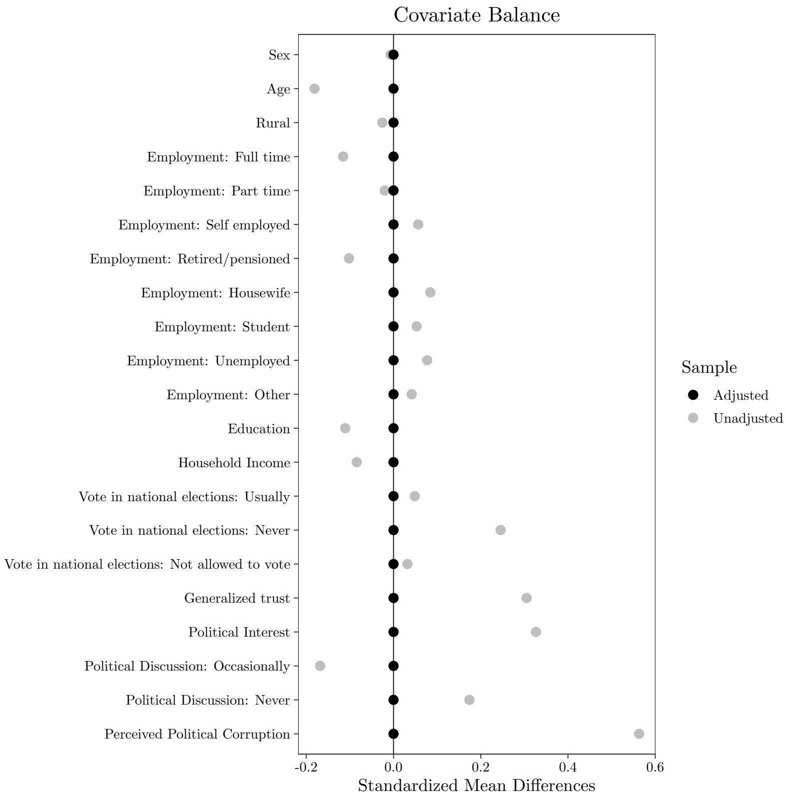


Figure C.1. Covariate balance before and after sample adjustment using direct exact matching.

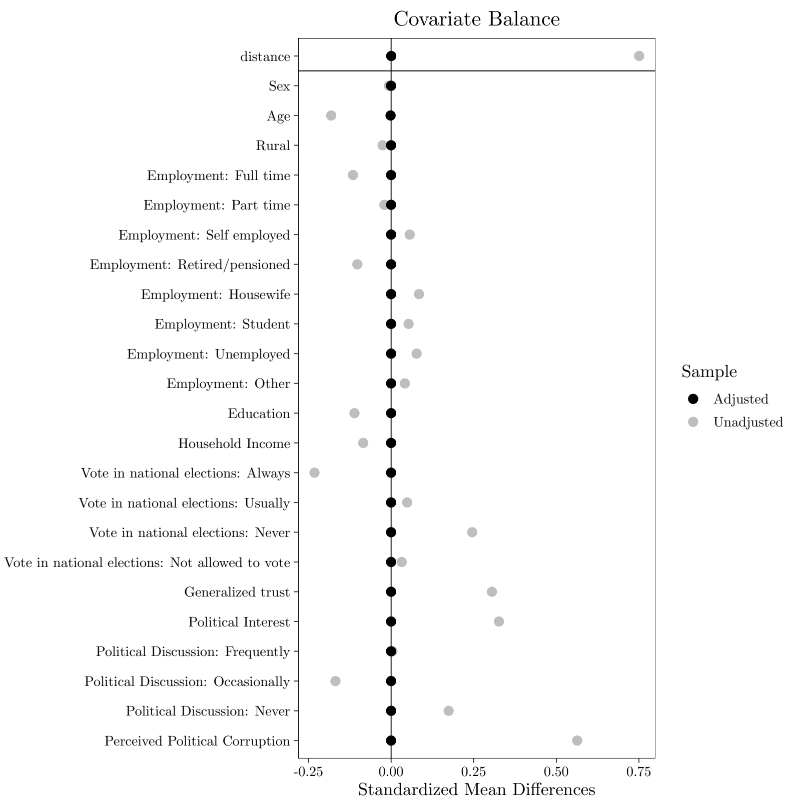


Figure C.2. Covariate balance before and after sample adjustment using coarsened exact matching.

C.2 Details on Case Selection

In the post-World War II era, Mexico underwent a gradual and often described as pendular democratization process (Cantú and García-Ponce 2015; Hiskey and Bowler 2005). Since the end of the Mexican Revolution in 1917, elections were held regularly in six-year intervals and political opposition was granted passive voting rights. Yet the authoritarian rule of the Institutional Revolutionary Party (PRI) effectively consolidated a hegemonic one-party party system (Sartori 1976). Popular distrust in the legitimacy of Mexican elections roots in the experience of PRI's one-party rule, which was notoriously engaged in attempts of electoral manipulation against parties from both sides of the political spectrum. PRI's strategies in balancing out authoritarianism with competitive elections manifested in unrecognized victories of the right-wing National Action Party (PAN) in a multitude of subnational elections in the 1980s and 1990s (Greene 2007; Cantú and García-Ponce 2015), systematic repression against the left-winged Democratic Revolution Party's (PRD) candidates (Greene 2007), and election-day fraud such as the manual alteration of vote tallies in multiple regional and national-level contests (c.f. Cantú 2019b). It was not before the 1980s that electoral competition led to more inclusive electoral contests which produced changing majorities. First, recognized opposition victories occurred in state-level and local elections and only recently culminated in the first national-level contest since the Mexican Revolution of 1920 being decided in favor of the political challenger when the National Action Party's (PAN) candidate defeated PRI's Francisco Labastida in July 2000. Today, Mexico's political party system shows a remarkable level of institutionalization when compared to other new democracies and locates the country on the upper end on the scale of party system stability in Latin America (Greene and Sánchez-Talanquer 2018). Notably, the historical baggage of electoral maladministration and attempts of manipulation persists and reaches forward up until the country's most recent electoral events (Cantú 2014; Cantú 2019a).

Colombia's history of democratization holds several paradoxes. Formally, regular multi-party elections are held since the 1830s. In practice, electoral events consistently trigger fraud accusations both from the citizenry and academic literature (Duque Daza 2019) and candidates, politicians and journalists are regularly assassinated (Bejarano and Pizarro 2005). These and other observations have led scholars of the Colombian case to describe the country as a 'besieged' (Bejarano and Pizarro 2005) or 'fraudulent' electoral democracy (Duque Daza 2019), in which the institutional design and democratic practice of the country diverge strongly, which resembles the Mexican case to a large extent.

In Russia, we study a context of institutionalized authoritarian rule. After the dissolution of the Soviet Union in 1991, meaningful opposition has effectively been banned since the beginning of Vladimir Putin's administration in 1999 and several observers note that election-day fraud has metasized since in the earlier 2000s

(Myagkov, Ordeshook, and Shakin 2009). Consistently accompanied by widespread fraud allegations, Russian elections are also occasionally followed by anti-regime protests (Robertson 2017; Lankina and Tertytchnaya 2020). Election monitors routinely condemn Russian elections and a whole range of scholarly contributions focuses on highly unusual patterns in published voting returns that are hard to explain with processes other than manual alteration of vote counts (Rozenas 2017; Klimek et al. 2012; Myagkov, Ordeshook, and Shakin 2009; Jimenez, Hidalgo, and Klimek 2017; Kobak, Shpilkin, and Pshenichnikov 2016b; Kobak, Shpilkin, and Pshenichnikov 2018). Studies suggest that in the 2011 parliamentary elections, the vote share of the incumbent United Russia party was at least 11 percentage points lower than documented by official figures (Enikolopov et al. 2013). Today, only about 15.5% of Mexicans, 20.5% of Colombians, and about 39.8% of Russians say that election officials are fair and that votes are counted free and fairly.[1]

C.3 Survey Experiment

C.3.1 Descriptive Statistics

Tables C.3 and C.2 contain summary statistics for the survey responses in Russia and Latin America for the observations we used in the main analysis, for each treatment group. As per the pre-registration analysis plan, we exclude any unfinished interviews and cases where the interviews were finished in under 3 minutes to control for data quality. To represent the target population, i.e. potential voters, we also remove responses from respondents under 18 and non-citizens of target countries, which was also set as a restriction for participation at the recruitment stage.Our sample consists of primarily young, urban, middle-to-high educated people in both Russia and Latin America.

[1]Source: World Values Survey Wave 7 (Inglehart et al. 2020), 2017-2021.

		Fraud (N=222)		Control (N=218)		Punishment (N=216)		Judicial Punishment (N=216)	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Confidence in Armed Forced		2.3	0.9	2.5	0.9	2.2	0.9	2.3	0.9
Confidence in Police		1.8	0.7	1.9	0.7	1.8	0.8	1.9	0.8
Confidence in CEC		1.8	0.8	2.3	0.8	1.9	0.8	1.9	0.8
Confidence in Government		1.7	0.8	2.0	0.8	1.7	0.7	1.7	0.8
Confidence in Parties		1.6	0.7	1.8	0.7	1.6	0.6	1.6	0.6
Confidence in Pariament		1.7	0.7	2.0	0.8	1.7	0.7	1.8	0.8
Confidence in Courts		1.9	0.8	2.2	0.8	1.9	0.7	2.1	0.9
Confidence in President		1.7	0.8	2.0	1.0	1.7	0.9	1.8	0.9
Confidence in Elections		1.8	0.8	2.3	0.8	1.9	0.8	2.0	0.8
Confidence in Companies		2.2	0.8	2.3	0.9	2.2	0.8	2.3	0.9
Confidence in Banks		2.3	0.9	2.2	0.9	2.1	0.8	2.3	0.9
Confidence in Environmental Organizations		3.0	0.8	3.1	0.8	2.9	0.8	2.8	0.9
Confidence in UN		2.8	0.9	3.0	0.8	2.8	0.9	2.9	1.0
Confidence in World Bank		2.5	0.9	2.6	0.9	2.5	0.9	2.5	0.9
Confidence in WTO		2.6	0.8	2.7	0.8	2.5	0.8	2.5	0.8
Political Interest		3.0	0.8	2.9	0.7	3.1	0.7	3.0	0.8
Age		26.3	7.4	27.1	7.9	27.6	8.8	27.1	8.0
Pol. Corruption		9.1	1.5	9.2	1.4	9.1	1.5	9.2	1.1
Involvement		2.8	0.6	2.9	0.5	2.8	0.6	2.8	0.5
		N	%	N	%	N	%	N	%
Generalized Trust	Most people can be trusted	29	13.1	19	8.7	29	13.4	25	11.6
	Need to be very careful	192	86.5	194	89.0	184	85.2	190	88.0
Sex	Female	101	45.5	95	43.6	106	49.1	95	44.0
	Male	121	54.5	123	56.4	110	50.9	121	56.0
Education	Lower	3	1.4	2	0.9	5	2.3	7	3.2
	Middle	151	68.0	141	64.7	131	60.6	137	63.4
	Higher	66	29.7	71	32.6	72	33.3	68	31.5
Empl. Status	Paid employment	71	32.0	99	45.4	83	38.4	74	34.3
	Retired/pensioned	0	0.0	1	0.5	2	0.9	1	0.5
	Housewife	13	5.9	14	6.4	16	7.4	19	8.8
	Student	78	35.1	68	31.2	68	31.5	67	31.0
	Unemployed	51	23.0	29	13.3	33	15.3	42	19.4
	Other	9	4.1	7	3.2	14	6.5	12	5.6
Empl. Sector	Government or public institution	13	5.9	18	8.3	14	6.5	19	8.8
	Private business or industry	119	53.6	121	55.5	119	55.1	108	50.0
	Private non-profit organization	16	7.2	14	6.4	15	6.9	15	6.9
Savings: Last year	Saved money	37	16.7	36	16.5	33	15.3	41	19.0
	Just got by	92	41.4	92	42.2	97	44.9	95	44.0
	Spent some savings	43	19.4	40	18.3	47	21.8	38	17.6
	Spent savings and borrowed money	49	22.1	50	22.9	38	17.6	42	19.4
Set. Type	Rural	17	7.7	16	7.3	9	4.2	15	6.9
	Urban	205	92.3	202	92.7	207	95.8	201	93.1
Opponent	No	56	25.2	54	24.8	53	24.5	56	25.9
	Yes	166	74.8	164	75.2	163	75.5	160	74.1

Table C.2. Summary statistics of key variables, survey data for Latin America

		Fraud (N=297)		Control (N=310)		Punishment (N=309)		Judicial Punishment (N=307)	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Confidence in Armed Forced		2.7	1.0	2.8	1.0	2.8	1.0	2.9	0.9
Confidence in Police		2.3	0.9	2.3	0.9	2.3	0.9	2.4	0.9
Confidence in CEC		1.9	0.9	2.2	1.0	1.9	0.9	2.0	1.0
Confidence in Government		2.1	1.0	2.2	1.0	2.1	0.9	2.3	1.0
Confidence in Parties		1.9	0.8	2.1	0.8	2.0	0.9	2.1	0.8
Confidence in Parliament		2.0	0.9	2.1	0.9	2.0	0.9	2.2	0.9
Confidence in Courts		2.2	0.9	2.3	0.9	2.2	0.8	2.4	1.0
Confidence in President		2.4	1.1	2.5	1.1	2.4	1.0	2.6	1.1
Confidence in Elections		1.9	0.9	2.3	0.9	1.9	0.9	2.2	1.0
Confidence in Companies		2.1	0.9	2.2	0.8	2.2	0.8	2.3	0.9
Confidence in Banks		2.1	0.9	2.2	0.9	2.2	0.9	2.2	0.9
Confidence in Environmental Organizations		2.5	0.9	2.6	0.9	2.5	0.9	2.6	1.0
Confidence in UN		2.3	0.9	2.4	0.9	2.3	0.9	2.4	1.0
Confidence in World Bank		2.2	0.9	2.2	0.9	2.2	0.9	2.3	0.9
Confidence in WTO		2.2	0.9	2.2	0.9	2.2	0.8	2.3	0.9
Political Interest		2.6	0.8	2.6	0.8	2.7	0.8	2.6	0.8
Age		36.2	10.6	36.1	10.7	36.6	10.8	36.1	10.4
Pol. Corruption		7.9	2.1	8.2	1.9	8.1	1.9	8.0	1.9
Involvement		2.7	0.7	2.7	0.8	2.8	0.7	2.7	0.7
		N	%	N	%	N	%	N	%
Generalized Trust	Most people can be trusted	55	18.5	55	17.7	62	20.1	64	20.8
	Need to be very careful	227	76.4	244	78.7	240	77.7	228	74.3
Sex	Female	124	41.8	158	51.0	155	50.2	151	49.2
	Male	173	58.2	152	49.0	154	49.8	156	50.8
Education	Lower	7	2.4	10	3.2	6	1.9	7	2.3
	Middle	133	44.8	132	42.6	154	49.8	131	42.7
	Higher	153	51.5	166	53.5	147	47.6	167	54.4
Empl. Status	Paid employment	182	61.3	167	53.9	167	54.0	182	59.3
	Retired/pensioned	15	5.1	17	5.5	17	5.5	16	5.2
	Housewife	34	11.4	45	14.5	41	13.3	31	10.1
	Student	13	4.4	23	7.4	19	6.1	23	7.5
	Unemployed	38	12.8	42	13.5	48	15.5	33	10.7
	Other	15	5.1	16	5.2	17	5.5	21	6.8
Empl. Sector	Government or public institution	56	18.9	74	23.9	78	25.2	65	21.2
	Private business or industry	184	62.0	180	58.1	184	59.5	180	58.6
	Private non-profit organization	24	8.1	25	8.1	20	6.5	25	8.1
Savings: Last year	Saved money	46	15.5	55	17.7	48	15.5	66	21.5
	Just got by	148	49.8	150	48.4	164	53.1	157	51.1
	Spent some savings	59	19.9	63	20.3	57	18.4	52	16.9
	Spent savings and borrowed money	44	14.8	42	13.5	40	12.9	32	10.4
Set. Type	Rural	25	8.4	28	9.0	20	6.5	18	5.9
	Urban	272	91.6	282	91.0	289	93.5	289	94.1
Opponent	No	121	40.7	125	40.3	129	41.7	127	41.4
	Yes	176	59.3	185	59.7	180	58.3	180	58.6

Table C.3. Summary statistics of key variables, survey data for Russia.

Condition		Continued	Dropped Out	All
Fraud	N	243	3	246
	% row	98.78	1.22	100.00
Control	N	236	3	239
	% row	98.74	1.26	100.00
Punishment	N	235	0	235
	% row	100.00	0.00	100.00
Judicial Punishment	N	250	1	251
	% row	99.60	0.40	100.00
All	N	991	7	998
	% row	99.30	0.70	100.00

Table C.4. Dropout rates right after reading the treatment across experimental conditions, survey data in Latin America.

C.3.2 Data Quality

Since our target population of study are the voting-eligible population in the countries of Mexico, Columbia, and Russia, our sampling strategy implies using these characteristics as pre-requisites for participation in the survey. Hence, we excluded all the cases where the reported age of the respondent was below 18 years old, which still occurred despite the crowd-sourcing platform's official age restriction which excludes underage individuals, and the cases where the respondent claimed to not be a citizen of the respective country, also occurring despite the restrictions set on the crowd-sourcing platform.

We are limiting the analysis to complete cases only, which requires us to have a closer look at the dropout and sample response when discussing data quality. As we have relied on the crowd-sourcing platforms for participant recruitment, participants were paid for completing the survey. Given that participants only received the payment upon completion of the survey, we obtained a 95% completion rate in Russia and 92% in Latin American countries.

Apart from the absolute number of completed surveys, a valid concern regarding the dropout rate could be that it varies over our experimental conditions, introducing bias into the sampling procedure and completed cases. As evidenced by the tables below, there seem to be no systematic differences in the dropout rates across the experimental conditions, i.e. dropout rates right after reading the treatment condition.

Furthermore, as per the pre-analysis plan, we have excluded cases where the survey was completed in under 3 minutes, as reading and answering all the questions meaningfully in a shorter amount of time seems to be unrealistic. For reference, we provide summary statistics for the time required for completing the treatment text page for the finished interviews, which required summarizing the text, and statistics for institutional trust evaluation page. While small variations in completing the treatment page may be related to small differences in the length of

Condition		Continued	Dropped Out	All
Fraud	N	317	4	321
	% row	98.75	1.25	100.00
Control	N	342	5	347
	% row	98.56	1.44	100.00
Punishment	N	325	5	330
	% row	98.48	1.52	100.00
Judicial Punishment	N	323	4	327
	% row	98.78	1.22	100.00
All	N	1324	18	1342
	% row	98.66	1.34	100.00

Table C.5. Dropout rates right after reading the treatment across experimental conditions, survey data in Russia.

treatment conditions, we can conclude that the lower bound of 3 minutes seems reasonable for each condition for any of them.

We have also manually coded the summaries to differentiate the responses by the levels of engagement to allow for robustness checks. We have distinguished between *complete summaries*, which contained the condition-specific part of the treatment, *incomplete summaries*, where the responses described only a part of the treatment message but lack the crucial info for that condition; *responses* to the texts which allowed us to expect that the treatment text was read; *copy-paste* of one or multiple paragraphs from the treatment text in the questionnaire, and *unacceptable answer*, which does not allow us to observe any meaningful engagement with the treatment (e.g., a sequence of random characters). While we avoid removing any data based on this criterion in the main analysis due to potential introduction of post-treatment bias though case selection, we report the analysis on the restricted samples in robustness section, and the results show that our findings hold also when relying on a more restrictive approach to data preparation. Furthermore, for the cases where we received complete summaries, i.e. we can be certain that the treatment was read, the effects are particularly pronounced.

	Fraud (N=228)		Control (N=226)		Punishment (N=224)		Jud. Punishment (N=230)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Total	483.97	211.53	502.75	243.18	503.88	205.97	520.13	245.08
Treatment text	153.98	160.88	157.73	151.82	157.23	105.26	186.15	211.76
Trust eval.	80.54	41.68	93.14	92.87	96.40	138.83	92.97	122.23

Table C.6. Time required for completion in seconds (finished interviews), Survey Data for Latin America.

	Fraud (N=308)		Control (N=327)		Punishment (N=316)		Jud. Punishment (N=311)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Total	501.16	223.98	432.18	174.26	502.97	212.07	464.21	171.22
Treatment text	168.56	132.93	131.45	89.66	188.01	148.41	174.52	170.61
Trust eval.	85.50	49.52	82.72	52.99	89.13	49.01	85.41	83.13

Table C.7. Time required for completion in seconds (finished interviews), Survey Data for Russia.

Attention check	N	%
Summary	667	54.54
Incomplete	419	34.26
Response	101	8.26
Unacceptable	22	1.80
Copy-paste	14	1.14

Table C.9. Data quality categories based on manual coding for the Russian sample.

Attention check	N	%
Summary	376	43.12
Response	317	36.35
Incomplete	151	17.32
Unacceptable	25	2.87
Copy-paste	3	0.34

Table C.8. Data quality categories based on manual coding for the Latin American sample.

C.3.3 Raw Regression Results

Dependent Variable: Political Institutions

	Armed Forces	Police	Central EC	Government
Intercept_1	-1.44^*	-0.59^*	-0.34^*	-0.10
	$[-1.72; -1.18]$	$[-0.85; -0.33]$	$[-0.60; -0.09]$	$[-0.35; 0.16]$
Intercept_2	0.33^*	1.67^*	1.58^*	1.93^*
	$[0.07; 0.58]$	$[1.40; 1.97]$	$[1.29; 1.86]$	$[1.65; 2.23]$
Intercept_3	2.25^*	3.55^*	3.50^*	3.75^*
	$[1.95; 2.57]$	$[3.12; 4.02]$	$[3.11; 3.91]$	$[3.31; 4.22]$
Control	0.29	0.31	1.13^*	0.68^*
	$[-0.06; 0.63]$	$[-0.05; 0.67]$	$[0.78; 1.49]$	$[0.33; 1.04]$
Punishment	-0.30	-0.18	0.23	0.05
	$[-0.65; 0.05]$	$[-0.53; 0.19]$	$[-0.12; 0.60]$	$[-0.31; 0.40]$
Judicial Punishment	-0.07	0.12	0.24	0.18
	$[-0.42; 0.27]$	$[-0.24; 0.48]$	$[-0.11; 0.60]$	$[-0.18; 0.54]$
Observations	849	851	860	861
	Political Parties	Parliament	Courts	President
Intercept_1	-0.01	-0.25	-0.83^*	0.10
	$[-0.28; 0.24]$	$[-0.50; 0.00]$	$[-1.09; -0.57]$	$[-0.16; 0.35]$
Intercept_2	2.56^*	1.95^*	1.30^*	1.64^*
	$[2.23; 2.90]$	$[1.66; 2.25]$	$[1.03; 1.58]$	$[1.36; 1.92]$
Intercept_3	4.81^*	3.95^*	3.48^*	3.02^*
	$[4.16; 5.58]$	$[3.49; 4.46]$	$[3.08; 3.92]$	$[2.67; 3.39]$
Control	0.55^*	0.64^*	0.55^*	0.69^*
	$[0.18; 0.91]$	$[0.28; 1.00]$	$[0.20; 0.91]$	$[0.33; 1.05]$
Punishment	0.11	0.07	-0.06	0.07
	$[-0.25; 0.48]$	$[-0.28; 0.43]$	$[-0.41; 0.30]$	$[-0.28; 0.43]$
Judicial Punishment	0.16	0.34	0.42^*	0.21
	$[-0.21; 0.52]$	$[-0.02; 0.70]$	$[0.07; 0.78]$	$[-0.15; 0.56]$
Observations	863	858	850	857

* Null hypothesis value outside 95% credible interval. Reported are medians and 95% credible intervals.
Fraud treatment group serves as the baseline.

Table C.10. Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample.

	Armed Forces	Police	Central EC	Government
Intercept_1	-1.49^*	-1.04^*	-0.16	-0.48
	$[-2.19; -0.78]$	$[-1.76; -0.32]$	$[-0.84; 0.57]$	$[-1.16; 0.23]$
Intercept_2	-0.26	0.55	1.26^*	1.01^*
	$[-0.95; 0.45]$	$[-0.18; 1.26]$	$[0.57; 1.98]$	$[0.34; 1.72]$
Intercept_3	1.45^*	2.67^*	2.93^*	2.80^*
	$[0.76; 2.16]$	$[1.93; 3.39]$	$[2.23; 3.68]$	$[2.10; 3.53]$
Control	-0.07	0.31	0.33	-0.12
	$[-1.01; 0.90]$	$[-0.62; 1.27]$	$[-0.60; 1.31]$	$[-1.06; 0.82]$
Punishment	-0.42	-0.24	-0.86	-0.96
	$[-1.45; 0.62]$	$[-1.29; 0.83]$	$[-1.95; 0.20]$	$[-1.99; 0.09]$
Judicial Punishment	-0.21	0.42	0.32	-0.08
	$[-1.14; 0.75]$	$[-0.57; 1.42]$	$[-0.65; 1.29]$	$[-1.05; 0.90]$
Pol. Interest: Somewhat	0.11	0.51	0.36	0.36
	$[-0.63; 0.89]$	$[-0.26; 1.27]$	$[-0.37; 1.14]$	$[-0.39; 1.13]$
Pol. Interest: Not Very	0.27	0.24	0.17	0.29
	$[-0.46; 1.02]$	$[-0.53; 1.00]$	$[-0.57; 0.95]$	$[-0.45; 1.03]$
Pol. Interest: Not at All	0.26	-0.38	-0.08	0.07
	$[-0.58; 1.13]$	$[-1.26; 0.50]$	$[-0.95; 0.83]$	$[-0.78; 0.94]$
Control × Pol. Interest: Somewhat	0.37	-0.40	0.25	0.42
	$[-0.68; 1.41]$	$[-1.45; 0.62]$	$[-0.81; 1.28]$	$[-0.60; 1.44]$
Punishment × Pol. Interest: Somewhat	0.38	0.14	0.48	0.74
	$[-0.73; 1.51]$	$[-1.02; 1.25]$	$[-0.66; 1.64]$	$[-0.39; 1.87]$
Judicial Punishment × Pol. Interest: Somewhat	0.65	-0.47	-0.22	0.33
	$[-0.38; 1.66]$	$[-1.55; 0.60]$	$[-1.28; 0.85]$	$[-0.73; 1.37]$
Control × Pol. Interest: Not Very	0.26	-0.29	0.30	0.35
	$[-0.78; 1.25]$	$[-1.30; 0.71]$	$[-0.74; 1.32]$	$[-0.64; 1.37]$
Punishment × Pol. Interest: Not Very	0.80	0.35	1.05	1.24^*
	$[-0.29; 1.88]$	$[-0.78; 1.46]$	$[-0.07; 2.20]$	$[0.14; 2.35]$
Judicial Punishment × Pol. Interest: Not Very	0.83	0.00	-0.00	0.58
	$[-0.19; 1.82]$	$[-1.05; 1.07]$	$[-1.04; 1.04]$	$[-0.46; 1.60]$
Control × Pol. Interest: Not at All	0.20	-0.14	-0.07	0.08
	$[-1.00; 1.37]$	$[-1.30; 1.03]$	$[-1.29; 1.13]$	$[-1.10; 1.23]$
Punishment × Pol. Interest: Not at All	1.16	1.04	1.25	1.11
	$[-0.08; 2.43]$	$[-0.25; 2.32]$	$[-0.05; 2.59]$	$[-0.16; 2.38]$
Judicial Punishment × Pol. Interest: Not at All	0.40	0.35	-0.03	0.74
	$[-0.86; 1.64]$	$[-0.90; 1.62]$	$[-1.29; 1.24]$	$[-0.51; 1.99]$
Observations	1161	1180	1175	1179

	Political Parties	Parliament	Courts	President
Intercept_1	0.38	-0.38	-1.31^*	-0.59
	$[-0.38; 1.18]$	$[-1.06; 0.31]$	$[-1.96; -0.63]$	$[-1.32; 0.13]$
Intercept_2	2.09^*	1.25^*	0.27	0.48
	$[1.33; 2.90]$	$[0.57; 1.94]$	$[-0.37; 0.94]$	$[-0.26; 1.20]$
Intercept_3	4.59^*	3.46^*	2.37^*	1.85^*
	$[3.79; 5.45]$	$[2.76; 4.18]$	$[1.71; 3.04]$	$[1.11; 2.57]$
Control	1.02^*	0.42	-0.03	0.40
	$[0.03; 2.03]$	$[-0.51; 1.34]$	$[-0.91; 0.87]$	$[-0.55; 1.34]$
Punishment	-0.29	-0.92	-0.88	-0.36
	$[-1.45; 0.83]$	$[-1.96; 0.07]$	$[-1.85; 0.09]$	$[-1.40; 0.65]$
Judicial Punishment	0.82	0.76	0.34	0.20
	$[-0.24; 1.90]$	$[-0.23; 1.74]$	$[-0.61; 1.28]$	$[-0.77; 1.15]$
Pol. Interest: Somewhat	1.02^*	0.48	-0.17	0.53
	$[0.21; 1.88]$	$[-0.26; 1.22]$	$[-0.88; 0.58]$	$[-0.24; 1.31]$
Pol. Interest: Not Very	1.20^*	0.31	-0.28	0.35
	$[0.39; 2.04]$	$[-0.42; 1.02]$	$[-0.98; 0.42]$	$[-0.42; 1.12]$
Pol. Interest: Not at All	0.75	-0.22	-0.80	0.02
	$[-0.17; 1.69]$	$[-1.07; 0.62]$	$[-1.65; 0.05]$	$[-0.88; 0.92]$
Control × Pol. Interest: Somewhat	-0.63	-0.26	0.39	-0.17
	$[-1.74; 0.42]$	$[-1.22; 0.78]$	$[-0.58; 1.36]$	$[-1.17; 0.85]$
Punishment × Pol. Interest: Somewhat	0.27	0.67	0.91	0.11
	$[-0.92; 1.47]$	$[-0.40; 1.78]$	$[-0.14; 1.96]$	$[-0.98; 1.24]$
Judicial Punishment × Pol. Interest: Somewhat	-0.51	-0.47	0.00	0.02
	$[-1.67; 0.63]$	$[-1.53; 0.61]$	$[-1.02; 1.03]$	$[-1.01; 1.08]$
Control × Pol. Interest: Not Very	-0.74	0.02	0.39	-0.18
	$[-1.81; 0.33]$	$[-0.95; 1.04]$	$[-0.58; 1.34]$	$[-1.18; 0.80]$
Punishment × Pol. Interest: Not Very	0.36	1.27^*	1.11^*	0.54
	$[-0.84; 1.54]$	$[0.22; 2.36]$	$[0.09; 2.14]$	$[-0.53; 1.64]$
Judicial Punishment × Pol. Interest: Not Very	-0.52	-0.41	0.13	0.16
	$[-1.67; 0.60]$	$[-1.44; 0.65]$	$[-0.85; 1.15]$	$[-0.85; 1.18]$
Control × Pol. Interest: Not at All	-0.90	-0.20	0.03	-0.16
	$[-2.11; 0.31]$	$[-1.37; 0.97]$	$[-1.14; 1.20]$	$[-1.33; 1.05]$
Punishment × Pol. Interest: Not at All	0.99	1.32^*	0.94	0.95
	$[-0.33; 2.34]$	$[0.07; 2.57]$	$[-0.26; 2.18]$	$[-0.30; 2.23]$
Judicial Punishment × Pol. Interest: Not at All	-0.36	0.01	0.37	0.39
	$[-1.69; 0.96]$	$[-1.23; 1.25]$	$[-0.87; 1.61]$	$[-0.84; 1.68]$
Observations	1173	1170	1176	1165

* Null hypothesis value outside 95% credible interval.. Reported are medians and 95% credible intervals.

Table C.11. Ordinal logistic regression estimates of treatment effects for the Russian sample.

	Armed Forces	Police	Central EC	Government
Intercept$_1$	-2.19^*	-1.05^*	-0.33	-0.61^*
	$[-2.68; -1.70]$	$[-1.57; -0.54]$	$[-0.82; 0.17]$	$[-1.12; -0.10]$
Intercept$_2$	-0.37	1.24^*	1.59^*	1.48^*
	$[-0.83; 0.10]$	$[0.73; 1.76]$	$[1.09; 2.11]$	$[0.97; 2.01]$
Intercept$_3$	1.62^*	3.13^*	3.52^*	3.33^*
	$[1.13; 2.11]$	$[2.53; 3.77]$	$[2.95; 4.11]$	$[2.72; 3.97]$
Control	0.20	0.23	1.05^*	0.81^*
	$[-0.48; 0.89]$	$[-0.47; 0.94]$	$[0.35; 1.76]$	$[0.09; 1.53]$
Punishment	-0.47	-0.24	0.27	0.17
	$[-1.14; 0.21]$	$[-0.98; 0.48]$	$[-0.43; 0.97]$	$[-0.54; 0.89]$
Judicial Punishment	-0.26	0.04	0.16	0.37
	$[-0.91; 0.40]$	$[-0.68; 0.76]$	$[-0.54; 0.86]$	$[-0.34; 1.07]$
Opponent	-0.96^*	-0.60^*	0.02	-0.66^*
	$[-1.51; -0.42]$	$[-1.18; -0.03]$	$[-0.55; 0.59]$	$[-1.24; -0.07]$
Control × Opponent	0.16	0.11	0.11	-0.15
	$[-0.62; 0.96]$	$[-0.71; 0.92]$	$[-0.69; 0.91]$	$[-0.96; 0.69]$
Punishment × Opponent	0.23	0.08	-0.05	-0.16
	$[-0.55; 1.02]$	$[-0.75; 0.92]$	$[-0.86; 0.76]$	$[-0.99; 0.67]$
Judicial Punishment × Opponent	0.25	0.10	0.10	-0.25
	$[-0.51; 1.02]$	$[-0.72; 0.93]$	$[-0.70; 0.91]$	$[-1.07; 0.57]$
Observations	849	851	860	861

	Political Parties	Parliament	Courts	President
Intercept$_1$	-0.17	-0.53^*	-1.12^*	-0.88^*
	$[-0.68; 0.34]$	$[-1.03; -0.03]$	$[-1.62; -0.61]$	$[-1.36; -0.39]$
Intercept$_2$	2.43^*	1.70^*	1.03^*	0.82^*
	$[1.89; 2.99]$	$[1.19; 2.21]$	$[0.54; 1.53]$	$[0.34; 1.32]$
Intercept$_3$	4.69^*	3.71^*	3.21^*	2.33^*
	$[3.90; 5.57]$	$[3.08; 4.36]$	$[2.63; 3.82]$	$[1.81; 2.88]$
Control	0.94^*	0.76^*	0.28	1.05^*
	$[0.20; 1.68]$	$[0.04; 1.47]$	$[-0.44; 0.99]$	$[0.33; 1.77]$
Punishment	0.25	0.39	-0.10	0.62
	$[-0.47; 0.97]$	$[-0.31; 1.09]$	$[-0.79; 0.60]$	$[-0.07; 1.31]$
Judicial Punishment	-0.11	0.11	0.34	0.33
	$[-0.83; 0.61]$	$[-0.58; 0.82]$	$[-0.36; 1.04]$	$[-0.34; 1.02]$
Opponent	-0.21	-0.37	-0.38	-1.27^*
	$[-0.80; 0.39]$	$[-0.94; 0.19]$	$[-0.95; 0.20]$	$[-1.84; -0.69]$
Control × Opponent	-0.51	-0.14	0.37	-0.34
	$[-1.35; 0.33]$	$[-0.96; 0.69]$	$[-0.45; 1.20]$	$[-1.17; 0.49]$
Punishment × Opponent	-0.18	-0.42	0.06	-0.67
	$[-1.02; 0.65]$	$[-1.23; 0.39]$	$[-0.75; 0.87]$	$[-1.48; 0.14]$
Judicial Punishment × Opponent	0.36	0.30	0.13	-0.16
	$[-0.47; 1.19]$	$[-0.51; 1.11]$	$[-0.69; 0.95]$	$[-0.96; 0.65]$
Observations	863	858	850	857

* Null hypothesis value outside 95% credible interval.. Repoted are medians and 95% credible intervals.

Table C.12. Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample.

	Armed Forces	Police	Central EC	Government
Intercept_1	-2.39^*	-1.74^*	-1.09^*	-1.56^*
	$[-2.69; -2.09]$	$[-2.03; -1.46]$	$[-1.37; -0.81]$	$[-1.85; -1.28]$
Intercept_2	-1.11^*	-0.12	0.39^*	0.03
	$[-1.39; -0.83]$	$[-0.39; 0.15]$	$[0.11; 0.66]$	$[-0.24; 0.31]$
Intercept_3	0.68^*	2.07^*	2.15^*	1.92^*
	$[0.41; 0.96]$	$[1.77; 2.38]$	$[1.84; 2.46]$	$[1.62; 2.23]$
Control	0.24	0.13	0.57^*	0.31
	$[-0.14; 0.63]$	$[-0.25; 0.51]$	$[0.19; 0.96]$	$[-0.08; 0.68]$
Punishment	0.37	0.29	-0.27	-0.07
	$[-0.02; 0.75]$	$[-0.09; 0.68]$	$[-0.64; 0.11]$	$[-0.45; 0.30]$
Judicial Punishment	0.44^*	0.49^*	0.15	0.35
	$[0.06; 0.81]$	$[0.11; 0.89]$	$[-0.24; 0.52]$	$[-0.04; 0.73]$
Opponent	-1.07^*	-0.72^*	-1.16^*	-1.28^*
	$[-1.42; -0.72]$	$[-1.07; -0.36]$	$[-1.52; -0.80]$	$[-1.65; -0.93]$
Control × Opponent	-0.02	-0.17	-0.02	-0.17
	$[-0.54; 0.46]$	$[-0.66; 0.33]$	$[-0.51; 0.48]$	$[-0.66; 0.33]$
Punishment × Opponent	-0.19	-0.31	0.32	0.16
	$[-0.70; 0.30]$	$[-0.81; 0.18]$	$[-0.16; 0.82]$	$[-0.33; 0.65]$
Judicial Punishment × Opponent	0.02	-0.29	0.15	0.05
	$[-0.48; 0.52]$	$[-0.80; 0.19]$	$[-0.35; 0.65]$	$[-0.45; 0.56]$
Observations	1167	1186	1180	1185
	Political Parties	Parliament	Courts	President
Intercept_1	-1.43^*	-1.29^*	-1.81^*	-1.96^*
	$[-1.72; -1.13]$	$[-1.56; -1.01]$	$[-2.10; -1.52]$	$[-2.25; -1.68]$
Intercept_2	0.41^*	0.44^*	-0.19	-0.79^*
	$[0.13; 0.69]$	$[0.16; 0.71]$	$[-0.47; 0.09]$	$[-1.06; -0.52]$
Intercept_3	3.01^*	2.73^*	1.97^*	0.74^*
	$[2.66; 3.39]$	$[2.40; 3.07]$	$[1.66; 2.28]$	$[0.47; 1.02]$
Control	0.41^*	0.44^*	0.09	0.26
	$[0.03; 0.79]$	$[0.07; 0.81]$	$[-0.30; 0.48]$	$[-0.11; 0.63]$
Punishment	0.11	0.22	-0.23	-0.17
	$[-0.28; 0.49]$	$[-0.17; 0.60]$	$[-0.61; 0.14]$	$[-0.53; 0.20]$
Judicial Punishment	0.28	0.49^*	0.16	0.39^*
	$[-0.10; 0.67]$	$[0.11; 0.87]$	$[-0.23; 0.55]$	$[0.02; 0.77]$
Opponent	-1.23^*	-0.98^*	-1.26^*	-1.55^*
	$[-1.59; -0.86]$	$[-1.33; -0.63]$	$[-1.62; -0.90]$	$[-1.91; -1.19]$
Control × Opponent	-0.13	-0.22	0.31	0.00
	$[-0.63; 0.38]$	$[-0.71; 0.27]$	$[-0.18; 0.81]$	$[-0.49; 0.49]$
Punishment × Opponent	-0.08	-0.22	0.50^*	0.36
	$[-0.58; 0.43]$	$[-0.71; 0.29]$	$[0.01; 1.00]$	$[-0.12; 0.83]$
Judicial Punishment × Opponent	0.10	-0.14	0.47	-0.08
	$[-0.42; 0.60]$	$[-0.65; 0.35]$	$[-0.04; 0.97]$	$[-0.57; 0.41]$
Observations	1179	1176	1182	1171

* Null hypothesis value outside 95% credible interval.. Repoted are medians and 95% credible intervals.

Table C.13. Ordinal logistic regression estimates of treatment effects for the Russian sample.

Dependent Variable: Non-Political Institutions

We ask the respondents to evaluate their trust in non-political institutions as well as in political upon reading the treatment text, which allows us to evaluate the limitations of the spillover theory. As we can observe from the estimates in the tables below, the treatment groups do not differ significantly in their confidence in non-political institutions in both Latin America and Russia when only presented with the fraud information in comparison to the status quo outcome, while the only exception for the main effect seems to the Russian sample and the effect of court intervention as response to fraud in comparison to fraud alone (i.e., the baseline condition). The milder version of system response, i.e. responsible members' exclusion from the commissions, has no effect on trust lost after the fraud information was released. Judicial punishment information seems to increase trust in companies, banks, environmental organisations, and the World Trade Organisation, but not in the United Nations or the World Bank. One could argue that this could be an evidence of a spillover effect, which goes beyond the political sphere and exists for entities that are not portrayed negatively and "pro-Western" in the Russian media. The fact that we observe it specifically in Russia and only in response to the message about the intervention of another political institution could also be interpreted as evidence for the belief updating mechanism in evaluations. As this information may go against the general expectations of the systems' responses to election manipulation in an autocratic state, judicial interventions' consistent restorative effect on trust for various institutions indicates that respondents seem to have updated their beliefs rather than disregard the information that goes counter their current perceptions.

	Companies	Banks	Environmental Organizations
Intercept_1	-1.37^*	-1.28^*	-2.88^*
	$[-1.59; -1.16]$	$[-1.50; -1.07]$	$[-3.18; -2.59]$
Intercept_2	0.57^*	0.42^*	-1.19^*
	$[0.36; 0.77]$	$[0.21; 0.62]$	$[-1.40; -0.97]$
Intercept_3	2.71^*	2.49^*	0.90^*
	$[2.44; 2.99]$	$[2.23; 2.75]$	$[0.69; 1.11]$
Control	0.27	-0.03	0.09
	$[-0.01; 0.56]$	$[-0.32; 0.25]$	$[-0.19; 0.38]$
Punishment	0.04	-0.27	-0.19
	$[-0.24; 0.32]$	$[-0.54; 0.01]$	$[-0.48; 0.09]$
Judicial Punishment	0.21	0.09	-0.29^*
	$[-0.07; 0.50]$	$[-0.19; 0.37]$	$[-0.58; -0.01]$
Observations	856	854	852
	United Nations	World Bank	WTO
Intercept_1	-2.35^*	-1.86^*	-2.21^*
	$[-2.61; -2.09]$	$[-2.10; -1.63]$	$[-2.47; -1.95]$
Intercept_2	-0.67^*	-0.08	-0.39^*
	$[-0.87; -0.46]$	$[-0.28; 0.13]$	$[-0.60; -0.18]$
Intercept_3	1.10^*	1.92^*	1.95^*
	$[0.89; 1.31]$	$[1.68; 2.16]$	$[1.71; 2.19]$
Control	0.27	0.16	0.26
	$[-0.00; 0.55]$	$[-0.13; 0.45]$	$[-0.03; 0.56]$
Punishment	-0.13	-0.08	-0.17
	$[-0.41; 0.14]$	$[-0.36; 0.20]$	$[-0.46; 0.12]$
Judicial Punishment	0.18	-0.01	-0.25
	$[-0.10; 0.47]$	$[-0.29; 0.28]$	$[-0.54; 0.04]$
Observations	850	841	825

* Null hypothesis value outside 95% credible interval.
Fraud treatment group serves as the baseline.

Table C.14. Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample.

	Companies	Banks	Environmental Organizations
Intercept_1	-1.08^*	-0.97^*	-1.48^*
	$[-1.26; -0.89]$	$[-1.16; -0.78]$	$[-1.68; -1.29]$
Intercept_2	0.61^*	0.61^*	-0.14
	$[0.42; 0.79]$	$[0.43; 0.80]$	$[-0.32; 0.04]$
Intercept_3	3.11^*	3.04^*	2.06^*
	$[2.84; 3.38]$	$[2.78; 3.30]$	$[1.85; 2.27]$
Control	0.13	0.10	0.19
	$[-0.12; 0.37]$	$[-0.14; 0.35]$	$[-0.04; 0.43]$
Punishment	0.19	0.09	0.12
	$[-0.05; 0.43]$	$[-0.15; 0.34]$	$[-0.13; 0.36]$
Judicial Punishment	0.27^*	0.25^*	0.27^*
	$[0.03; 0.53]$	$[0.01; 0.50]$	$[0.02; 0.52]$
Observations	1151	1173	1166
	United Nations	World Bank	WTO
Intercept_1	-1.27^*	-0.97^*	-1.11^*
	$[-1.46; -1.07]$	$[-1.15; -0.78]$	$[-1.30; -0.92]$
Intercept_2	0.13	0.39^*	0.46^*
	$[-0.05; 0.31]$	$[0.21; 0.57]$	$[0.28; 0.64]$
Intercept_3	2.19^*	2.50^*	2.78^*
	$[1.97; 2.41]$	$[2.27; 2.73]$	$[2.54; 3.04]$
Control	0.06	0.10	0.12
	$[-0.18; 0.31]$	$[-0.15; 0.35]$	$[-0.13; 0.36]$
Punishment	-0.04	0.06	0.01
	$[-0.28; 0.21]$	$[-0.19; 0.30]$	$[-0.24; 0.26]$
Judicial Punishment	0.15	0.18	0.29^*
	$[-0.09; 0.40]$	$[-0.06; 0.44]$	$[0.03; 0.54]$
Observations	1137	1134	1114

* Null hypothesis value outside 95% credible interval.
Fraud treatment group serves as the baseline.

Table C.15. Ordinal logistic regression estimates of treatment effects for the Russian sample.

Robustness Checks

This section contains the results of various robustness checks and investigation of further observable implications.

Data Quality Restrictions

Our manual coding of attention checks allowed us to closely examine the cases where the respondents seemed to have read the treatment text closely and in full, as their responses include the treatment-specific scenario details. For such individuals, we should observe strongest effects should our theory hold, though this effect could be counteracted by the fact of smaller sample sizes. Still, we observe stronger effects for more political institutions than in the full sample. In fact, we find strong evidence for the spillover effects of fraud in all countries for this subsample.

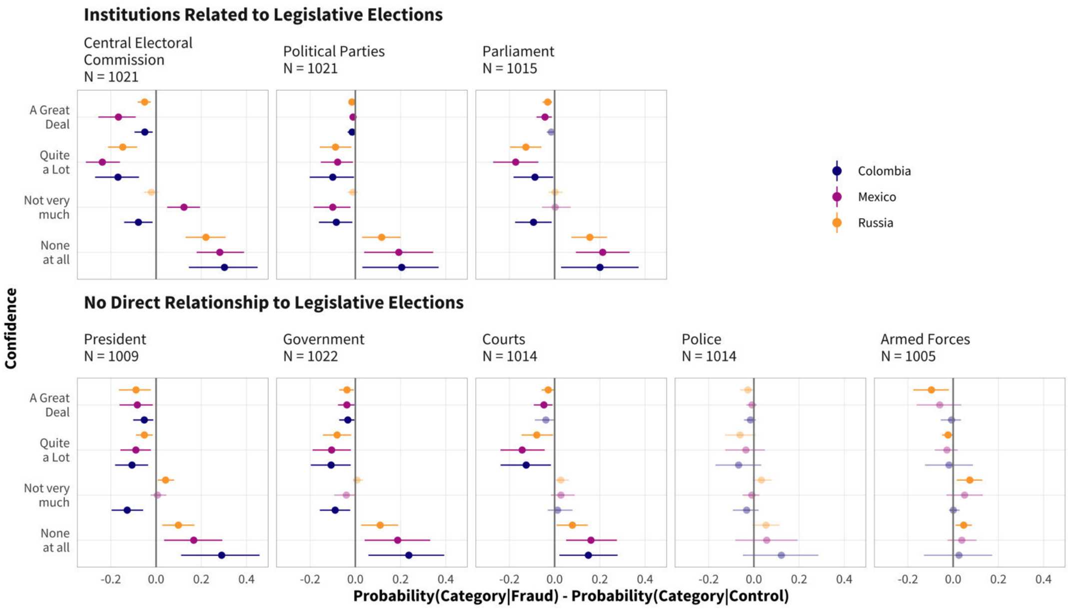


Figure C.3. The effects of exposure to fraud information on confidence in political institutions, only complete summaries of treatment texts. Plots depict medians and 83% (bold) and 95% (thin) highest-density intervals of differences in probabilities for choosing respective categories based on draws from the posterior predictive distributions. Transparent point ranges include zero in the 95% HDCI. Probabilities are calculated based ordered logit model estimates as lined out in Equation 1. The dashed line schematically depicts the hypothesized relationship between categories.

Subsample Analysis for Fraud Effect

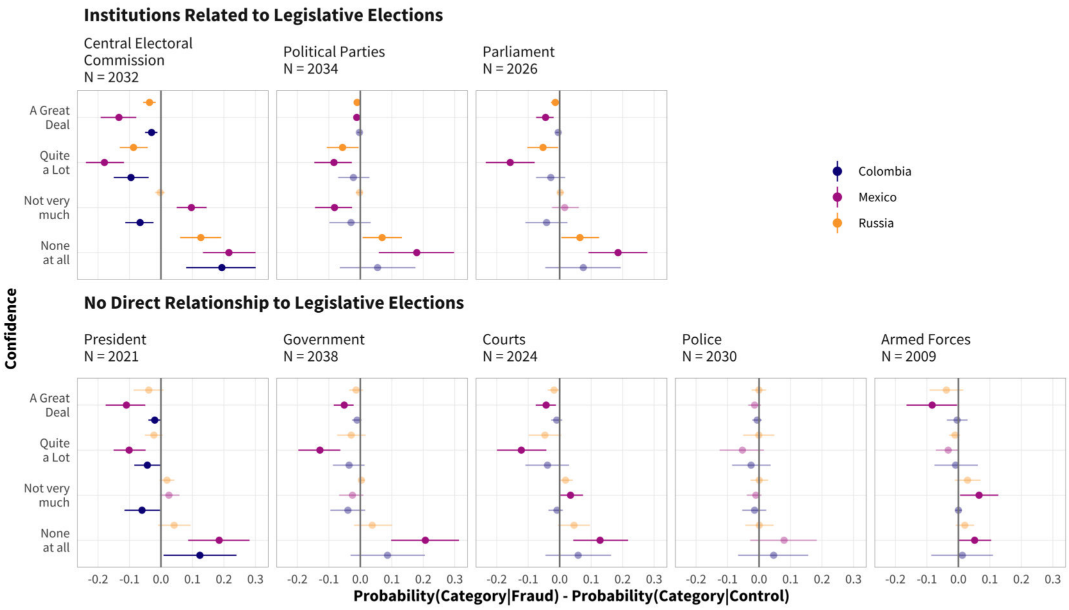


Figure C.4. The effect of exposure to fraud and punishment information on confidence in political institutions across countries. Plots depict medians and 83% (bold) and 95% (thin) highest-density intervals of differences in probabilities for choosing respective categories based on draws from the posterior predictive distributions. Transparent point ranges include zero in the 95% HDCI. Probabilities are calculated based ordered logit model estimates as lined out in Equation 1. The dashed line schematically depicts the hypothesized relationship between categories.

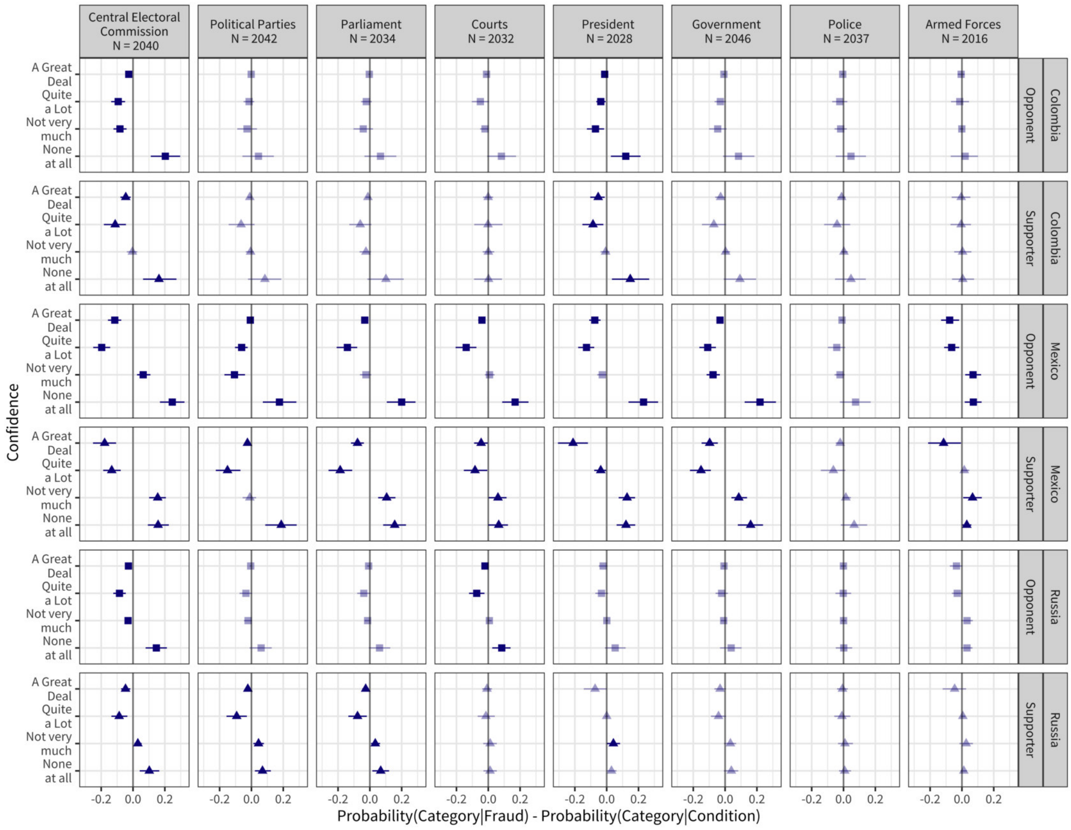


Figure C.5. The effects of exposure to fraud information on confidence in political institutions across countries, opponents and supporters. Plots depict medians and 83% (bold) and 95% (thin) highest-density intervals of differences in probabilities for choosing respective categories based on draws from the posterior predictive distributions. Transparent point ranges include zero in the 95% HDCI. Probabilities are calculated based ordered logit model estimates as lined out in Equation 1. The dashed line schematically depicts the hypothesized relationship between categories.

Mediation Analysis

In this section we further investigate the mechanisms of attitudes' updating via mediation analysis. The spillover theory implies that it is the changes in trust in elections and electoral process that are responsible for the differences in the confidence in institutions of the political system. We thus have included a separate question that accounts for trust in election system after the treatment (the exact phrasing is: *In this hypothetical scenario, how much confidence would you have in elections in [Mexico/Colombia/Russia]?*) We use answers to this question as a mediator to trace the effects of the fraud and punishment information on political attitudes. Figure below presents the argument graphically:

[mynode] (m)Confidence in Elections; [mynode,below left=of m](a) Fraud/Punishment Information; [mynode,below right=of m](b) Confidence in Institution; [-latex] (a.north) – node[auto,font=] (m.west); [-latex] (m.east) – node[auto,font=] (b.north); [-latex, dotted] (a.east) – node[below=3mm,font=,align=center] (b.west);

We estimate the ordered logit models specified as follows.
Outcome model:

$$ln\left(\frac{\Pr(y_i \leq j)}{\Pr(y_i > j)}\right) = \alpha_j - \underbrace{(\beta_1 \text{ Control}_i + \beta_2 \text{ Punishment}_i + \beta_3 \text{ Judicial Punishment}_i)}_{\text{Treatment Variables}} + \underbrace{\beta_4 \text{ Trust in Elections}_i}_{\text{Mediator}}, \tag{C.1}$$

Mediator model:

$$ln\left(\frac{\Pr(\text{Trust in Elections}_i \leq j)}{\Pr(\text{Trust in Elections}_i > j)}\right) = \gamma_j - (\delta_1 \text{ Control}_i + \delta_2 \text{ Punishment}_i + \delta_3 \text{ Judicial Punishment}_i) \tag{C.2}$$

where y_i is the level of diffuse support of an individual i with $(i = 1, ..., n)$ for an institution, and Control, Punishment, and Judicial Punishment are binary indicators for membership in the experimental groups (2), (3), and (4) [2]. Trust in Elections$_i$ is the level of trust in electoral process in the country and is on the same scale as the main outcome variable, institutional trust. As in the main analysis, we analyze Russian and Latin American samples separately, using all available observations that fulfill the basic data-quality criteria, such as completing the questionnaire in lower than 3 minutes.

For each political institution, we estimate an ordered logit model for trust in elections and model institutional trust using the estimates from the ordered logit in the previous step. It is our expectation that *direct effect* of treatment on institutional trust, once the model includes trust in elections, is close to zero, while the indirect effect, the *average causal mediation effect*, would be different from zero and positive: as trust categories are both measured on the same scale of 1 to 4, we should expect a positive relationship between them, and in comparison to fraud alone, status quo

[2] Individuals who only received fraud information serve as the reference category in our analysis.

information and the punishment would be expected to raise the trust. Figure C.6 presents the results of the analysis for Latin American and Russian samples.

When we look at the effect of fraud information alone, we can observe that across all institutions, indirect effect, *ACME*, is significantly different from zero and positive, which is in line with our expectations: increased (decreased) trust in elections is associated with increased (decreased) trust in political institutions. At the same time, the direct effect, i.e. the effect of treatment alone on confidence in institutions, is for most institutions, not different from zero, meaning that most of the changes in institutional confidence in response to treatment seem to be driven by the declines in trust in the elections. For the punishment information, we only observe the effects of treatment once the courts are reported to intervene , and the effect again is primarily existing via trust in elections. For punishment information alone, with no judicial intervention, we observe only the relationship between the trust in elections and trust in

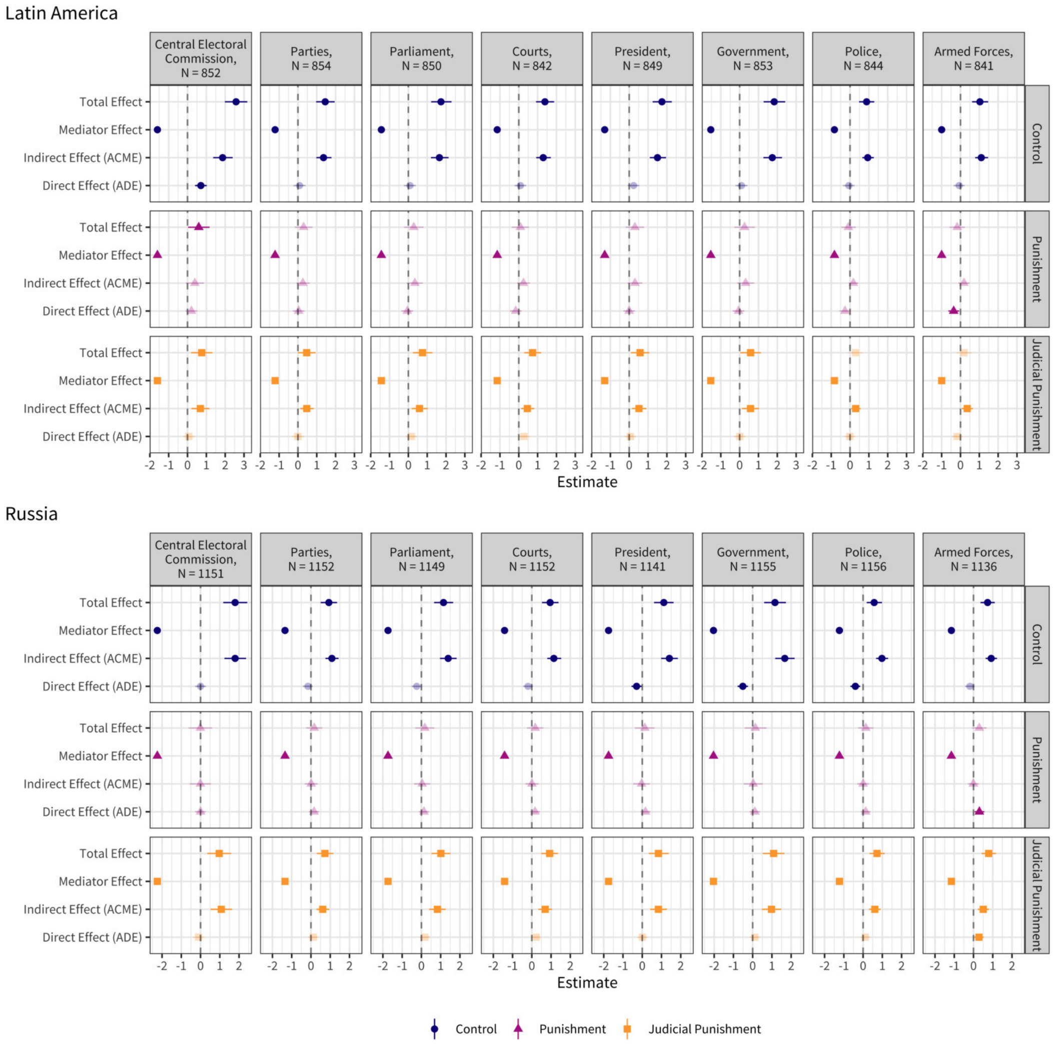


Figure C.6. The results of mediation analysis. Depicted are the *direct effect* (median value and 95% highest-density continuous interval (HDCI) of posterior samples from treatment of the institutional trust model), *mediator effect* (median value and 95% HDCI of posterior samples from mediator, trust in elections, of the institutional trust model), *indirect effect* (median value and 95% HDCI of the multiplication of the posterior samples from mediator, trust in elections, of the institutional trust model and the posterior samples from treatment of the trust in elections' model) and the *total effect* (median value and 95% HDCI of sums of posterior samples used for the direct and indirect effect).

Models with Controls

This section contains tables with estimates for the models from our survey experiment, but this time with control variables (tables C.16 and C.17). While random assignment allows us to drop the controls in our main analysis (for balance assessment across groups see tables C.2 and C.3), we replicate the results using control variables as well. We use sociodemographic (measured post-treatment) and political attitudes (measured pre-treatment) variables. We have attempted to include the same variables as in our matching analysis for the sake of uniformity and, to a certain extent, comparability. We thus control for

- *generalized trust*, as it may be directly related to confidence in political system institutions;
- *political interest*, as we may expect a relationship between trust and investment into the topic with causal effects pointing in either direction;
- *political affiliation*, as opposition to the regime may decrease trust in its institutions as of itself;
- *age (logged)*, as we may expect younger respondents to show less trust to political institutions;
- *education*, as it may proxy the critical thinking skills and potentially, differences in degrees of sophistication in the evaluations;
- *employment status*, as it may impact the overall government performance evaluation as well as impact the socialisation and information channels available to respondents;
- *employment sector*, as working for the government may be associated with changes in political attitudes;
- *savings*, as economic security and income are known to impact the performance of (and, potentially, trust in) government institutions;
- *political corruption*, as perceptions of corruption are likely directly related to trust in political institutions.

None of these variables are expected to be systematically related to the treatment variables due to random assignment. As a result, the estimates differ marginally with their significance and signs, and follow the patterns we observe in the main analysis.

	Armed Forces	Police	Central EC	Government
Intercept_1	-7.43^*	-4.53^*	-7.89^*	-6.00^*
	$[-10.76; -4.12]$	$[-7.82; -1.27]$	$[-11.18; -4.67]$	$[-9.42; -2.59]$
Intercept_2	-5.45^*	-2.31	-6.45^*	-4.06^*
	$[-8.77; -2.17]$	$[-5.55; 0.94]$	$[-9.72; -3.25]$	$[-7.41; -0.66]$
Intercept_3	-3.31^*	0.33	-4.22^*	-1.82
	$[-6.61; -0.05]$	$[-2.95; 3.64]$	$[-7.46; -1.00]$	$[-5.19; 1.56]$
Control	0.69^*	0.48	1.58^*	1.16^*
	$[0.20; 1.18]$	$[-0.02; 0.99]$	$[1.08; 2.09]$	$[0.65; 1.68]$
Punishment	-0.35	-0.59^*	0.23	-0.12
	$[-0.86; 0.15]$	$[-1.10; -0.08]$	$[-0.28; 0.73]$	$[-0.66; 0.41]$
Judicial Punishment	-0.28	0.24	0.59^*	0.44
	$[-0.78; 0.23]$	$[-0.28; 0.77]$	$[0.06; 1.11]$	$[-0.09; 0.97]$
Opponent	-0.58^*	-0.28	0.51^*	-1.05^*
	$[-0.96; -0.21]$	$[-0.66; 0.10]$	$[0.13; 0.90]$	$[-1.44; -0.66]$
Political Interest	-0.10	-0.17	0.09	-0.01
	$[-0.36; 0.15]$	$[-0.43; 0.09]$	$[-0.16; 0.34]$	$[-0.28; 0.25]$
General Trust	-0.52	-0.09	-0.32	-1.07^*
	$[-1.08; 0.03]$	$[-0.66; 0.46]$	$[-0.86; 0.20]$	$[-1.66; -0.50]$
Age (log)	-1.11^*	-0.88	-2.29^*	-0.80
	$[-2.02; -0.19]$	$[-1.79; 0.04]$	$[-3.21; -1.38]$	$[-1.71; 0.11]$
Male	0.11	0.35	0.18	0.46^*
	$[-0.27; 0.49]$	$[-0.06; 0.75]$	$[-0.20; 0.56]$	$[0.06; 0.87]$
Education	-0.08	0.02	0.17	0.11
	$[-0.48; 0.31]$	$[-0.37; 0.41]$	$[-0.20; 0.54]$	$[-0.27; 0.49]$
Employment: retired	-0.11	1.35	3.47^*	1.75
	$[-3.28; 3.00]$	$[-1.69; 4.48]$	$[0.35; 6.65]$	$[-1.34; 4.84]$
Employment: housewife	-0.50	-0.24	-0.35	-0.91^*
	$[-1.21; 0.21]$	$[-1.02; 0.52]$	$[-1.07; 0.34]$	$[-1.71; -0.12]$
Employment: student	0.01	-0.52	-0.96^*	-0.20
	$[-0.56; 0.59]$	$[-1.11; 0.07]$	$[-1.52; -0.40]$	$[-0.78; 0.38]$
Employment: unemployed	-0.05	0.19	-0.39	0.52
	$[-0.61; 0.53]$	$[-0.38; 0.75]$	$[-0.97; 0.19]$	$[-0.09; 1.10]$
Employment: other	-1.18^*	-0.42	-0.79	-0.92
	$[-2.16; -0.21]$	$[-1.39; 0.51]$	$[-1.76; 0.13]$	$[-2.01; 0.09]$
Sector: Private Business	0.51	-0.34	-0.32	-0.38
	$[-0.09; 1.11]$	$[-0.93; 0.26]$	$[-0.87; 0.26]$	$[-0.95; 0.18]$
Sector: Non-profit	0.98^*	-0.26	0.35	0.14
	$[0.14; 1.80]$	$[-1.11; 0.54]$	$[-0.46; 1.18]$	$[-0.66; 0.94]$
Savings	-0.10	-0.01	0.11	-0.24^*
	$[-0.29; 0.09]$	$[-0.21; 0.18]$	$[-0.08; 0.29]$	$[-0.44; -0.05]$
Observations	296	295	301	298

* Null hypothesis value outside 95% credible interval.. Repoted are medians and 95% credible intervals.
Respective baseline categories are Fraud, Supporter, Employment: paid employment, Sector: Government or public institution

Table C.16. Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample.

	Political Parties	Parliament	Courts	President
Intercept_1	−10.39*	−8.50*	−6.18*	−7.62*
	[−13.98; −6.96]	[−11.83; −5.19]	[−9.52; −3.02]	[−10.89; −4.25]
Intercept_2	−8.16*	−6.22*	−4.07*	−5.79*
	[−11.69; −4.76]	[−9.50; −2.95]	[−7.37; −0.94]	[−9.02; −2.45]
Intercept_3	−5.21*	−3.73*	−1.46	−3.98*
	[−8.78; −1.70]	[−6.99; −0.45]	[−4.75; 1.69]	[−7.21; −0.64]
Control	0.88*	1.33*	1.08*	1.36*
	[0.35; 1.41]	[0.82; 1.87]	[0.58; 1.59]	[0.84; 1.88]
Punishment	−0.31	−0.07	0.04	−0.14
	[−0.86; 0.25]	[−0.58; 0.45]	[−0.48; 0.54]	[−0.66; 0.37]
Judicial Punishment	0.17	0.74*	1.01*	0.26
	[−0.37; 0.71]	[0.20; 1.29]	[0.47; 1.55]	[−0.27; 0.79]
Opponent	−0.37	−0.46*	−0.31	−1.88*
	[−0.75; 0.02]	[−0.85; −0.07]	[−0.68; 0.07]	[−2.30; −1.47]
Political Interest	−0.22	−0.07	−0.10	−0.22
	[−0.49; 0.04]	[−0.33; 0.18]	[−0.36; 0.15]	[−0.48; 0.04]
General Trust	−0.94*	−0.51	−0.21	−0.69*
	[−1.49; −0.38]	[−1.07; 0.04]	[−0.74; 0.32]	[−1.24; −0.13]
Age (log)	−2.06*	−1.55*	−1.37*	−0.87
	[−3.04; −1.09]	[−2.47; −0.62]	[−2.30; −0.48]	[−1.78; 0.04]
Male	−0.05	0.05	0.17	0.51*
	[−0.47; 0.36]	[−0.34; 0.44]	[−0.21; 0.55]	[0.11; 0.91]
Education	−0.19	−0.19	0.16	−0.27
	[−0.59; 0.20]	[−0.58; 0.20]	[−0.22; 0.54]	[−0.66; 0.12]
Employment: retired	2.94	1.62	1.25	1.22
	[−0.28; 6.11]	[−1.57; 4.82]	[−1.85; 4.35]	[−1.94; 4.23]
Employment: housewife	−0.22	−0.82*	0.05	−0.50
	[−0.96; 0.52]	[−1.58; −0.08]	[−0.67; 0.77]	[−1.29; 0.29]
Employment: student	−1.18*	−0.92*	−0.14	−0.35
	[−1.83; −0.54]	[−1.50; −0.35]	[−0.72; 0.43]	[−0.91; 0.20]
Employment: unemployed	0.04	−0.11	0.10	−0.06
	[−0.54; 0.64]	[−0.68; 0.46]	[−0.48; 0.68]	[−0.66; 0.54]
Employment: other	−0.13	−1.28*	−0.95	−1.17*
	[−1.12; 0.82]	[−2.33; −0.29]	[−1.96; 0.02]	[−2.33; −0.10]
Sector: Private Business	−0.37	−0.55	−0.50	−0.39
	[−0.97; 0.23]	[−1.14; 0.05]	[−1.04; 0.06]	[−0.98; 0.18]
Sector: Non-profit	−0.09	−0.26	−0.35	0.07
	[−0.96; 0.79]	[−1.09; 0.54]	[−1.13; 0.45]	[−0.71; 0.86]
Savings	−0.09	−0.09	0.01	−0.16
	[−0.29; 0.11]	[−0.28; 0.11]	[−0.18; 0.20]	[−0.35; 0.04]
Observations	299	297	296	295

* Null hypothesis value outside 95% credible interval. Repoted are medians and 95% credible intervals.
Respective baseline categories are Fraud, Supporter, Employment: paid employment, Sector: Government or public institution

Ordinal logistic regression estimates of treatment effects for the Latin American pooled sample. (cont.)

	Armed Forces	Police	Central EC	Government
Intercept$_1$	-2.02^*	-3.71^*	-0.66	-3.14^*
	$[-3.74; -0.25]$	$[-5.47; -1.95]$	$[-2.39; 1.06]$	$[-4.92; -1.44]$
Intercept$_2$	-0.71	-2.03^*	0.91	-1.48
	$[-2.42; 1.06]$	$[-3.79; -0.28]$	$[-0.80; 2.61]$	$[-3.24; 0.23]$
Intercept$_3$	1.22	0.30	2.76^*	0.50
	$[-0.48; 3.00]$	$[-1.45; 2.06]$	$[1.02; 4.46]$	$[-1.25; 2.19]$
Control	0.23	-0.04	0.64^*	0.25
	$[-0.03; 0.49]$	$[-0.31; 0.23]$	$[0.38; 0.91]$	$[-0.02; 0.52]$
Punishment	0.26^*	0.04	-0.04	0.03
	$[0.00; 0.52]$	$[-0.22; 0.31]$	$[-0.31; 0.23]$	$[-0.23; 0.29]$
Judicial Punishment	0.47^*	0.31^*	0.24	0.32^*
	$[0.21; 0.74]$	$[0.05; 0.59]$	$[-0.03; 0.52]$	$[0.06; 0.60]$
Opponent	-1.08^*	-0.87^*	-1.04^*	-1.25^*
	$[-1.28; -0.88]$	$[-1.06; -0.68]$	$[-1.24; -0.84]$	$[-1.45; -1.05]$
Political Interest	-0.13^*	0.02	0.01	-0.08
	$[-0.26; -0.00]$	$[-0.11; 0.14]$	$[-0.12; 0.13]$	$[-0.21; 0.04]$
General Trust	-0.47^*	-0.58^*	-0.58^*	-0.49^*
	$[-0.70; -0.24]$	$[-0.81; -0.35]$	$[-0.82; -0.34]$	$[-0.72; -0.25]$
Age (log)	0.62^*	0.12	0.78^*	0.21
	$[0.22; 1.05]$	$[-0.29; 0.54]$	$[0.36; 1.20]$	$[-0.22; 0.62]$
Male	0.31^*	-0.20	0.04	-0.07
	$[0.11; 0.52]$	$[-0.40; 0.01]$	$[-0.17; 0.25]$	$[-0.28; 0.14]$
Education	-0.16	-0.21	-0.29^*	-0.15
	$[-0.39; 0.07]$	$[-0.44; 0.01]$	$[-0.50; -0.07]$	$[-0.36; 0.07]$
Employment: retired	0.32	-0.24	0.38	0.54^*
	$[-0.17; 0.81]$	$[-0.71; 0.24]$	$[-0.10; 0.86]$	$[0.05; 1.03]$
Employment: housewife	0.02	-0.07	-0.13	-0.23
	$[-0.29; 0.31]$	$[-0.37; 0.24]$	$[-0.45; 0.19]$	$[-0.54; 0.09]$
Employment: student	-0.00	0.55	0.26	-0.40
	$[-0.77; 0.75]$	$[-0.16; 1.26]$	$[-0.61; 1.11]$	$[-1.21; 0.40]$
Employment: unemployed	-0.28	-0.18	0.47^*	0.13
	$[-0.59; 0.02]$	$[-0.49; 0.11]$	$[0.17; 0.77]$	$[-0.17; 0.43]$
Employment: other	0.43	0.38	0.52^*	0.69^*
	$[-0.04; 0.88]$	$[-0.06; 0.85]$	$[0.06; 0.97]$	$[0.22; 1.17]$
Sector: Private Business	-0.10	-0.32^*	-0.01	-0.13
	$[-0.31; 0.13]$	$[-0.54; -0.09]$	$[-0.23; 0.21]$	$[-0.35; 0.10]$
Sector: Non-profit	-0.44^*	-0.61^*	-0.03	-0.27
	$[-0.80; -0.06]$	$[-1.00; -0.23]$	$[-0.41; 0.35]$	$[-0.64; 0.11]$
Savings	-0.12^*	-0.13^*	-0.32^*	-0.34^*
	$[-0.23; -0.02]$	$[-0.23; -0.03]$	$[-0.43; -0.21]$	$[-0.45; -0.23]$
Observations	1021	1033	1032	1034

* Null hypothesis value outside 95% credible interval.. Repoted are medians and 95% credible intervals.
Respective baseline categories are Fraud, Supporter, Employment: paid employment, Sector: Government or public institution

Table C.17. Ordinal logistic regression estimates of treatment effects for the Russian sample.

	Political Parties	Parliament	Courts	President
Intercept$_1$	-4.09^*	-4.13^*	-4.68^*	-1.10
	$[-5.89; -2.33]$	$[-5.92; -2.38]$	$[-6.41; -2.98]$	$[-2.82; 0.62]$
Intercept$_2$	-2.19^*	-2.31^*	-3.03^*	0.12
	$[-3.96; -0.45]$	$[-4.07; -0.57]$	$[-4.76; -1.35]$	$[-1.61; 1.83]$
Intercept$_3$	0.47	0.09	-0.76	1.77^*
	$[-1.33; 2.22]$	$[-1.68; 1.83]$	$[-2.47; 0.93]$	$[0.04; 3.50]$
Control	0.34^*	0.33^*	0.32^*	0.31^*
	$[0.07; 0.61]$	$[0.06; 0.59]$	$[0.06; 0.59]$	$[0.04; 0.58]$
Punishment	0.12	0.09	0.12	-0.02
	$[-0.14; 0.39]$	$[-0.18; 0.35]$	$[-0.14; 0.38]$	$[-0.28; 0.23]$
Judicial Punishment	0.31^*	0.36^*	0.47^*	0.28^*
	$[0.03; 0.58]$	$[0.09; 0.63]$	$[0.19; 0.74]$	$[0.02; 0.55]$
Opponent	-1.11^*	-1.07^*	-0.93^*	-1.43^*
	$[-1.31; -0.91]$	$[-1.27; -0.88]$	$[-1.13; -0.73]$	$[-1.63; -1.23]$
Political Interest	-0.16^*	0.02	0.10	0.00
	$[-0.29; -0.03]$	$[-0.11; 0.14]$	$[-0.02; 0.23]$	$[-0.12; 0.13]$
General Trust	-0.43^*	-0.58^*	-0.49^*	-0.41^*
	$[-0.68; -0.19]$	$[-0.82; -0.34]$	$[-0.72; -0.25]$	$[-0.64; -0.17]$
Age (log)	-0.14	-0.07	-0.48^*	0.71^*
	$[-0.56; 0.29]$	$[-0.49; 0.34]$	$[-0.89; -0.08]$	$[0.29; 1.12]$
Male	-0.25^*	-0.16	-0.16	-0.11
	$[-0.46; -0.04]$	$[-0.36; 0.06]$	$[-0.37; 0.05]$	$[-0.31; 0.10]$
Education	-0.23^*	-0.23^*	-0.10	-0.06
	$[-0.45; -0.01]$	$[-0.45; -0.00]$	$[-0.32; 0.13]$	$[-0.27; 0.16]$
Employment: retired	0.36	0.26	-0.04	0.23
	$[-0.12; 0.84]$	$[-0.20; 0.72]$	$[-0.49; 0.43]$	$[-0.25; 0.72]$
Employment: housewife	-0.07	-0.10	-0.17	-0.05
	$[-0.39; 0.24]$	$[-0.41; 0.22]$	$[-0.47; 0.15]$	$[-0.35; 0.26]$
Employment: student	0.50	-0.05	0.06	0.26
	$[-0.27; 1.27]$	$[-0.79; 0.70]$	$[-0.65; 0.78]$	$[-0.47; 1.00]$
Employment: unemployed	-0.17	0.15	-0.01	0.19
	$[-0.47; 0.13]$	$[-0.14; 0.45]$	$[-0.31; 0.28]$	$[-0.11; 0.49]$
Employment: other	0.15	0.40	0.38	0.77^*
	$[-0.29; 0.60]$	$[-0.06; 0.86]$	$[-0.08; 0.82]$	$[0.29; 1.26]$
Sector: Private Business	0.17	-0.12	-0.21	-0.06
	$[-0.06; 0.40]$	$[-0.34; 0.11]$	$[-0.43; 0.01]$	$[-0.27; 0.16]$
Sector: Non-profit	0.11	-0.20	-0.50^*	-0.36^*
	$[-0.28; 0.50]$	$[-0.57; 0.18]$	$[-0.88; -0.13]$	$[-0.73; -0.00]$
Savings	-0.19^*	-0.33^*	-0.09	-0.31^*
	$[-0.30; -0.08]$	$[-0.44; -0.22]$	$[-0.20; 0.01]$	$[-0.42; -0.20]$
Observations	1028	1027	1032	1025

* Null hypothesis value outside 95% credible interval. Repoted are medians and 95% credible intervals.
Respective baseline categories are Fraud, Supporter, Employment: paid employment, Sector: Government or public institution

Ordinal logistic regression estimates of treatment effects for Russian sample. (cont.)

Bibliography

AFP (2019). “Morales Opponent Rejects Poll Results in Bolivia”. In: *Agence France Presse, October 22*.

Alvarez, Michael R. and Ines Levin (2021). “Uncertain Neighbors: Bayesian Propensity Score Matching for Causal Inference”. In: *Unpublished Manuscript*.

Alvarez, R. Michael, Thad E. Hall, and Susan D. Hyde, eds. (2008). *Election Fraud*. Washington D.C.: Brookings Institution Press.

ANF (2019). “Al Menos 40 Autoridades del MAS Renunciaron en Medio de la Crisis Política”. In: *Agencia de Noticias Fides, November 10*.

Anria, Santiago and Jennifer Cyr (2019). “Is Bolivia’s Democracy in Danger? Here’s What’s Behind the Disputed Presidential Election”. In: *Washington Post, October 30*.

Awoyemi, John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadare (2017). “Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis”. In: *2017 International Conference on Computing Networking and Informatics (ICCNI)*, pp. 1–9.

Barié, Cletus G. (2022). “Representation of Indigenous Peoples in Times of Progressive Governments: Lessons Learned From Bolivia”. In: *Latin American and Caribbean Ethnic Studies* 17.2, pp. 167–192.

Bartneck, Christoph et al. (2015). “Comparing the Similarity of Responses Received From Studies in Amazon’s Mechanical Turk to Studies Conducted Online and With Direct Recruitment”. In: *PLoS ONE* 10.4, pp. 1–23.

Beber, Bernd and Alexandra Scacco (2012). “What the Numbers Say: A Digit-Based Test for Election Fraud”. In: *Political Analysis* 20.2, pp. 211–234.

Becker, Richard A., Chris Volinsky, and Allan R. Wilks (2010). “Fraud Detection in Telecommunications: History and Lessons Learned”. In: *Technometrics* 52.1, pp. 20–33.

Bejarano, Ana María and Eduardo Pizarro (2005). “From ’Restricted’ to ’Besieged’: The Changing Nature of the Limits to Democracy in Colombia”. In: *The Third Wave of Democratization in Latin America: Advances and Setbacks*. Ed. by Frances Hagopian and Scott P. Mainwaring. Cambridge: Cambridge University Press, pp. 235–260.

Benford, Frank (1938). “The Law of Anomalous Numbers”. In: *Proceedings of the American Philosophical Society* 78, pp. 551–572.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz (2012). “Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk”. In: *Political Analysis* 20.3, pp. 351–368.

Birch, Sarah (2011). *Electoral Malpractice*. Oxford: Oxford University Press.

Bolton, Richard J. and David J. Hand (2002). “Statistical Fraud Detection: A Review”. In: *Statistical Science* 17.3, pp. 235–255.

Borghesi, Christian and Jean-Philippe Bouchaud (2010). “Spatial Correlations in Vote Statistics: A Diffusive Field Model for Decision-Making”. In: *European Physical Journal B* 75.3, pp. 395–404.

Bowler, Shaun and Jeffrey A. Karp (2004). “Politicians, Scandals, and Trust in Government”. In: *Political Behavior* 26.3, pp. 271–287.

Bown, Francis (1994). “Influencing the House of Lords: The Role of the Lords Spiritual 1979–1987”. In: *Political Studies* 42.1, pp. 105–119.

Bullock, John G. (2009). “Partisan Bias and the Bayesian Ideal in the Study of Public Opinion”. In: *Journal of Politics* 71.3, pp. 1109–1124.

Bush, Sarah Sunn and Lauren Prather (2018). “Who’s There? Election observer identity and the local credibility of elections”. In: *International Organization* 72.3, pp. 659–692.

Callen, Michael and James D. Long (2015). “Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan”. In: *American Economic Review* 105.1, pp. 354–381.

Cantú, Francisco (2014). “Identifying Irregularities in Mexican Local Elections”. In: *American Journal of Political Science* 58.4, pp. 936–951.

— (2019a). “Groceries for Votes: The Electoral Returns of Vote Buying”. In: *Journal of Politics* 81.3, pp. 790–804.

— (2019b). “The Fingerprints of Fraud: Evidence From Mexico’s 1988 Presidential Election”. In: *American Political Science Review* 113.3, pp. 710–726.

Cantú, Francisco and Omar García-Ponce (2015). “Partisan Losers’ Effects: Perceptions of Electoral Integrity in Mexico”. In: *Electoral Studies* 39, pp. 1–14.

Cantú, Francisco and Sebastián M. Saiegh (2011). “Fraudulent Democracy? An Analysis of Argentina’s Infamous Decade Using Supervised Machine Learning”. In: *Political Analysis* 19.4, pp. 409–433.

Carter Center (2007). “Final Report on Ecuador´s September 30, 2007, Constituent Assembly Elections”. In: *Election Observation Report*.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik (2020). *Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge: Cambridge University Press.

Chipman, Hugh A., Edward I. George, and Robert E. McCulloch (2010). "BART: Bayesian Additive Regression Trees". In: *Annals of Applied Statistics* 4.1, pp. 266–298.

Ciria, Alberto (1974). *Parties and Power in Modern Argentina*. Albany, NY: State University of New York Press.

Citrin, Jack (1974). "Comment: The Political Relevance of Trust in Government". In: *American Political Science Review* 68.3, pp. 973–988.

Clifford, Scott, Ryan M. Jewell, and Philip D. Waggoner (2015). "Are Samples Drawn from Mechanical Turk Valid for Research on Political Ideology?" In: *Research & Politics* 2.4, pp. 1–9.

Coppock, Alexander (2019). "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach". In: *Political Science Research and Methods* 7.3, pp. 613–628.

Curriel, John A., Charles Stewart III, and Jack Williams (2021). "One Shift, Two Shifts, Red Shift, Blue Shift: Reported Election Returns in the 2020 Election". In: *Unpublished Manuscript—MIT Election Data and Science Lab*.

Daxecker, Ursula E (2012). "The Cost of Exposing Cheating: International Election Monitoring, Fraud, and Post-Election Violence in Africa". In: *Journal of Peace Research* 49.4, pp. 503–516.

Deckert, Joseph, Mikhail Myagkov, and C. Ordeshook (2011). "Benford's Law and the Detection of Election Fraud". In: *Political Analysis* 19.03, pp. 245–268.

Dellmuth, Lisa Maria and Jonas Tallberg (2015). "The Social Legitimacy of International Organisations: Interest Representation, Institutional Performance, and Confidence Extrapolation in the United Nations". In: *Review of International Studies* 41.3, pp. 451–475.

Duque Daza, Javier (2019). "Democracia Electoral Fraudulenta. La Trashumancia Electoral Como Estrategia Para Ganar Elecciones Locales en Colombia." In: *Estudios Políticos (Medellín)* 55, pp. 61–86.

Easton, David (1965). *A Systems Analysis of Political Life*. New York: Wiley.

— (1975). "A Re-Assessment of the Concept of Political Support". In: *British Journal of Political Science* 5.4, pp. 435–457.

Ekin, Tahir, Luca Frigau, and Claudio Conversano (2021). "Health Care Fraud Classifiers in Practice". In: *Applied Stochastic Models in Business and Industry* 37.6, pp. 1182–1199.

El Espectador (2015). "A Dos Días de Elecciones, Sorprenden a Yahír Acuña Con $200 Millones". In: *El Espectador, October 24*.

Enikolopov, Ruben et al. (2013). "Field Experiment Estimate of Electoral Fraud in Russian Parliamentary Elections". In: *PNAS* 110.2, pp. 448–452.

Escobari, Diego and Gary Hoover (2019). “Evo Morales and Electoral Fraud in Bolivia: A Natural Experiment Estimate”. In: *Unpublished Manuscript*.

EU, European Union (2002). “Final Report on the Presidential and Parliamentary Elections, Ecuador”. In: *Election Observation Report*.

— (2009). “Final Report on Presidential and Parliamentary Elections 26 April, 2009, Ecuador”. In: *Election Observation Report*.

Fan, Jianqing and Irene Gijbels (1996). *Local Polynomial Modelling and Its Applications*. Boca Raton: Chapman & Hall.

Ferrufino, Cambara and Pablo Cesar (2019). “Costas Atribuye a Una ‘Impericia’ la Paralización de la Transmisión de Votos del TREP”. In: *El Deber, October 23*.

Foley, Edward B. and Charles Stewart III (2020). “Explaining the Blue Shift in Election Canvassing”. In: *Journal of Political Institutions and Political Economy* 1.2, pp. 239–265.

Frye, Timothy and Ekaterina Borisova (2019). “Elections, Protest, and Trust in Government: A Natural Experiment from Russia”. In: *The Journal of Politics* 81.3, pp. 820–832.

Geddes, Barbara and John Zaller (1989). “Sources of Popular Support for Authoritarian Regimes”. In: *American Journal of Political Science* 33.2, pp. 319–347.

Gehlbach, Scott (2012). “Electoral Fraud in Russia: Report from the Russian Blogosphere”. In: *The Monkey Cage, January 27*.

Gelman, Andrew et al. (2004). *Bayesian Data Analysis*. Boca Raton: Chapman & Hall/CRC.

Greene, Kenneth F. (2007). *Why Dominant Parties Lose: Mexico's Democratization in Comparative Perspective*. Cambridge: Cambridge University Press.

Greene, Kenneth F. and Mariano Sánchez-Talanquer (2018). “Authoritarian Legacies and Party System Stability in Mexico”. In: *Party Systems in Latin America: Institutionalization, Decay, and Collapse*. Ed. by Scott Mainwaring. Cambridge: Cambridge University Press, pp. 201–226.

Haerpfer, C. et al. (2022). *World Values Survey: Round Seven - Country-Pooled Datafile Version 3.0*. Madrid, Spain and Vienna, Austria: JD Systems Institute and WVSA Secretariat.

Ham, van Carolien (2015). “Getting Elections Right? Measuring Electoral Integrity”. In: *Democratization* 22.4, pp. 714–737.

Harshvardhan, GM et al. (2020). “A Comprehensive Survey and Analysis of Generative Models in Machine Learning”. In: *Computer Science Review* 38, p. 100285.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning—Data Mining, Inference and Prediction*. New York: Springer.

Hicken, Allen and Walter R. Mebane (2017). “A Guide to Elections Forensics - Research and Innovation Grants Working Paper Series”. In: *Research Report*, pp. 1–43.

Hilas, Constantinos S. and John N. Sahalos (2005). "User Profiling for Fraud Detection in Telecommunication Networks". In: *5th International Conference on Technology and Automation*, pp. 1–6.

Hill, Seth J. (2017). "Learning Together Slowly: Bayesian Learning About Political Facts". In: *Journal of Politics* 79.4, pp. 1403–1418.

Hill, Theodor (1995). "Base-Invariance Implies Benford's Law". In: *Proceedings of the American Mathematical Society* 123.3, pp. 887–895.

— (1996). "A Statistical Derivation of the Significant-Digit Law". In: *Statistical Science* 10.4, pp. 354–363.

Hiskey, Jonathan T. and Shaun Bowler (2005). "Local Context and Democratization in Mexico". In: *American Journal of Political Science* 49.1, pp. 57–71.

Hooghe, Marc and Sofie Marien (2012). "A Comparative Analysis of the Relation Between Political Trust and Forms of Political Participation in Europe". In: *European Societies* 15.1, pp. 131–152.

Hyde, Susan D. (2007). "The Observer Effect in International Politics: Evidence from a Natural Experiment". In: *World Politics* 60, pp. 37–63.

— (2011). "Catch Us If You Can: Election Monitoring and International Norm Diffusion". In: *American Journal of Political Science* 55.2, pp. 356–369.

Idrobo, Nicolás, Dorothy Kronick, and Francisco Rodríguez (2020). "Do Shifts in Late-Counted Votes Signal Fraud? Evidence From Bolivia". In: *Unpublished Manuscript*, pp. 1–48.

— (2022). "Do Shifts in Late-Counted Votes Signal Fraud? Evidence From Bolivia". In: *Journal of Politics* 84.4, Online Only.

Inglehart, R. et al. (2020). "World Values Survey: Round Seven - Country-Pooled Datafile 2017-2020". In: *Madrid: JD Systems Institute*.

IRI, International Republican Institute (2003). "Republic of Ecuador, National Elections October 20 November 24, 2002 - Report and Recommendations". In: *Election Observation Report*.

Itri, Bouzgarne et al. (2019). "Performance Comparative Study of Machine Learning Algorithms for Automobile Insurance Fraud Detection". In: *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*, pp. 1–4.

Jimenez, Raúl, Manuel Hidalgo, and Peter Klimek (2017). "Testing for Voter Rigging in Small Polling Stations". In: *Science Advances* 3.6, pp. 1–7.

Johnston, Jake and David Rosnick (2020). "Observing the Observers: The OAS in the 2019 Bolivian Elections". In: *Centre for Economic and Policy Research Report*, pp. 1–92.

Jullum, Martin et al. (2020). "Detecting Money Laundering Transactions With Machine Learning". In: *Journal of Money Laundering Control* 23.1, pp. 173–186.

Kelley, Judith G. (2012a). *Monitoring Democracy*. Princeton and Oxford: Princeton University Press.

— (2012b). "The Good, the Bad, and the Ugly: Rethinking Election Monitoring". In: *International Institute for Democracy and Electoral Assistance*.

Kerr, Nicholas and Michael Wahman (2021). "Electoral Rulings and Public Trust in African Courts and Elections". In: *Comparative Politics* 53.2, pp. 257–290.

King, Gary and Richard Nielsen (2019). "Why Propensity Scores Should Not Be Used for Matching". In: *Political Analysis* 27.4, pp. 435–454.

Klimek, Peter et al. (2012). "Statistical Detection of Systematic Election Irregularities". In: *PNAS* 109.41, pp. 16469–16473.

Kobak, Dmitry, Sergey Shpilkin, and Maxim S. Pshenichnikov (2016a). "Integer Percentages as Electoral Falsification Fingerprints". In: *Annals of Applied Statistics* 10.1, pp. 54–73.

— (2016b). "Statistical Fingerprints of Electoral Fraud?" In: *Significance* 13.4, pp. 20–23.

— (2018). "Putin's Peaks: Russian Election Data Revisited". In: *Significance* 15.3, pp. 8–9.

Kurmanaev, Anatoly and Cesar Del Castillo (2019). "How an Unknown Female Senator Came to Replace the Bolivian President Evo Morales". In: *New York Times, November 27*.

Lankina, Tomila and Katerina Tertytchnaya (2020). "Protest in Electoral Autocracies: A New Dataset". In: *Post-Soviet Affairs* 36.1, pp. 20–36.

Levin, Ines, Julia Pomares, and R Michael Alvarez (2016). "Using Machine Learning Algorithms to Detect Election Fraud". In: *Computational Social Science*. Cambridge: Cambridge University Press, pp. 266–294.

Li, Jing et al. (2008). "A Survey on Statistical Methods for Health Care Fraud Detection". In: *Health Care Management Science* 11.3, pp. 275–287.

Li, Yimeng, Michelle Hyun, and R. Michael Alvarez (2022). "Why Do Election Results Change after Election Day? The "Blue Shift" in California Elections". In: *Political Research Quarterly* 75.3, pp. 860–874.

Lindberg, Staffan I (2006). "Tragic Protest: Why Do Opposition Parties Boycott Elections?" In: *Electoral Authoritarianism: The Dynamics of Unfree Competition*. Ed. by Andreas Schedler. Lynne Rienner Publishers.

Los Tiempos (2019). "TSE Dice Que Se Paralizó el TREP Para 'Evitar la Confusión' Con Resultados de Departamentos". In: *Los Tiempos, October 20*.

Mack, Verena and Lukas F. Stoetzer (2019). "Election Fraud, Digit Tests and How Humans Fabricate Vote Counts - An Experimental Approach". In: *Electoral Studies* 58, pp. 31–47.

Madrid, Raúl (2012). *The Rise of Ethnic Politics in Latin America*. Cambridge: Cambridge University Press.

Mauk, Marlene (2019). "When Does Cheating Matter? Electoral Malpractice, Perceptions of Electoral Fairness, and Political Trust". In: *Unpublished Manuscript*.

Mebane, Walter R. (2006). "Election Forensics: Vote Counts and Benford's Law". In: *Unpublished Manuscript*, pp. 1–48.

— (2008). “Election Forensics: The Second-Digit Benford’s Law Test and Recent American Presidential Elections”. In: *Election Fraud: Detecting and Deterring Electoral Manipulation*. Ed. by R. Michael Alvarez, Thad E. Hall, and Susan Hyde. Washington, D.C.: Brookings Institution Press, pp. 162–181.

— (2011). “Comment on “Benford’s Law and the Detection of Election Fraud””. In: *Political Analysis* 19.03, pp. 269–272.

— (2013). “Using Vote Counts’ Digits to Diagnose Strategies and Frauds: Russia”. In: *Unpublished Manuscript*, pp. 1–59.

— (2020). “Inappropriate Applications of Benford’s Law Regularities to Some Data from the 2020 Presidential Election in the United States”. In: *Unpublished Manuscript*.

Medzihorsky, Juraj (2015). “Election Fraud: A Latent Class Framework for Digit-Based Tests”. In: *Political Analysis* 23, pp. 506–517.

Mishler, William and Richard Rose (2001). “What Are the Origins of Political Trust?: Testing Institutional and Cultural Theories in Post-communist Societies”. In: *Comparative Political Studies* 34.1, pp. 30–62.

Montgomery, Jacob M et al. (2015). “An Informed Forensics Approach to Detecting Vote Irregularities”. In: *Political Analysis* 23.4, pp. 488–505.

Morgan, Stephen L. and Christopher Winship (2007). *Counterfactual and Causal Inference—Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

Myagkov, Mikhail, Peter C. Ordeshook, and Dimitri Shakin (2009). *The Forensics of Election Fraud: Russia and Ukraine*. Cambridge: Cambridge University Press.

Nadeau, Richard, Jean-François Daoust, and Ruth Dassonneville (2021). “Winning, Losing, and the Quality of Democracy”. In: *Political Studies*, Online Only.

Neunhoeffer, Marcel, Zhiwei Steven Wu, and Cynthia Dwork (2021). “Private Post-GAN Boosting”. In: *ICLR International Conference on Learning Representations Conference Paper*.

Newcomb, Simon (1881). “Note on the Frequency of Use of the Different Digits in Natural Numers”. In: *American Journal of Mathematics* 4, pp. 39–40.

Newman, John (2020). “The OAS Conclusions about the Election Integrity of the Bolivian Election Are Correct”. In: *Unpublished Manuscript*.

Nisbett, Richard E. and Timothy DeCamp Wilson (1977). “The Halo Effect: Evidence for Unconscious Alteration of Judgements”. In: *Journal of Personality and Social Psychology* 35.4, pp. 250–256.

Norris, Pippa (2014). *Why Electoral Integrity Matters*. Cambridge: Cambridge University Press.

— (2019). “Do Perceptions of Electoral Malpractice Undermine Democratic Satisfaction? The US in Comparative Perspective”. In: *International Political Science Review* 40.1, pp. 5–22.

Norris, Pippa and Ronald Inglehart, eds. (2019). *Cultural Backlash: Trump, Brexit and Authoritarian Populism*. New York: Cambridge University Press.

OAS (1998). "Electoral Observation in Ecuador 1998". In: *Election Observation Report*.

— (2006). "Informe de la Misión de Observación Electoral de la OEA en la República del Ecuador". In: *Election Observation Report*.

— (2008). "Informe de la Misión de Observación Electoral de la OEA en la Rrepública del Ecuador". In: *Election Observation Report*.

— (2019a). "Electoral Integrity Analysis General Elections in the Plurinational State of Bolivia—FINAL REPORT". In: *Organization of American States*.

— (2019b). "Informe Preliminar de la Misión de Observación Electoral en Bolivia". In: *Organization of American States*.

Palmer, Carl L. and Rolfe D. Peterson (2016). "Halo Effects and the Attractiveness Premium in Perceptions of Political Expertise". In: *American Politics Research* 44.2, pp. 353–382.

Pericchi, Luis and David Torres (2011). "Quick Anomaly Detection by the Newcomb-Benford Law, With Applications to Electoral Processes Data From the USA, Puerto Rico and Venezuela". In: *Statistical Science* 26.4, pp. 502–516.

Przeworski, Adam, Susan C. Stokes, and Bernard Manin (1999). *Democracy, Accountability, and Representation*. Cambridge: Cambridge University Press.

Queirolo, Rosario (2013). *The Success of the Left in Latin America: Untainted Parties, Market Reforms, and Voting Behavior*. Notre Dame: University of Notre Dame Press.

Reuter, Ora John and David Szakonyi (2021). "Electoral Manipulation and Regime Support: Survey Evidence from Russia". In: *World Politics* 73.2, 275–314.

Reuters (2016). "Russian Local Election Officials Fired Over Vote Irregularities". In: *Reuters, November 2*.

— (2020). "Fact Check: Deviation From Benford's Law Does Not Prove Election Fraud". In: *Reuters, November 10*.

Robertson, Graeme (2017). "Political Orientation, Information and Perceptions of Election Fraud: Evidence from Russia". In: *British Journal of Political Science* 47.3, pp. 589–608.

Rozenas, Arturas (2017). "Detecting Election Fraud from Irregularities in Vote-Share Distributions". In: *Political Analysis* 25, pp. 41–56.

Sartori, Giovanni (1976). *Parties and Party Systems*. New York: Cambridge University Press.

Schedler, Andreas (2002). "The Menu of Manipulation". In: *Journal of Democracy* 13.2, pp. 36–50.

Schneider, Irena (2017). "Can We Trust Measures of Political Trust? Assessing Measurement Equivalence in Diverse Regime Types". In: *Social Indicators Research* 133.3, pp. 963–984.

Schwirtz, Michael and David M. Herszenhorn (2011). "Voters Watch Polls in Russia, and Fraud is What They See". In: *New York Times, December 5*.

Shikano, Susumu et al. (2019). "Detection of Election Fraud". In: *The Oxford Handbook of Public Choice, Volume 2*. Oxford University Press, pp. 879–895.

Stan Development Team (2020). *RStan: the R interface to Stan*. URL: `http://mc-stan.org/`.

Sudjianto, Agus et al. (2010). "Statistical Methods For Fighting Financial Crimes". In: *Technometrics* 52.1, pp. 5–19.

The Carter Center (2005). *Observing the Venezuela Presidential Recall Referendum*. Tech. rep. URL: `https://www.cartercenter.org/documents/2020.pdf`.

The Guardian (2011). "Russians Come Out In Force to Protest Against Alleged Electorarl Fraud". In: *The Guardian, December 10*.

Thorndike, E.L. (1920). "A Constant Error in Psychological Ratings". In: *Journal of Applied Psychology* 4.1, pp. 25–29.

Torcal, Mariano and Pablo Christmann (2019). "Congruence, national context and trust in European institutions". In: *Journal of European Public Policy* 26.12, pp. 1779–1798.

Torre, Carlos de la (2015). *De Velasco a Correa: Insurrecciones, populismos y elecciones en Ecuador, 1944-2013.* Quito, Ecuador: Corporación Editora Nacional - Unversidad Andina Simón Bolívar.

Vickery, Chad and Erica Shein (2012). "Assessing Electoral Fraud in New Democracies: Refining the Vocabulary". In: *International Foundation for Electoral Systems—White Paper Series*, pp. 1–13.

Warner, Zach et al. (2021). "Hidden in Plain Sight? Irregularities on Statutory Forms and Electoral Fraud". In: *Electoral Studies* 74, pp. 1–10.

Williams, Jack and John Curiel (2020). "Analysis of the Quick Count in the 2019 Bolivia Election". In: *Unpublished Manuscript*.

Williamson, Scott (2021). "Elections, Legitimacy, and Compliance in Authoritarian Regimes: Evidence from the Arab World". In: *Democratization* 28.8, pp. 1483–1504.

Yang, Hung Chia et al. (2015). "Assessment of household appliance surveys collected with Amazon Mechanical Turk". In: *Energy Efficiency* 8.6, pp. 1063–1075.

Zaller, John R (1992). *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zhang, Mali, R. Michael Alvarez, and Ines Levin (2019). "Election Forensics: Using Machine Learning and Synthetic Data for Possible Election Anomaly Detection". In: *PLoS ONE* 14.10, pp. 1–14.