Title
Electoral Corruption in Developing Democracies

Author
Cantu, Francisco

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Electoral Corruption in Developing Democracies

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Political Science

by

Francisco Cantú

Committee in charge:

Professor Sebastián Saiegh, Chair
Professor Stephan Haggard, Co-Chair
Professor R. Michael Alvarez
Professor Scott Desposato
Professor Alberto Díaz-Cayeros
Professor Peter Smith

2013

The dissertation of Francisco Cantú is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____
Co-Chair

____
Chair

University of California, San Diego

2013

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

In the course of planning, writing, and finishing the dissertation, I became indebted to many people. First and foremost I would like to express my deepest gratitude to Sebastián Saiegh, who backed this project unconditionally, even in its most challenging stages. I cannot repay his mentorship, commitment, and advising. Scott Desposato also played a key role during my years in graduate school, and I thank him for his open door and kind help. I am grateful to the Co-Chair of my dissertation, Steph Haggard, for constantly setting the bar high to improve both my research and teaching skills. I am honored to have Peter Smith on my committee, a person from whom I always learned something new, even in the briefest of conversations. Mike Alvarez shared with me his knowledge and experience on the topic of election forensics, and his work was a source of inspiration for this dissertation. Alberto Díaz-Cayeros's provided me vital observations and comments to improve this project.

This work is also the scholarship of many scholars with whom I have had the opportunity to interact with over this years. At UC San Diego, I owe my gratitude to Marisa Abrajano, Nageeb Ali, Amy Bridges, Gary Cox, Clark Gibson, Sam Kernell, Peter Galderisi, Gerry Mackie, Mat McCubbins, Megumi Naoi, Kaare Strøm, Randy Willoughby, and Langche Zeng. I wish to warmly thank Federico Estévez and Eric Magar for their encouragement to apply to graduate school and their long-distance support. Finally, Allyson Benton, Mario Cortina, and Pablo Mateos provided some of the data and key comments in the final stage of writing Chapter 4.

I also feel very fortunate to be part of a great cohort of student peers, from whom I have learned a lot. I would especially like to thank Saul Cunow, Matt Kearney, Alex Ruiz, and Devesh Tiwari for their friendship. I also acknowledge the moral and intellectual support of Cesi Cruz, Dimitar Gueorguiev, Alex Hughes, Ryan Jablonski, Don Lee, Mike Rivera, Paul Schuler, Martín Tobal, and Madeline Woo. Special thanks to Vero Hoyo and Ale Ríos, who offered me their uninterested advice to survive the first few years in San Diego.

On a personal note, I am thankful, in no particular order, to Jim and Ann Inouye, the Sprat's, José Cruz, David Cantú, the García's, my parents, the Beacon's Beach gang, and Pedrito Riera. Each of them knows why. Last, but not least, this work could not have been done without Lucero Gonzalez Franco, who has been a partner in this adventure and my everyday reminder to keep it real.

***

I recognize the generous financial support from Mexico's National Council for Science and Technology (CONACYT) and a complimentary scholarship from the Mexican Ministry of Education (SEP).

Chapter 4 is based on a forthcoming article in the *American Journal of Political Science*. The dissertation author was the primary investigator and author of this paper.

VITA

2013	Ph. D. in Political Science. University of California, San Diego.
2009	M. A. in Political Science. University of California, San Diego.
2004	*Licenciatura en Ciencia Política*. Instituto Tecnológico Autónomo de México.

ABSTRACT OF THE DISSERTATION

Electoral Corruption in Developing Democracies

by

Francisco Cantú

Doctor of Philosophy in Political Science

University of California, San Diego, 2013

Professor Sebastián Saiegh, Chair
Professor Stephan Haggard, Co-Chair

This dissertation concerns the ways to detect electoral manipulation, even when we are not able to rely on the observable characteristics of a particular election. This dissertation explains how political parties disrupt the electoral process and proposes different ways to identify electoral manipulation. Chapter 1 provides the motivation of the project and the literature upon which it is built. Chapter 2 models the allocation of resources for corruption in competitive elections. The results of the model illustrate the conditions for parties to engage in corruption and

the situations in which electoral manipulation does define the outcome. Chapter 3 discusses the limitations of the most common forensic tools. Given the growing use of these techniques to evaluate the quality of the elections, I propose a way to analyze both when and how each technique is found beneficial. Chapter 4 introduces a novel technique to detect electoral irregularities while analyzing the 2010 gubernatorial elections in Mexico. I propose a novel empirical strategy to detect electoral fraud, which exploits a feature of the country's electoral code: within each electoral precinct, voters are assigned to polling stations according to their childhood surnames; consequently, the only difference between the voters at contiguous polling stations should be their last names. Chapter 5 extends the analysis of chapters 3 and 4 to evaluate both the 2008 recall referendum in Bolivia and the 2012 presidential elections in Mexico and Venezuela. These elections have been surrounded by suspicion of electoral fraud, yet the existing evidence supporting those beliefs has been inconclusive. The findings for Mexico and Venezuela show no strong evidence of electoral manipulation. Meanwhile, the results in Bolivia confirm what both the media reports and the political opposition claim.

Chapter 1

Introduction

The paradox of electoral manipulation is that it takes place as a side effect of the secret ballot, which protects voters from intimidation, yet conceals the voting preferences of the citizens. In this environment of incomplete information, politicians are tempted not only to manipulate the election but also to avoid revealing the truth: perpetrators of fraud want to keep their activities as hidden as possible (Lehoucq, 2003). The alleged victims of fraud, in turn, may want to claim the existence of electoral manipulation even if it did not occur (Magaloni, 2007; Eisenstadt and Poiré, 2006). Therefore, the necessary condition of the secret ballot in contemporary democracies creates the contradictory effect of protecting citizens' rights and frustrating an accurate representation of voting preferences.

This project concerns the ways to detect electoral manipulation, even when we are not able to rely on the observable characteristics of a particular election. To fulfill politicians' double goal of keeping their hold on power and their wish to claim that they are following democratic norms (Hyde, 2011, p. 159),[1] electoral manipulation often appears out of sight of both observers and voters, allowing the most blatant fraud to occur in those places with insufficient resources to detect it (Hartlyn and McCoy, 2006). This dissertation explains how political parties disrupt the electoral process and proposes different ways to identify electoral manipulation.

My theory seeks to bridge conflicting theories on the observability and purpose of fraud. On the one hand, studies consider fraud as a hidden tactic to modify electoral results (Lehoucq, 2003; Schedler, 2002*c*). In these studies, the benefits of manipulation are outcome-driven and include not only winning an election but also obtaining a majority in congress (Magaloni, 2006)Additionally, they could discourage voters of casting ballots for the opposition in future elections. (Donno and Roussias, 2012). Following this logic, manipulation only exists when the benefits outweigh the potential retributions of observing the fraud, which range from losing international legitimacy (Hyde, 2011; Kelley, 2012) to being ousted from power after local fraud allegations (Gandhi and Przeworski, 2009; Tucker, 2007). These potential penalties warn perpetrators to keep their activities concealed from both observers and citizens.

On the other hand, there is another group of studies which challenges the hidden nature of electoral manipulation and argues that most of the benefits of fraud are seen only when the perpetrators are outwardly recognized. These theories focus on the empirical pattern of observing fraud in the elections where winners would

[1] Summarized by Przeworski et al. (2000, p. 26) as “politicians are just PRIstas by nature.”

have won even if they had abstained from corruption. According to this argument, the main benefits of electoral manipulation do not come from the electoral results but rather from the demoralizing effects felt by the opposition, which potentially prevents participation in future elections (Simpser, 2005; Little, 2011; Kalinin and Mebane, 2011; Donno and Roussias, 2012).

Building on these works, I propose that the observability and the benefits of fraud for a party depend on the competitiveness of the political system, which is determined by the distribution of "resources, political power, and institutional advantages" between the incumbent and opposition parties (Simpser, 2013, p. 15). My argument suggests that the distribution of resources within a political system determines not only the degree of electoral corruption but also the identity of the perpetrators of fraud.

In addition, this work contributes to the empirical evaluation of electoral integrity. Similar to the theoretical debates described above, the empirical literature on fraud is organized by the different types of evidence showing electoral manipulation.[2] On one side, some scholars collect and analyze observable evidence of electoral corruption such as surveys (Cornelius, 2002; Schedler, 2002*a*; Alvarez and Hall, 2008), legal denunciations of fraud (Lehoucq and Molina, 2002; Ziblatt, 2009; Kuo, Teorell and Ziblatt, 2011; Mares and Zhu, 2012), and records of post-electoral protests (Afshari and Underwood, 2009; Eisenstadt, 2004; Tucker, 2007). Given that defeated candidates have reasons to claim fraud even when no deception oc-

[2] For an exhaustive review of the literature about electoral fraud, see (Lehoucq, 2003).

curred, a limitation of this type of evidence is that perceptions can overestimate the real level of electoral manipulation. Therefore, observational evidence is insufficient to evaluate the quality of an election.

Instead, this project is associated with a second collection of research that assumes the hidden nature of electoral manipulation and provides identification strategies such as natural experiments (Fukumoro and Horiuchi, 2011), field experiments (Collier and Vicente, 2008; Vicente and Wantchekon, 2009; Hyde, 2007; Ichino and Schundeln, 2012; Callen and Long, 2012), or forensic analyses (King, 2001; Myagkov, Ordeshook and Shakin, 2009; Mebane, 2006; Levin et al., 2009; Cantú and Saiegh, 2011; Beber and Scacco, 2012), in order to identify electoral irregularities. My contributions to this field are two. First, I propose a method to detect electoral irregularities in a context where fraud occurs in a decentralized and concealed manner. Here, I illustrate the advantages of this technique in the analysis of the Mexican local elections in 2010. Second, I provide a survey of the extant forensic tools and evaluate the conditions in which each technique yields relevant information. Given the growing use of these techniques to evaluate the quality of the elections, I propose a way to analyze when and how each technique is found beneficial.

Lastly, this dissertation also engages with the literature on developing democracies (Diamond, 1999; O'Donnell, 1999; Schedler, 2006; Smith, 2011; Levitsky and Way, 2010; Gibson, 2010; Rebolledo, 2012). Many of these studies propose a residual category, in order as to include those regimes that are neither long-established democracies nor full autocracies. Acknowledging the problems that most of this studies face, this work proposes to clear out the "gray zone" where scholars situate most developing democracies and, instead, refocus the discussion on the ways in which parties hold and manipulate elections.[3]

Roadmap

The dissertation is organized as follows. Chapter 2 models the allocation of resources for corruption in competitive elections. The extant literature is inconclusive about the observability and the benefits of election tampering. The model proposed here includes the level of political competition to predict the characteristics of fraud. Also, given that many studies assume that electoral manipulation goes hand in hand with victory for the perpetrators, instances in which the cheaters lose have rarely been studied. If electoral corruption does not always affect the result, why do parties still engage in these practices? The results of the model illustrate the conditions for parties to engage in corruption and the situations in which electoral manipulation does define the outcome.

Chapter 3 discusses the limitations of the most common forensic tools. Given the growing use of these techniques to evaluate the quality of the elections, I propose a way to analyze when and how each technique is found beneficial. The goal of this chapter is to be a guideline for scholars and election administrators about the correct

[3] Quote from (Carothers, 2002, p. 9).

use of each technique to evaluate the integrity of any election.

Chapter 4 introduces a novel technique to detect electoral irregularities while analyzing the 2010 gubernatorial elections in Mexico. A generalized distrust in local elections has raised the question of whether electoral corruption has vanished in Mexico today, or if it remains a prevalent practice in the country. To answer this question, I exploit a feature of the country's electoral system: within each electoral precinct, voters are assigned to polling stations according to their childhood surnames. Consequently, the only difference between voters in contiguous polling stations should be their last names. Given that political preferences are seldom correlated with voters' names, I use suspicious differences in turnout levels across contiguous polling stations to identify fraudulent practices.

Chapter 5 extends the analysis of the previous chapter to evaluate both the 2008 recall referendum in Bolivia and the 2012 presidential elections in Mexico and Venezuela. These elections have been surrounded by suspicion of electoral fraud, yet the existing evidence supporting those beliefs has been inconclusive. In the case of Bolivia, this chapter represents the first forensic analysis of its most recent elections and, consistent with anecdotal evidence, the results provide substantiation of electoral manipulation in Bolivia. On the other hand, the evidence is insufficient to support a similar hypothesis of electoral manipulation in Venezuela and Mexico, and I demonstrate that most of the irregularities are a product of non-deliberate errors that do not affect any of the candidates in a significant way.

Chapter 2

Cheating and Losing: Corruption in Competitive Elections

Contrary to common knowledge, electoral corruption does not assure political victory. As illustrations of this fact, consider the three following cases in which incumbents lost elections despite their corrupt activities. First, the 2002 presidential elections in Kenya marked the end of power for the Kenya African National Union, whose leader held the presidential seat for 25 years (Los Angeles Times, December 27, 2002, p. A-8). Second, evidence of both tampered ballots and voter registration irregularities were found during the 2008 presidential elections in Ghana (Jockers, Kohnert and Nugert, 2009; Ichino and Schundeln, 2012), yet these events were not enough to prevent the incumbent's defeat in the close run-off election. Finally, tactics of vote-buying, electoral fraud, and the incarceration of political rivals were insufficient to prevent the end of local autocracy under the Institutional Revolutionary Party (PRI) in Oaxaca, Mexico in the summer of 2010 (Gibson, 2005; Rebolledo, 2012).

Given that many studies assume that electoral manipulation goes hand in hand with victory for the perpetrators, instances in which the cheaters lose have rarely been studied. One must ask, then, if electoral corruption does not always affect the result, why do parties still engage in fraud?

This paper explores parties' strategic behavior of manipulation in competitive elections. The extant literature on electoral corruption is divided between those who argue for the hidden and outcome-defining nature of electoral fraud, and those who contend that fraud is an observable event that transcends the winning of an election. The model proposed here seeks to solve this dichotomy. I argue that the level of political competition determines the characteristics of fraud. Once partisan resources are taken into account, electoral corruption is found to be an inefficient, yet indispensable activity in many competitive elections

The institutional conditions that allow incumbents to lose elections have been analyzed in the past (Hyde and Marinov, 2012). Yet, if we already know which elections can be lost, we might ask to what length incumbents are willing to try to prevent defeat. The findings of the chapter have two main implications. First, I provide a theoretical framework to understand when, and how, incumbents are willing to try to prevent defeat. The results build on the extant work that analyzes the conditions in which incumbents may lose elections (Hyde and Marinov, 2012) Second, I demonstrate that a competitive environment pushes parties to be engaged in electoral corruption even when the outcome of the election barely changes from what expected in a clean election. This result explains the events of electoral corruption during the political development of Germany (Mares and Zhu, 2012), Costa Rica (Lehoucq and Molina, 2002), and the U.S. (Cox and Kousser, 1981).

This chapter is organized as follows. Section 1 describes competing theories concerning the degree of fraud and the role of political competition to better explicate the issues. Section 2 proposes a model of the role of political competition in electoral corruption in two different environments. Section 3 explores some predictions of the model using comparative statistics. Finally, Section 4 proposes possible future directions for research.

2.1 Literature Review

Almost every analysis of electoral fraud agrees on two basic assumptions: (1) corrupt activities generate costs, and (2) when successfully accomplished, corruption brings benefits not available through standard political competition. Conceptual differences of these two assumptions lead to contrasting explanations of when and how fraud appears. This section describes the two main perspectives of electoral manipulation.

On one side of the debate, fraud appears as an instrument only used to modify the electoral result. This argument assumes that perpetrators are willing to bear the costs of corruption only when there is uncertainty in an electoral outcome (Lehoucq, 2003; Schedler, 2002*c*). These studies consider fraud as a way not only to win an election but also to attain other outcome-based benefits, such as obtaining a majority in congress (Magaloni, 2006). In these cases, the benefits are based on the electoral outcome and corruption decreases their uncertainty. Moreover, the costs of electoral manipulation, when detected, warn perpetrators to keep their activities hidden. Deterrents of committing obvious fraud range from losing international legitimacy (Hyde, 2011; Kelley, 2012) to being ousted from power after fraud allegations (Gandhi and Przeworski, 2009; Tucker, 2007). In sum, as the recent wave of studies on electoral forensics show, fraudulent activities "are things that only its victims want publicized (Lehoucq, 2003, p. 235)."

The other side of the debate considers the empirical pattern of observing fraud, even when both the perpetrators and the victims can accurately predict the electoral result ex-ante. These explanations highlight the outward manipulation of the electoral process, which Simpser (2013, p. 5) identifies as the "indirect effects" of electoral manipulation. Examples of these effects are the discouragement of opposition parties to participate in future elections (Simpser, 2005; Donno and Roussias, 2012), the incitement of conflict within the opposition against the incumbent (Magaloni, 2006, 2007), and the impression of dominant authority by the incumbent party (Little, 2011; Kalinin and Mebane, 2011). When these benefits are available, the perpetrators' priority is not the electoral result per se, but rather, what different types of manipulation convey to voters and the opposition. For example, after the 1975 presidential election in Tanzania, President Julius Nyerere was worried about the 6 percent of votes cast against him and the implication of these opposing votes as a possible roadblock to his mobilization strategy (Martin, 1978). In contrast to outcome-driven fraud theories, these theories generally challenge the argument of the hidden nature of electoral manipulation: the intended effects of the illicit practices materialize only after the perpetrators make the fraud known (Simpser, 2005; Magaloni, 2006).

What we deduce from this debate is that the existence of fraud depends upon the expected benefits and costs of an election. I argue that the prevalence of a particular type of fraud is a function of the degree of competitiveness in a political system, which Simpser (2013) defines as the distribution of "resources, political power, and institutional advantages" among the main parties (p. 19). The level of competition in the political system will determine not only the degree of corruption available to the incumbent and the opposition parties but also the degree to which a party can neutralize its rival's effort.

Therefore, the degree of competitiveness helps us to understand when and how corruption is likely to appear. Moreover, it suggests dividing the empirical cases into two groups. On the one hand, there are those cases where the resources and the decision to engage in manipulation are exclusive to the incumbent party, providing opportunities for blatant and non-outcome driven fraud. Examples of this type of corruption are in Argentina during the 1930s (Alston and Gallo, 2010; Ciria, 1974), Chile until 1960 (Drake, 2009; Baland and Robinson, 2008), Mexico until the late 1980s (Magaloni, 2006; Domínguez and McCann, 1996), and contemporary Russia

(Myagkov, Ordeshook and Shakin, 2009). On the other hand, there are those cases where the resources and opportunity for fraud are available to more than one group, thus increasing the possibility that the perpetrators will be exposed. This type of corruption can be found in the historical cases of political development in Germany (Mares and Zhu, 2012), Costa Rica (Lehoucq and Molina, 2002), and the U.S. (Cox and Kousser, 1981).

With a few exceptions (Fearon, 2011; Magaloni, 2007), contemporary studies of electoral manipulation in competitive environments focus on the empirical identification of these events.Gonzalez-Ocantos et al. (2012) show evidence of vote buying in Nicaragua by two rival political machines, and Hidalgo (2011) shows how local political machines in Brazil affected electoral results via the use of spoiled votes. Finally, Fukumoro and Horiuchi (2011) show another way of influencing the election using the mobilization of voters across districts in Japan. Unlike analyses for elections in which only the incumbent party is willing to commit fraud (Myagkov, Ordeshook and Shakin, 2009; Mebane, 2009; Hausmann and Rigobón, 2011), the studies above demonstrate how fraudulent opportunities are available to more than one contender in an election; further, they allow us to see the creativity on the part of the perpetrators in keeping their activities hidden from view.

The model discussed below builds on two types of analyses. First, there is a recent wave of formal models that study the decision of politicians to engage in electoral fraud (Simpser, 2005, 2012*b*; Hyde, 2011; Little, 2011; Cox, 2009; Fearon, 2011; Gandhi and Przeworski, 2009). Most of these works consider electoral corruption as an exclusive decision of the incumbent party. My model contributes to this literature, incorporating the consequences of fraud within a competitive environment. Second, this model also builds on the theoretical work concerning resource allocation in electoral campaigns (Snyder, 1989; Tofias, 2007). Given that the nature of elections is to institutionalize the uncertainty of results (Przeworski, 1991), I explore how unknown electoral results affect the decision of parties to engage in electoral corruption.

2.2 Model

To explore how the uncertainty of electoral results and the resources that politicians have at their disposal affect the electoral process, I borrow the preliminary settings of the probabilistic voting model framework (Persson and Tabellini, 2000; Besley, 2006). I first describe the settings of the core model, solve the game by backward induction, and then compare the predictions of the model given the number of precincts in the election.

2.2.1 Settings

Consider a gubernatorial election with two parties competing, $i \in \{1, 2\}$, and the winner is chosen using a simple plurality rule. I label the incumbent party as Party 1 and the challenger as Party 2. The votes each party gets are normalized around 0 and over the range $V_{i\in\{1,2\}} \in [-\frac{1}{2}, \frac{1}{2}]$. A positive value for V_1 means that

Party i has at least one half of the votes, while a negative value corresponds to those cases in which the proportion of votes for the party is less than one half. For simplicity, I assume that the number of voters is odd, which rules out the possibility of a tie in the election.

The votes that each party gets are determined by two different groups of voters. First, there is a fraction $\omega \in (0,1)$ of partisan voters. Within this group, a proportion $\frac{1}{2}+\eta$ will vote for Party 1 while the reminder $\frac{1}{2}-\eta$ will do so for Party 2; therefore, $\eta \in [-\frac{1}{2}, \frac{1}{2}]$. Using a similar definition of 'core voters' discussed by Cox and McCubbins (1986), the electoral preferences of this group are commonly known to the parties and not expected to change in the short term, so parties avoid persuasion to this group on the election day (Dixit and Londregan, 1996).

The second group includes the fraction $(1-\omega)$ of non-partisan voters, whose preferences are revealed at the end of the election as an idiosyncratic shock (δ) with an expected value $E(\delta)=0$ over the interval $\delta \in \left[-\frac{1}{2\xi}, \frac{1}{2\xi}\right]$, where $\xi \in [1,\infty)$ represents the stability of the preferences of the non-partisan voters. That is, a highly predictable electorate corresponds to larger values of ξ, narrowing the interval in which the exogenous shock δ varies. As the value of ξ gets larger, the preferences of the non-partisan voters resemble those from their partisan counterpart; otherwise, their preferences are highly volatile and easier to manipulate through corruption.

Given these settings, Party 1 gets the majority of votes if the following holds:

$$V_1 : \omega\eta + (1-\omega)\delta > 0 \tag{2.1}$$

Otherwise, Party 2 wins the election as $V_2 = -V_1 > 0$ or:

$$V_2 = -\omega\eta - (1-\omega)\delta > 0 \tag{2.2}$$

Therefore, the expected probability for each party to win the election, $\pi_{i\in\{1,2\}}$, is determined by the proportion of the segment $[\omega\eta - \frac{1}{2\xi}, \omega\eta + \frac{1}{2\xi}]$ that falls above 0. Formally:

$$\pi_1 = \begin{cases} 0 & \text{if} \quad \omega\eta + \frac{1-w}{2\xi} < 0 \\ \frac{1}{2} + \frac{\xi\omega\eta}{1-\omega} & \text{if} \quad \omega\eta + \frac{1-w}{2\xi} > 0 \\ \\ 1 & \text{if} \quad \omega\eta - \frac{1-w}{2\xi} > 0 \end{cases} \tag{2.3}$$

$$\pi_2 = \begin{cases} 0 & \text{if} \quad \omega\eta - \frac{1-w}{2\xi} > 0 \\ 1 - \frac{1}{2} + \frac{\xi\omega\eta}{1-\omega} & \text{if} \quad \omega\eta + \frac{1-w}{2\xi} > 0 \\ \\ 1 & \text{if} \quad \omega\eta - \frac{1-w}{2\xi} < 0 \end{cases} \tag{2.4}$$

Electoral Corruption

The probabilities above can be affected when at least one of the parties engages in electoral corruption. In that case, each party may exert an effort level (e_i) to illegally earn the fraction of votes $(1-\omega)\frac{1}{2\xi}$.[1] This effort represents the resources

[1] I assume that a party can illegally get that fraction of votes when its effort is positive, increasing their effort with a decreasing rate, so $e_i(0) = 0$, $e_i^{'} > 0$, and $e_i^{''} < 0$. The results from a model with convex technology will hold as long as the rate of increasing returns is smaller than 2 (Szidarovszky

that the party spends on the ground on election day to illegally distort the results. These manipulations include voter intimidation, vote buying, ballot stuffing, or extortion of poll workers (Schedler, 2002*c*). If a party engages in corrupt activities, its likelihood of success in affecting the electoral result depends on the party's own effort (e_i) relatively to the effort made by the rival party (e_{-i}) as well as the capability of the electoral administration to prevent any kind of manipulation in the district (S). Therefore, the likelihood that a party successfully disrupts the electoral process is:

$$p_{i\in\{1,2\}} = \frac{e_i}{e_i + e_{-i} + S} \tag{2.5}$$

Equation (2.5) modifies the Tullock (1980) rent-seeking model to include the term S, making it possible that neither party could successfully modify the vote count in the district through illegal means (Dasgupta and Nti, 1998; Jia, Skaperdas and Vaidya, Forthcoming). This term represents the resources of the election administration to prevent any type of manipulation. Therefore, the chances of electoral corruption for either party decrease with a more resourceful electoral administration, and I assume S as a fixed parameter before the election day. By construction, the probability that no party disrupts the vote count is given by the following:

$$p_S = 1 - p_1 - p_2 = \frac{S}{e_1 + e_2 + S} \tag{2.6}$$

and Okuguchi, 1997).

If a party succeeds in disrupting the electoral process, it gets a fraction $(1-\omega)\frac{1}{2\xi}$ of the votes in the district plus the votes from its core supporters. In a case that neither party succeeds in disrupting the election, the votes from the non-partisan citizens are distributed between the parties after the realization of δ. Therefore, if parties decide to exert an effort level $e_{i\in\{1,2\}} > 0$, their expected votes are:

$$E(V_1) = \omega\eta + \frac{1-\omega}{2\xi}(p_1 - p_2) \tag{2.7}$$

and

$$E(V_2) = -\omega\eta + \frac{1-\omega}{2\xi}(p_2 - p_1) \tag{2.8}$$

By engaging in corrupt activities, a party incurs a cost $c_i(e_i) = \frac{e_i}{t_i}$, where $t_i > 0$ represents the party's available resources.[2] We can think of these resources as the strength of a party's political machine in the district (Stokes, 2005; Langston, 2012), the access to additional resources to mobilize voters (Calvo and Murillo, 2004), or the partiality of the prosecutors who adjudicate complaints regarding fraudulent activities (Crespo, 1996; Harberes and Ingram, 2012). Therefore, higher values of t_i correspond to elections where parties have efficient corruption technologies.

In principle, both parties could have the same resources for electoral corruption, $\alpha_H = \frac{t_2}{t_1} = 1$. To incorporate the resource advantage of the incumbent party, I assume that the governor of a state can provide additional public resources to the

[2]In this model t_i is similar to the efficiency technology of the players denoted by Cornes and Hartley (2005).

incumbent party, generating an asymmetry represented by $\alpha_L = \frac{t_2}{t_1} \in (0,1]$. This asymmetry can be interpreted as the degree of accountability or transparency in the state, which I assume as exogenous in the model. The governor can only decide whether to provide additional resources to her party; she cannot interfere, however, with the decision of the parties to engage in corruption.

If the governor provides additional resources to Party 1, she incurs in a legitimacy cost, which inversely decreases with α.[3] On the other hand, if Party 1 wins the election, the governor gets a payoff of R, while Party 1 gets a payoff of K and Party 2 gets nothing; otherwise, Party 2 receives the payoff of K, while the governor and Party 1 get nothing. Therefore, the utility functions for parties $i \in \{1,2\}$ and the governor (G) are:

$$U_{i\in\{1,2\}} = \begin{cases} R - e_i & \text{if} \quad V_i > 0 \\ -e_i & \text{if} \quad V_i < 0 \end{cases}$$

$$U_G = \begin{cases} K\alpha & \text{if} \quad V_1 > 0 \\ 0 & \text{if} \quad V_2 < 0 \end{cases}$$

2.2.2 Analysis

The sequence of this game is as follows:

[3]The governor can get different rewards for surpassing distinct electoral thresholds $\tau(V_1)$. She can get an utility U for winning the election, $\tau(V_1) = 0$ and an utility $U^{'} > U$ for reaching a specific threshold $\tau^{'}(V_1) > 0$. The possibility for τ to take different values allows one to consider what Simpser (2005) illustrates for non-competitive elections, where the goal for the incumbent during elections goes beyond the plurality of votes, pushing the incumbent to cheat even when they are certain to win the election (Magaloni, 2006; Kalinin and Mebane, 2011). For simplicity, I only explore in this paper the case in which the governor is only concerned about the electoral victory.

1. As election day approaches, the asymmetry in the resources (α_L), the fraction of partisan voters (ω), the distribution of the partisan votes between the parties (η), and the resources of the electoral board to prevent fraud (S) are common knowledge to both the parties and the governor.

2. The governor decides whether or not to interfere in the election. If there is no interference, $\alpha_H = 1$; otherwise, $\alpha_L = \frac{t_2}{t_1} < 1$.

3. Parties observe their technology and estimate their probabilities of winning the election when they engage in corrupt activities.

4. If parties engage in electoral corruption, they deploy an effort level ($e_{i\in\{1,2\}}$) that maximizes the ability to the vote count.

5. Once the election is over, the idiosyncratic shock (δ) is revealed, and the winner of the election is announced.

I will first analyze the behavior of the parties when they maximize their probability of winning the election goal in a unique precinct. Then I extend the strategies of the parties when the election is defined in multiple precincts.

Maximization of the probability of winning

Given that the parties are only concerned with winning the election, there is an uncertainty of the electoral outcome only when $0 < \pi_{i\in\{1,2\}} < 1$, or when the interval of the expected number of votes includes 0. In that case, each party will achieve the amount of electoral corruption that increases its possibility of winning the election at the lowest cost. Under electoral corruption, the expected probabilities of an electoral victory to the parties change from equations (2.3) and (2.4) to:

$$\pi_1 = p_1 + (1 - p_1 - p_2)[\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega}] \tag{2.9}$$

and

$$\pi_2 = p_2 + (1 - p_1 - p_2)[\frac{1}{2} - \frac{\xi\omega\eta}{1-\omega}] \tag{2.10}$$

Therefore, the expected utilities for the parties are:

$$EU_1(e_1, e_2) = \pi_1(R - \frac{e_1}{t_1}) + \pi_2(-\frac{e_1}{t_1}) \tag{2.11}$$

$$EU_2(e_1, e_2) = \pi_2(R - \frac{e_2}{t_2}) + \pi_1(-\frac{e_2}{t_2}) \tag{2.12}$$

Using Equation (2.5), the expected utilities can be rewritten as:

$$EU_1(e_1, e_2) = [p_1 + p_S(\frac{1}{2} + \frac{\xi\eta\omega}{1-\omega})](R - \frac{e_1}{t_1}) + (1 - [p_1 + p_S(\frac{1}{2} + \frac{\xi\eta\omega}{1-\omega})])(-\frac{e_1}{t_1}) \tag{2.13}$$

$$EU_2(e_1,e_2) = [p_2+p_S(\frac{1}{2}-\frac{\xi\eta\omega}{1-\omega})](R-\frac{e_2}{t_2})+(1-[p_2+p_S(\frac{1}{2}-\frac{\xi\eta\omega}{1-\omega})])(-\frac{e_2}{t_2}) \quad (2.14)$$

After making explicit the effort levels of each party on the probabilities to manipulate the election, the first order conditions for an interior Nash equilibrium are:

$$\frac{\partial EU_1}{\partial e_1} = R\Big[\frac{e_2+S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})}{(e_1+e_2+S)^2}\Big]-\frac{1}{t_1} \quad (2.15)$$

$$\frac{\partial EU_2}{\partial e_2} = R\Big[\frac{e_1+S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})}{(e_1+e_2+S)^2}\Big]-\frac{1}{t_2} \quad (2.16)$$

Equations (2.15) and (2.16) show that the level of electoral corruption that the party exerts not only depends not only on their resources but also on the preferences and volatility of the electorate.[4] Rewriting the first order conditions for the maximization problem leads to:

$$t_1[\frac{e_2+S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})}{(e_1+e_2+S)^2}] = 1 \quad (2.19)$$

[4]Rearranging the first order conditions, the best response functions for the parties are:

$$\hat{e}_1(e_2) = max\{0, \sqrt{Rt_1[e_2+S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})]}-e_2-S\} \quad (2.17)$$

$$\hat{e}_2(e_1) = max\{0, \sqrt{Rt_2[e_1+S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})]}-e_1-S\} \quad (2.18)$$

and

$$t_2[\frac{e_1+S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})}{(e_1+e_2+S)^2}]=1 \tag{2.20}$$

The intersection of the two conditions above denotes the effort level of each party with respect to their technologies as follows:

$$\frac{e_1+S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})}{t_1}=\frac{e_2+S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})}{t_2} \tag{2.21}$$

Rearranging equation (2.21), $e_2=\frac{t_2}{t_1}[e_1+S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})]-S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})$, which can be replaced in equation (2.19) as the first order condition for Party 1:

$$t_1R\left[\frac{\frac{t_2}{t_1}[e_1+S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})]-S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})+S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})}{e_1+\frac{t_2}{t_1}[e_1+S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})+S}\right]=1 \tag{2.22}$$

This result sets the conditions that affect the equilibrium effort level for the parties as follows:

$$e_1=\frac{Rt_1^2t_2}{(t_1+t_2)^2}-S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega}) \tag{2.23}$$

and

$$e_2=\frac{Rt_1t_2^2}{(t_1+t_2)^2}-S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega}) \tag{2.24}$$

Equations (2.23) and (2.24) denote the equilibrium effort level when both parties engage in corruption. A value e_2 lower than 0 prevents party 2 from exerting a positive effort of electoral manipulation. In response, Party 1 updates its estimations and abstains from corruption if:

$$\frac{S}{Rt_1} < \frac{\xi\omega\eta}{1-\omega} + \frac{1}{2} \tag{2.25}$$

Observe how the left side of equality (2.25) increases with S, reducing the likelihood for Party 1 to engage in electoral corruption. This condition is equivalent for Party 2 when $e_1 < 0$. Therefore, the equilibrium effort levels for the parties are mapped as follows:

$$e_1^*, e_2^* = \begin{cases} e_1^* = \frac{Rt_1^2 t_2}{(t_1+t_2)^2} - S(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega}) & \text{and} \quad e_2^* = \frac{Rt_1 t_2^2}{(t_1+t_2)^2} - S(\frac{1}{2} - \frac{\xi\omega\eta}{1-\omega}) \quad \text{if} \quad \frac{Rt_1 t_2^2}{(t_1+t_2)^2} > S(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega}) \\ & \text{and} \quad \frac{Rt_1^2 t_2}{(t_1+t_2)^2} > S(\frac{1}{2} - \frac{\xi\omega\eta}{1-\omega}) \\ e_1^* = \sqrt{Rt_1 S(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega})} - S & \text{and} \quad e_2^* = 0 \quad \text{if} \quad \frac{1}{2} + \frac{\xi\omega\eta}{>} \frac{S}{Rt_1} \\ & \text{and} \quad \frac{Rt_1^2 t_2}{(t_1+t_2)^2} < S(\frac{1}{2} - \frac{\xi\omega\eta}{1-\omega}) \\ e_1^* = 0 & \text{and} \quad e_2^* = \sqrt{Rt_1 S(\frac{1}{2} - \frac{\xi\omega\eta}{1-\omega})} - S \quad \text{if} \quad \frac{Rt_1 t_2^2}{(t_1+t_2)^2} < S(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega}) \\ & \text{and} \quad \frac{1}{2} - \frac{\xi\omega\eta}{>} \frac{S}{Rt_2} \\ e_1^* = 0 & \text{and} \quad e_1^* = 0 \quad \text{if} \quad \frac{1}{2} + \frac{\xi\omega\eta}{<} \frac{S}{Rt_1} \\ & \text{and} \quad \frac{1}{2} - \frac{\xi\omega\eta}{<} \frac{S}{Rt_2} \end{cases}$$

In this case, the proportion and distribution of partisan voters, as well as the volatility of non-partisan voters, help to determine the decision of a party to engage in electoral corruption. Even when the resources are not evenly distributed, either party can be defeated if the exogenous shock overcomes the votes that the a party would gain through corruption. I use the equilibrium effort levels, e_1^* and e_2^*, to calculate the probabilities for parties to disrupt the electoral process.

$$e_1^*, e_2^* = \begin{cases} p_1^* = 0 & \text{and} \quad p_2^* = 0 & \text{if} \quad e_1^* = 0 \\ & \text{and} \quad e_2^* = 0 \\ p_1^* = \frac{t_1}{t_1+t_2} - \frac{S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})(t_1+t_2)}{Rt_1t_2} & \text{and} \quad p_2^* = \frac{t_1}{t_1+t_2} - \frac{S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})(t_1+t_2)}{Rt_1t_2} & \text{if} \quad e_1^* > 0 \\ & \text{and} \quad e_2^* > 0 \\ p_1^* = \frac{\sqrt{Rt_1S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})}-S}{\sqrt{Rt_1S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})}} & \text{and} \quad p_2^* = 0 & \text{if} \quad e_1^* > 0 \\ & \text{and} \quad e_2^* = 0 \\ p_1^* = 0 & \text{and} \quad p_2^* = \frac{\sqrt{Rt_2S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})}-S}{\sqrt{Rt_1S(\frac{1}{2}-\frac{\xi\omega\eta}{1-\omega})}} & \text{if} \quad e_1^* = 0 \\ & \text{and} \quad e_2^* < 0 \end{cases}$$

In this case, parties with an electoral advantage will invest less effort to disrupt the electoral process, while the highest levels of electoral corruption come from those parties with enough resources but an electoral disadvantage among partisan voters.

Turning to the stage in which the governor decides whether or not to interfere in an election, she estimates the winning probabilities for Party 1 when there are equal resources for both parties ($\alpha_H = 1$) and when Party 1 has additional resources to create an asymmetry ($\alpha_L \in (0,1)$). She chooses a resource asymmetry in the election only if it gives her a higher expected utility. In this case, the governor has incentives to interfere in the election when the change in the expected probability to disrupt in the election overshadows the legitimacy cost that she will pay for if she does interfere, or $K[\pi_1|\alpha_H] > \alpha_L K[\pi_1|t_1 = \frac{t_2}{\alpha_L})]$. The next proposition summarizes the conditions in where the governor does interferes in the election. All proofs are in the Appendix.

Proposition 1. *Consider a plurality election in a unique precinct. The governor will provide additional resources to her party and create an asymmetry ($\alpha_L < 1$) only when the change in the probability that her party wins the election is greater than the legitimacy cost of interfering in the process, or:*

$$\frac{p_{1|\alpha_H} + p_{S|\alpha_H}(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega})}{p_{1|\alpha_L} + p_{S|\alpha_L}(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega})} < \alpha_L$$

Proposition 1 shows the role of the electoral administration to prevent irregularities in an election. When the resources of the electoral administration to prevent corruption are scarce, the rate of probabilities for the parties to disrupt the election will be higher. In contrast, when the election administration is strong enough to prevent large-scale fraud, the role of the non-partisan voters determines the electoral outcome, thus encouraging the parties to make a positive effort on their behalf in order to reduce the uncertainty of the outcome.

Allocation of resources in different precincts

I now extend the model to a case in which votes are distributed across different precincts (d), $d \in \{1, ..., D\}$. Each precinct contributes a fraction γ_d to the aggregate number of votes in the election, $\sum_{d=1}^{D} \gamma_d = 1$. While the resources of the parties and the electoral administration are constant across precincts, each precinct may vary in the proportion and distribution of partisan votes between the parties.

Winning a precinct is irrelevant to the overall picture: instead, parties are concerned about maximizing the fragment of votes in the precinct that contributes to the aggregate number of votes $V_{i\in\{1,2\}} = \sum_{d=1}^{D} \gamma_d v_{i,d}$.[5] Therefore, Party i wins the election if

$$V_i = \sum_{d=1}^{D} \gamma_d [\omega_d \eta_d + (1-\omega_d)\delta] > 0 \tag{2.26}$$

Similar to the analysis above, if there is uncertainty about reaching the electoral goal, parties will try to maximize the probability of reaching the electoral target, and be as efficient as possible with their resources to manipulate the vote counts. Therefore, parties' expected payoffs are given by:

$$EU_1(e_{1,d}, e_{2,d}, S) = R\sum_{d=1}^{D} \gamma_d \left[\frac{1}{2} + \omega_d \eta_d + (1-\omega_d)\frac{1}{2\xi}(p_{1,d} - p_{2,d})\right] - \sum_{d=1}^{D} \frac{e_{1,d}}{t_1} \tag{2.27}$$

$$EU_2(e_{1,d}, e_{2,d}, S) = R\sum_{d=1}^{D} \gamma_d \left[\frac{1}{2} - \omega_d \eta_d + (1-\omega_d)\frac{1}{2\xi}(p_{2,d} - p_{1,d})\right] - \sum_{d=1}^{D} \frac{e_{1,d}}{t_1} \tag{2.28}$$

Where R represents the reward to the winning party. Equations (2.27) and (2.28) can be rewritten to make explicit the effects of the efforts of each party on

[5] This assumption contrasts with other analyses for the resource allocation in elections (Snyder, 1989, 1990; Tofias, 2007)

the probabilities to manipulate the election in every precinct as follows:

$$u_1(e_{1,d}, e_{2,d}, S) = R\sum_{d=1}^{D}\gamma_d\left[\frac{1}{2} + \omega_d\eta_d + (1-\omega_d)\frac{1}{2\xi}\left(\frac{e_{1,d}-e_{2,d}}{e_{1,d}+e_{2,d}+S}\right)\right] - \sum_{d=1}^{D}\frac{e_{1,d}}{t_1} \tag{2.29}$$

$$u_2(e_{1,d}, e_{2,d}, S) = R\sum_{d=1}^{D}\gamma_d\left[\frac{1}{2} - \omega_d\eta_d + (1-\omega_d)\frac{1}{2\xi}\left(\frac{e_{2,d}-e_{1,d}}{e_{1,d}+e_{2,d}+S}\right)\right] - \sum_{d=1}^{D}\frac{e_{2,d}}{t_2} \tag{2.30}$$

One can observe that when a party increases its effort, it produces a positive effect on its expected number of votes, but at the same time, it also has a negative effect by means of the costs involved in that effort. With these settings, the first order conditions for an interior Nash equilibrium are as follows:

$$\frac{\partial u_1}{\partial e_{1,d}} = R\gamma_d\left[(1-\omega_d)\frac{1}{2\xi}\frac{2e_{2,d}+S}{(e_{1,d}+e_{2,d}+S)^2}\right] - \frac{1}{t_1} \tag{2.31}$$

$$\frac{\partial u_2}{\partial e_{2,d}} = R\gamma_d\left[(1-\omega_d)\frac{1}{2\xi}\frac{2e_{1,d}+S}{(e_{1,d}+e_{2,d}+S)^2}\right] - \frac{1}{t_2} \tag{2.32}$$

Equations (2.31) and (2.32) show that electoral corruption is more profitable in elections with a volatile electorate (low ξ). That is, when the electorate is highly predictable, the scope of votes available to be manipulated is small, reducing the utility of fraud. In contrast, when the preferences of the non-partisan voters are unclear, the manipulation or coercion of voters' choices becomes a more attractive alternative. To make the notation easier, I substitute $R\gamma_d(1-\omega_d)\frac{1}{2\xi}$ for the term W_d, which represents the marginal utility for the party to disrupt the vote count in the precinct. Different values of W_d correspond to differences in the number of votes in the precinct in relation to the aggregate number of votes, γ_d, and the proportion of non-partisan voters, $(1-\omega_d)$.[6]

Rearranging the first order conditions and introducing the term W_d the best response function for either party is:

$$\hat{e}_{i,d}(e_{-i,d}) = max\{0, \sqrt{W_d t_{i,d}(2e_{-i,d} + S)} - e_{-i,d} - S\} \tag{2.33}$$

Figure 2.1 shows the reaction function given by equation (2.33) when $S = 3$, $R\gamma_d = 10$, and $\xi = 1$. The blue and red lines show the reaction functions for each of the parties when the efficiency technologies are evenly distributed between the parties, $t_1 = t_2 = 1$. The pink line shows the reaction function for Party 1 for an efficiency technology of $t_1 = 0.75$. As the functions are not monotonic, any party could have either an increasing or decreasing reaction to the opponent's effort. Also, as the difference between the red and the pink lines shows, a decrease in the efficiency technologies reduces the efficient frontier for the player's function.

Given that the effort level cannot be negative ($e_{1,d} \geq 0$) and the amount

[6]Although standard contest functions introduce asymmetries of the players by differentiating the valuation of the prize (see Konrad (2009), for example), there are no reasons to consider different values for U between the players in this model.

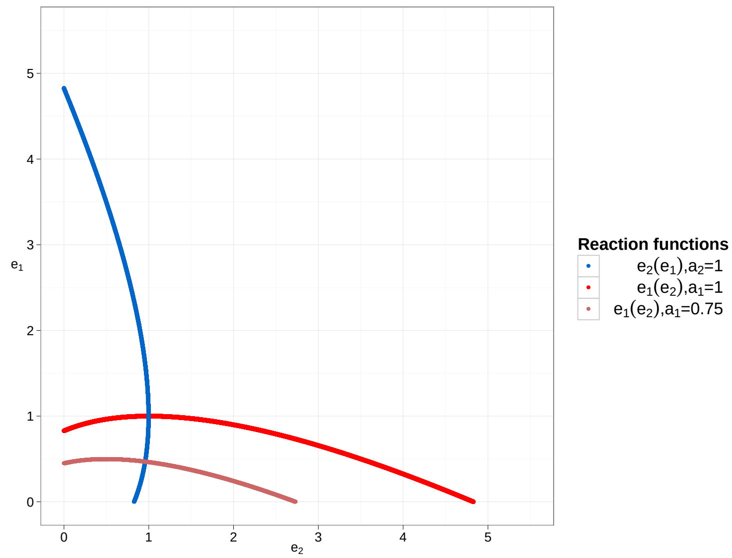


Figure 2.1: Reaction functions for $\hat{e}_{1,d}(e_{2,d})$ and $\hat{e}_{2,d}(e_{1,d})$ when $S = 3$, $R\gamma_d = 10$, and $\xi = 1$.

of resources for the election officials to be always positive ($S > 0$), the functions of the parties intersect only once. Also, when both parties are equally efficient in their use of their resources, $t_1 = t_2$, the effort level made by the parties will be the same $e_1 = e_2$, reaching the maximum level of aggregated effort given both reaction functions.

I summarize these results in the following comment:

Comment 1. *Let the equation* (2.33) *be the best response function for parties* 1 *and* 2. *If both parties have the same efficiency technology, they will reach the same effort level, which is the maximum level of aggregate effort given both reaction functions.*

Rewriting the first order conditions (2.31) and (2.32) for the maximization problem leads to:

$$\frac{2e_{2,d}+1}{e_{1,d}+e_{2,d}+S}W_d t_1 = 1 \tag{2.34}$$

and

$$\frac{2e_{1,d}+1}{e_{1,d}+e_{2,d}+S}W_d t_2 = 1 \tag{2.35}$$

The intersection of the two conditions above denotes the effort level of each party with respect to their efficiency technologies as follows:

$$\frac{2e_{1,d}+S}{t_1} = \frac{2e_{2,d}+S}{t_2} \tag{2.36}$$

Therefore, the player with greater efficiency in the use of her resources ($t_1 > t_2$) spends more effort in equilibrium, $e_{1,d} > e_{2,d}$.[7] Rearranging equation (2.36), $e_{2,d} = (2e_{1,d}+S)\frac{t_2}{2t_1} - \frac{S}{2}$, which can be replaced in equation (2.34) as the first order condition for Party 1:

$$\frac{2\left[(2e_{1,d}+S)\frac{t_2}{2t_1} - \frac{S}{2}\right]+1}{\left(e_{i,S} + \left[(2e_{1,d}+S)\frac{t_2}{2t_1} - \frac{S}{2}\right] + S\right)^2}W_d t_1 = 1 \tag{2.37}$$

Simplifying the equation above leads to:

$$\frac{W t_2(S+2e_{1,d})}{\left(\frac{(t_1+t_2)(2e_{1,d}+1)}{2t_1}\right)^2} = 1 \tag{2.38}$$

[7] This observation follows what Nti (1999) and Konrad (2009) find for games with unequal prize valuations.

This result sets the conditions to the equilibrium value for Party 1, $\tilde{e}_{1,d}$, as follows:

$$e_{1,d} = \frac{2W_d t_1^2 t_2}{(t_1 + t_2)^2} - \frac{S}{2} \tag{2.39}$$

To obtain the equilibrium effort level for Party $-i$, consider again equation (2.36),

$$\frac{2e_{1,d} + S}{t_1} = \frac{2e_{2,d} + S}{t_2} = \frac{4t_1 t_2 W_d}{(t_1 + t_2)^2} \tag{2.40}$$

which leads to,

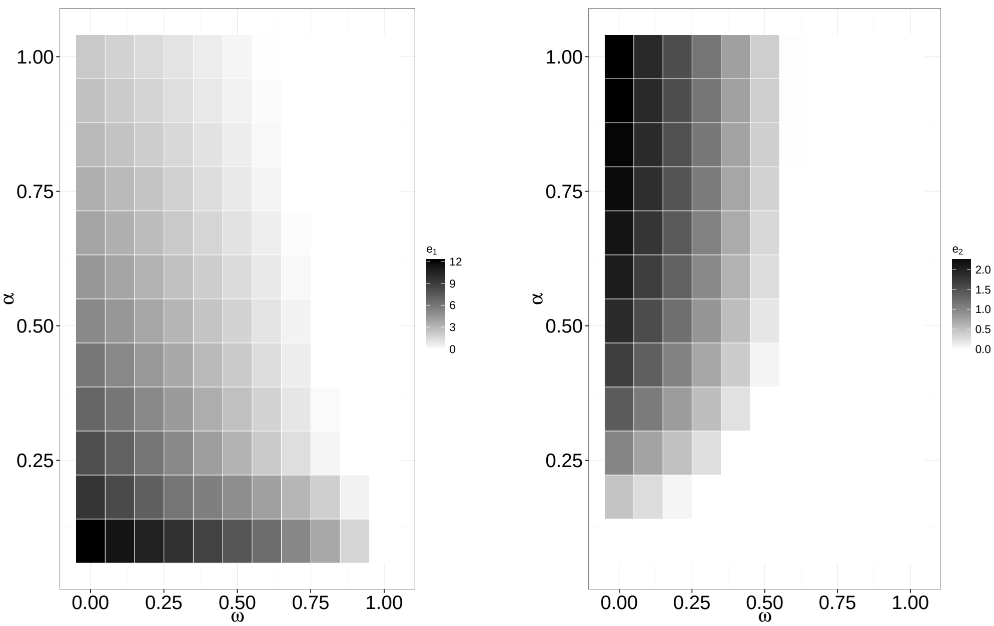
$$e_{2,d} = \frac{2W_d t_1 t_2^2}{(t_1 + t_2)^2} - \frac{S}{2} \tag{2.41}$$

Similar to the previous subsection, in the case that $e_{1,d} < 0$, Party 1 abstains of investing resources in the precinct. In response to this, Party 2 re-estimates its optimal effort level in the precinct setting its reaction function to $\hat{e_1}(e_2 = 0)$. In this case, Party 2 exerts a positive effort only when $\sqrt{W_d t_2 S} - S > 0$ or $W_d > \frac{S}{t_2}$. This result is comparable to the estimation for Party 1 when Party 2 refrains from corruption. Therefore the equilibrium effort levels for the parties in every district are mapped as follows:

$$e^*_{1,d}, e^*_{2,d} = \begin{cases} e^*_{1,d} = \frac{2W_d t_1^2 t_2}{(t_1+t_2)^2} - \frac{S}{2} & \text{and} \quad e^*_{2,d} = \frac{2W_d t_1 t_2^2}{(t_1+t_2)^2} - \frac{S}{2} \quad \text{if} \quad \frac{2W_d t_1^2 t_2}{(t_1+t_2)^2} > \frac{S}{2} \quad \text{and} \quad \frac{2W_d t_1 t_2^2}{(t_1+t_2)^2} > \frac{S}{2} \\ e^*_{1,d} = \sqrt{W_d t_1 S} - S & \text{and} \quad e^*_{2,d} = 0 \quad \text{if} \quad W_d > \frac{S}{t_1} \quad \text{and} \quad \frac{2W_d t_1 t_2^2}{(t_1+t_2)^2} < \frac{S}{2} \\ e^*_{1,d} = 0 & \text{and} \quad e^*_{2,d} = \sqrt{W_d t_2 S} - S \quad \text{if} \quad \frac{2W_d t_1^2 t_2}{(t_1+t_2)^2} < \frac{S}{2} \quad \text{and} \quad W_d > \frac{S}{t_2} \\ e^*_{1,d} = 0 & \text{and} \quad e^*_{1,d} = 0 \quad \text{if} \quad W_d < \frac{S}{t_1} \quad \text{and} \quad W_d < \frac{S}{t_2} \end{cases}$$

Figure 2.2 illustrates the allocation of the resources of the parties and the probability that a polling station remains uncorrupted given the number of non-partisan voters and the distribution of efficiency technologies between the parties. Each cell represents a particular combination of an asymmetry level, $\alpha \in [\frac{1}{10}, 1]$, and the proportion of partisan voters $\omega \in [0, 1]$. The volatility level of the non-partisan voters is fixed at $\xi = 2$, thus variations on the marginal gain to disrupt an electoral district W_d only vary with the proportion of non-partisan voters, $(1 - \omega)$.

The left panel of Figure 2.2 shows the effort exerted by Party 1 as ω varies on the horizontal axis and α varies on the vertical axis. In this case, the effort level of Party 1 is larger at those polling stations where there are more non-partisan voters (low ω), and they have a relative advantage with respect to the efficiency technologies for Party 2 (low α). The resource allocation for Party 2 is reciprocal to the one observed for Party 1. In this case, there is a higher effort at those polling stations with a large number of non-partisan voters (low ω) when its technology is similar to the existent for Party 1 (α is closer to 1). Consequently, the larger the effort level with respect to that which the rival chooses, the higher the probabilities of cheating at the polling station. In sum, the polling stations with the higher probabilities for



(a) Party 1 (b) Party 2

Figure 2.2: Effort made by parties 1 and 2 in an election with different values for the proportion of partisan voters (ω) and the distribution of the technologies between the parties (α) in an plurality election defined in multiple precincts.

Party 1 are in the lower left corner on the grid while those for Party 2 are in the upper left corner.

Having the equilibrium effort levels, e_1^* and e_2^*, I calculate the probabilities that Party i will seek to disrupt the electoral process in precinct d as follows:

$$p^*_{i\in\{1,2\},d} = \begin{cases} 0 & \text{if} \quad e_i^* = 0 \\ \frac{t_i}{t_i+t_{-i}} - \frac{S(t_i+t_{-i})}{4W_d t_i t_{-i}} & \text{if} \quad e_i^* > 0 \quad \text{and} \quad e_{-i}^* > 0 \\ 1 - \frac{S}{\sqrt{W_d t_i S}} & \text{if} \quad e_i^* > 0 \quad \text{and} \quad e_{-i}^* = 0 \end{cases} \tag{2.42}$$

Observe that if both parties have the same technology, $t_1 = t_2$, then the equilibrium effort level will be $e^*_{1,d} = e^*_{2,d}$, leading to the same probabilities to tamper the electoral process in the district, $p^*_{1,d} = p^*_{2,d}$.[8] Equations (2.43) and (2.44) illustrate how the probabilities to disrupt the electoral process change with respect of the resource asymmetries. In this case, I include the term $\alpha = \frac{t_2}{t_1}$ to represent the ratio of the technology of Party 2 with respect to the technology of Party 2.

$$p^*_{1,d} = \frac{1}{1+\alpha_d} - \frac{S(1+\alpha_d)}{4Wt_1\alpha_d} \tag{2.43}$$

$$p^*_{2,d} = \frac{1}{1+\alpha_d} - \frac{S(1+\alpha_d)}{4Wt_2\alpha_d} \tag{2.44}$$

These expression show how the probability of disrupting the electoral process in the district is determined by the distribution of the efficiency technologies between the parties. For the case of $t_1 = t_2$, such that $\alpha_d = 1$. Then the probability for Party i to manipulate the vote process in the district is: $p^*_{1,d}(\alpha_d = 1) = \frac{1}{2} - \frac{S}{2W_d t_1}$. If the election administration is weak, each player has a probability of benefiting from corruption in the district closer to $p_1 = \frac{1}{2}$, and this probability decreases as the ratio between S and $W_d t_1$ becomes larger.

Substituting equilibrium probability (2.42) into the expected utility function (2.11), the expected payoff in equilibrium for either party is:

[8] As an additional observation, as S becomes larger, moving this equilibrium away from those expected probabilities in the standard contest game. For example, suppose that $S = 0$. Then the equilibrium probabilities become $p^*_{1,d} = \frac{t_1}{t_1+t_2}$ and $p^*_{2,d} = \frac{t_2}{t_1+t_2}$.

$$EU_1^* = R\sum_{d=1}^{D}\left[\omega_d\eta_d + (1-\omega_d)\frac{1}{2\xi}\frac{t_1-t_2}{t_1+t_2}\right] - \sum_{d=1}^{D}\left[\frac{2W_d a_{1,d}^2 t_2}{(t_1+t_2)^2} - \frac{S}{2t_1}\right] \quad (2.45)$$

Observe that when the efficiency technologies are the same for both parties ($t_1 = t_2$), the change in the expected votes is 0, and the utility of the parties declines as the level of the efficiency technologies increases. This proposition presents a counterintuitive result. When the technologies are evenly distributed between the parties, the expected number of votes garnered by fraud is the same as when the party is not involved in corrupt activities. Nevertheless, each party has incentives to cheat when the opposite party refrains from corruption. That is, parties might be better off not to engage in fraud on election day, yet no party can trust the other to abstain from deceit.

Proposition 2. *If both parties are equally efficient in their use of resources, the expected utility of the parties for corruption will decrease as the efficiency of their technology increases.*

I have explained the incentives for both parties to manipulate the election. I now turn to the stage in which the governor decides between not interfering in the election ($\alpha_H = 1$) or interfering to create asymmetries in the efficient technologies between the parties ($\alpha_L < 1$). At this point, she estimates the winning probabilities in both situations and chooses the one that generates the higher expected utility. In this case, the governor keeps out of the electoral process if $K[\pi_1|t_1 = t_2)] >$

$\hat{\alpha}K[\pi_1|t_1 = \frac{t_2}{\hat{\alpha}})]$. The next proposition summarizes the result.

Proposition 3. *Consider a plurality rule election across different precincts. The governor will opt to interfere in the election and create an asymmetry ($\alpha_L < 1$) only when:*

$$\frac{\sum_{d=1}^{D} \gamma_d (1-\omega_d) \frac{1}{2\xi} (p_{1,d|\alpha_H} - p_{2,d|\alpha_H})}{\sum_{d=1}^{D} \gamma_d (1-\omega_d) \frac{1}{2\xi} (p_{1,d|\alpha_L} - p_{2,d|\alpha_L})} < \alpha_L$$

Proposition 3 suggests that the governor will intervene in the election only if the expected success for electoral corruption under resource asymmetry is greater than the legitimacy costs of interfering in the election. This change is determined only by the equilibrium effort levels for both parties in every precinct given the two resource distributions $\alpha_H = 1$ and $\alpha_L < 1$.

In sum, in a competitive election, a party exerts a positive effort only when it knows that fraud does not have a decisive effect on the electoral outcome. When the institutional settings to prevent fraud are relatively weak, the probability that a precinct is disrupted will be high, leaving the electoral outcome to be determined by the distribution of the efficiency technologies between the parties rather than by the behavior of the non-partisan voters. This result reduces the uncertainty of the result to the point that parties are no longer willing to make an effort. In contrast, when the election administration is strong enough to prevent large-scale fraud, the role of the non-partisan voters is crucial to determining the electoral outcome, encouraging parties to make a positive effort and reduce the uncertainty of the outcome in its favor. Finally, the governor may provide additional resources to her party only when the expected loss of legitimacy for winning a biased election is low.

2.3 Comparative Statics

This section explores the predictions of the model with different values for the key parameters. In particular, I show specific changes in the effort level for electoral corruption given different levels of resources for the electoral administration, efficiency asymmetries between the parties, and partisan preferences under the two variations of a majority election described in Section 2.

The resources available for the electoral administration affect the expected success of electoral corruption and, consequently, affect the decision of the parties to exert a positive effort. This relationship holds regardless of the asymmetries in the technologies of the two parties. To analyze the effects of electoral administration on deterring corruption, I simulate how the expected probabilities of disruption Party 1 to disrupt the election change with different resource levels for the electoral administration has to prevent corruption. For each setting, I simulate 1,000 elections in which the efficient technology for Party 1 is fixed at $t_1 = 1$ while the efficient technology for Party 2 is determined by the function $t_2 = \alpha t_1$, in which α is a randomly drawn from a uniform distribution $\alpha \sim U[0.5, 1]$.[9]

Figure 2.3 shows the probability distribution for Party 1 to disrupt the election at $S = 1$ and $S = 3$. All else being equal, Party 1's probability of success

[9]The settings of the simulation are for a precinct with a marginal profit of $W = \frac{1-\omega}{2\xi} = 5$.

decreases as the resource level for the electoral administration increases. Also, this comparison illustrates the role of electoral observers and their negative effect on electoral corruption (Hyde, 2011). Regardless of the willingness of a party for electoral corruption, logistical improvements, such as electronic voting or post-election audits, may help prevent irregularities (Alvarez, Atkenson and Hall, 2013).

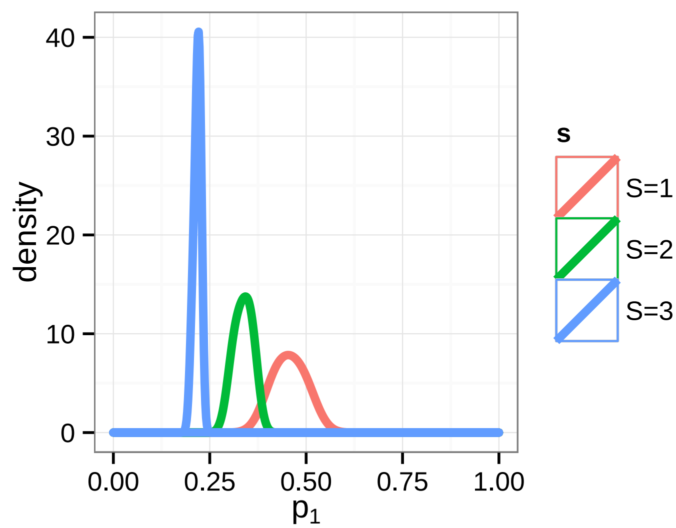


Figure 2.3: Probability densities for Party 1 to disrupt the electoral process given at different resource levels for the electoral administration.

I use Monte Carlo simulations to explore other responses of the parties to different relevant parameters. Tables 2.1 and 2.2 summarize the results for different combinations of $T = t_1 + t_2$, $\alpha = \frac{t_2}{t_1}$, and η, across 1,000 different elections per permutation. The resource level for the electoral administration is fixed at $t_1 = 3$, while every simulation takes a different value for ω and ξ, given $\omega \sim U(0,1)$, and $\xi \sim U(1,4)$. Table 1 shows the equilibrium effort and probabilities for disrupting the election when the goal of the parties is to win the election in a unique precinct. In contrast, Table 2 explores the equilibrium values for the parties when the election is defined in more than one precinct. In this case, the settings of the simulations are the same as those for Table 1 with the exception that the election is defined in three different precincts with the same number of votes, therefore $\gamma_d = \frac{1}{3}$.

For each table, Column (1) shows the proportion of simulated events in which Party 1 wins the election. Columns (2) and (3) show the proportion of times that each party exerted a positive effort level. Finally, columns (4) and (5) present the average effort level that parties exert with each permutation of the settings.

The results of the comparative statistics are summarized as follows. First, corruption is more likely to appear in centralized elections. As the gains from disruption at a polling station decrease, parties are less willing to exert a positive effort for corruption. This result explains the divergence in effort in both Tables 1 and 2, in which the only difference in the settings is whether the distribution of votes is found in one or three precincts. Since the expected profits for disrupting a polling station in Table 2 decrease with the fraction of votes at the polling station relative to the total number of votes (γ), a corruption effort is only attractive when their technologies are efficient enough to expect some positive profit. As in the cases of Mexico before the electoral reform of 1996 (Preston and Dillon, 2004) and Venezuela in 2004 (Hausmann and Rigobón, 2011), corruption is more likely to occur when the results can be manipulated in a centralized way.

Second, centralized corruption is more likely to appear in close elections.

As expected, fraud becomes an attractive enterprise when it could help define the outcome of a particular contest. Even when parties have the majority of partisan votes, if that advantage does not solidify a victory, parties could exert a positive effort level not only to get the votes from the non-partisan voters but also to prevent the other party's success in disrupting the election. As Table 1 shows, the elections with the highest effort level from the parties are those in which η is close to 0, or those in which the partisan votes are evenly distributed between the parties. This result does not rule out corruption in elections with a wide electoral (dis)advantage, but it considers these events as special cases.

In the case of decentralized elections, the distribution of the partisan voters' preferences has only a marginal effect on the estimations that parties make for their optimal effort level. Rather, it is the absolute and relative levels of efficiency between the parties that affect their decisions in a critical way. Given that parties want to maximize their expected number of votes at every polling station, the distribution of partisan votes is irrelevant to their estimations. As Table 2 shows, the average effort each party exerts is quite similar across the different distributions of partisan votes (η). Moreover, as the efficient technology levels differ, we observe the largest changes. Therefore, dissuading corruption when votes are distributed in different precincts requires structural modifications such as a more resourceful electoral administration (S) or an even distribution of efficient technologies (α).

Finally, regardless of the effort level that each party exerts, a centralized election creates a setting where both parties could engage in electoral corruption.

Consider columns (2) and (3) on Table 1, where the frequency of corruption that both parties engage in is found identical. In contrast, a decentralized election allows for different settings in which the parties could potentially engage in electoral corruption. It is only the case when both parties have the same resources, $\alpha_H = 1$, that parties match each other's probabilities of having a positive effort.

2.4 Conclusion

This paper considers the competition level between the incumbent and rival parties to estimate not only the amount of resources each party will invest but also the allocation and success of those resources to disrupt an election. In contrast with most of the extant analyses, where electoral corruption is a comprehensive category that includes many types of manipulation, the dynamics of political competition and the quality of the electoral administration help to explain the variety of contexts in which electoral corruption appears in non-consolidated democracies.

In addition to the consequences of the different resource levels for both the electoral administration and the various parties, it is the symmetry of resources and the institutional settings of the election that affect different parties' decisions to engage in electoral corruption. Exploring different frameworks enables us to better understand the motivations and observable behavior of political machines. A natural extension of this research is to empirically test the prediction of the models and compare the results with the comparative statics. The model can be empirically

Table 2.1: Monte Carlo Simulations for electoral corruption when the election is defined in a unique polling station.

$t_1 + t_2$	$\alpha = \frac{t_1}{t_2}$	η	(1) Proportion of victories, $P(V_1 > 0)$	(2) Party 1's average effort	(3) Party 2's average effort	(4) Frequency of $e_1 > 0$	(5) Frequency of $e_2 > 0$
2	0.5	-0.25	0.110	0.304	0.000	0.286	0.000
		0	0.646	1.000	0.000	1.472	0.000
		0.25	0.952	0.455	0.441	0.406	0.252
	0.75	-0.25	0.118	0.296	0.260	0.410	0.088
		0	0.638	1.000	1.000	1.299	0.599
		0.25	0.948	0.425	0.425	0.337	0.470
	1	-0.25	0.138	0.372	0.372	0.520	0.224
		0	0.591	1.000	1.000	1.000	1.000
		0.25	0.924	0.365	0.365	0.212	0.518
3	0.5	-0.25	0.174	0.376	0.270	0.930	0.113
		0	0.721	1.000	1.000	2.944	0.722
		0.25	0.949	0.460	0.460	1.090	0.597
	0.75	-0.25	0.221	0.468	0.468	1.537	0.497
		0	0.650	1.000	1.000	2.698	1.649
		0.25	0.936	0.467	0.467	0.988	1.042
	1	-0.25	0.201	0.463	0.463	1.306	0.777
		0	0.637	1.000	1.000	2.250	2.250
		0.25	0.936	0.428	0.428	0.712	1.214
4	0.5	-0.25	0.262	0.438	0.438	2.168	0.411
		0	0.728	1.000	1.000	4.426	1.463
		0.25	0.965	0.461	0.461	1.755	0.960
	0.75	-0.25	0.220	0.446	0.446	2.086	0.945
		0	0.681	1.000	1.000	4.098	2.698
		0.25	0.944	0.452	0.452	1.597	1.474
	1	-0.25	0.208	0.468	0.468	1.929	1.347
		0	0.655	1.000	1.000	3.500	3.500
		0.25	0.930	0.463	0.463	1.351	1.890
5	0.5	-0.25	0.307	0.465	0.465	3.023	0.749
		0	0.749	1.000	1.000	5.907	2.204
		0.25	0.946	0.475	0.475	2.540	1.313
	0.75	-0.25	0.244	0.456	0.456	2.785	1.430
		0	0.716	1.000	1.000	5.497	3.748
		0.25	0.964	0.456	0.456	2.226	1.989
	1	-0.25	0.199	0.432	0.432	2.291	1.813
		0	0.642	1.000	1.000	4.750	4.750
		0.25	0.938	0.480	0.480	2.010	2.550
6	0.5	-0.25	0.311	0.472	0.472	3.756	1.121
		0	0.759	1.000	1.000	7.389	2.944
		0.25	0.963	0.489	0.489	3.325	1.728
	0.75	-0.25	0.261	0.446	0.446	3.337	1.878
		0	0.690	1.000	1.000	6.897	4.797
		0.25	0.947	0.456	0.456	2.871	2.462
	1	-0.25	0.206	0.444	0.444	2.923	2.405
		0	0.686	1.000	1.000	6.000	6.000
		0.25	0.931	0.459	0.459	2.483	3.025

tested in two different ways. First, it can compare different levels of electoral corruption in different countries or regions, specifically accounting for the distribution of resources among the parties and the quality of the electoral administration. Sec-

Table 2.2: Monte Carlo Simulations for electoral corruption when the election is defined in three polling stations.

t_1+t_2	$\alpha=\frac{t_1}{t_2}$	η	(1) Proportion of victories, $P(V_1>0)$	(2) Party 1's average effort	(3) Party 2's average effort	(4) Frequency of $e_1>0$	(5) Frequency of $e_2>0$
2	0.5	-0.25	0.008	0.000	0.000	0.000	0.000
		0	0.505	0.000	0.000	0.000	0.000
		0.25	0.990	0.000	0.000	0.000	0.000
	0.75	-0.25	0.010	0.000	0.000	0.000	0.000
		0	0.490	0.000	0.000	0.000	0.000
		0.25	0.988	0.000	0.000	0.000	0.000
	1	-0.25	0.007	0.000	0.000	0.000	0.000
		0	0.496	0.000	0.000	0.000	0.000
		0.25	0.989	0.000	0.000	0.000	0.000
3	0.5	-0.25	0.008	0.000	0.000	0.000	0.000
		0	0.497	0.000	0.000	0.000	0.000
		0.25	0.990	0.000	0.000	0.000	0.000
	0.75	-0.25	0.015	0.000	0.000	0.000	0.000
		0	0.494	0.000	0.000	0.000	0.000
		0.25	0.983	0.000	0.000	0.000	0.000
	1	-0.25	0.011	0.000	0.000	0.000	0.000
		0	0.505	0.000	0.000	0.000	0.000
		0.25	0.989	0.000	0.000	0.000	0.000
4	0.5	-0.25	0.025	0.121	0.000	0.050	0.000
		0	0.501	0.093	0.000	0.036	0.000
		0.25	0.995	0.110	0.000	0.041	0.000
	0.75	-0.25	0.011	0.084	0.000	0.016	0.000
		0	0.527	0.073	0.000	0.015	0.000
		0.25	0.995	0.071	0.000	0.014	0.000
	1	-0.25	0.014	0.039	0.039	0.003	0.003
		0	0.473	0.047	0.047	0.004	0.004
		0.25	0.986	0.045	0.045	0.003	0.003
5	0.5	-0.25	0.029	0.216	0.000	0.122	0.000
		0	0.533	0.229	0.000	0.131	0.000
		0.25	0.991	0.199	0.000	0.118	0.000
	0.75	-0.25	0.018	0.189	0.055	0.069	0.007
		0	0.520	0.184	0.057	0.069	0.006
		0.25	0.988	0.176	0.055	0.068	0.007
	1	-0.25	0.014	0.159	0.159	0.040	0.040
		0	0.507	0.119	0.119	0.031	0.031
		0.25	0.986	0.131	0.131	0.039	0.039
6	0.5	-0.25	0.053	0.346	0.000	0.255	0.000
		0	0.554	0.343	0.000	0.242	0.000
		0.25	0.994	0.335	0.000	0.253	0.000
	0.75	-0.25	0.030	0.281	0.127	0.144	0.033
		0	0.516	0.291	0.135	0.156	0.035
		0.25	0.993	0.290	0.121	0.149	0.034
	1	-0.25	0.009	0.197	0.197	0.084	0.084
		0	0.484	0.214	0.214	0.087	0.087
		0.25	0.987	0.226	0.226	0.093	0.093

ond, the model can be used to compare the evolution of electoral corruption in a single case over time. As the model predicts, electoral corruption fails not because of a change in the motivations of the parties, but rather, through the redistribution of political resources among the parties.

The normative implications of this research concern ways to prevent electoral corruption. The success of disrupting an election for an incumbent decreases with the presence of electoral observers. This result is supported in the literature on non-competitive elections, when it argues for the relationship between fraud and electoral observers (Hyde, 2011; Little, 2011; Gandhi and Przeworski, 2009). Also, it is the "democratization" of the resources among the political parties that makes fraud inefficient. That is, it is when all parties are able to commit the same level of fraud that electoral corruption becomes a costly activity. The opportunity for more than one party to engage in fraud might temporarily raise the level of electoral corruption, but it also represents the beginning of the end for this mechanism.

Chapter 3

Comparative Electoral Forensics

3.1 Introduction

Since the U.S. presidential election in 2000, the use of statistical methods to evaluate the integrity of elections has become standard practice. These studies are coined "election forensics," and have been used to analyze contemporary elections in Iran (Mebane, 2009; Roukema, 2013), Germany (Breunig and Goerres, 2011), Kenya (Long, Kanyinga and Ndii, 2010), Afghanistan (Worden, 2010), Brazil (Hidalgo, 2011), and the Philippines (Pe Lero, 2010). Most of the statistical studies, however, tend to disregard ancillary evidence to confirm their findings and to ignore the assumptions behind their tests. Therefore, the exclusive use of statistical tools without consideration of the assumptions behind each test provides an incomplete analysis to distinguish electoral manipulation from "normal politics" (Mebane, 2010, p. 2)

This chapter discusses the limitations of the most common forensic tools and highlights the assumptions behind them. My work builds on the existent literature of election forensics that use multiple tests to evaluate the integrity of elections (Mebane, 2009; Mebane and Kalinin, 2009; Levin et al., 2009; Jiménez, 2011; Klimek et al., 2012). In contrast to the existing literature, I provide a survey approach of precisely when and how each forensic technique should be utilized.

This chapter is organized as follows. First, I describe the different forensic tools and examine their limitations and assumptions. Next, I discuss the expectations of fraud and illustrate the usefulness of each tool for the analysis of the 2012 presidential elections in Mexico. Finally, I discuss possible directions for further research.

3.2 Methods

Election forensics employ statistical techniques to detect irregularities in vote counts (Mebane, 2008*b*, p. 162). The popularity of these tools has increased with the availability of electoral data worldwide; however, a lack of understanding of the assumptions behind each statistical technique has led to misguided analyses and often incorrect conclusions. This section discusses the potential advantages and limitations of the most common electoral forensics and identifies the conditions in which each technique provides information of value to cross-validate with other metrics, such as observers' reports or poll workers' surveys (Alvarez, Atkenson and

Hall, 2013).

3.2.1 Digit Analyses

One type of forensic technique, digit analysis, finds irregularities by looking at the digit distribution of vote returns. In the absence of manipulation, particular digits in the vote counts should follow a specific distribution, and divergences between the electoral data and the expected distribution suggest that political manipulation exists. The type of test used to analyze the vote counts depends on the specific order of the digits. Below, I discuss two variations of these tests to check for inconsistencies in the first and last digits of vote returns.

First and Second Digits

First digits analysis of vote counts checks if the first significant digits converge to the logarithmic distribution. This property, known as "Benford's law," predicts that lower first significant digits are more likely to appear than higher significant digits (Fewster, 2009; Ciofalo, 2009). Formally, the probability of a leading digit (d) to appear as first digit is $P(d) = log_{10}(\frac{1+d}{d})$ for $d \in 1, ..., 9$. Although Benford's law does not appear in every random and unbiased distribution, this property holds even with variations in the scale, power, or numerical basis of the data (Fewster, 2009; Ciofalo, 2009; Leemis, Schmeiser and Evans, 2000).[1] In general, first digits follow

[1] Ciofalo (2009) shows that the conditions for a distribution that follows Benford's law are: (1) a distribution with a hyperbolic probability density function $p(B)$ $1/B$; (2) a geometric progression (i.e., a sequence of numbers where each term after the first if found by multiplying the previous one by a fixed non-zero number; and (3) an exponential rank-size distribution $B(k)$ $exp(-k/kn)$. In

Benford's law in large datasets that combine different distributions, and where the mean is greater than the median (Cho and Gaines, 2007; Janvresse and de la Rue, 2004). Mebane (2006, 2007) proposes this method as a forensic tool for elections. He suggests, however, that the second digit of the vote returns be examined. In this case, the relative frequency that the number d appears as second digit follows $q_{B2i} = \sum_{k=0}^{9} log_{10}(1+(10k+d)^{-1})$ (Pericchi and Torres, 2011, p. 505), and the statistic for a second digit Benford's law is: $\chi^2 = \sum_{i=10}^{9} \frac{(d_{2i}-d_2 q_{B2i})^2}{d_2 q_{B2i}}$, where $d_2 = \sum_{i=0}^{9} d_{2i}$. (Mebane, 2006, p. 7).

Since this test does not require any other data other than election returns, the analysis of the second digit became a standard tool to audit elections where fraud was possible. There are Second-Digit analyses of elections where fraud presumably took place during the compilation of results from voting machines, such as Florida in 2000 (Mebane, 2006) or Venezuela in 2004 (Pericchi and Torres, 2011). Similar analyses exist for cases where fraud has been blatant and affected the majority of the polling stations, such as the examples of Iran in 2009 (Mebane, 2009; Roukema, 2013), the U.S. Gilded Age (Buttorf, 2008), and contemporary Russia (Mebane, 2008*a*). Finally, this test has also been used to examine cases where fraud presumably occurred in a decentralized way such as Mexico in 2006 (Mebane, 2007) or Germany between 1990 and 2005 (Breunig and Goerres, 2011).

Despite its extensive use, there is general skepticism about using Benford's

general terms, the distribution must cover several orders of magnitude and is not "artificially" biased in favor of any particular value.

law to audit electoral data. The most common critique is that the alleged irregularities found in many of these analyses are not supported by a solid theoretical framework. First digits' disparities from Benford's law can appear even in the absence of electoral manipulation (Deckert, Myagkov and Ordeshook, 2011). Therefore, this tool cannot distinguish the underlying process that generates these deviant observations (Jiménez, 2011, p. 569-570). Moreover, Deckert, Myagkov and Ordeshook (2011) evaluate the accuracy of this test in both clean and manipulated elections, showing the poor capacity of this technique to classify the quality of elections. In sum, First and Second digits analyses are vulnerable to both false positives and false negatives, without providing enough information to distinguish those events. The usefulness of this test, therefore, is very limited.[2]

Last Digit

Based on experimental evidence, the Last Digit test expects that non-manipulated vote counts resemble a uniform distribution in their last digit (Beber and Scacco, 2012, p. 218-220). If the voting records deviate from this pattern, it raises a red flag of potential manipulation of the vote returns. The test that the last digit d_L occurs frequently is: $\chi^2_L = sum^9_{i=0} \frac{(d_{Li} - d_L/10)^2}{d_L/10}$, where $d_L = \sum_{i=0}^{9} d_{Li}$.

The Last Digit test works with electoral data, which has two very general

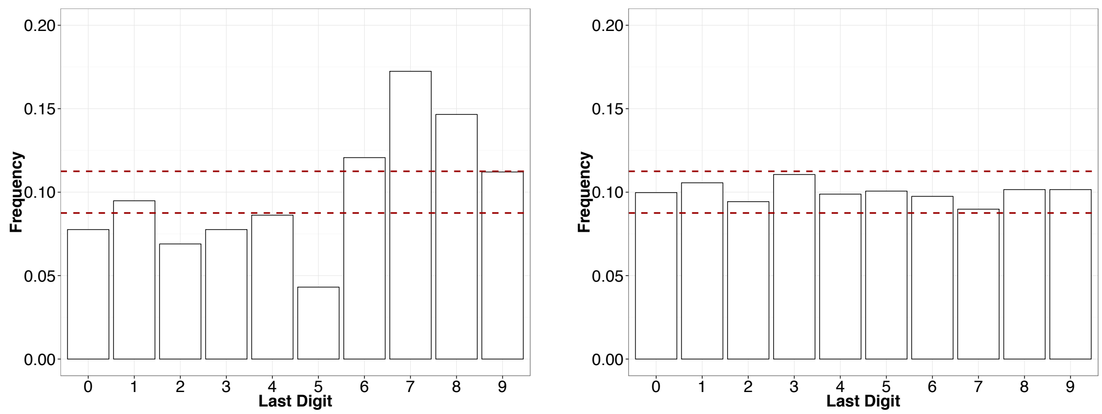
[2]As a way to overcome the limitations of this test, Cantú and Saiegh (2011) check the first digit distribution for the Argentinian elections in the 1930s preprocessing a collection of simulated data that resemble the particular characteristics of the election in the analysis. This refinement, however, restricts the set of elections to analyze with this tool only to the instances in which there exist information about a clean election in order to calibrate the model.

conditions (Beber and Scacco, 2012). First, vote counts should vary within a large range of numbers. This condition rules out an analysis of elections in which the number of registered voters or votes cast for a candidate is very similar across polling stations. Second, when counting the votes, the mean should be relatively higher than the standard deviation. Consequently, this condition excludes this test from vote returns for a particular candidate or party with a small number of votes at most polling stations.

This forensic technique has been used to conduct post-election audits in cases where manipulation is allegedly carried out during the tabulation of votes. Beber and Scacco (2012, p. 223) apply this tool to analyze the 2003 elections in Nigeria, where anecdotal evidence suggests that vote tallies were manipulated in the state collation centers. The authors also find evidence of electoral manipulation for the 2007 election in Senegal. Levin et al. (2009) use a similar test for the 2006 Venezuelan presidential election, as well as for the 2007 and 2009 constitutional referenda in that country. The authors do not find significant deviations of the vote counts' last-digits from the expected distributions in any of the cases. Myagkov and Ordeshook (2010) use a similar analysis for the 2010 elections in Ukraine. The authors do not find evidence of fraud in this case.

On the other hand, some analyses of the 2006 presidential election in Mexico have used this technique and found an irregular pattern of last digits for the National Action Party's (PAN) vote returns in the state of Guerrero (Mochán, 2011; Gallardo, 2009, 2012). However, a close analysis of these reports suggests otherwise. First, the replication of this test for each candidate in every state indicates that the uneven findings for that PAN in Guerrero may have been due to chance. Second, the authors' conclusions omit the fact the PAN had a poor performance in the southern part of the country, particularly in the state of Guerrero. For the 2006 election in particular, the average number of votes for the PAN in Guerrero at the polling station level was 35.7 votes with a standard deviation of 36.55. Given the low vote count of this party, it is natural to expect a deviation from a uniform distribution in the state.

The shortcomings of the Last Digit test are two. First, the analysis of aggregate vote tallies is vulnerable to false positives. As an illustration, the left panel of Figure 3.1 replicates the analysis of Beber and Scacco of the 2009 presidential election in Iran Beber and Scacco (2009). The authors check for the last digit distribution of the four candidates in every province and conclude that "(f)ewer than four in a hundred non-fraudulent elections would produce such numbers." However, this conclusion does not hold up when the data is analyzed at the township level, which is the immediate lower aggregation level. As the right panel of Figure 3.1 shows, there is no significant deviation from a uniform distribution when the analysis is made at the lower level. A second potential limitation may arise with the introduction of electronic voting machines into the electoral process. As Levin et al. (2009, p. 8) suggest, electronic manipulation of the results may overcome this test with a random generation of the last digits, which would make this tool an obsolete way to detect irregularities.



(a) By Province (b) By Township

Figure 3.1: Last digit distribution for the 2009 presidential elections in Iran.

3.2.2 Turnout analysis

If political participation is randomly distributed across polling stations, turnout rates should follow a normal distribution and affect all of the candidates or parties in a proportional way. The following techniques exploit these assumptions to seek evidence for electoral anomalies.

Turnout Distribution

When voting turnout rates vary for reasons other than the geography of support for a particular candidate, those rates should follow a normal distribution. Therefore, bimodal and skewed distributions suggest turnout inflation in a subset of polling stations. One of the modes would correspond to the turnout rate among the uncontaminated ballot boxes and the second mode would correspond to the turnout in tampered observations (Myagkov, Ordeshook and Shakin, 2009, p. 32).

A bimodal distribution, however, does not necessarily indicate manipulation. Myagkov, Ordeshook and Shakin (2009, p. 43) highlight the fact that bimodal distributions can also be the product of combining observations with covariates that correlate with turnout (i.e., income or education). To rule out explanations other than electoral manipulation, the authors suggest accounting for the homogeneity of the covariates and, if possible, comparing the distributions over time to detect non-uniform changes in the turnout levels.

This technique has been used in several electoral analyses. Myagkov, Ordeshook and Shakin (2009) identify artificial inflations in turnout rates in Russia, Ukraine, and the United States. Similarly, Levin et al. (2009) compare turnout distributions in their analysis of the 2009 referendum in Venezuela. Here, the authors do not find any significant pattern with the one exception of the state of Vargas. Using a variation of this technique, Klimek et al. (2012) compare the vote rates for winning parties in a sample of countries, finding large deviations in both Uganda and Russia.

Turnout and votes for a candidate

Another expectation is that increases in turnout should neither harm nor benefit a candidate in a disproportionate way. The test for this hypothesis was developed by Sobyanin for the 1993 constitutional referendum in Russia.[3] Specifically, the author looks at the number of votes for the two alternatives in the referendum

[3]For a preliminary study, see Sobyanin, Gelman and Kaiunov (1994).

(V_{yes} and V_{no}) and divides them by the registered voters (E) in every district.[4] Under normal conditions, variations of these ratios should correspond with the turnout level in the district (T). In other words, when regressing V/E on T the coefficient should fall in the interval $[0,1]$, whereas coefficients outside this interval would represent evidence of electoral manipulation.

Formally, consider two precincts A, B with the same electorate size (E) and uncontaminated turnout rates of T_A and T_B, where $T_A > T_B$. Suppose that among those who vote, a candidate has a similar proportion of votes (Z) in each precinct, $0 < Z < 1$. Therefore, when fraud does not exist, regressing votes for the incumbent as a proportion of the electorate (V/E) on turnout (T) leads to $V/E = \beta_0 + \beta_1 T$. In this case, $\beta_1 = \frac{T_B - T_A}{ZT_B - ZT_A}$, which would be a value between 0 and 1. Now, suppose that in precinct A, the candidate is able to inflate the turnout rate to be 100%, receiving all the irregular votes $(1-X)$. Consequently, the estimation of β_1 changes to $\beta_1 = \frac{ZT_A + (1-T_A) - ZT_B}{1-T_B}$, which is a value greater than 1 and, therefore, would raise the alarm of potential fraud.

Myagkov, Ordeshook and Shakin (2009, p. 38) expand Sobyanin's study and describe the conditions in which the values of β_1 outside the $[0,1]$ range suggest potential manipulation of the results. First, in order to justify the comparison of observations across different units of analysis, an accurate interpretation of the results should have information regarding the level of aggregation at which fraud occurs.

[4]The author groups the data in *rayons*, which are the administrative entities in the Post-Soviet countries.

Second, this test works only under the assumption of homogeneous data, or after ruling out unobserved covariates of the vote count and turnout. Therefore, the authors constrain the interpretation of the results to a high R^2, which ensures that the regression explains most of the variance in the rates of support for the candidate.

Despite the straightforward intuition behind this test, the commonly lacking of data of voting rolls at polling stations limits the use of this tool. Beyond the analyses of the elections in Russia and Ukraine (Myagkov, Ordeshook and Shakin, 2009; Myagkov and Ordeshook, 2010), Levin et al. (2009) apply this technique for the 2009 referendum in Venezuela, without finding any suspicious results and attributing the rare observations to the heterogeneity of the data.

3.2.3 Flow of Votes

The Flow of Votes test expects that the share of votes that a candidate receives in a previous election is a percentage between 0 and 100 percent of the votes, and thus anything outside this percentage rate proportion would be a red flag indicating potential electoral manipulation. The most common way to estimate these shares is to regress the proportion of votes in each precinct between two elections and analyze the coefficients for each alternative. Although this technique is vulnerable to ecological inference problems, Myagkov, Ordeshook and Shakin (2009) argue that coefficients with values that are either negative or over one suggest potential cases of electoral manipulation when controlling for aggregation issues.

Formally, consider two consecutive elections in $i = 1, ..., n$ precincts where voters decide between three alternatives: a vote for the incumbent party (I), a vote for the challenger party (C), or abstention (A). The flow of votes for each alternative is then estimated using Goodman regressions as follow (Myagkov, Ordeshook and Shakin, 2009, p.47-53; Levin et al., 2009, p. 5-6):

$$y_{Ii} = \beta_{II} x_{Ii} + \beta_{IC} x_{Ci} + \beta_{IA}(1 - x_{Ii} - x_{Ci}) \tag{3.1}$$

$$y_{Ci} = \beta_{CI} x_{Ii} + \beta_{CC} x_{Ci} + \beta_{CA}(1 - x_{Ii} - x_{Ci}) \tag{3.2}$$

$$y_{Ai} = \beta_{AI} x_{Ii} + \beta_{AC} x_{Ci} + \beta_{AA}(1 - x_{Ii} - x_{Ci}) \tag{3.3}$$

Where each β_{jk} represents the proportion of votes for an alternative k in a previous election shared with the alternative j in the most recent election. Therefore, the expectation is that all of the coefficients will lie between 0 and 1, adding up to 3 for the β's with the same x_i and to 9 across the three equations.

Consistent with other approaches discussed in this paper, aggregation errors may produce incorrect conclusions. Myagkov, Ordeshook and Shakin (2009) illustrate some problems with ecological inference when analyzing electoral results with different support levels for a particular candidate across regions. While the standard procedure to reduce aggregation problems would be to restrict the coefficients to certain values, Myagkov, Ordeshook and Shakin (2009) allow the values for β to fall outside the $[0, 1]$ interval after clustering observations of similar regions to correct for aggregation issues. Alternatively, Levin et al. (2009) consider a similar approach using Goodman regressions and include random effects for different regions. Another potential problem with this tool is that it relies on information of a previous election that is presumably uncorrupted. If fraud occurs in sequential elections in the same polling places, the coefficients are likely to fall between the normal $[0,1]$ interval and thus manipulation will go undetected with this technique (Levin et al., 2009, p. 6).

3.2.4 Outlier Analysis

This technique includes additional information available (e.g., election polls or results from previous elections) to forecast the vote returns in order to find significant differences between the expected and the observed election results, which under certain conditions could be considered as evidence of fraud. The idea behind this approach is to generate an unbiased estimate of voters' preferences in every geographical unit. As the accuracy of the forecast increases, outlier observations can be interpreted as evidence of vote tampering. This methodology thus requires accurate information to simulate the electoral results and account for heterogeneity of the data.

Depending on the type of data available, this test has different variations. Wand et al. (2001) use the 1996 proportion of votes for both the Republican and the Reform candidates to forecast the number of votes for Pat Buchanan in Palm Beach County during the 2000 presidential elections. After identifying the residuals for each state, the authors find the effects for the butterfly ballot when comparing vote returns across counties, precincts, and ballot types. Mebane and Sekhon (2004) extend this analysis and propose a robust estimator for count data, which accounts for heteroskedasticity in the vote counts, overcoming the interference of outlier observations in their regression results.

Similarly, Alvarez and Katz (2008) employ this technique for the analysis of the 2002 state elections in Georgia. The authors first estimate the vote results for the 1998 state elections using the results from the 1996 presidential elections at the county level. Then, they use the estimated coefficients to forecast the 2002 election results, comparing the differences between the predictions and the observed data for each county. Levin et al. (2009) also use the same technique for the analysis of the 2009 referendum in Venezuela. The authors first regress the share of *Yes* votes in the 2007 referendum on the proportion of votes for Chávez in the 2006 presidential election. In the second stage, they use the estimated coefficients to forecast the proportion of *Yes* votes in the 2009 election. Their analysis does find an underestimation of the model at those polling places that exhibit a share of *Yes* votes beyond 80 percent, yet the authors do not find a particular group of polling places clustering away from the rest of the data.

Hausmann and Rigobón (2011) examine the 2004 Recall Referendum in Venezuela using two imperfect predictors of voters' preferences: the proportion of signatures in the 2003 recall petition and the proportion of collected *Yes* votes in an independent exit poll. Assuming that both variables are independent proxies of voters' intentions, the authors interpret the correlated errors from both measurements as evidence of fraud, concluding that manipulation in the referendum occurred in about 70 percent of the voting centers.

Given that the ability to accurately forecast voters' preferences depends on unbiased estimators, the main limitation of this approach is the quality of available data. As an illustration, in the specific case of the 2004 exit poll in Venezuela,[5] Carriquiry (2011) highlights the lack of information concerning the non-respondents which might alter the results. Similarly, Jiménez (2011) notes the fact that voters in the sample were stratified by sociodemographic variables and not by voting center, thus the results may not be representative of the voters' characteristics at the precinct level.

Polling place comparisons

The methodology that I use, described in Chapter 3, is a variation of the outlier analyses and combines two techniques described above. First, similar to Myagkov, Ordeshook and Shakin (2009), it analyzes the relationship between turnout and the number of votes for a party or candidate. However, it also estimates the amount of support for the candidates by dividing the votes for a candidate by the total number of ballots cast rather than by the number of eligible voters. Second, similar to Alvarez and Katz (2008) and Mebane and Sekhon (2004), it identifies observations that are significantly separate from the rest of the data. However, this tool does not use data other than the vote returns and exploits a feature of the electoral

[5] Also used by Prado and Sansó (2011)

administration that is common to many places in Latin America: the quasi-random assignment of voters to polling stations. Finally, contrary to Mebane and Sekhon (2004) and Wand et al. (2001), my research looks first for outliers in the turnout distribution and then compares their vote levels.

The limitation of this methodology is also one of its advantages when compared to other possible tools: it accounts only for local and decentralized fraud. If the electoral irregularities are committed at the aggregate level, this model would not be able to identify outliers between the expected and the observed number of votes. In the case of the analysis of Mexican elections in the previous chapter, had electoral manipulation occurred at all the polling stations within a precinct, the methodology would undercount those events.

3.2.5 Modeling Fraud

Modeling fraud is an alternative forensic tool used when a researcher has specific knowledge of the type of corruption that took place during an election. This method uses the information about the way in which fraud occurred simulate a subset of "clean" and "contaminated" elections.

For example, suppose that there is anecdotal evidence suggesting that fraud occurred at most polling stations by taking a quarter of the votes and adding them to the incumbent's vote counts. Moreover, suppose that there is available information about turnout and vote support from a previous election, which is presumably clean. The methodology then creates a series of elections that replicates the turnout rate and vote support similar to a clean election. Under the assumption that voters' behavior is similar in both elections, a subset of manipulated elections is simulated; in this case, 25 percent of the votes for the challenger at each polling station shift to the incumbent's vote count. Finally, the results from the real elections are compared with the two subsets of artificial data to evaluate the resemblance of the results with both the clean and manipulated figures.

When anecdotal evidence is available, this technique is a useful way to assess the integrity of an election. Myagkov, Ordeshook and Shakin (2009) simulate different elections to illustrate their forensics tools. However, they do not assess the similarity of the real data with their simulations. Cantú and Saiegh (2011) propose this technique for the analysis of the elections in Argentina in the 1930s. Briefly, the authors simulate 10,000 elections and "tamper" with half of them, using their prior regarding how fraud was carried out, which included inflating the incumbent votes while decreasing the opposition votes. Then they compare their quantities of interest between clean and fraudulent simulated elections and then between the real and simulated data.

Klimek et al. (2012) make a cross-national analysis of elections to test for two different types of fraud: incremental fraud, in which the ballots of non-voters and the opposition are removed from a few polling stations; and extreme fraud, where corruption occurs at nearly all polling stations. The authors find only the significant similitude of extreme fraud in the Russian elections of the last ten years –2003, 2007, 2011, and 2012– as well as for Uganda in 2012.

Using this tool requires a strong basis of knowledge of how manipulations typically occur, as well as a familiarity with the case study to evaluate the existence of alternative explanations. For example, Klimek et al. (2012) find statistical resemblance to an incremental fraud in Canada and Finland, yet the authors rule out electoral corruption in these cases once they account for the different voting behaviors of voters in Quebec and the high voter mobilization in the Helsinki region (p. 5).

3.3 Conclusion

Given the growing use of these techniques to evaluate the quality of the elections, I propose a way to analyze when and how each technique is found beneficial. This chapter surveys the extant forensic tools and shows that, as electoral manipulation comprises a variety of technologies, the use of different forensic tools should be a part of comprehensive analysis of where and how fraud can occur. Table 3.1 summarizes all of the assumptions and statistical techniques behind for each of these forensic tools. The turnout analyses produce interesting results, but their conclusions are subject to the homogeneity of the electoral data. Digit analysis is a useful way to detect fraud when there is no other information available except the election results; yet, it is important to acknowledge the conditions in which potential evidence for fraud is an artifact of the specific dynamics in the election. The analysis of the flow of votes is an intuitive way to study elections, yet some of the results outside of the $[0, 1]$ interval are equivocal as they can be caused by aggregation problems or electoral manipulation. Finally, outlier analyses have the advantage of identifying individual observations but they require information beyond the vote counts to confirm that the findings are not due to unintentional errors. As there are different technologies for electoral manipulation, each forensic tool may detect a specific type of fraud. Therefore, the findings of election forensics should not be considered as final and conclusive evidence in assessing the integrity of a given election.

Table 3.1: Summary of the Forensics Techniques.

	Methods (Reference)	**Underlying Assumptions**	**Test**	**Expected type of fraud**
Digit Analysis	First and Second Significant Digits (Mebane 2006)	First and second significant digits follow a Benford distribution	Chi-square test	Imprecise
	Last Significant Digit (Beber and Scacco 2012)	Last Significant Digit follow a uniform distribution	Chi-square test and confidence intervals	Alteration of vote tallies
Turnout	Turnout distribution (Myagkov et al. 2009)	Turnout rates follow a normal distribution with uniform variations across elections	Finding variations to unimodal distributions	Inflation of turnout or stuffed ballot boxes
	Regressing votes for a candidate on turnout (Myagkov et al. 2009)	Increases in turnout should add to the number of votes of every candidate in a proportional way	Regressing votes for a candidate on turnout	Inflation of turnout, mobilization.
Flow of votes (Myagkov et al. 2009, Levin et al. 2009)		The share of votes that a candidate receives from a given party or candidate in a previous election is a proportion between 0 and 100% of the votes	Goodman regressions with random effects for regions	Overwhelming change of votes in particular regions
Outlier analysis	Forecast (Wand et al. 2001, Hausmann and Rigobón 2011, Alvarez and Katz 2008)	Covariate information can be used as unbiased estimator of voters' preferences	Instrumental variables or Robust regression	Targeted manipulation in a few polling places
	Polling place comparisons (Hausmann and Rigobón 2011, Cantú 2013)	Voters assignment to polling stations is orthogonal to voters' preferences	Confidence intervals in the polling stations with a relative high turnout rate; t-test	Manipulation at the polling station level
Model fraud (Cantú and Saiegh 2011, Klimek et al. 2012)		Electoral manipulation can be modeled and simulated	Logit analysis and distribution comparison	Widespread fraud

Chapter 4

Finding Irregularities in Mexican Local Elections

The defeat of the Institutional Revolutionary Party (PRI) in the 2000 presidential election marked a watershed moment in Mexican politics. To some extent, the outcome was the result of a series of electoral reforms adopted in the 1990s. The scope of these reforms, however, was uneven across Mexican sub-national governments. As such, popular distrust in the integrity of elections is still quite common at the local level (Moreno, 2012; Hiskey and Bowler, 2005). It is unclear, however, whether popular protests reflect authentic fraud allegations or are simply a political tool for electoral losers.

Evaluating the integrity of elections is a complicated task because neither "winners" nor "losers" may have incentives to reveal the truth: perpetrators of fraud usually want to keep their activities as hidden as possible (Lehoucq, 2003). The alleged victims of fraud, in turn, may be willing to claim the existence of electoral manipulation even when no irregularities are recorded (Magaloni, 2007; Eisenstadt and Poiré, 2006). To overcome this challenge, I propose a novel empirical strategy to detect electoral fraud. I focus on the case of Mexico and exploit a feature of the country's electoral code: within each electoral precinct, voters are assigned to polling stations according to their childhood surnames; consequently, the only difference between the voters at contiguous polling stations should be their last names. Because political preferences are seldom correlated with voters' last names, I identify fraudulent practices by finding unexpected differences in turnout levels and partisan votes across contiguous polling stations. I illustrate this methodology by evaluating the 2010 gubernatorial elections held in twelve states and provide evidence that candidates in three of the states benefitted from electoral irregularities. In fact, some of the polling stations that are flagged as suspicious using this method were notorious for violent disruptions or duplicated ballots.

This chapter makes two contributions to the research on electoral fraud. First, the findings indicate that nondemocratic enclaves that actively obstruct the completion of Mexico's democratic transition still remain. Second, from a methodological standpoint, the proposed approach contributes to the growing literature that uses statistical tools to evaluate the quality of elections (Myagkov, Ordeshook and Shakin, 2009; Mebane, 2006; Levin et al., 2009; Beber and Scacco, 2012; Fukumoro and Horiuchi, 2011). Specifically, my method focuses on identifying outlier observations where fraud can be observed (Wand et al., 2001; Powell, 1989).

This chapter is organized as follows. Section 1 reviews the challenges of measuring the democratization process in Mexico at the sub-national level. Section 2 describes the foundations of the empirical model and explains the methodology I propose for measuring electoral fraud. Section 3 presents the results. Finally, Section 4 discusses the implications of this research.

4.1 Local Elections in Mexico

During most of the twentieth century, elections in Mexico failed to function as a legitimate process for selecting public officials. Electoral manipulation included the modification of final vote counts (Castañeda, 2000, p.231-239; de la Madrid, 2004, p. 814-824), the alteration of ballot boxes (Langston, 2012, p. 19-22), the certification of the results by a biased legislature (Lehoucq, 2002), the inflation of voter registration lists (Preston and Dillon, 2004; Molinar Horcasitas, 1987; Gillingham, 2012; Simpser, 2012*a*), and the intimidation of opposition supporters (Craig and Cornelius, 1995).[1] Many of these practices took place in a decentralized manner: local brokers manipulated the ballot boxes in their own regions in exchange for political favors from the winning candidates (Langston, 2012).[2]

Following the electoral reforms of the 1990s, political parties became unable to carry out electoral manipulation at the federal level (Magaloni, 2006, p. 36-38).[3]

[1] Although vote-buying is a common issue in many instances (Cornelius, 2002; Gibson, 2005), this project does not include this type of electoral manipulation.

[2] Larreguy (2012) provides evidence of how local party brokers mobilize clientelistic networks for local elections in Mexico.

[3] For a detailed description of each electoral reform see Ochoa-Reza (2004) and Craig and Cor-

In the case of local elections, however, the scope of reforms was uneven across Mexican provinces. Although the Constitution provides a minimum set of rules that local elections must follow, each state freely interprets its right to organize the elections for governors, local legislators, and municipal councils, leaving the quality of these elections dependent on the extra-institutional dynamics of each province (Peschard, 2010). Consequently, there is still distrust in the fairness of local elections and the impartiality of state electoral institutions (Moreno, 2012).

A strategy known as *fraude hormiga* may be prevalent.[4] This form of electoral fraud refers to "the illegal introduction or subtraction of a very few votes in order not to affect the outcome in the polling station –and avoiding its potential nullification– but enough to affect the final outcome in the aggregate (Crespo, 2006, p. 128-129)." In other words, political machines, when they have the opportunity, change the vote counts of the polling stations by an amount unlikely to get noticed but considerable enough to be decisive in the aggregate count. Mexico's electoral code prescribes the nullification of vote tallies at polling stations where the number of irregular votes, either by accident or fraud, is greater than the difference between the two leading candidates. However, these grounds for nullification are not applied to the overall result, even in cases where the aggregate number of irregular votes is greater than the difference between the two main political competitors. This limitation of the electoral legislation creates an opportunity for this type of manipulation, which

nelius (1995).

[4]The closest translation of this term in English would be "ant fraud."

becomes more attractive in close elections (Crespo, February 2, 2008.).

"Buying" poll workers and replacing them with partisan agents facilitates this type of fraud. Political parties typically offer money or violent threats in exchange for a poll worker's absence on the election day (Raphael, 2007; The New York Times, July 4, 2010). Installing partisans as poll workers not only allows irregularities to be tolerated at the polling stations but also facilitates altering the vote count in favor of a particular candidate.[5]

4.1.1 2010 Gubernatorial Elections

On July 4 of 2010, voters in twelve out of the thirty-two states in Mexico elected a new governor. These elections represented a decisive phase in defining the political strength of each party in anticipation of the 2012 presidential election. When the Institutional Revolutionary Party (PRI) lost the presidency in 2000, many of its members sought consolation in their remaining bastions of political power: the sub-national governments (Dresser, 2003). Ten years later, the PRI had positioned itself to be in control of not only those regions still controlled by the old guard *priístas* but also the states run by other political parties. In most of these states, the elections pitted the PRI candidates against an unusual alliance of many opposition parties. The coalition, principally formed by the conservative National Action Party (PAN), and the leftist Democratic Revolutionary Party (PRD), was a response to the risk of an overwhelming outcome in favor of the PRI. This strategy was controver-

[5] Author's interview with PRI and PRD party brokers. Oaxaca, July 2010.

sial, even among the members of the various coalition parties, but it represented a pragmatic response to the common practices of local political machines.[6]

Figure 4.1: States with governor elections in July 2, 2010.

The overall results of the election were mixed. While the coalition won in Oaxaca, Sinaloa, and Puebla, the PRI retook Aguascalientes, Tlaxcala, and Zacatecas. Aguascalientes and Tlaxcala were were formerly governed by the PAN, whereas Zacatecas had an outgoing governor from the PRD. Although both the PAN and the PRD celebrated seizing three of the PRI's historical bastions, they contested the elections in Durango, Hidalgo, and Veracruz, claiming fraud. Upon review, the Federal Electoral Court did not find substantive evidence to support the claim that the

[6]The PRI's leader in the Senate, Manlio Fabio Beltrones, commented that "alliances between enemies that don't respect each other are contrary to nature," while the Interior Minister of the Federal Administration, the PAN member Fernando Gomez Mont, said that that kind of coalition "could end up leading to fraud." (Los Angeles Times, February 01, 2010)

elections were rigged, and thus certified the legitimacy of these results.[7]

4.2 Empirical Analysis

Previous studies in Mexico have focused on detecting electoral irregularities (Crespo, 2006; Mebane, 2006; Instituto Federal Electoral, 2010) and determining whether these inconsistencies affect the electoral outcome (Aparicio, 2006, 2009; Poiré and Estrada, 2006). This paper's proposed methodology has two advantages over previous analyses. First, the method I use distinguishes between random and systematic effects at the polling-station level. This feature reduces the possibility of confusing cases of deliberate manipulation at the polling station with cases that should be classified as accidental errors. Second, unlike Aparicio (2009) and Poiré and Estrada (2006), the methodology in this paper takes into account the high variance in electoral behavior at the precinct level (Navia, 2000) and does not categorize the observations according to the winning candidate in the district.

4.2.1 Identification Strategy

A *sección*, or electoral precinct, is the smallest electoral subunit in Mexico and it groups voters into units of 50 to 1,500 (*Código Federal de Institutciones y Preocedimientos Electorales*, 2012, art 155). Within each precinct, there must be one

[7] See Prats (August 8, 2010). Still, a post-electoral poll showed that the proportion of respondents who lacked confidence in the electoral processes in their respective states ranged from approximately 25% in Oaxaca to more than 50% in Durango (Gabinete de Comunicación Estratégica, 2010).

polling station for every 750 voters. However, due to demographic changes after a precinct has been drawn, the number of voters in a precinct can be greater than 1,500; consequently, some precincts may have more than two polling stations. The first polling station is called the *casilla básica*, while the subsequent polling stations are called *casilla contigua* (e.g., *casilla contigua 1*, *casilla contigua 2*, etc.). There is a *casilla contigua* for each additional group of 750 in the precinct. The assignment of precinct voters to a particular polling station is strictly alphabetical; that is, voters are distributed among the polling stations according to their last names (*Código Federal de Institutciones y Preocedimientos Electorales*, 2012, art 152). If possible, all of a precinct's polling stations must be located in the same building; otherwise, polling stations must be in adjacent locations in order to provide similar transportation access to all of the voters (*Código Federal de Institutciones y Preocedimientos Electorales*, 2012, art 239).

As an illustration, we can focus on the electoral geography of the state of Aguascalientes, which is shown in Figure 4.2. The white lines define the boundaries of the 584 precincts in that state, which are delineated by geographical and demographic characteristics. There are a total of 1,313 polling stations throughout the state. Consider precinct 129, which is marked in red. By the time of the local election in 2010, it had 1,416 registered voters assigned by alphabetical order to two different polling stations, each allowing 708 voters on their voting list. Suppose that there is a household in the precinct with two registered voters, and that the last names of these voters are Abasolo and Zurita. Regardless of the fact that they share the same address, voter Abasolo would be assigned to a different polling station than the one assigned to voter Zurita. In short, the only condition for the allocation of voters to precincts is their home address and within each precinct, the assignment of voters to a particular polling station depends only upon their last name.[8]

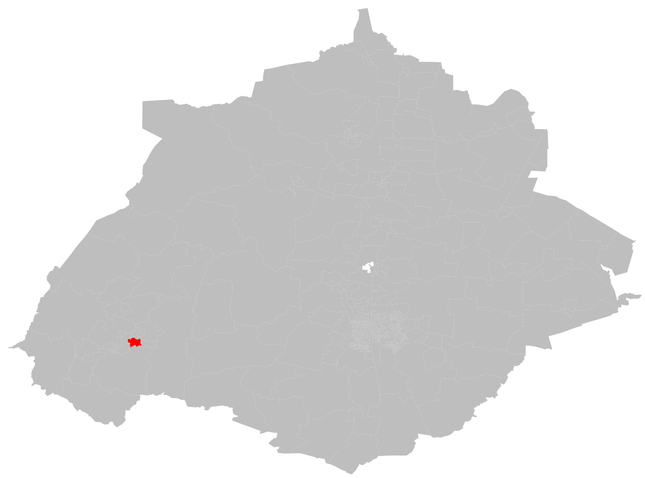


Figure 4.2: Electoral precincts in Aguascalientes.

Therefore, I identify polling stations with potential irregularities by comparing the turnout rates in each unit with those at the other polling stations within a particular precinct. After classifying each polling situation given its turnout rate relative other polling stations in the precinct, I measure the change in vote returns

[8]There are two other types of polling stations that have additional requirements than those labeled as either *básica* or *contigua*. There are "special polling stations," or *casillas especiales* that receive the ballot papers of voters who are temporarily situated outside of their assigned precinct (COFIPE, Art. 270). Similarly, there are "extraordinary polling stations," or *casillas extraordinarias*, which are designed for those precincts whose sociocultural or geographic conditions make it difficult for all voters to commute to the same place (*Código Federal de Institutciones y Preocedimientos Electorales*, 2012, art 239). These polling stations represent less than 1% of the observations and are not taken into account in the analysis.

for each candidate.

Last Names and Political Behavior

If last names are not correlated with voting behavior, then each polling station should be an unbiased sample of the precinct's voters. Therefore, it is important to ensure that sorting voters by their last names is orthogonal to their voting behavior. Due to the Mexican legislation to protect public records of voter turnout and the identification of respondents' last name in political surveys, I use two indirect approaches to ensure that there are no differences in sociodemographic variables when grouping individuals by their last names. Detailed results of this analysis and the databases for this subsection are available in the Appendix.

The first approach uses a national monthly survey of TV viewership conducted by the Social Research Institute. The poll was conducted in July, 2011 with a sample size of 1,499 respondents. Given that this poll contains the names of the interviewed individuals, I could identify the first letter of their last names to determine whether there are differences in age and gender, and the frequency of those who watch TV news, which represents the main source of political information in the country (Moreno, 2012). However, I did not find any significant differences in any of these variables when the respondents were sorted by their last names and grouped into halves, thirds, and quartiles.

The second approach tests whether there are socioeconomic differences when grouping voters by their last names. I analyzed the last names of the children who received benefits from the cash-conditioned transfer program *Oportunidades*.[9] This program seeks to improve health and education among poor families and currently serves more than five million households across the country. The only condition for a family to be targeted by *Oportunidades* is that its income falls below the poverty line established by the program (Levy, 2006). Figure 4.3 shows the distribution of the first letter of the children's last name in the twelve states with local elections in 2010. A comparison across states shows only two outlier observations: first, the high proportion of children with the last name *Hernández* in Hidalgo; and second, the density of last names with Mayan origin and beginning with letter 'C' in Quintana Roo.

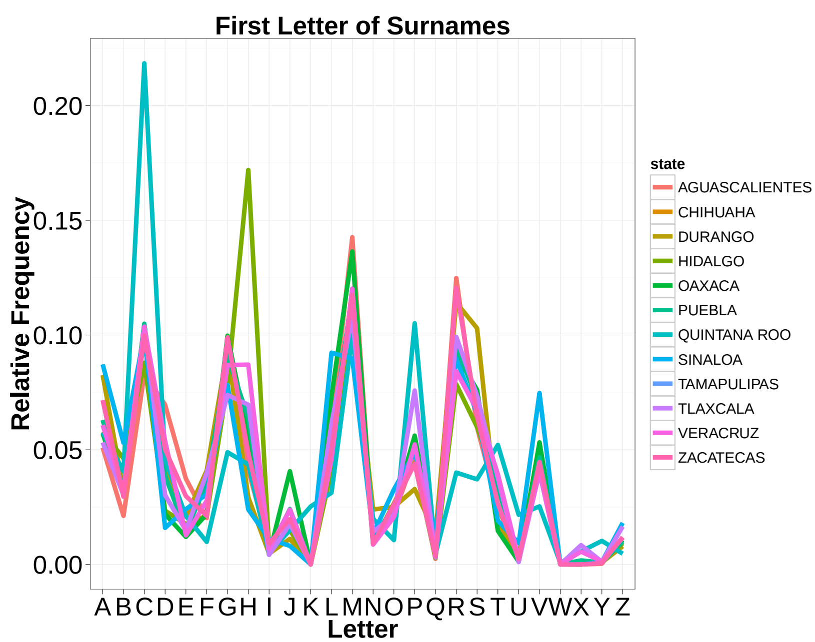


Figure 4.3: Distribution of the first letter in the last name of children benefitted by *Oportunidades* in the twelve states with governor elections in 2010.

[9]The information is available online at http://www.oportunidades.gob.mx/

To determine whether the population under the poverty line is clustered by a few surnames, the electronic Appendix shows the results at state and municipality levels using Yule's K and Simpson's D, which are two different diversity measurements to show the probability of randomly selecting two individuals with the same last name.[10] The results by state show that the highest probability of randomly choosing two individuals with the same surname is given by the Yule's K in Hidalgo, representing a probability of 0.03.

Additionally, I compare the frequency and order of the most common last names on the *Oportunidades*'s list with the most common last names registered in the electoral roll in each state. Figure 4.4 shows the order list for the most common last names on the voting list and the *Oportunidades* list in Aguascalientes. As the information of the electoral roll is truncated to show the 100 most common names in every state, I rank the last names in both lists and compare both rankings using a Spearman correlation analysis. The correlation coefficient for each state is greater 0.75 in eleven out of the twelve states.

The Appendix also includes the heat maps for the first letter of the last names

---

[10] Yule's K is given by:

$$K = 10^4 \frac{1}{N^2}(\sum_{r=1}^{v} r^2 V_r - N) \tag{4.1}$$

where N in the size of the sample, v is the highest occurring frequency of the last name, and $V_r = r = 1, 2, ..., v$ is the number of different surnames with frequency r. The lowest values of K mean that almost every individual in the population has a unique last name, while $K = 10,000$ means that all the individuals share the same last name (McElduff et al., 2008, p.189). Simpson's D is given by:

$$D = \frac{1}{\sum_{i=1}^{v} p_i^2} \tag{4.2}$$

or the proportion of last names i relative to the total number of last names, p_i (Hunter and Gaston, 1988).

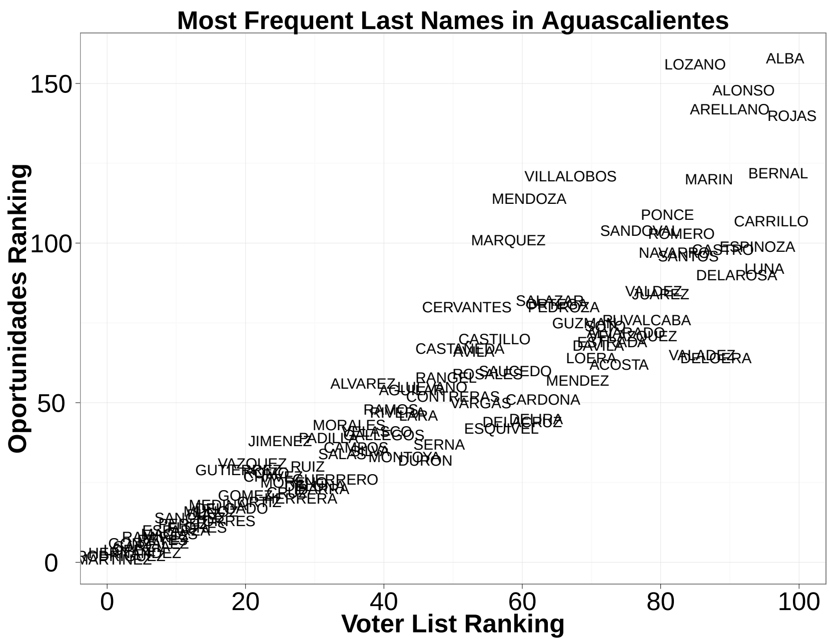


Figure 4.4: List order of the most common last names in the voting list and the *Oportunidades* list in Aguascalientes.

by state. The figures at the left and the top of each plot are the graphic representations of the cluster analysis, which indicates how similarly any two clusters were joined. As the dendrograms for each state show, the clusters do not sort letters in alphabetical order.

Finally, the Appendix shows an additional test. Given the classification of the polling stations, which I will explain in the following subsection, I explore the probability that the observations classified by the methodology as “suspicious,” depending on the way voters in a precinct are assigned to the polling stations. Given that it is possible to sort polling stations by the alphabetical order voters were assigned, I can compare whether, for example, the first polling station is more prone to be classified as suspicious than a *contigua* polling station.

4.2.2 Methodology

Figure 4.5 graphically represents the procedure for the analysis of each of the twelve elections held on July 2, 2010. The unit of analysis is the polling station, and the variables are the total number of valid votes (i.e., the number of votes at the polling station that are neither spoiled nor cast for a nonregistered candidate), the number of registered voters, and the number of votes for each candidate.

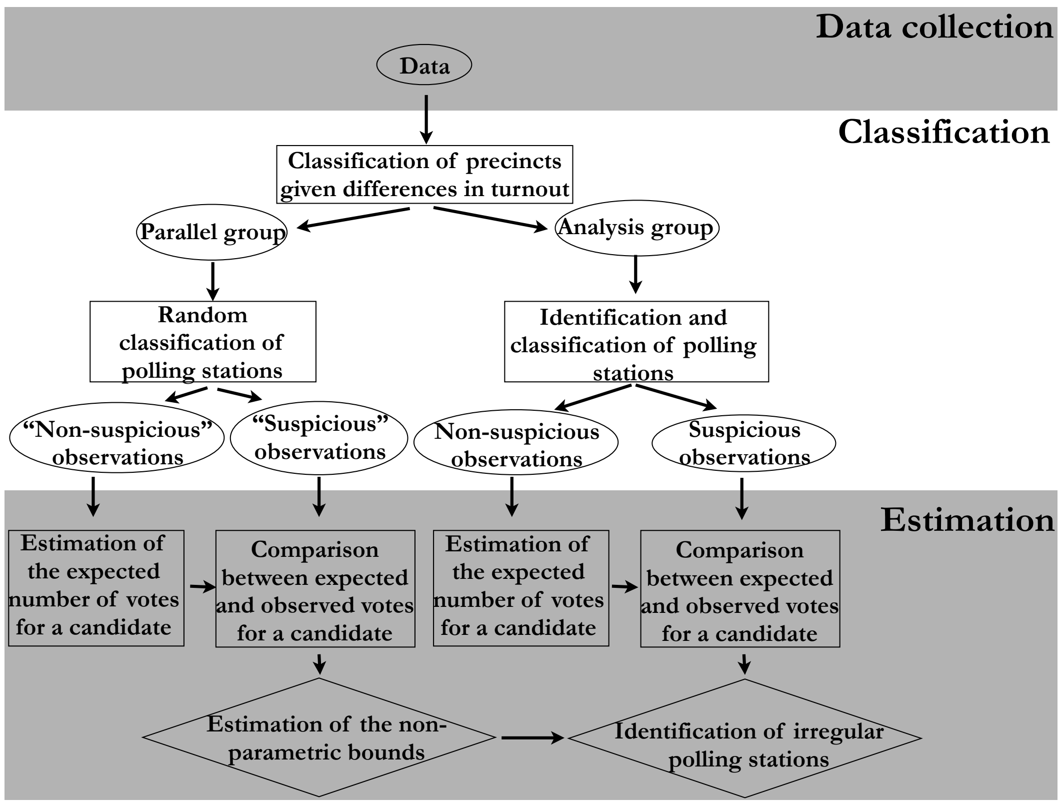


Figure 4.5: Diagram of the procedure for detecting irregular observations.

Classification

If electoral behavior is unrelated to the assignment of voters to particular polling stations within the precincts, I do not expect to find significant differences in the turnout levels among the polling stations in the same precinct. Therefore, I consider the general dispersion of the observations in each precinct as follows.[11] Consider an election with K different precincts. In every precinct A, I measure the differences in turnout between a polling station i and the rest of the polling stations $j \neq i$. The turnout estimation, t, divides the number of valid votes by the number of registered voters in the polling station. The differences in turnout rates between polling stations are estimated as follows:

$$L_A = |t_i - t_j|; \quad i, j \in X, i \neq j$$

That is, the absolute difference is calculated between the turnout of polling station i and the turnout of polling station j in precinct A. I consider an observation L_A significantly high whenever the difference in turnout between two polling stations in the same precinct is higher than the quantile 0.95 in the distribution of differences $L_{A=1,\ldots,K}$ in the state, or whenever $T_A > Q_{0.5}(L)$, where $Q_{0.5}(L)$ stands

[11] This study only considers observations in the database that are free of inconsistencies identified by the electoral authorities. That is, I remove those observations that were marked as ambiguous by the election officials due to the illegibility of the numbers on the official form, the presence of obvious errors made when filling out the form, or the submission of the form without the official envelope. These observations represent less than 0.5% of the total observations. I also exclude both extraordinary and special polling stations, which allow voters from outside the precinct to vote or are designed for voters living in communities distant from the regular polling stations but in an insufficient number to constitute a precinct.

for the quantile 0.95 in the distribution of L. As an illustration of the methodology, Table 4.1 shows the vote returns for four electoral precincts in Aguascalientes of which precinct 129 is one of the four. To calculate the differences in voter turnout in precinct 129, I get the absolute value of the difference of turnout rates observed between polling stations $129-A$ and $129-B$, which is $d_{129} = |0.64 - 0.55| = 0.09$. The same procedure is used for each precinct in the state, where the value of the 95^{th} percentile is $Q_{0.5}(L) = 0.078$.

After each observation is classified according to its relative difference in the turnout rates, the data are divided into two sets of observations. The first subset contains all polling stations without significant differences among turnout levels between the precincts; this subset is labeled the *parallel group*. The second subset contains polling stations in precincts where there is at least one observation with $T_A > Q_{0.5}(L)$; this subset is labeled the *analysis group*. In the case of Aguascalientes, the analysis group contains precincts 129 and 434, where the differences in turnout, $d_{129} = 0.09$ and $d_{461} = 0.10$, are greater than the critical value $Q_{0.5}(L)$.

In the analysis group, it is necessary to classify each polling station as either a *suspicious* or a *non-suspicious* observation. Because the quantity to be estimated is the additional number of votes for a particular party, a *suspicious* observation must be in the analysis group, and the difference in turnout with another polling station in the precinct must to be greater than $Q_{0.5}(L)$. The rest of the polling stations in the analysis group are labeled as *non-suspicious* observations.

In the example described above, while precincts 377 and 546 are assigned to

Table 4.1: Vote returns for four different precincts in Aguascalientes, 2010.

Precinct-Polling Station	PRI	PAN	Total Votes (Turnout)
129-A	154 (55%)	87 (31%)	278 (55%)
129-B	142 (42%)	138 (44%)	323 (64%)
377-A	159 (48%)	76 (23%)	331 (45%)
377-B	156 (47%)	63 (19%)	331 (45%)
434-A	128 (43%)	138 (46%)	294 (70%)
434-B	105 (41%)	109 (43%)	243 (60%)
546-A	113 (49%)	105 (46%)	227 (57%)
546-A	109 (43%)	98 (48%)	227 (57%)

the parallel group, precincts 129 and 434 are part of the analysis group. In the latter group, polling stations $129-B$ and $434-A$ are classified as suspicious observations because their turnout rates are higher than those of the other observations in their precinct.

Estimation

The next step is to determine whether the turnout differences also affect the distribution of votes among the candidates. If the turnout differences are caused by unintentional errors during vote counting, the consequences of these errors should affect the political parties in an unbiased way. In other words, errors made while counting the votes should not affect the differences in the estimated proportion of votes for each candidate.

If this condition holds true, consider a precinct A with a suspicious polling station i and a non-suspicious polling station j. If voters can choose between alternatives x and y; $v(x)$ and $v(y)$ are the votes for each alternative and $V = v(x) + v(y)$ are total votes cast in the polling station. The expected votes for the alternative x in the polling station i, $e(v_i(x))$, is :

$$e(v_i(x)) = \frac{v_j(x)}{V_j} * V_i; \quad i, j \in A, i \neq j$$

Regardless of the differences in the total number of votes between the polling stations, the proportion of votes for any given party should be the same at different polling stations in the same precinct.[12]

Recall the example described in Table 4.1. The expected number of votes for the PAN in polling station $129-B$ is $e(v_{129-B}(PAN)) = \frac{87}{278} * 323 = 101.08$, while the expected vote tally polling station $434-A$ is: $e(v_{434-A}(PAN)) = \frac{109}{243} * 294 = 131.9$. Comparing the number of expected votes for the PAN with the observed number of votes in these polling stations, a difference of more than 36 votes in polling station $129-B$ and less than 7 votes in $434-A$ may be observed.

To evaluate the random error for the expected number of votes, I consider the dispersion of votes for a party within the parallel group, where observations

[12]In the case that the number of non-suspicious polling stations J in the precinct is greater than one, the expected number of votes for the treated section is

$$e(v_i(x)) = \frac{\sum_{j=1}^{J} v_j(x)}{\sum_{j=1}^{J} V_j} * V_i; \quad i, j \in A, i \neq j$$

In other words, the estimation of the expected number of votes averages the controlled observations within the precinct.

lack significant differences in the total number of votes within the precincts, and I assume no irregularities can be observed. In order to estimate the expected number of votes for each polling station in the precinct, I randomly select a group of polling stations in the parallel group and assign them a “placebo treatment.” In this step, I treat each observation in the subset as a Bernoulli trial with two possible outcomes and a probability $p = 0.5$ of being a treated or control observation in the precinct.

Figure 4.6 shows the results for the PAN in Aguascalientes. The left plot shows the difference between the number of expected and observed votes in the parallel group. The plot displays only the observations that were randomly assigned as suspicious, and the expected number of votes is estimated using the unselected polling stations. For example, polling stations $377-A$ and $546-B$ are randomly considered “suspicious” in the parallel group, and each station’s expected number of votes for each party is estimated by using the vote returns in polling stations $377-B$ and $546-A$.

A perfect correlation between the number expected and observed votes would imply that the proportion of votes for any given candidate is the same across all polling stations in the same precinct, so all observations would fall over the dashed line. Because the measurement is subject to random events that may affect the number of votes at any polling station, I estimate the continuous quantiles that include 95% of the observations. In Figure 4.6, the solid lines in the plot show the Bayesian quantile regression for quantiles 0.025 and 0.975. In other words, these two lines represent the bounds where the comparison between the expected and observed votes should be located 95 out of 100 times.[13]

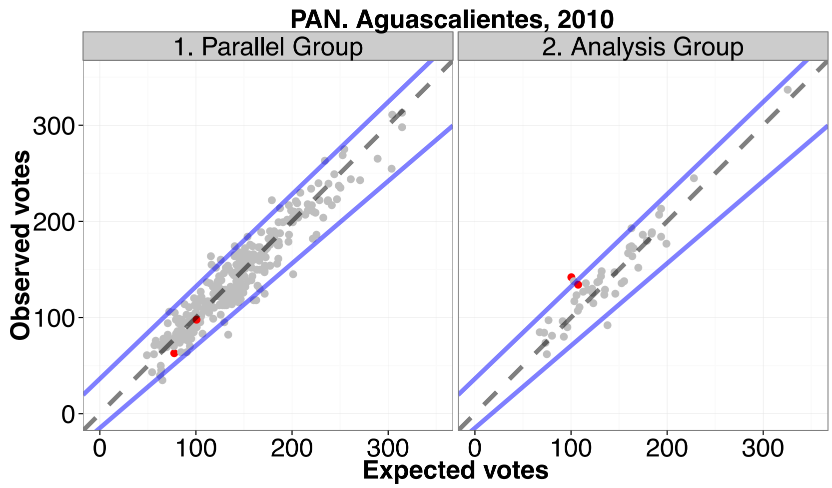


Figure 4.6: Expected and Observed votes in Aguascalientes, 2010.

The final step is to use the bounds from the parallel group to analyze the treated polling stations that belong to the analysis group. If turnout differences can be explained by unintentional factors, then the proportion of votes that a party receives should correspond to the polling stations of the precinct (i.e., the observed number of votes lies within the 95% confidence interval of the expected vote estimation). Otherwise, it is plausible that the irregularities of the turnout levels observed at the polling station disproportionally affect a particular political party.

The right panel of Figure 4.6 shows the difference between the expected and

[13]A quantile regression follows a model given by $y_i = x_i^{'}\beta_\tau + \epsilon_i$, where ϵ is the error term restricted to have the τ-th quantile equal to zero. Thus the objective is to estimate β_τ by minimizing $\sum_{i=1}^{n} \rho_\tau(y_i - x_i^{'}\beta_\tau)$, where ρ is the tilted absolute value function that yields at τ-th sample quantile as its solution (Koenker and Hallock, 2001; Kozumi and Kobayashi, 2009).

observed numbers of votes in the analysis group. I use the same bounds from the left plot to consider the variation when no differences in the total number of valid votes were observed. The suspicious observations described in the example are the red crosses on the right panel. The observed number of votes for the PAN in polling station $434 - A$ falls within the solid green lines; that is, despite the difference in the turnout rates between the polling stations of precinct number 434, the number of votes for the candidates are still similar. In contrast, polling station $125 - B$ displays not only a greater number of valid votes compared with other observations in the precinct but also a higher proportion of votes for the PAN. This difference falls outside the bounds and would be considered an electoral irregularity.

4.3 Results

The graphical analyses for the incumbent party and the most significant challenger in each of the twelve elections examined in this paper are located in the Appendix. In this section, I discuss three of the cases represented in Figures 4.7 to 4.10. First, consider Figure 4.7, which displays the PRI's electoral returns in the state of Hidalgo. In this case, the number of observations outside the upper bound is very similar for both the analysis and parallel groups. Consequently, although the election results were contested by the opposition, the evidence produced by this methodology does not support the claim of fraud.

Second, in the case of Oaxaca, illustrated in Figure 4.8, the proportion of observations above the upper bound is evidently larger in the analysis group than in the parallel group, which suggests that most of the irregularities benefitted the PRI's candidate. Some of the irregularities that the algorithm detects are in the municipality of Tututepec, where citizens filmed a meeting between people from the local electoral institution and PRI supporters, in which the former group received paper ballots and other electoral supplies.[14]

Finally, the graphical analysis for Durango, illustrated in Figure 4.10 shows that most of the electoral irregularities have unusual voting returns for the PRI; for example, there are only two observations in the parallel group that depict vote returns above 250 for the PRI. The results identify specific events that occurred on election day. Most of the detected irregular observations in Durango occurred in the city of Gómez Palacio, where a group of armed people disrupted the electoral process at the polling stations in precinct 447, representing the observations in the cluster. This one event resulted in two dead policemen and caused voters and poll workers to flee the polling stations. Despite the evidence and the allegations from the challenging coalition, the local electoral institute counted the votes from the eighteen polling stations located in this precinct (El Siglo de Durango, July 15, 2010, p. 6; La Jornada, August 16, 2010, p. 31).

When the irregularities lack anecdotal evidence to assess their causes, I propose an additional step to evaluate whether the irregularities in an election are the product of innocent accidents or have a systematic bias for a particular candidate.

[14]http://www.youtube.com/watch?v=Mt3Azm0OtuI

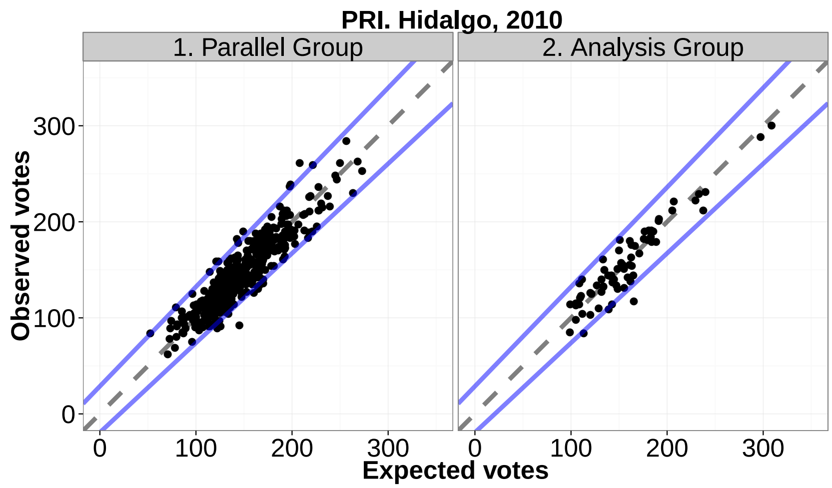


Figure 4.7: Observed and expected votes for the incumbent coalition in Hidalgo, 2010.

The following technique is based on the test proposed by Jiménez (2011) in his analysis of the 2004 referendum in Venezuela. Let us consider as our quantity of interest the difference D of comparing the expected $(e(v))$ and observed votes (v) for an alternative x in every polling station i. In the context of this analysis, the alternative x_i is the incumbent party or coalition in every state. If K is the total number of polling stations in both parallel and analysis group, then $D = \frac{\sum_{i=1}^{K}(v_{x_i} - e(v_{x_i}))}{K}$.

Out of this population, there is a sample S of size k, which quantity of interest is the average difference between the expected and the observed votes for the alternative x, denoted by $d_k = \frac{\sum_{i \in S}^{k}(v_{X_{\in S}} - e(v_{X_{\in S}}))}{k}$. Jiménez (2011) proposes to vary the size and composition of this sample according to different confidence levels. In contrast, the sample in this analysis considers the observations in the analysis group. If

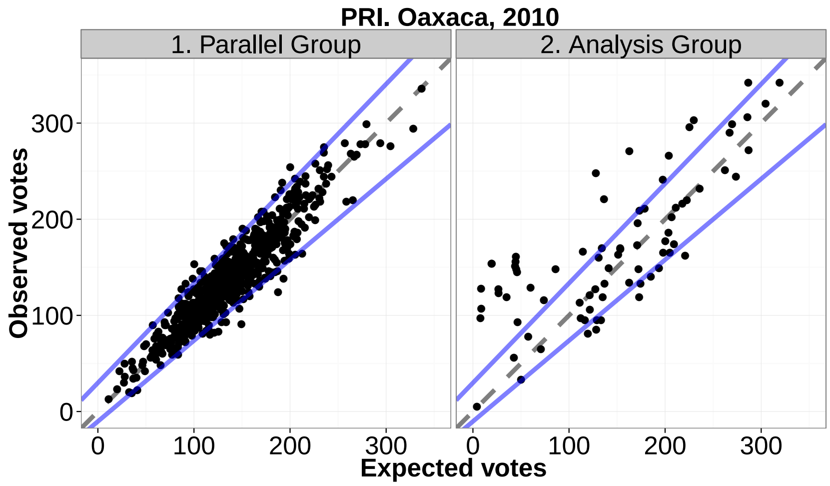


Figure 4.8: Estimation of expected votes for the incumbent coalition in Oaxaca, 2010.

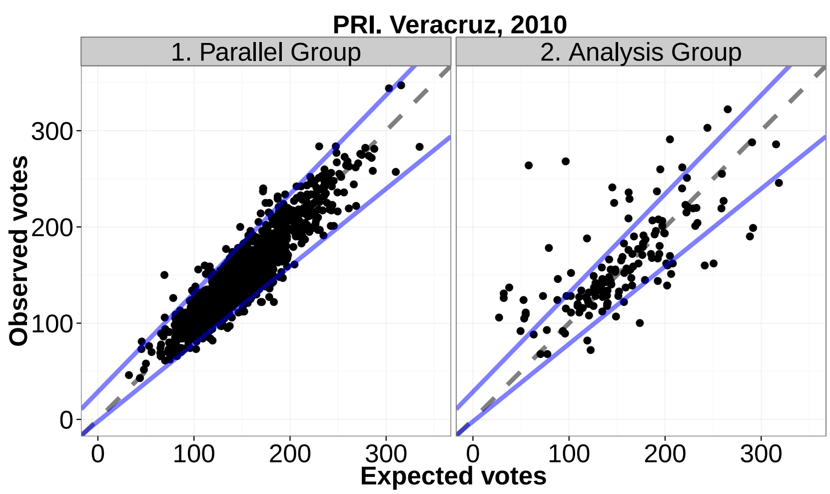


Figure 4.9: Estimation of expected votes for the incumbent coalition in Veracruz, 2010.

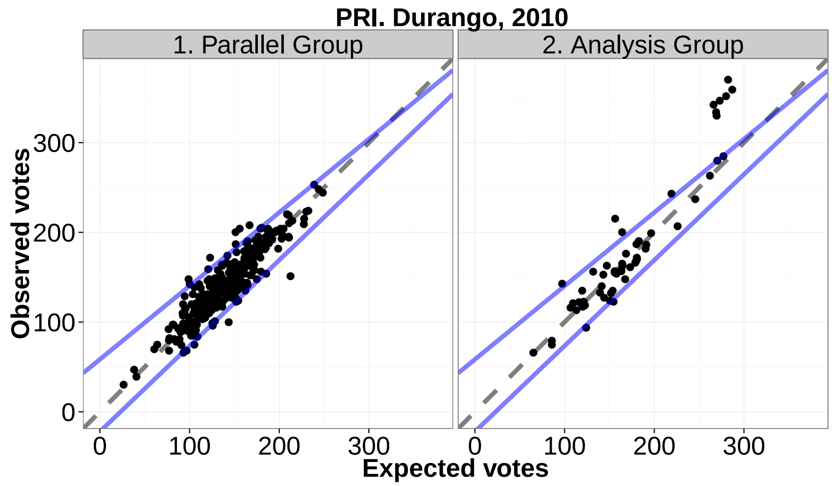


Figure 4.10: Estimation of expected votes for the incumbent coalition in Durango, 2010.

the irregularities are innocent or affect both the incumbent and opposition parties in a similar way, d_k should be similar to a random sample ratio. Moreover, if K is large, the variance of the estimator d_k should approximate to: $Var(d_k) = (1 - \frac{k}{K})\frac{s^2}{n}$ with $s^2 = \frac{1}{k-1}\sum_{i \in S_k}(d_k)^2$.[15]

Therefore, to test the hypothesis that errors are accidental, under the assumption that k and K are large enough:

$$\zeta_k = \frac{d_k - R}{S_k^2} \tag{4.3}$$

which is expected to be distributed as a standard normal distribution. In this

[15] See (Lohr, 1999).

case, a positive and large value of ζ_k means that the observations in the analysis group have a bias in favor of the incumbent party, while a negative and large value implies that the opposition party benefitted by the irregularities.

Table 4.2 shows the values of ζ_k across the twelve elections analyzed in this chapter. The values outside the 95% confidence interval (-1.96, 1.96) are for the elections in Oaxaca and Veracruz. In both cases, it is the candidate of the incumbent party which benefits from the irregularities. There is no reason to assume that random error would benefit the incumbent party, but the distribution of political capital among the candidates makes it plausible that non-accidental irregularities were triggered by the manipulation of votes in favor of candidates supported by the local government and the PRI. Given that the irregularities in Durango only occurred in one precinct, the value for ζ_k in Durango falls within the confidence intervals, and thus shows that this test provides robust results when fraud occurs in several locations.

Table 4.2: Statistical summary of the analysis of irregularities in the Analysis group. Values of ζ in bold show those scores outside the 95% confidence interval.

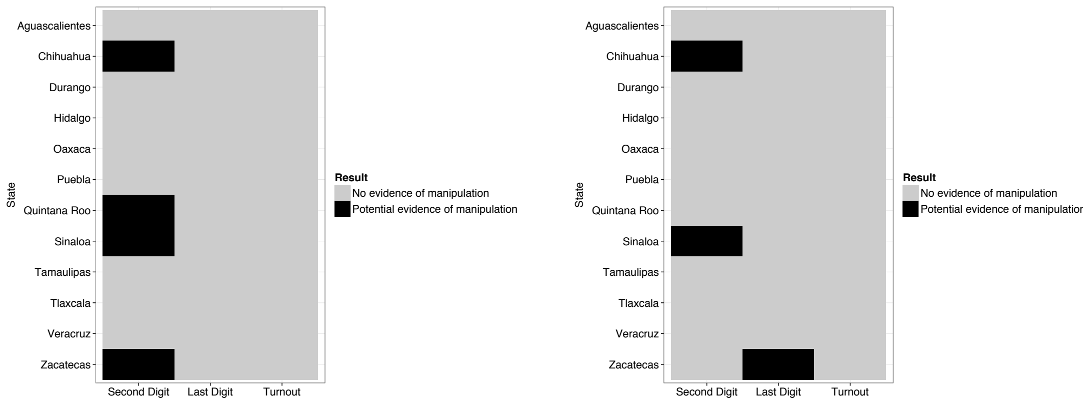
	K	k	D	d	ζ
Aguascalientes	603	550	0.01	0.01	1.00
Chihuahua	1976	86	0.00	0.01	0.89
Durango	1000	59	0.01	0.02	0.92
Hidalgo	1039	70	-0.01	-0.00	0.23
Oaxaca	1812	85	0.01	0.05	**2.61**
Puebla	2209	190	-0.00	-0.00	-0.78
Quintana Roo	284	30	-0.01	-0.01	-0.17
Sinaloa	2099	40	-0.00	0.01	1.35
Tamaulipas	1753	148	0.00	0.01	0.76
Tlaxcala	489	33	-0.00	0.00	0.71
Veracruz	3335	168	0.01	0.06	**4.25**
Zacatecas	1148	43	-0.00	0.00	0.23

4.4 Cross-Validation

The methodology described above finds other type of fraud that is often overlooked using other forensic tools. To demonstrate this claim, I compare the results from my analysis with some of the standard techniques typically used in electoral fraud analysis and discuss the differences found in this comparison.

Although these forensic tools are discussed in detail in Chapter 4, I will briefly summarize the three techniques I am analyzing here. The second-digit Benford's Law test employs a chi-square goodness-of-fit test to establish conformity with Benford's Law given the relative frequency of the second significant digit in the vote counts for any given candidate (Mebane, 2006). The Last Digit test detects an artificial vote record when the last digit does not resemble a uniform distribution, is serially ordered, and has low frequencies of distant numerals and consecutive digits, (Beber and Scacco, 2012, p. 218-220). Finally, the turnout analysis expects that increases in turnout should neither harm nor benefit a candidate in a disproportionate way. In particular, when regressing the number of votes (V) for an alternative as a share of the registered voters in every unit (E) on the turnout level (T), coefficients should fall in the $[0, 1]$ interval (Myagkov, Ordeshook and Shakin, 2009).

The way in which each of the tests provides evidence for manipulation depends upon its specific quantities of interest. For the Second Digit and Last Digit tests, the null hypothesis is rejected when the value of the χ^2 for each test is greater than $\chi^2_{0.05,8}$ for the second digit or $\chi^2_{0.05,9}$ for the last digit. In the case of the turnout analysis, the method raises a red flag when regressing V/E on T, and the coefficient falls outside the interval $[0, 1]$.



(a) Incumbent (b) Challenger

Figure 4.11: Analysis of local elections in Mexico by other fraud detection algorithms .

Figure 4.11 summarizes the analysis of other fraud detection algorithms, and Tables B.7 and B.8 in Appendix 2 show the exact quantities of interest. The second digit test finds irregularities in Chihuahua, Quintana Roo, Sinaloa, and Zacatecas, yet it is difficult to assess how many of these results are false positives, as (Deckert, Myagkov and Ordeshook, 2011) show. The last digit test finds a significant deviation from the expected uniform distribution in the vote counts of the PRI in Zacatecas. Finally, the turnout analysis does not detect any irregularities. None of these techniques finds irregularities in the cases highlighted by the methodology of this paper. As such, the evidence suggests that the methodology proposed in this chapter is a new an technique to find electoral irregularities, which complements the extant research tools.

4.5 Conclusion

This research has mixed implications for Mexico. For what it was a common event in Mexico only a few decades ago, the findings here only show a limited number of electoral irregularities in a few states in 2010, thus indicating that fraud is currently the exception rather than the rule. Nonetheless, the findings also suggest that there are still regions where the interests of local powers are in conflict with national-level regulations in Mexico. Specifically, I provide evidence of electoral irregularities that favored the PRI in Durango, Oaxaca, and Veracruz, which represent a quarter of the electoral contests.

This paper introduces a novel method to uncover electoral manipulation, a characteristic practice of "brown areas," identified by O'Donnell (1998) as those regions in a country where the interests of local powers might conflict with laws that regulates affairs at the national level. Yet, as recent studies on sub-national authoritarianism in Argentina, Brazil, Russia and the Philippines indicate, the empirical identification of those regions is a challenging task (Gervasoni, 2008; Borges, 2008; Sidel, 2004; Gel'Man and Ross, 2010). As long as voters are assigned to polling stations in a way that is not correlated with their electoral behavior, this method can be used fruitfully to analyze electoral contests in other countries.[16]

It should be noted, however, that there are a couple of caveats that require

[16]Examples of elections in which the assignment of voters within electoral wards occurs via family names or voters' identification number are Bolivia, Jamaica, Pakistan, Venezuela, and the U.S. state of Alabama, (Figueros, 1985; Electoral Commission of Pakistan, 2002; Hausmann and Rigobón, 2011; Alabama League of Municipalities, 2011).

further investigation. First, since the identification of suspicious observations is seen empirically at the polling station level, the methodology I propose may not identify fraud when it occurs in all the observations of the precinct. A plausible extension of this analysis could detect manipulated precincts by comparing turnout differences among those precincts with similar observable characteristics. Second, this paper does not evaluate another common type of fraud, namely, the subtraction of votes by nullifying paper ballots. Although this analysis excludes those observations with an atypical number of spoiled votes, it is possible to adapt the methodology to detect differences in voided votes, rather than looking at differences in turnout levels.

Overall, together with an emerging wave of studies on electoral integrity, this paper represents a first step in reducing the complexity of fraud denials and allegations in non-consolidated democracies. On the one hand, this research helps to detect instances in which incumbents might try to "reap the fruits of electoral legitimacy without running the risks of democratic uncertainty" (Schedler, 2002*b*, p.37). On the other hand, this methodology can evaluate fraud allegations in light of the available evidence and can assess the legitimacy of discontent on the part of the electoral losers. Moreover, this methodology provides a tool to help observers and authorities evaluate the strengths and weaknesses of their electoral administration.

Chapter 4 is based on a forthcoming article in the *American Journal of Political Science*. The dissertation author was the primary investigator and author of this paper.

Chapter 5

A Cross-National Evaluation

I expand the analysis of the previous two chapters to explore two elections surrounded by suspicions of electoral fraud: the 2012 presidential elections in Venezuela and Mexico, as well as the 2008 recall referendum in Bolivia. Consistent with anecdotal testimony, my analysis provides evidence of electoral manipulation in Bolivia. In contrast, the results are less concrete and, in fact, insufficient to support a similar hypothesis of fraud in Mexico and Venezuela. The substantive contribution of this chapter is the exhaustive forensic analyses of both elections. In the case of the Bolivian referendum, this is the first concrete research of electoral irregularities in this country. Each case study provides a description of the election and the allegations of fraud made by opposition parties, citizens, and international observers. It also includes a forensic analysis of the results and concludes by discussing those results together with what the findings of the previous chapter.

5.1 Mexico, 2012

On the night of July 1, 2012, after the polling stations closed and the president of the Federal Electoral Institute (IFE) congratulated citizens and poll workers for the success of the federal election, rumors of electoral manipulation began to spread via social media. Pictures showing burned paper ballots, videos revealing that marks on paper ballots were easy to erase, and blogs discussing inconsistencies between online results and those at the ballot boxes became viral in the days following the election.

Election observers and NGOs stated that the 2012 election was characterized by "calmness, respect, and order" (Organization of American States, August 11, 2008). Voters, however, felt differently. Mexicans had reasons to be skeptical of the integrity of this election. Throughout most of the twentieth century, elections in Mexico failed to function as a legitimate process for selecting public officials. Events such as the 1987 gubernatorial election in Chihuahua (Preston and Dillon, 2004, ch. 4) and the 1988 presidential election (Castañeda, 2000) reinforced citizens' beliefs that elections were a mockery of true political competition between parties.

The gradual democratization of the electoral process in Mexico led to electoral reforms in the mid-1990s (Ochoa-Reza, 2004), delegating the federal electoral administration to the IFE, which included a six-member council elected by a two-thirds vote in Congress and proposed by the three main political parties (Magaloni, 2006, p. 38). This institution not only brought professionalization to the electoral administration, but also increased the credibility regarding its impartiality of the institution (Lehoucq, 2002). After the Institutional Revolutionary Party's (PRI) defeat in 2000, the IFE was among the most trusted public institutions in the country with nearly seven out of 10 citizens expressing a positive opinion about this regulatory body (Mendizabal and Moreno, 2010).

However, the general opinion about the efficiency and impartiality of the electoral process vanished in anticipation of the 2006 election. In 2003, only the PRI and PAN took part in the renewal of the IFE's council, leaving the Party of the Democratic Revolution (PRD) outside of the decision. Furthermore, a year before the election, both the PRI and PAN attempted to impeach Mexico City's major, Andrés Manuel López Obrador, a PRD politician who was leading the presidential polls at that time. Public perception was that the impeachment was an illegitimate attempt to eliminate López Obrador from the contest, and the impeachment was ultimately withdrawn (Lawson, 2007, p. 46). The 2006 results declared PAN's Felipe Calderón the winner of the presidential election with a margin of victory of 0.56% against López Obrador. López Obrador rejected the results until a "full recount" of the votes was conducted, which was turned down by the electoral court. The polarization of the campaigns and the unresolved inconsistencies of the electoral process harmed the perception of the IFE. Negative opinions of this institution went from six percent to 20 percent between June and September of 2006 (Eisenstadt and Poiré, 2006; Pastor, 2006).

General skepticism regarding the integrity of the electoral administration continued throughout the 2012 presidential election. This time, PRI's Enrique Peña Nieto won the election by a 6.5%, yet López Obrador pronouncements against the IFE continued the distrust of the impartiality of the process (The New York Times, July 6, 2012, p. A5). Five weeks later, a poll by a national newspaper showed that 39 percent of citizens believed that elections were not clean (El Universal, August 13, 2012).

The analyses for the 2006 election are divided between those who claim deliberate fraud (Raphael, 2007; Gallardo, 2009; Mebane, 2007; Mochán, 2011) and those who explain irregularities as human errors (Crespo, 2006; Poiré and Estrada, 2006; Aparicio, 2006). Forensics of the 2012 election are scarce and fall short in explaining the assumptions behind the statistical techniques they use (Mochán, 2012). As a way to provide an alternative analysis to the 2012 Mexican presidential election, I analyze it using the election forensic tools discussed in the previous section.

5.1.1 Results

The results for the federal elections in 2006 and 2012 are available at the Electoral Federal Institute's (IFE) website.[1] With the exception of the flow of votes test, the analyses in this section uses data at the polling station level. I measured turnout by dividing the number of votes cast by the number of registered voters at every polling station. As Figure 5.1 shows, the turnout distributions across states are very similar for the last two presidential elections. No bimodal distributions

[1] http://www.ife.org.mx

appeared, and the significant increases of turnout in Chiapas, Guerrero, Yucatan, and Zacatecas occurred in a uniform way across the board at all polling stations for all these cases.

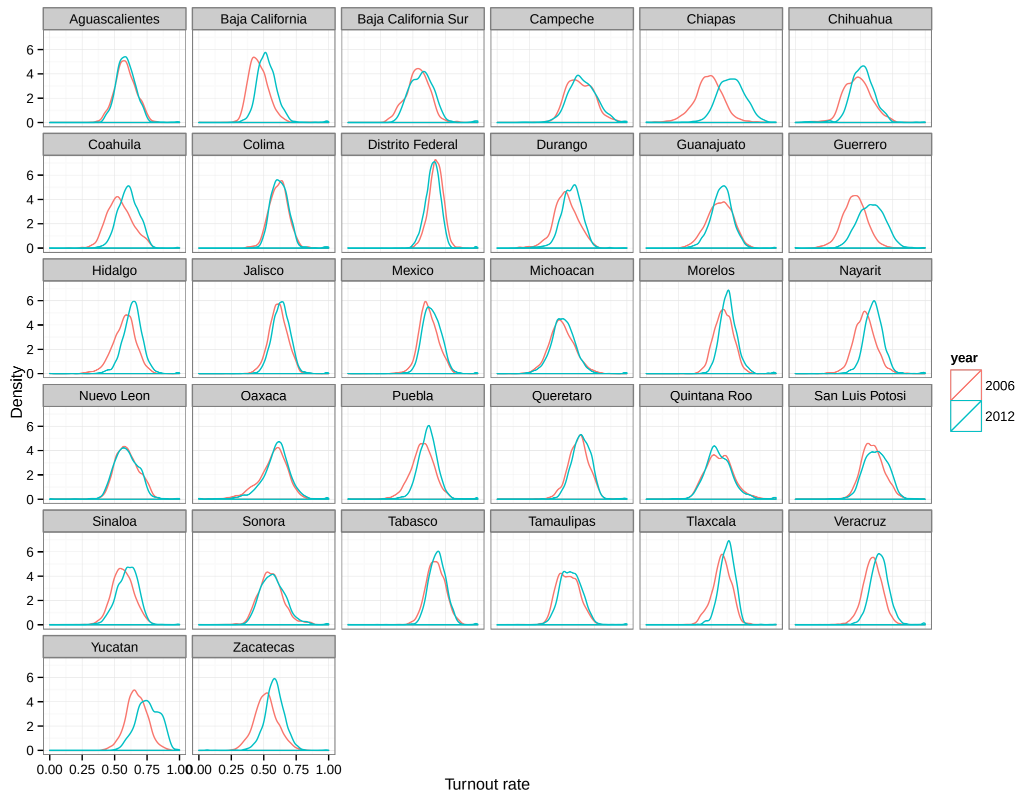


Figure 5.1: Turnout distribution by state for the 2006 and 2012 presidential elections in Mexico.

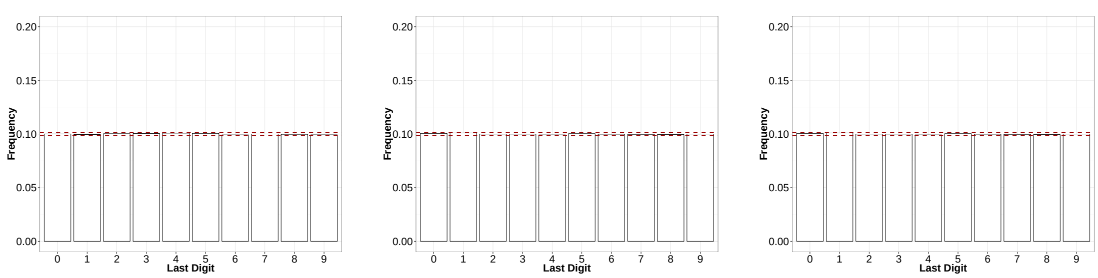
To assess whether variations in turnout rates benefited a candidate in a disproportionate way, I regress the turnout level on the share of votes for each of the three most important candidates. None of the state regressions for the three most important candidates in Table 5.1 shows a value for β above 1, which suggests that increments in turnout benefitted the candidates in a proportional way. Moreover, the regressions for the López Obrador in Campeche and Yucatan are the only two cases in which the coefficient for β is negative; yet, as Myagkov, Ordeshook and Shakin (2009) discuss, a negative sign can be a product of high heterogeneity of the data. The comparison of the regressions for the PAN and PRI in both states show relatively high values for the R^2, which may explain that the relationship between turnout and the PRD vote is almost nonexistent.

Table 5.1: Regression Results from Turnout versus Votes for each Candidate.

	$\beta_{V_{PRI}}$	R^2_{EPN}	$\beta_{V_{PRD}}$	R^2_{PRD}	$\beta_{V_{PAN}}$	R^2_{PAN}
Aguascalientes	0.04	0.00	0.38	0.31	0.58	0.40
Baja California	0.24	0.20	0.25	0.18	0.48	0.52
Baja California Sur	0.47	0.38	0.26	0.20	0.21	0.15
Campeche	0.62	0.62	-0.02	0.00	0.34	0.31
Chiapas	0.70	0.37	0.19	0.04	0.06	0.01
Chihuahua	0.47	0.34	0.09	0.03	0.40	0.43
Coahuila	0.13	0.03	0.26	0.17	0.59	0.40
Colima	0.37	0.18	0.28	0.14	0.35	0.25
Distrito Federal	0.22	0.15	0.12	0.01	0.64	0.36
Durango	0.35	0.11	0.25	0.13	0.37	0.24
Guanajuato	0.28	0.12	0.16	0.08	0.52	0.34
Guerrero	0.38	0.24	0.46	0.29	0.10	0.05
Hidalgo	0.37	0.15	0.39	0.15	0.20	0.07
Jalisco	0.12	0.02	0.28	0.11	0.56	0.29
Mexico	0.45	0.20	0.05	0.00	0.48	0.29
Michoacan	0.40	0.33	0.23	0.10	0.31	0.24
Morelos	0.31	0.17	0.50	0.31	0.17	0.06
Nayarit	0.43	0.21	0.28	0.08	0.27	0.13
Nuevo Leon	0.00	0.00	0.21	0.19	0.78	0.56
Oaxaca	0.43	0.27	0.41	0.20	0.16	0.06
Puebla	0.31	0.10	0.25	0.06	0.41	0.19
Queretaro	0.12	0.01	0.25	0.10	0.63	0.35
Quintana Roo	0.54	0.46	0.16	0.10	0.27	0.18
San Luis Potosí	0.45	0.24	0.20	0.07	0.29	0.15
Sinaloa	0.40	0.19	0.26	0.11	0.31	0.14
Sonora	0.39	0.36	0.24	0.13	0.32	0.25
Tabasco	0.35	0.13	0.52	0.18	0.12	0.05
Tamaulipas	0.38	0.26	0.20	0.12	0.39	0.31
Tlaxcala	0.40	0.19	0.39	0.13	0.19	0.07
Veracruz	0.48	0.24	0.17	0.02	0.34	0.09
Yucatan	0.56	0.38	-0.11	0.02	0.56	0.39
Zacatecas	0.29	0.14	0.22	0.10	0.48	0.30

Figure 5.2 shows the histograms of the last digit distribution for the three candidates in 2012. While the dashed lines display the 95% confidence intervals

for the frequency of each digit when the expected density is 0.10. The black bars indicate observations falling outside those intervals. Figure 5.3 shows the p-values for the chi-square tests in every state. Departures of the last digit from an uniform distribution only appear in Chihuahua and Guerrero for the PRI's candidate and in Hidalgo and Guerrero for the PAN's candidate. For the case of Guerrero, these deviations are not due to manipulation, but rather to heterogeneity of the data once one considers that both candidates received some of their worst results in that state. Figures C.1 to C.3 in the Appendix show the histograms of the last digit distribution by state. Out of the 320 observations in the analysis for each candidate (i.e., 10 densities for each of the 32 states), there are 20, 17, and 18 frequencies falling outside the confidence intervals for Peña Nieto, López Obrador, and Vazquez Mota. These observations represent approximately five percent of total observations, and I cannot reject the hypothesis that these outliers are purely generated by chance.



(a) Peña Nieto (b) López Obrador (c) Vazquez Mota

Figure 5.2: Last digit distributions for the candidates' vote counts at national level.

As an alternative way to evaluate the last digits in the vote counts, Figure 5.4 shows the frequency of the last two digits in the vote returns for each candidate, with

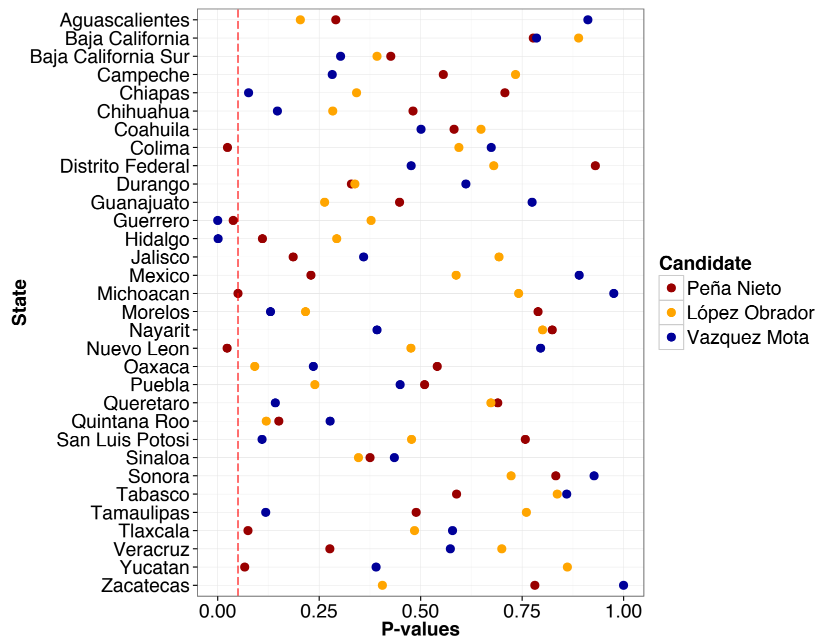
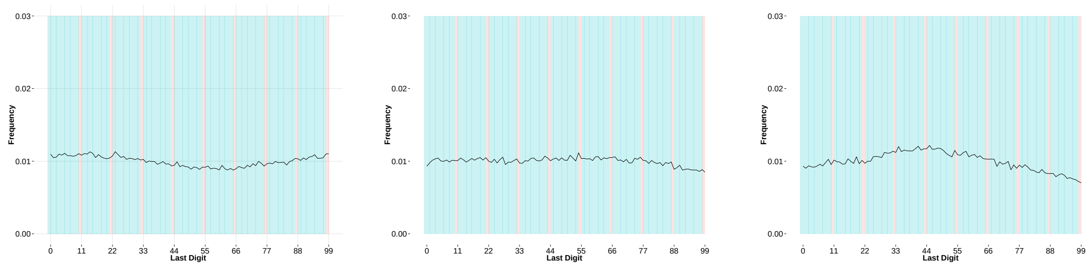


Figure 5.3: P-values for the Chi-squared tests for the Last digit distribution by state and candidate. Observations to the left of the red line are for those p-values below 0.05.

the shaded regions indicating the location for the density of the repeated digits (e.g., 11, 22, 33). There is not a clear pattern in the distribution of the last two digits, and I cannot confirm that the density for this group of numbers is more different than for others.

Finally, table 5.5 shows the Chi-squared test for departures from the Benford distribution. Departures are more frequent for the PRI candidate, but there is no pattern suggesting how manipulation occurred. It is worth noticing that the second-digit distributions for all three candidates are different from those expected in Mexico City, where we should expect fraud to be less prevalent. The results of the analysis of the second digit are, therefore, inconclusive.



(a) Peña Nieto (b) López Obrador (c) Vazquez Mota

Figure 5.4: Last two digit distributions for the candidates' vote counts at national level.

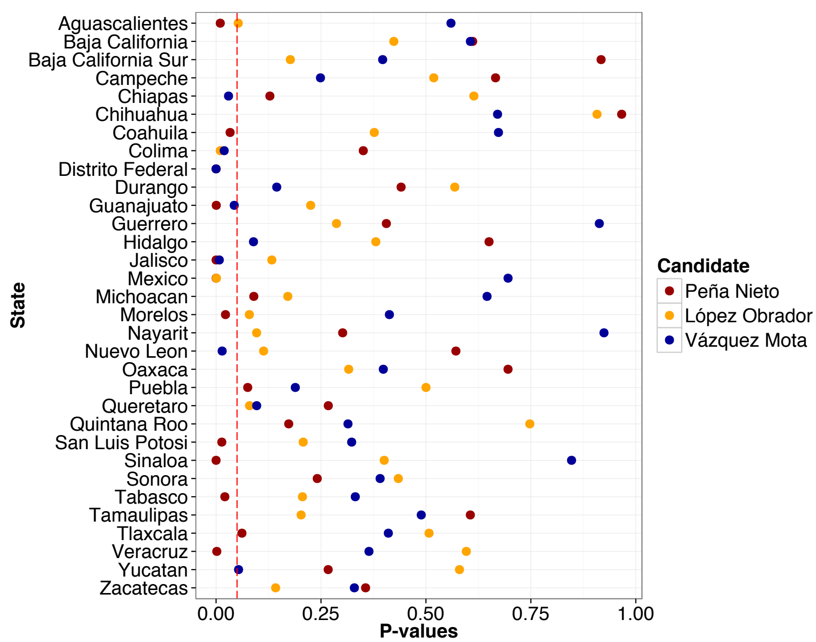


Figure 5.5: P-values for chi-squared tests for the second digit distribution by candidate and state. Observations to the left of the red line are for those p-values below 0.05.

For the analysis of the flow of votes, Table 5.2 shows the results of the Goodman regressions using random effects for municipalities. Unless fraud occurred in the same municipalities in both elections, the comparison of the election results in 2006 and 2012 does not find anything unexpected. Almost all the coefficients adding up to values closer to one and the negative coefficients suggesting a negative correlation between the shares of vote in both elections (Myagkov, Ordeshook and Shakin, 2005). To find outlier deviations from the mean effects by municipality, Figure 5.6 shows the random effects when the dependent variable is the proportion of votes for PRI's Enrique Peña Nieto. Out of the 2,204 municipalities in the analysis, only 52 showed flows in the PRI votes from 2006 to 2012 greater than 100%. The biggest outlier is for the municipality of Belisario Dominguez in Chihuahua, where the proportion of votes for PRI's Roberto Madrazo in 2006 received by Peña Nieto in 2012 is equal to 402 percent, which is a disturbing outcome as it falls far away from the rest of the estimations in the analysis. The remaining outliers are in very populated municipalities such as Ixtapaluca and Naucalpan in the State of Mexico, where the heterogeneity of the population may explain those results. In contrast, when the dependent variable is the proportion of votes for López Obrador, the biggest outliers are in the municipalities of San Miguel Ejutla and Santa Cruz Mixtepec, both in the state of Oaxaca, with a flow of votes in 2012 of 151% and 125% with respect to what he got in 2006. These municipalities are under the "Usos and Costumbres" denomination, which implications extend to state and federal elections and may explain the overwhelming change of the votes in the municipalities between the two elections (Benton, 2013).[2]

[2] The "Usos and Costumbres" system is a legal recognition of indigenous customary laws to select local leaders via practices that go from community assemblies to council of elders meetings. The system exists in 418 municipalities in the state of Oaxaca, from which the secret vote is only guaranteed in 11 percent of them (Eisenstadt, 2007).

Table 5.2: Flow of votes between the 2006 and 2012 presidential elections in Mexico.

	PRI 06	PRD 06	PAN 06	OTRO 06	ABST 06
PRI 12	0.77	0.15	0.21	0.24	0.15
PRD 12	-0.04	0.65	0.06	0.13	0.09
PAN 12	0.10	-0.05	0.53	0.04	0.12
PANAL 12	0.00	0.01	0.02	0.10	0.01
NULL 12	0.02	0.00	0.00	0.13	0.02
ABST 12	0.12	0.22	0.15	0.35	0.63
Column Sum	0.98	0.98	0.98	1.00	1.02

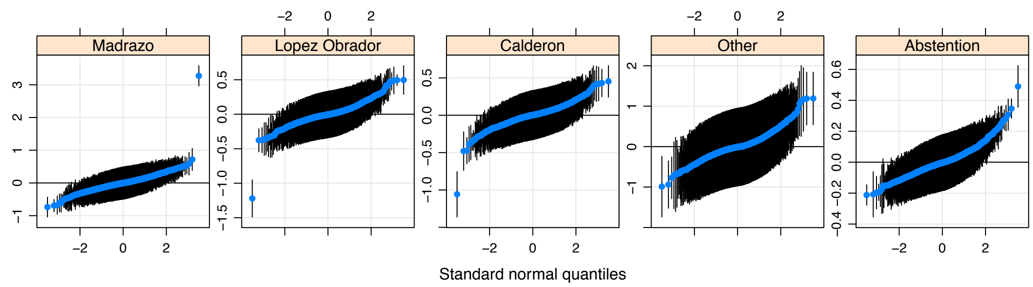


Figure 5.6: Goodman regression coefficients by municipality with votes for Peña Nieto as the dependent variable.

Finally, I apply the methodology described in the previous chapter to find outlier observations given the election turnout levels and the share of votes for the candidates. The analysis groups the observation by state, and the complete results are found in the Appendix. This methodology identifies those polling stations with a relative high turnout and strong support for any of the candidates. As an additional step to understand the reasons behind the irregularities found using this method, I take advantage of the scanned copies of all the ballot acts that were available online the night after the election.[3]

Figures 5.8 and 5.9 show the outlier analysis for Peña Nieto and López

[3]To show the advantages of the methodology, I use the preliminary results published the day after the elections and available at https://prep2012.ife.org.mx/prep/

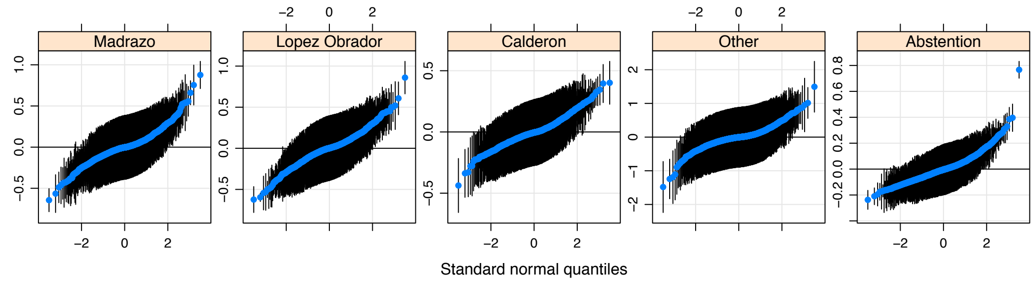


Figure 5.7: Goodman regression coefficients by municipality with votes for López Obrador as the dependent variable.

Obrador in the State of Mexico and Mexico City, which are their respective regional strongholds. The outliers in every case do not affect the results of any candidate in a significant way. Most of the observations outside of the confidence intervals in the analysis group are due to three different types of errors. First, the most common irregularity is due to errors during the vote count process, in which poll workers wrote down the total number of votes for a candidate in every field for each party coalition. Consider the example of Figure 5.10, which shows the scanned copy from a polling station in Guanajuato. In this case, the poll workers filled out the 212 votes for Peña Nieto and 51 votes for López Obrador in each combination of their respective party coalitions (i.e., PRI, PVEM, and PRI-PVEM for Peña Nieto and PRD, PT, Movimiento Ciudadano, PRD-PT-Movimiento Ciudadano, PRD-PT, PRD-Movimiento Ciudadano, and PT-Movimiento Ciudadano for López Obrador).

A second type of irregularity come from mistakes made while writing the number of votes in full. For example, Figure 5.11 shows the case in which the votes for both PRI and PVEM is written down as "two hundred and twenty two," while the

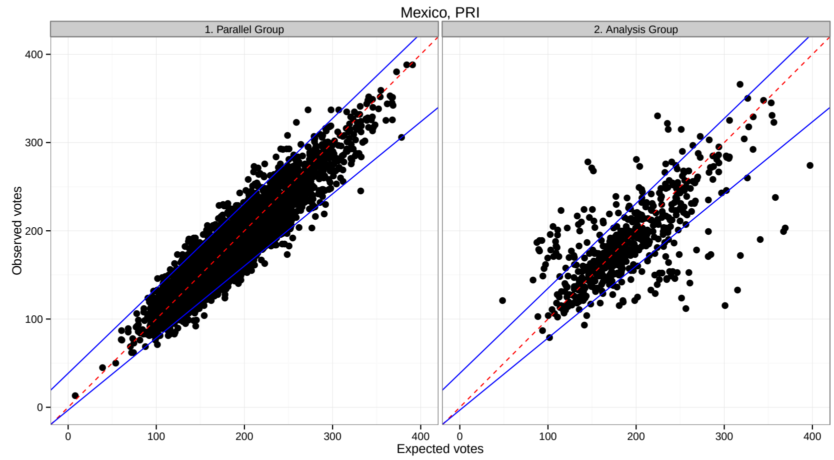


Figure 5.8: Outlier analysis for Peña Nieto in the State of Mexico.

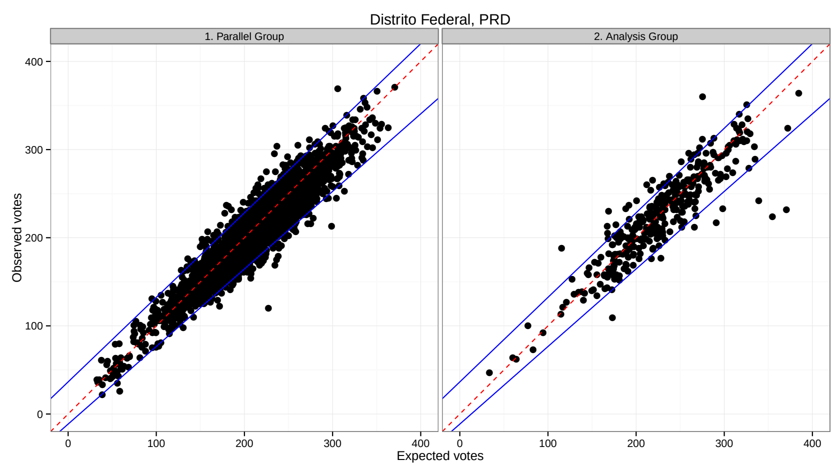


Figure 5.9: Outlier analysis for López Obrador in Mexico City.

number of votes is twenty-two. This type of error was very common during the vote count process without evidence that significantly affected the results for a particular party.

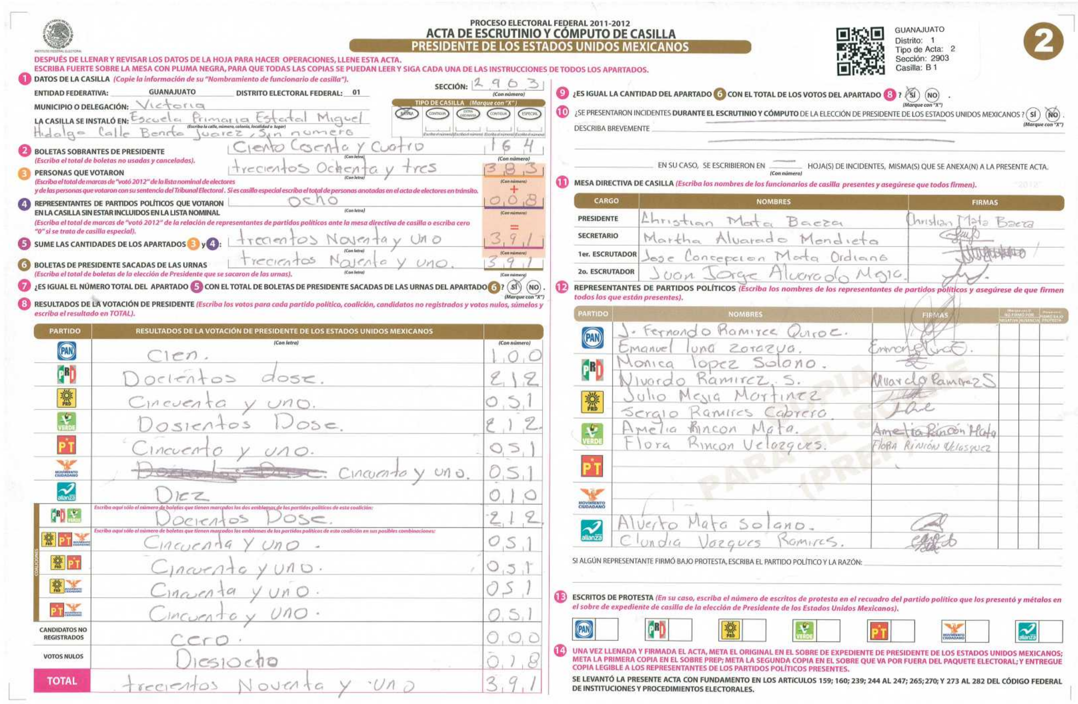
Finally, a less frequent type of incident come from typos by IFE officials when recording the results of the polling stations into the database. Consider the case in Figure 5.12, which shows the results in the ballot act and the website for a polling station in Oaxaca. In this case, the website shows that the number of votes for the PRD is 92, while the votes counted in the polling station is 192. These errors were marginal and did not affect any of the candidates in a significant way.

Figures C.4 to C.9 in the Appendix show the results after the recount conducted four days latter after the election. Most of the errors found during the precount process were eliminated, and the remaining irregularities are typing errors made while processing the final results and do not significantly affect a particular candidate.[4]

5.2 Venezuela

On the night of October 7, 2012, the National Electoral Court (CNE) announced the victory of Hugo Chávez in his re-election. Everyone expected a speech from the opposition candidate, Henrique Capriles, to challenge the validity of the results. Despite the overwhelming campaign resources and media coverage in fa-

[4]The list of remaining irregularities in the database have been already sent to the IFE.



PROCESO ELECTORAL FEDERAL 2011-2012
ACTA DE ESCRUTINIO Y CÓMPUTO DE CASILLA
PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS

GUANAJUATO
Distrito: 1
Tipo de Acta: 2
Sección: 2903
Casilla: B 1

DESPUÉS DE LLENAR Y REVISAR LOS DATOS DE LA HOJA PARA HACER OPERACIONES, LLENE ESTA ACTA.
ESCRIBA FUERTE SOBRE LA MESA CON PLUMA NEGRA, PARA QUE TODAS LAS COPIAS SE PUEDAN LEER Y SIGA CADA UNA DE LAS INSTRUCCIONES DE TODOS LOS APARTADOS.

1 DATOS DE LA CASILLA
ENTIDAD FEDERATIVA: GUANAJUATO — DISTRITO ELECTORAL FEDERAL: 01 — SECCIÓN: 2 9 0 3
MUNICIPIO O DELEGACIÓN: Victoria
LA CASILLA SE INSTALÓ EN: Escuela Primaria Estatal Miguel Hidalgo Calle Benito Juarez Sin numero

2 BOLETAS SOBRANTES DE PRESIDENTE: Ciento Sesenta y Cuatro — 1 6 4
3 PERSONAS QUE VOTARON: trecientos Ochenta y tres — 3 8 3
4 REPRESENTANTES DE PARTIDOS POLÍTICOS QUE VOTARON EN LA CASILLA SIN ESTAR INCLUIDOS EN LA LISTA NOMINAL: ocho — 0 0 8
5 SUME LAS CANTIDADES DE LOS APARTADOS 3 y 4: trecientos Noventa y uno — 3 9 1
6 BOLETAS DE PRESIDENTE SACADAS DE LAS URNAS: trecientos Noventa y uno — 3 9 1
7 ¿ES IGUAL EL NÚMERO TOTAL DEL APARTADO 5 CON EL TOTAL DE BOLETAS DE PRESIDENTE SACADAS DE LAS URNAS DEL APARTADO 6? SÍ NO
8 RESULTADOS DE LA VOTACIÓN DE PRESIDENTE

PARTIDO	RESULTADOS DE LA VOTACIÓN (Con letra)	(Con número)
PAN	Cien.	1 0 0
PRI	Docientos dose.	2 1 2
PRD	Cincuenta y uno.	0 5 1
VERDE	Dosientos Dose.	2 1 2
PT	Cincuento y uno.	0 5 1
MOVIMIENTO CIUDADANO	~~Docientos Dose~~ Cincuenta y uno.	0 5 1
NUEVA ALIANZA	Diez.	0 1 0
PRI-VERDE	Docientos Dose.	2 1 2
PRD-PT-MC	Cincuenta y uno.	0 5 1
PRD-PT	Cincuenta y uno.	0 5 1
PRD-MC	Cincuenta y uno.	0 5 1
PT-MC	Cincuenta y uno.	0 5 1
CANDIDATOS NO REGISTRADOS	Cero.	0 0 0
VOTOS NULOS	Diesiocho	0 1 8
TOTAL	trecientos Noventa y uno	3 9 1

9 ¿ES IGUAL LA CANTIDAD DEL APARTADO 6 CON EL TOTAL DE LOS VOTOS DEL APARTADO 8? SÍ NO
10 ¿SE PRESENTARON INCIDENTES DURANTE EL ESCRUTINIO Y CÓMPUTO DE LA ELECCIÓN DE PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS? SÍ NO
11 MESA DIRECTIVA DE CASILLA

CARGO	NOMBRES	FIRMAS
PRESIDENTE	Christian Mata Baeza	Christian Mata Baeza
SECRETARIO	Martha Alvarado Mendieta	
1er. ESCRUTADOR	Jose Concepcion Mata Ordiano	
2o. ESCRUTADOR	Juan Jorge Alvarado Mejia	

12 REPRESENTANTES DE PARTIDOS POLÍTICOS

PARTIDO	NOMBRES	FIRMAS
PAN	J. Fernando Ramirez Quiroz.	
	Emanuel luna zarazua.	Emanuel
PRI	Monica lopez Solano.	
	Alvarado Ramirez S.	Alvarado Ramirez S
PRD	Julio Mejia Martinez	
	Sergio Ramires Cabrera	
VERDE	Amelia Rincon Mata.	Amelia Rincon Mata
	Flora Rincon Velazques.	Flora Rincon Velazquez
PT		
MOVIMIENTO CIUDADANO		
NUEVA ALIANZA	Alverto Mata Solano.	
	Claudia Vazques Ramires.	

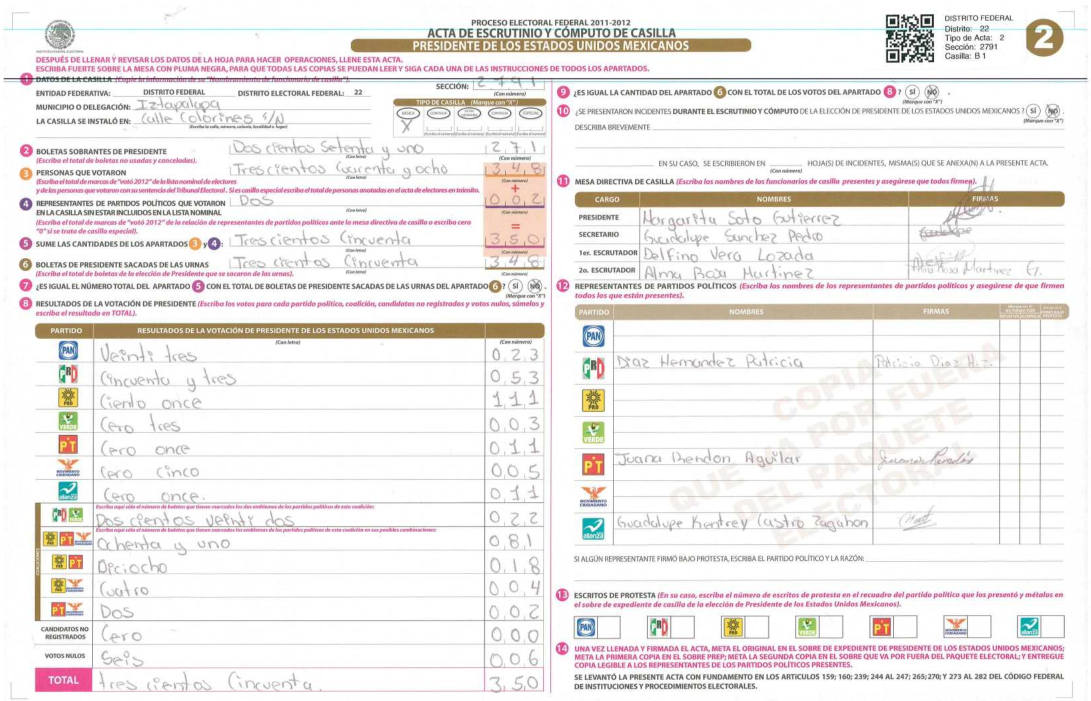
13 ESCRITOS DE PROTESTA
14 UNA VEZ LLENADA Y FIRMADA EL ACTA, META EL ORIGINAL EN EL SOBRE DE EXPEDIENTE DE PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS; META LA PRIMERA COPIA EN EL SOBRE PREP; META LA SEGUNDA COPIA EN EL SOBRE QUE VA POR FUERA DEL PAQUETE ELECTORAL; Y ENTREGUE COPIA LEGIBLE A LOS REPRESENTANTES DE LOS PARTIDOS POLÍTICOS PRESENTES.
SE LEVANTÓ LA PRESENTE ACTA CON FUNDAMENTO EN LOS ARTÍCULOS 159; 160; 239; 244 AL 247; 265; 270; Y 273 AL 282 DEL CÓDIGO FEDERAL DE INSTITUCIONES Y PROCEDIMIENTOS ELECTORALES.

Figure 5.10: Example of poll workers errors during the counting process.

vor of Chávez, polls showed a very close election between the two candidates. The competitiveness of the election raised alarms from the opposition and international observers. “If Capriles wins by a narrow margin of victory, it is very likely that the government will manipulate the outcome in its favor” predicted the Nobel Prize, and former presidential candidate of Peru, Mario Vargas Llosa (El País, October 7, 2012, p. 4). Yet, to the surprise of many, Capriles accepted the result and rejected any accusations of electoral manipulation, “I say it clearly, there was no fraud. (...) Had there been anything affecting the results.”[5]

Historically, the debate regarding the integrity of Venezuelan elections began

[5] (Noticias24, 2012). http://www.noticias24.com/venezuela/noticia/130475/en-video-aqui-no-hubo-fraude-lo-digo-clarito-asegura-capriles-ante-rumores/.


PROCESO ELECTORAL FEDERAL 2011-2012
ACTA DE ESCRUTINIO Y CÓMPUTO DE CASILLA
PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS

DISTRITO FEDERAL
Distrito: 22
Tipo de Acta: 2
Sección: 2791
Casilla: B 1

2

DESPUÉS DE LLENAR Y REVISAR LOS DATOS DE LA HOJA PARA HACER OPERACIONES, LLENE ESTA ACTA.
ESCRIBA FUERTE SOBRE LA MESA CON PLUMA NEGRA, PARA QUE TODAS LAS COPIAS SE PUEDAN LEER Y SIGA CADA UNA DE LAS INSTRUCCIONES DE TODOS LOS APARTADOS.

1 DATOS DE LA CASILLA (Copie la información de su "Nombramiento de funcionario de casilla").

SECCIÓN: 2791 (Con número)

ENTIDAD FEDERATIVA: DISTRITO FEDERAL DISTRITO ELECTORAL FEDERAL: 22

MUNICIPIO O DELEGACIÓN: Iztapalapa

LA CASILLA SE INSTALÓ EN: Calle Colorines S/N (Escriba la calle, número, colonia, localidad o lugar)

TIPO DE CASILLA (Marque con "X"): BÁSICA X | CONTIGUA | EXTRAORDINARIA | CONTIGUA | ESPECIAL

2 BOLETAS SOBRANTES DE PRESIDENTE (Escriba el total de boletas no usadas y canceladas). Dos cientos Setenta y uno (Con letra) 2 7 1 (Con número)

3 PERSONAS QUE VOTARON (Escriba el total de marcas de "votó 2012" de la lista nominal de electores y de las personas que votaron con su sentencia del Tribunal Electoral. Si es casilla especial escriba el total de personas anotadas en el acta de electores en tránsito.) Trescientos Curenta y ocho (Con letra) 3 4 8 (Con número)

\+

4 REPRESENTANTES DE PARTIDOS POLÍTICOS QUE VOTARON EN LA CASILLA SIN ESTAR INCLUIDOS EN LA LISTA NOMINAL (Escriba el total de marcas de "votó 2012" de la relación de representantes de partidos políticos ante la mesa directiva de casilla o escriba cero "0" si se trata de casilla especial). Dos (Con letra) 0 0 2 (Con número)

=

5 SUME LAS CANTIDADES DE LOS APARTADOS 3 y 4: Trescientos Cincuenta (Con letra) 3 5 0 (Con número)

6 BOLETAS DE PRESIDENTE SACADAS DE LAS URNAS (Escriba el total de boletas de la elección de Presidente que se sacaron de las urnas). Tres cientos Cincuenta (Con letra) 3 4 8 (Con número)

7 ¿ES IGUAL EL NÚMERO TOTAL DEL APARTADO 5 CON EL TOTAL DE BOLETAS DE PRESIDENTE SACADAS DE LAS URNAS DEL APARTADO 6? SÍ NO [X] (Marque con "X")

8 RESULTADOS DE LA VOTACIÓN DE PRESIDENTE (Escriba los votos para cada partido político, coalición, candidatos no registrados y votos nulos, súmelos y escriba el resultado en TOTAL).

PARTIDO	RESULTADOS DE LA VOTACIÓN DE PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS (Con letra)	(Con número)
PAN	Veinti tres	0 2 3
PRI	Cincuenta y tres	0 5 3
PRD	Ciento once	1 1 1
VERDE	Cero tres	0 0 3
PT	Cero once	0 1 1
MOVIMIENTO CIUDADANO	Cero cinco	0 0 5
alianza	Cero once.	0 1 1
COALICIONES: PRI + VERDE	Escriba aquí sólo el número de boletas que tienen marcados los dos emblemas de los partidos políticos de esta coalición: Dos cientos veinti dos	0 2 2
PRD + PT + MOVIMIENTO CIUDADANO	Escriba aquí sólo el número de boletas que tienen marcados los emblemas de los partidos políticos de esta coalición en sus posibles combinaciones: Ochenta y uno	0 8 1
PRD + PT	Deciocho	0 1 8
PRD + MOVIMIENTO CIUDADANO	Cuatro	0 0 4
PT + MOVIMIENTO CIUDADANO	Dos	0 0 2
CANDIDATOS NO REGISTRADOS	Cero	0 0 0
VOTOS NULOS	Seis	0 0 6
TOTAL	tres cientos Cincuenta.	3 5 0

9 ¿ES IGUAL LA CANTIDAD DEL APARTADO 6 CON EL TOTAL DE LOS VOTOS DEL APARTADO 8? SÍ NO [X] (Marque con "X")

10 ¿SE PRESENTARON INCIDENTES DURANTE EL ESCRUTINIO Y CÓMPUTO DE LA ELECCIÓN DE PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS? SÍ NO [X] (Marque con "X")
DESCRIBA BREVEMENTE

EN SU CASO, SE ESCRIBIERON EN ____ (Con número) HOJA(S) DE INCIDENTES, MISMA(S) QUE SE ANEXA(N) A LA PRESENTE ACTA.

11 MESA DIRECTIVA DE CASILLA (Escriba los nombres de los funcionarios de casilla presentes y asegúrese que todos firmen).

CARGO	NOMBRES	FIRMAS
PRESIDENTE	Margarita Soto Gutierrez	[signature]
SECRETARIO	Guadalupe Sanchez Pedro	[signature]
1er. ESCRUTADOR	Delfino Vera Lozada	[signature]
2o. ESCRUTADOR	Alma Rosa Martinez	Alma Rosa Martinez G.

12 REPRESENTANTES DE PARTIDOS POLÍTICOS (Escriba los nombres de los representantes de partidos políticos y asegúrese de que firmen todos los que están presentes).

PARTIDO	NOMBRES	FIRMAS
PAN		
PRI	Diaz Hernandez Patricia	Patricia Diaz H.
PRD		
VERDE		
PT	Juana Rendon Aguilar	[signature]
MOVIMIENTO CIUDADANO		
alianza	Guadalupe Kentrey Castro Zaguhon	[signature]

SI ALGÚN REPRESENTANTE FIRMÓ BAJO PROTESTA, ESCRIBA EL PARTIDO POLÍTICO Y LA RAZÓN:

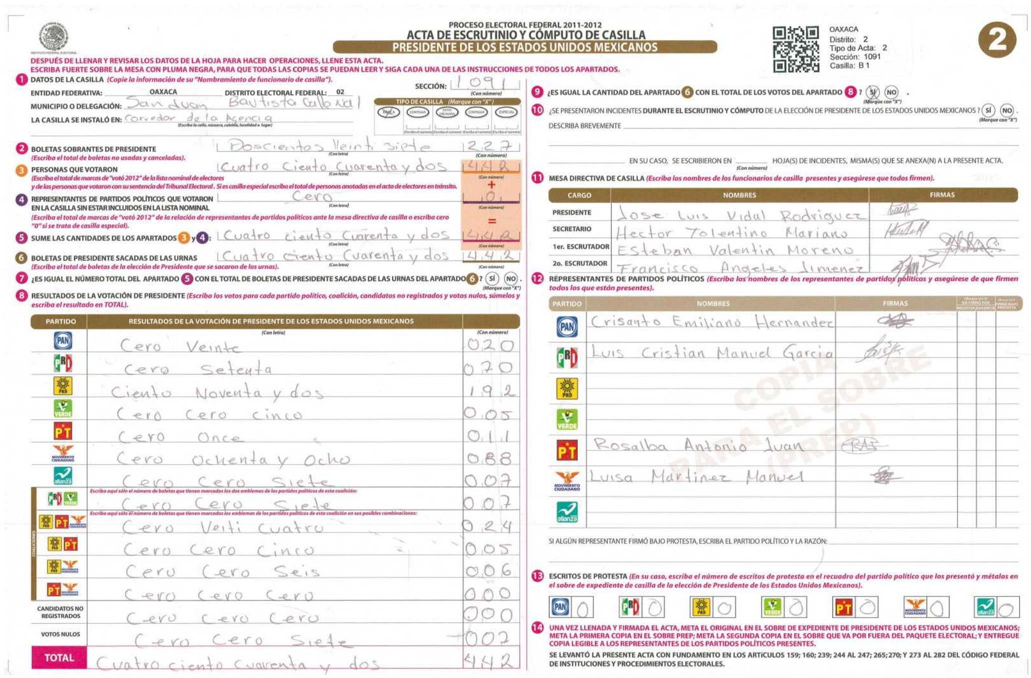
13 ESCRITOS DE PROTESTA (En su caso, escriba el número de escritos de protesta en el recuadro del partido político que los presentó y métalos en el sobre de expediente de casilla de la elección de Presidente de los Estados Unidos Mexicanos).

14 UNA VEZ LLENADA Y FIRMADA EL ACTA, METE EL ORIGINAL EN EL SOBRE DE EXPEDIENTE DE PRESIDENTE DE LOS ESTADOS UNIDOS MEXICANOS; META LA PRIMERA COPIA EN EL SOBRE PREP; META LA SEGUNDA COPIA EN EL SOBRE QUE VA POR FUERA DEL PAQUETE ELECTORAL; Y ENTREGUE COPIA LEGIBLE A LOS REPRESENTANTES DE LOS PARTIDOS POLÍTICOS PRESENTES.

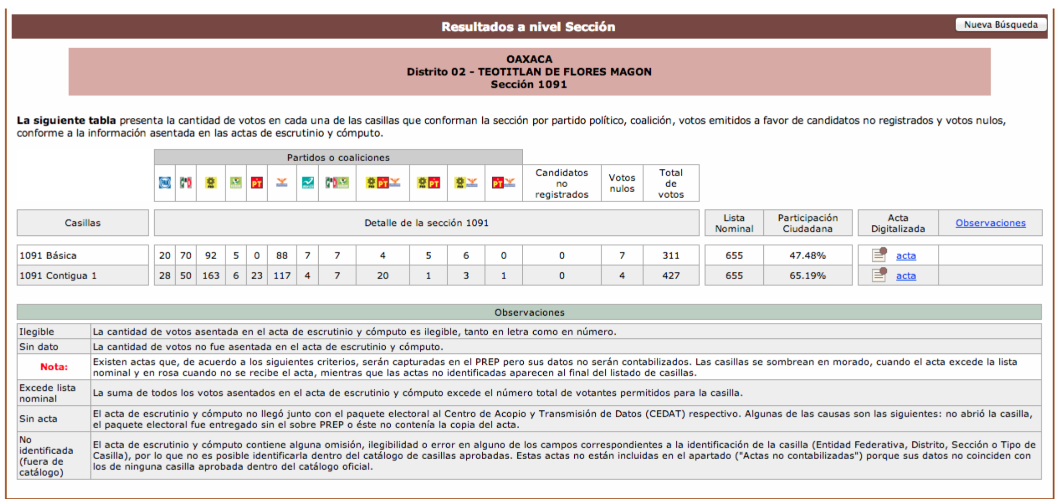
SE LEVANTÓ LA PRESENTE ACTA CON FUNDAMENTO EN LOS ARTÍCULOS 159; 160; 239; 244 AL 247; 265; 270; Y 273 AL 282 DEL CÓDIGO FEDERAL DE INSTITUCIONES Y PROCEDIMIENTOS ELECTORALES.

COPIA QUE VA POR FUERA DEL PAQUETE ELECTORAL

Figure 5.11: Example of incongruences made while writing the number of votes in full.



(a) Ballots



Resultados a nivel Sección

Nueva Búsqueda

OAXACA
Distrito 02 - TEOTITLAN DE FLORES MAGON
Sección 1091

La siguiente tabla presenta la cantidad de votos en cada una de las casillas que conforman la sección por partido político, coalición, votos emitidos a favor de candidatos no registrados y votos nulos, conforme a la información asentada en las actas de escrutinio y cómputo.

Casillas	Partidos o coaliciones												Candidatos no registrados	Votos nulos	Total de votos	Lista Nominal	Participación Ciudadana	Acta Digitalizada	Observaciones
1091 Básica	20	70	92	5	0	88	7	7	4	5	6	0	0	7	311	655	47.48%	acta	
1091 Contigua 1	28	50	163	6	23	117	4	7	20	1	3	1	0	4	427	655	65.19%	acta	

Observaciones	
Ilegible	La cantidad de votos asentada en el acta de escrutinio y cómputo es ilegible, tanto en letra como en número.
Sin dato	La cantidad de votos no fue asentada en el acta de escrutinio y cómputo.
Nota:	Existen actas que, de acuerdo a los siguientes criterios, serán capturadas en el PREP pero sus datos no serán contabilizados. Las casillas se sombrean en morado, cuando el acta excede la lista nominal y en rosa cuando no se recibe el acta, mientras que las actas no identificadas aparecen al final del listado de casillas.
Excede lista nominal	La suma de todos los votos asentados en el acta de escrutinio y cómputo excede el número total de votantes permitidos para la casilla.
Sin acta	El acta de escrutinio y cómputo no llegó junto con el paquete electoral al Centro de Acopio y Transmisión de Datos (CEDAT) respectivo. Algunas de las causas son las siguientes: no abrió la casilla, el paquete electoral fue entregado sin el sobre PREP o éste no contenía la copia del acta.
No identificada (fuera de catálogo)	El acta de escrutinio y cómputo contiene alguna omisión, ilegibilidad o error en alguno de los campos correspondientes a la identificación de la casilla (Entidad Federativa, Distrito, Sección o Tipo de Casilla), por lo que no es posible identificarla dentro del catálogo de casillas aprobadas. Estas actas no están incluidas en el apartado ("Actas no contabilizadas") porque sus datos no coinciden con los de ninguna casilla aprobada dentro del catálogo oficial.

(b) Website

Figure 5.12: Example of typing errors in the official results.

at the start at the beginning of Chávez's regime in 1999. On the one hand, Venezuela introduced a biometric election system in 2004, which is recognized as one of the best systems in the world to accurately register voters' preferecnes (The Carter Center, October 2012). Voters in every precinct are assigned to polling stations based on the last digits of their national identification card. Using voters' fingerprints, electronic machines register the votes and give the voter a receipt which he or she then places in a separate ballot box. Once the election is over, the machines send the results to the CNE, which automatically uploads the results to its website. After that, there is a random post-electoral audit to double-check the results from the machines with those from the ballot boxes. Because of these checks, international electoral experts state that the system is "reliable, allows for oversight and monitoring by the opposition, and rules out the possibility of a massive fraud that would go undetected" (Arriagada and Woldenberg, 2012, p. 1).

On the other hand, Chávez's regime had strengthened its authoritarian control using a variety of methods. First, Chávez limited the political rights of the opposition. For the 2008 election, the government blacklisted citizens from the opposition from running for office (Corrales, 2010, p. 371). Second, after the electoral results, Chávez forced into exile the former presidential candidate Manuel Rosales and several elected candidates of the opposition were stripped of their offices. One of them was the Governor of Miranda, Henrique Capriles (Corrales, 2010, p. 380). Third, regular harassment and systematic aggression tactics forced various television and radio stations to be kept off the air (Levitsky and Loxton, 2012). Finally, Chávez issued a law obligating broadcasters to air institutional public announcements for at least 10 minutes each day (Arriagada and Woldenberg, 2012, p. 1). Similar to other competitive authoritarian regimes (Levitsky and Way, 2010), Chávez's government held onto its power using unfair political practices and stymying the competition, without needing to manipulate the results on the election day.

To cope with the media bias, the opaque campaign financing, and the strong partisan affinities of the CNE council, the opposition participated in the 2012 process through a uniform front, the Democratic Unity Table (MUD), endorsing Governor Henrique Capriles as its candidate. As Capriles recognized, his campaign was not only against Chávez but also against the State (The Miami Herald, September 11, 2012). After the electoral results were announced, certain claims of fraud were made by Venezuelan citizens, yet Capriles did not support their view.

Regarding extant election forensics, Venezuelan elections have been among the most analyzed in the world. The debate centering on the irregularities in the 2004 constitutional referendum deserved a special issue devoted to it in *Statistical Analysis*. Different scholars compared the proportion of *Yes* votes with the signatures gathered in support of the referendum (Delfino and Salas, 2011), analyzed the first and second digit distributions in the vote counts (Pericchi and Torres, 2011), contrasted differences between exit polls and the official results (Prado and Sansó, 2011; Hausmann and Rigobón, 2011), studied the size and types of transmitted information between the voting centers and the CNE (Martín, 2011), and examined the distribution of votes within voting centers (Jiménez, 2011). For the 2009 referendum, Levin et al. (2009) conducted a survey analysis of the different forensic tools used in order to find any potential irregularities during the election, and many of the techniques employed in that paper are replicated here in my analysis of the 2012 election.

5.2.1 Results

There are 34,186 and 38,352 polling stations for the 2009 referendum and for the 2012 presidential election, which were downloaded from the National Electoral Court's website.[6] The turnout level in 2012 was of 80.6%, which is over ten percent greater than that for the 2009 referendum. The comparison of turnout distributions by state in Figure 5.13 shows unimodal distributions and uniform turnout increments in most of the cases. There are small bumps showings at the left of the distributions of Tachira, Vargas, and Nueva Esparta, which also appear in the 2009 distributions. For the case of Delta Amacuro, the bimodal distribution in 2009 disappears for 2012.

When analyzing the relationship between turnout and share of votes for Chávez and Capriles, most of the results for the states are consistent with the hypothesis in which both candidates get increases in their vote support proportional to the turnout rate. As Table 5.3 shows, the exceptions to this expectation were in Distrito Capital, Miranda, and Vargas. In the first case, the regression of this state explains a quarter of the variance in the data ($R^2 = 0.24$), and the coefficient for Chávez

[6] http://www.cne.org.ve. I thank Inés Levin for sharing the 2009 web-scrapped data with me.

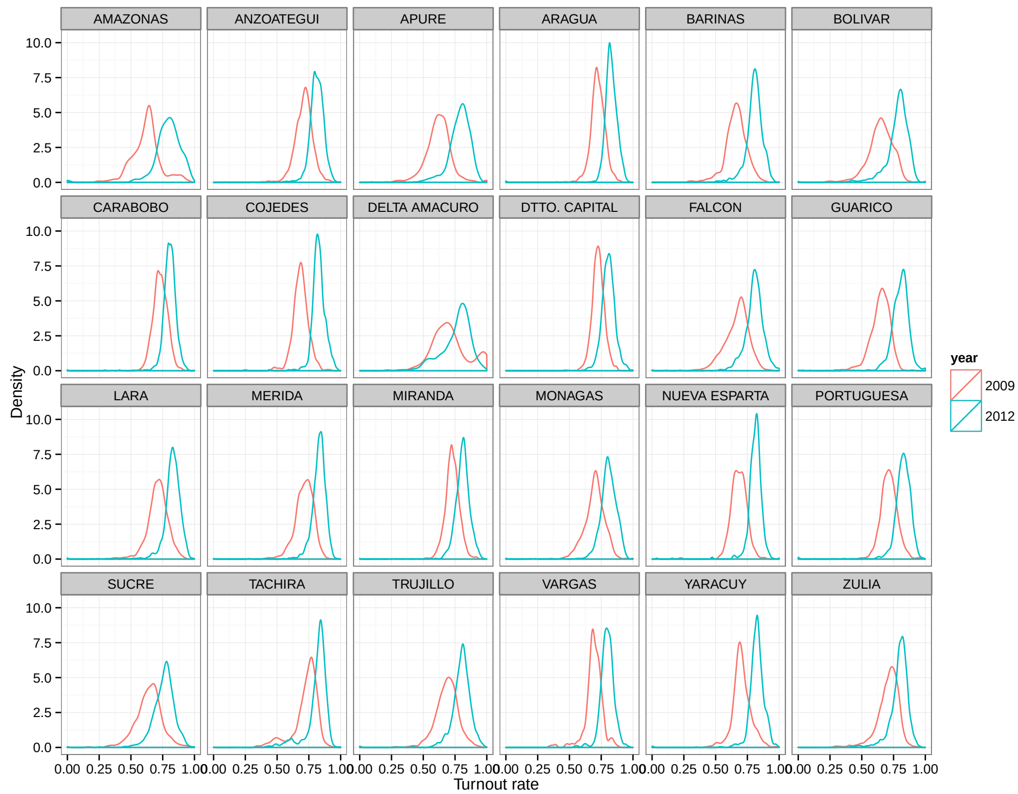


Figure 5.13: Turnout distribution by state for the 2009 referendum and 2012 presidential elections in Venezuela.

falls far away from the $[0,1]$ interval ($\beta = 1.50$). To reduce the heterogeneity of the data, I run the same regression only in only those observations where the *Yes* alternative won the 2009 referendum. The results are in Table C.13 in the Appendix and show that the coefficients for Distrito Capital, Cojedes, and Vargas are significantly larger than 1 for the regressions on Chávez's vote. Still, the regression for this subset of observations have values for the R^2 no greater than 0.38. Therefore, the deviations from the $[0,1]$ interval could, in fact, be due to data heterogeneity.

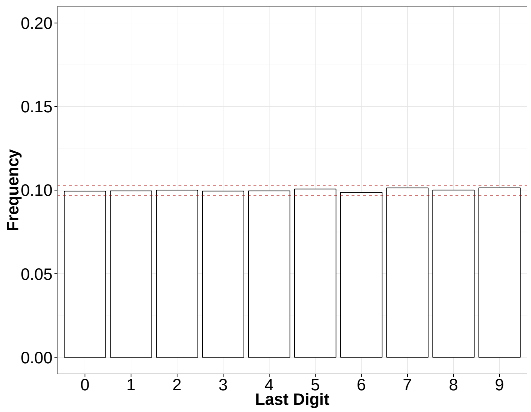
For the digit analysis of the vote returns, Figures 5.14 shows the distribution

Table 5.3: Regression Results from Turnout versus Votes for each Candidate.

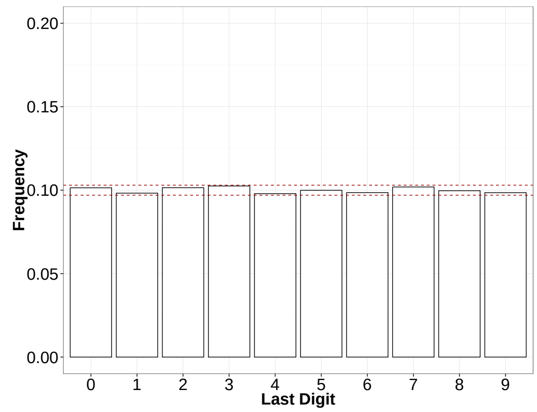
	Chávez		Capriles	
	β_V	R^2	β_V	R^2
Amazonas	0.33	0.09	0.17	0.03
Anzoategui	0.19	0.01	0.80	0.13
Apure	0.62	0.24	0.34	0.09
Aragua	0.27	0.01	0.48	0.04
Barinas	0.49	0.06	0.47	0.06
Bolivar	0.10	0.00	0.79	0.22
Carabobo	-0.10	0.00	1.00	0.12
Cojedes	0.99	0.16	0.06	0.00
Delta Amacuro	0.32	0.09	0.62	0.30
Distrito Capital	1.50	0.25	-0.54	0.04
Falcon	0.48	0.08	0.33	0.04
Guarico	0.29	0.04	0.52	0.12
Lara	0.06	0.00	0.79	0.11
Merida	0.08	0.00	0.89	0.15
Miranda	1.09	0.11	-0.19	0.00
Monagas	0.31	0.04	0.55	0.10
Nueva Esparta	0.47	0.07	0.51	0.07
Portuguesa	0.42	0.06	0.28	0.03
Sucre	0.39	0.10	0.55	0.19
Tachira	0.43	0.09	0.50	0.11
Trujillo	0.22	0.02	0.64	0.14
Vargas	1.01	0.15	-0.02	0.00
Yaracuy	0.51	0.05	0.30	0.02
Zulia	0.32	0.03	0.62	0.12

of the last digit for the two main candidates in the contest. Figure 5.15 summarizes the results of the last digit test in every state, where the only distributions significantly different from the expected distribution are in the votes for Chávez in Portuguesa.[7] Regarding the analysis for the second digit, Figure C.12 shows the only distribution significantly different from the Benford distribution is in the vote returns for Capriles in Carabobo. The rest of the vote counts comply with the expected digit distribution. Finally, Figures C.13 and C.14 in the Appendix show the distribution of the last two digits in the vote counts for both Chávez and Capriles. Similar to the Mexican case, there is no a particular pattern in their density. As Levin et al. (2009) recognize, the use of electronic voting machines in Venezuela

[7] Figures C.10 and C.11 in the Appendix show the digit distribution in each state.



(a) Chávez



(b) Capriles

Figure 5.14: Last digit distributions for the candidates' vote counts at national level.

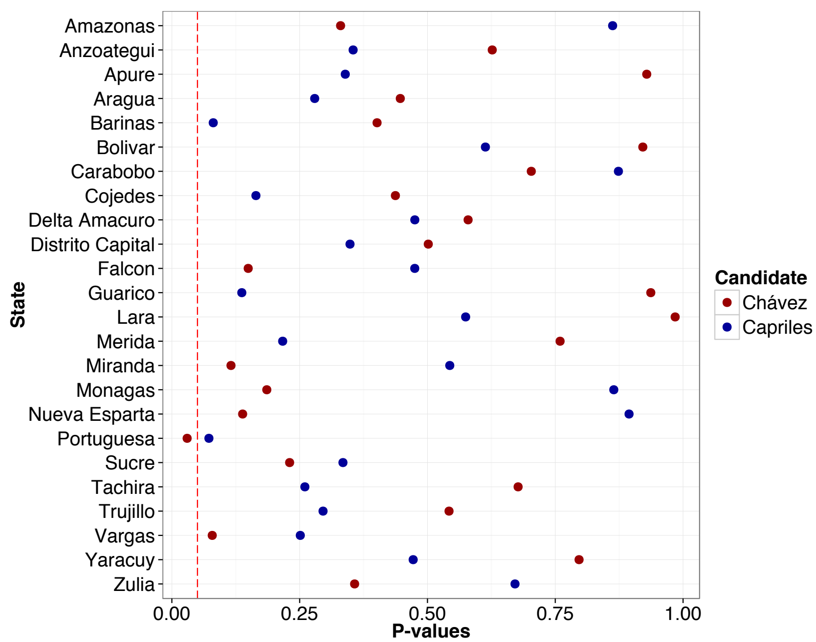


Figure 5.15: P-values for chi-squared tests for the last digit distribution by candidate and state. Observations to the left of the red line are for those p-values below 0.05.

may overcome this type of manipulation.

For the analysis of the flow of votes, Table 5.4 shows the results of the Good-

Table 5.4: Flow of votes between the 2009 referendum and 2012 presidential elections in Venezuela.

	Yes 09	No 09	Abst 09
Chávez 12	0.93	-0.06	0.29
Capriles 12	0.02	1.05	0.09
Otro 12	0.00	0.00	0.01
Abst 12	0.04	0.00	0.61
Column Sum	1.00	1.01	1.00

man regressions with random effects by municipality. Most of the results thus are in line the expectations: the coefficients fall in the $[0,1]$ interval. This was a very polarized election where most of the voters for the *No* alternative in 2009 supported Capriles in 2012. The first row of Table 5.4 suggests that 30% of those who did not vote in 2009 did vote for Chávez in 2012 while only 10% of the same group voted for Capriles. The negative sign for Chávez among those who voted *No* in 2009 is consistent with the flow analyses for the elections of 2006, 2007, and 2009 (Levin et al., 2009), and suggests a negative correlation in the strongholds for the opposition.

Figure 5.16 shows the random effects by region and their differences from the average effects when the dependent variable is the proportion of votes for Chávez in 2012. Most of the highest deviations from the average effects are in the states of Zulia and Miranda, yet no observation stands far from the rest of the municipalities. In the municipality of Boliviano Guajira, in the state of Zulia, the proportion of the 2009 *Yes* votes received by Chávez in 2012 is 115%, representing the the highest deviation from the average effects. When the dependent variable is the share of votes for Capriles, the highest deviations from the average effects are

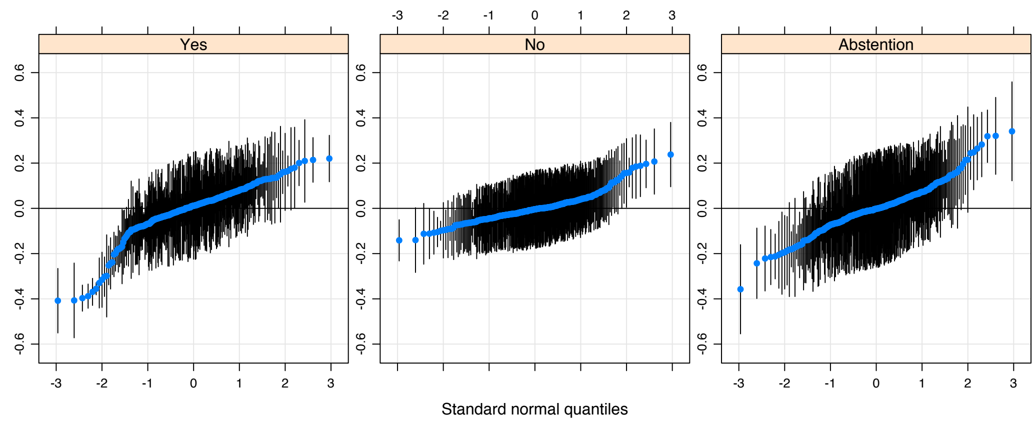


Figure 5.16: Goodman regression coefficients by municipality with votes for Chávez as the dependent variable.

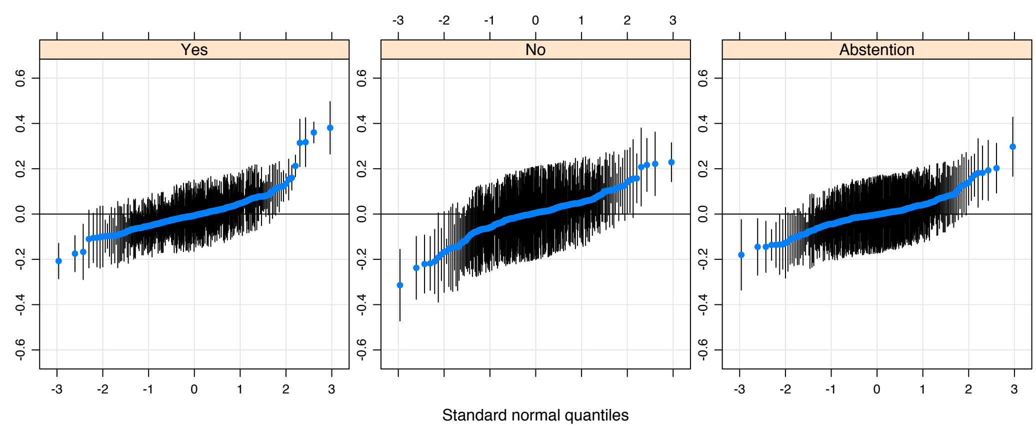


Figure 5.17: Goodman regression coefficients by municipality with votes for Capriles as the dependent variable.

located in the state of Bolivar and the southern municipalities in Lara. In particular, the municipalities of Moran and Andrés Blanco show shares for Capriles from the proportion of *No* votes in 2009 greater than 133 percent.

For the outlier analysis in Venezuela, I use the fact that the allocation of voters to polling stations is based on the last digits of their ID numbers, an assignment that is independent of the electoral process (Jiménez, 2011, p.565). The results do not find support for the existence of any kind of manipulation, and almost all the observations in the analysis sets fall within the 95% confidence interval. We can thus conclude that observations with a relative high turnout in the polling place do not affect to any of the candidates in a significant way. The results for both Chávez and Capriles are in Figures C.19 to C.22 in the Appendix. Since the voter allocation is similar in Venezuela and Mexico, the comparison of the results suggests that an electronic voting system avoids the types of errors described in the Mexican case.

5.3 Bolivia

On December 18, 2005, the Movement Toward Socialism (MAS) won the Bolivian presidential election by a majority of votes and took control of both legislative houses, a situation that had not occurred since the 1952 national revolution. Despite its victory, MAS claimed that the electoral process suffered from logistical problems and fraud. In particular, it complained of a large number of citizens who were unable to vote. Although Bolivian electoral law deprives voting rights to those citizens who had not voted in previous election and had not re-registered, the elected president, Evo Morales, ironically stated during his acceptance speech that "it is the National Electoral Court (CNE) that needs to be purged" rather than the citizens (Muñoz-Pogossian, 2008, p. 178).

From the moment of taking power, Morales found a country polarized into Pro-MAS and Anti-MAS regions. Pro-MAS support is found in the western highland region of the country, which covers four departments: La Paz, Oruro, Potosí, and Cochabamba. This region comprises about half of the national population, and, with the exception of La Paz, mining is the main commercial activity, along with the coca leaf production in Cochabamba (Romero Ballivián, 2003, p. 19-22). In contrast, the Anti-MAS forces are concentrated in the eastern part of the country, known as the *Media Luna* (Half Moon) region, which includes the departments of Santa Cruz, Tarija, Beni, and Pando (Romero Ballivián, 2003, p. 23-27). Regional conflicts and autonomy movements in the country come from the nineteenth century when the seat of government was moved from Sucre to La Paz, and the campaign in Santa Cruz to tax its oil exports. Chuquiasca joined to the *Media Luna* region for the autonomy demand in 2006, after explicitly stating its intention to return the capital of the country back to Sucre (Centellas and Buitrago, 2009, p. 4-5).

The political polarization of the country reached its highest point during the 2008 recall referendum, which required the president and department governors (prefects) to receive at least the same number of votes as when they were originally elected in order to remain in office. This referendum was a response to

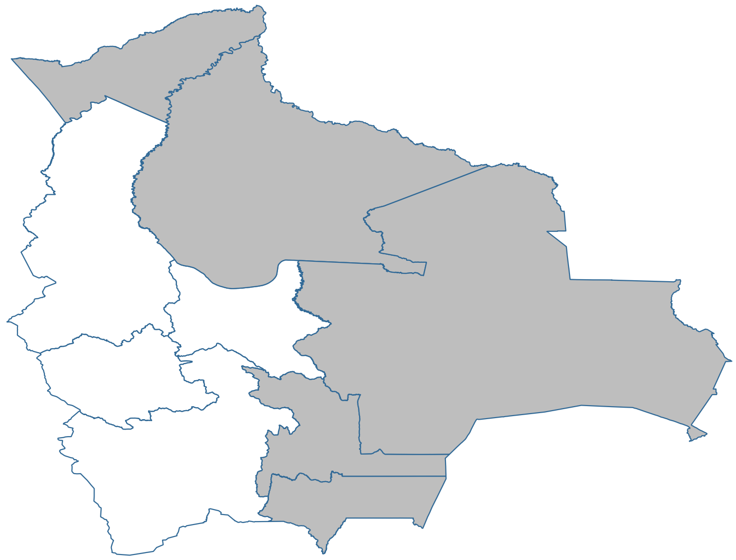


Figure 5.18: Administrative departments in Bolivia. Departments in black color represent those in the *Media Luna* region plus Chuquiasca.

the autonomy-seeking referenda in the *Media Luna* departments earlier that year (Lehoucq, 2010, p. 358). During the months leading up the referendum, Evo Morales's supporters and detractors polarized the campaign and public opinion. While the opposition presented a racist discourse against President Morales (Skidmore, Smith and Green, 2010, p. 183), Morales' administration filed criminal charges against five previous presidents and several opposition prefects. The charges were presumably based on political, rather than judicial, grounds (Madrid, 2011, p. 52).

The referendum reignited the polemic about the voting rolls. Opposition legislators and the media claimed that the system was inefficient. They highlighted the unexpected increase in the voting rolls, which grew by nine percent with respect to the 2006 election. This polemic intensified when the CNE president recognized that approximately 30 percent of citizens who presented their army cards as identification to register could have then used a different identification document to register a second time (Peñarada and Candia, 2010, p. 41-42). News agencies reported that voter lists in the *Media Luna* region increased during the two months prior to the referendum more than at any other time over the course of the previous two years (La Razón, January 6, 2009*a*, p. A1). Further, the local electoral court in Beni recognized that around one third of the names on the voting lists did not appear in the department's civil registry (Peñarada and Candia, 2010, p. 50). The controversy surrounding the voting rolls pushed electoral authorities to depurate them after the referendum, decreasing the number of voters by four percent in January of 2009 (Peñarada and Candia, 2010, p. 52).

The results of the referendum supported Evo Morales, who won with 63 percent of the vote. While the proportion of *Yes* votes in the western departments was 75 percent, the proportion *Yes* votes in the *Media Luna* region and Chuquiasca was only 40 percent. Meanwhile, opposition prefects in Cochabamba, La Paz, and Oruro lost their seats. Polarization and violence continued after the election, as the losing prefects refused to recognize the results. Morales organized protests and social mobilizations to put pressure on the opposition prefects to resign (Madrid, 2011, p. 252; Mayorga, 2009, p. 113). Meanwhile, the eastern departments called strikes and cut-off communication with the central government. The agitation on both sides culminated with the deaths of ten peasants and two civic committee members in Pando on September 10 (Gray Molina, 2010, p. 70).

Electoral forensics of Bolivian elections are non-existent. *La Razón* newspaper published a series of articles early in 2009 on its research on the voter lists, finding that in the coca regions, there was an unexpected increase in voter turnout, with many polling stations registering a turnout level over 95 percent (La Razón, January 7, 2009*b*, p. A1). Meanwhile, the mission from the Organization of American States reported that the elections were accomplished "under totally normal conditions"[8]. The following analysis represents is the first recent Bolivian election study to use forensic tools.

5.3.1 Results

To this date, the website of the CNE has not published electoral results, so the data I use in this analysis is available at the *Democracia en Bolivia*'s website.[9] I audited the results from the database at the department level with the official results and the data are consistent in every case. The database includes the total number of registered voters, and the votes for the *Yes* and *No*, as well as the null votes for the 2008 referendum, culled from 21,974 polling stations distributed in 3,906 polling places across nine Bolivian departments. It also includes the votes for MAS and the turnout rates for the 2006 elections.

Compulsory voting produced very high turnout levels in Bolivia by interna-

[8](Organization of American States, August 11, 2008).

[9]https://sites.google.com/a/democraciaenbolivia.org/www/Home

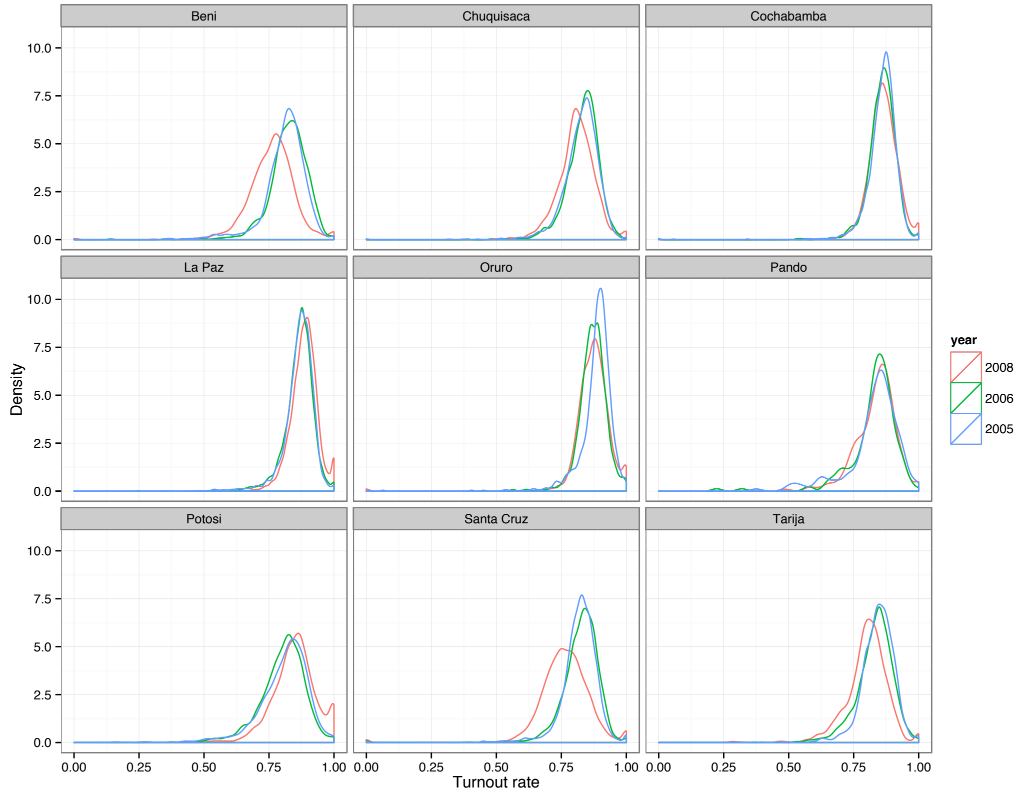


Figure 5.19: Turnout distribution by department for the 2008 recall referendum in Bolivia.

tional standards. The overall turnout level for the 2005 presidential elections and the 2006 constitutional referendum was 84 percent, while for the 2008 referendum was 83.2 percent. Nevertheless, there are changes in turnout levels across states. Figure 5.19 shows a decrease in the turnout levels for the *Media Luna* states, particularly in the departments of Beni, Chuquiasca, Santa Cruz, and Tarija. At the same time, some departments in the western part of the country show unexpected bimodal distributions for the 2008 referendum, with the second mode reaching 100 percent of voter turnout. This pattern is significantly high in the departments of

Potosí and La Paz. In the former case, there are 209 polling stations with turnout rates equal of above 95 percent, representing one-tenth of the observations in that department. This figure is significantly higher when compared with the 2006 election, when only 44 polling stations had turnout levels equal to or greater than 95 percent.

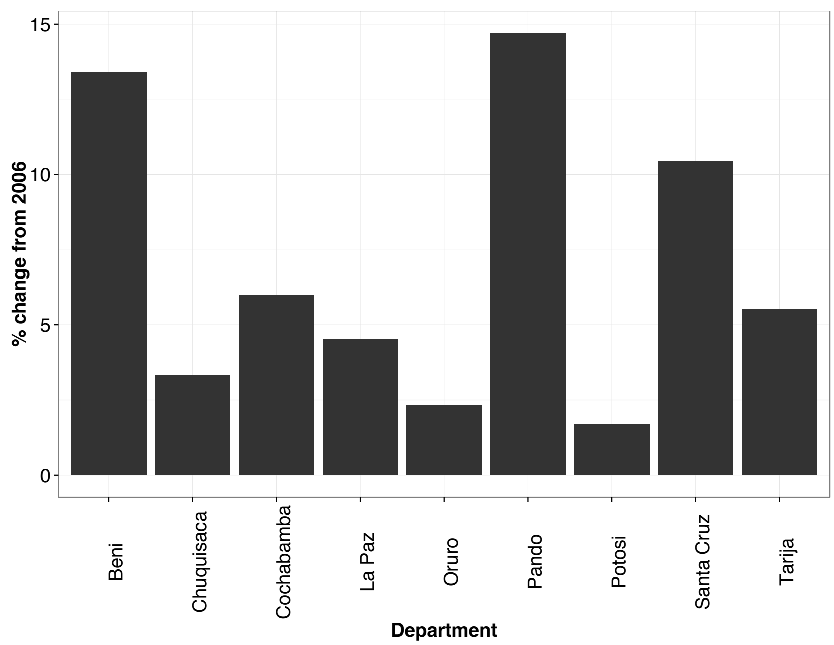


Figure 5.20: Change in the Registered Voters between 2006 to 2008 by department.

A possible explanation for the disparities in the turnout rates across the departments is the irregular change in the number of registered voters. As Figure 5.20 shows, the highest increase in the number of registered voters are in three specific *Media Luna* departments. Beni, Pando, and Santa Cruz, all experienced an increase of over ten percent in the size of the voters' list with respect to the 2006 referendum. Therefore, the stability of the turnout rates at national level is a mixture of

Table 5.5: Regression Results from Turnout versus Votes for each alternative. 2008.

	Yes		*No*	
	β_V	R^2	β_V	R^2
Beni	0.24	0.03	0.65	0.16
Chuquisaca	0.44	0.02	0.59	0.03
Cochabamba	1.27	0.12	-0.21	0.00
La Paz	1.34	0.18	-0.34	0.01
Oruro	0.98	0.26	-0.02	0.00
Pando	0.64	0.09	0.28	0.02
Potosí	1.07	0.30	-0.02	0.00
Santa Cruz	0.22	0.02	0.71	0.15
Tarija	0.32	0.03	0.63	0.07

Table 5.6: Regression Results from Turnout versus Votes for each alternative in 2008 for those polling places where the MAS votes above the average.

	Yes		*No*	
	β_V	R^2	β_V	R^2
Beni	0.72	0.39	0.15	0.04
Chuquisaca	1.00	0.12	0.00	0.00
Cochabamba	1.58	0.35	-0.50	0.06
La Paz	1.26	0.40	-0.25	0.03
Oruro	1.05	0.49	-0.07	0.01
Pando	0.66	0.12	0.06	0.00
Potosí	1.16	0.66	-0.12	0.04
Santa Cruz	1.06	0.49	-0.12	0.02
Tarija	0.95	0.41	-0.01	0.00

two phenomena. While the unusual increase in the voters' list across the *Media Luna* departments pushed down their turnout rates, the increase of observations with 100 percent of turnout pulled up the turnout rates in those departments supporting MAS candidates.

To explain how the variations in the turnout across polling stations affect the results, Table 5.5 illustrates the β coefficients and R^2 of regressing the share of *Yes* and *No* votes in relation to the turnout rates in every department. Although the results show coefficients outside the $[0, 1]$ interval in Cochabamba, La Paz, and

Table 5.7: Regression Results from Turnout versus Votes for MAS. 2006.

	MAS	
	β_V	R^2
Beni	0.06	0.00
Chuquisaca	0.50	0.06
Cochabamba	0.32	0.01
La Paz	0.81	0.12
Oruro	0.79	0.11
Pando	0.41	0.09
Potosí	0.52	0.08
Santa Cruz	0.06	0.00
Tarija	-0.13	0.00

Potosí, the regressions do not fully explain the variance of the data. To reduce data heterogeneity, I replicate the analysis only for those polling places with a proportion of votes for MAS in 2006 greater than or equal to 0.41, which was the mean share for that party in that election. As Table 5.6 shows, the regressions with a coefficient for *Yes* above 1 are in Cochabamba, La Paz, Oruro, Potosí, and Santa Cruz. The values of the coefficients and the percent of the variance explained by the regressions flag some manipulations during the referendum, especially when comparing the results to the analyses in Mexico and Venezuela. These results are also exceptional after replicating the same analysis for the 2006 congressional election, where, as Table 5.7 shows, all but one of the coefficients fall inside the $[0,1]$ interval. The results with this methodology provide evidence that the increase in turnout rates benefitted the *Yes* vote for President Evo Morales in a disproportionate and significant way.

Regarding the different tests for the vote count's digit, the last digit densities for *Yes* and *No* votes differ significantly from the expected uniform distribution. The P-values for the Chi-squared tests by department are displayed in Table 5.22.[10] Sig-

[10]Figures C.15 and C.16 in the Appendix show the histogram of the digits by department.

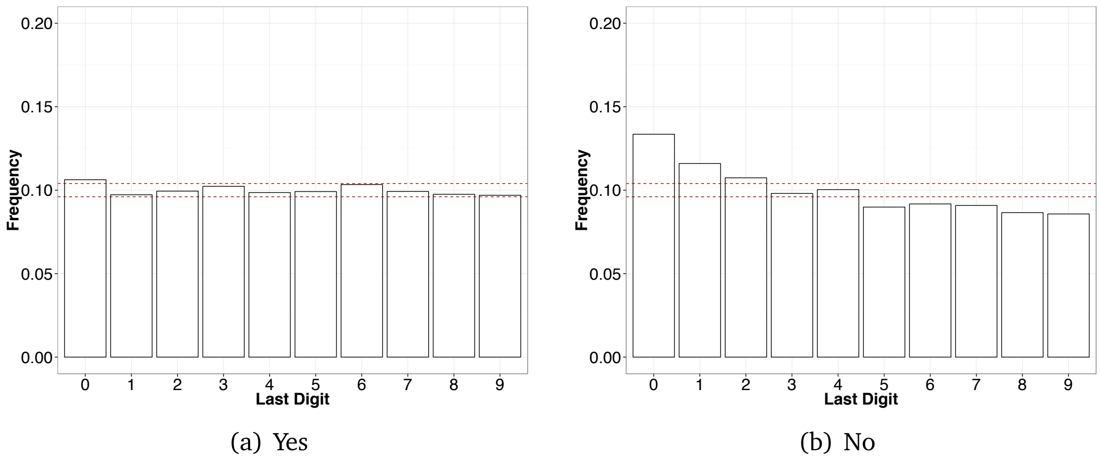


Figure 5.21: Last digit distributions for the alternatives in the recall referendum at national level.

nificant departures from the uniform distribution appear in Chiquiasca and Oruro. On the other hand, the distribution for *No* votes are significantly different in all the departments outside the *Media Luna* region. This may be due to the electoral polarization of the regions, where each alternative is expected to have vote counts clustered around either 0 or 100 percent, and voting rolls in every polling stations truncated to 300 voters. The results may be a result of the electoral dynamics in the western departments of the country where the *No* votes were very low.

For the analysis of the second digit, Figure 5.23 shows that most of the distributions follow the Benford law with the exception of Tarija, where the results are significant at the 95 percent for both *Yes* and *No* votes. Finally, Figures C.17 and C.18 in the Appendix show the distribution of the last two digits for the country, which does not indicate any significant pattern for the repeated digits.

The analysis of the flow of votes does not show any significant outlier re-

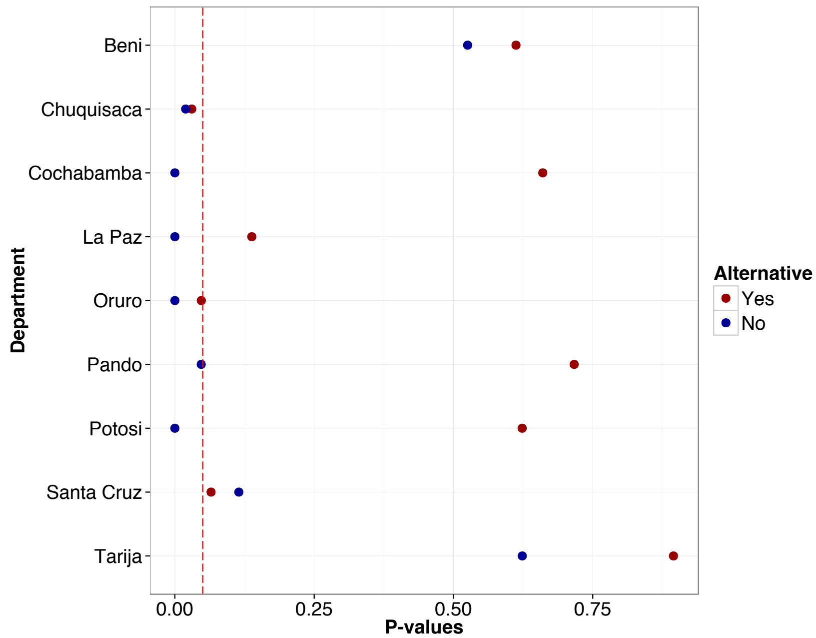


Figure 5.22: P-values for the last digit tests by department. Observations to the left of the red line are for those p-values below 0.05.

sult. The share of the 2006 votes for MAS received by the *Yes* choice in 2008 is 100 percent, and is consistent with what Centellas and Buitrago (2009) find for a low volatility of MAS partisans across elections. Moreover, half of the voters who abstained in 2006 did the same in 2008, while 90 percent of those voters who went to the polling station voted for the *Yes* alternative. Figure 5.24 shows deviations by municipality using the average estimations for *Yes* as the dependent variable. The highest deviation is in the municipality of Cercado in Oruro, where the proportion of the 2006 MAS vote share received by the *Yes* alternative in 2008 is equal to 118% percent. An observation relatively large when compared to other regions.

To analyze the results when the *No* votes as the dependent variable, one should examine the observation at the top right of the panel for 'Other' alternative

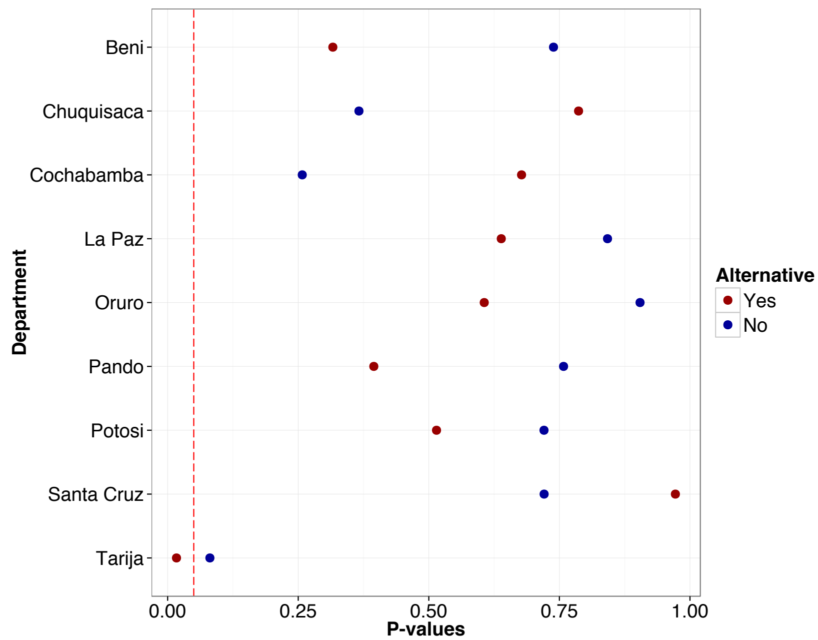


Figure 5.23: P-values for the second digit tests by department. Observations to the left of the red line are for those p-values below 0.05.

Table 5.8: Flow of votes between the 2009 referendum and 2012 presidential elections in Venezuela.

	MAS 06	OTHER 06	ABST 06
YES 12	1.00	0.42	0.46
NO 12	-0.08	0.31	0.05
ABST 12	0.07	0.27	0.50
Column Sum	0.99	1.01	1.01

in Figure 5.25. That observation is for the municipality in Oropeza in the Oruro department, where the proportion of 2006 share of votes for a party other than MAS receiving a *No* vote is 131%, which is significantly outside the normal range and very different from the rest of the regions. The rest of the regional random effects fall within the expected values.

Finally, I test the methodology described in the previous chapter to see sig-

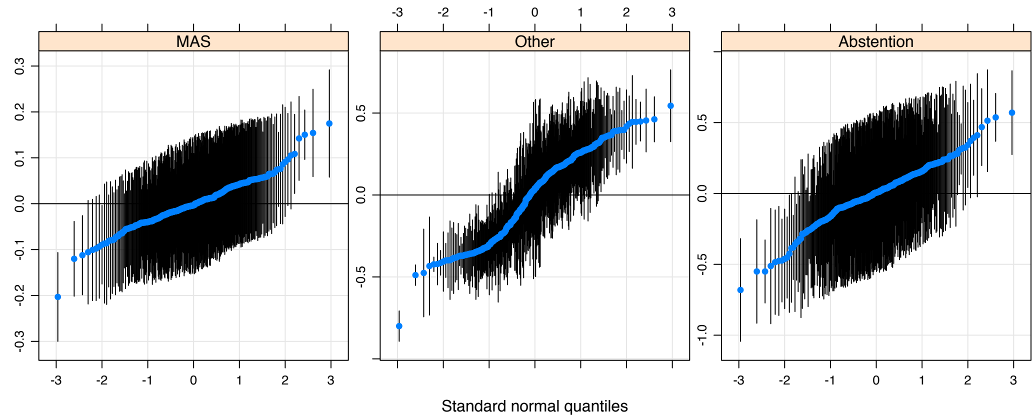


Figure 5.24: Goodman regression coefficients by municipality with votes for YES as the dependent variable.

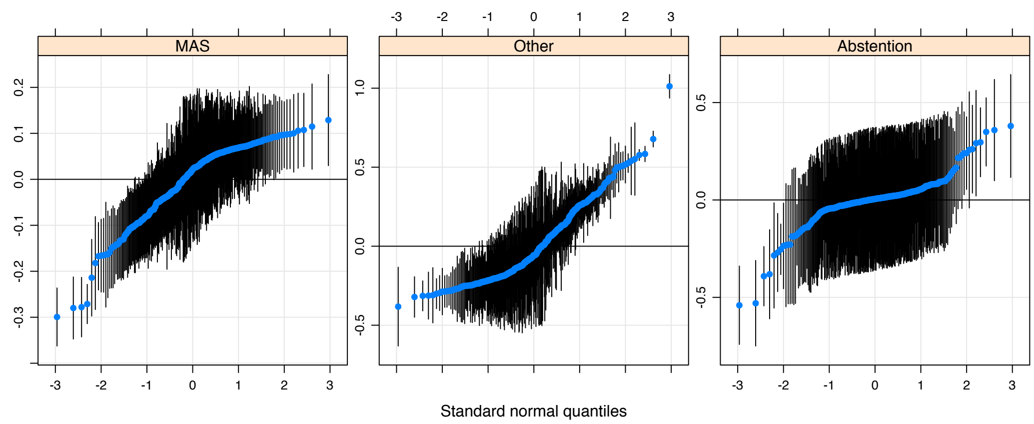
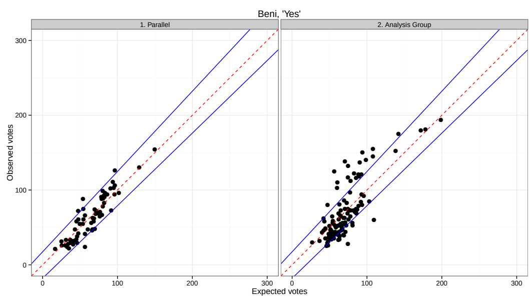
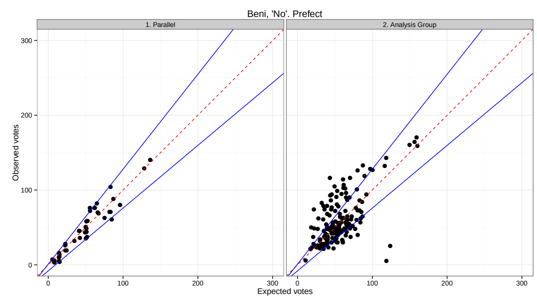


Figure 5.25: Goodman regression coefficients by municipality with votes for NO as the dependent variable.

nificant differences in the shares of votes within polling places. The procedure in this case differs from the analysis of Mexico and Venezuela. It includes additional information regarding how manipulation occurred. As shown above, voter turnout is relatively stable across time and variations across departments are due to changes in the voter lists. Therefore, I vary the methodology described in Chapter 3 by including in the analysis group those polling stations where the voting roll at least doubled in size from 2006 to 2008. Similar to the Mexican case, voters were assigned to polling station using their last names; therefore, we should expect that changes in the voting rolls vary by region and not by voters' name. The rest of the methodology follows the procedure described in the previous chapter, and the full results are in Figures C.23 and C.24 in the Appendix.

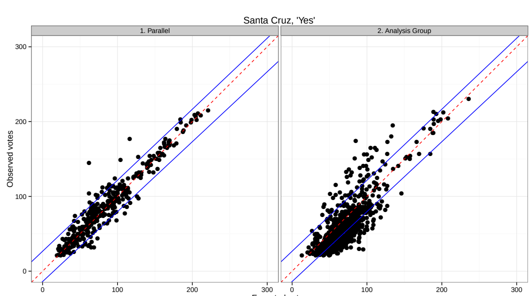


(a) *Yes* votes for president

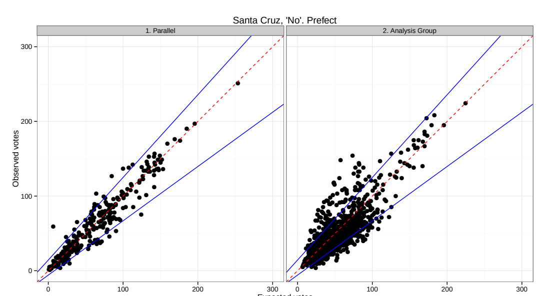


(b) *No* votes for prefect

Figure 5.26: Outlier analysis for the recall referendum in Beni.



(a) *Yes* votes for president



(b) *No* votes for prefect

Figure 5.27: Outlier analysis for the recall referendum in Santa Cruz.

Consider the cases of two departments in the the *Media Luna* region, Beni and Santa Cruz. For the recall of the president, there are many polling stations with a significant increase in the voter list showing a higher share of *Yes* votes compared to the rest of the observations at the same polling place. In this case, those places with an unusual increase in their voting lists had overwhelming support for president Morales. The results are consistent when analyzing the same observations for the revocation of the prefects, where the observations have less than expected proportion of *No* votes. Results are similar for other departments in the *Media Luna* region such as Tarija.

The results show evidence that those polling stations with an unexpected increase in the voters were overwhelmingly in support of President Morales and against the opposition prefects. Unlike the cases of Venezuela and Mexico, it is hard to argue that the observations outside the bounds are due to accidental or unintentional factors. One of the advantages of this tool is that it is possible to identify the polling stations with an odd increase in its voting roll and support levels for Evo Morales. Further research should look for qualitative evidence to confirm the irregularities in those places.

5.4 Conclusions

This chapter uses a variety of forensic tools to evaluate the 2012 presidential elections in Venezuela and Mexico, as well as the 2008 recall referendum in Bolivia.

Table 5.9 summarizes the findings and compares the efficiency of each forensic tool. For the Mexican case, and in contrast to previous analyses (Gallardo, 2012; Mochán, 2012), I show that the evidence available to assess this election as fraudulent is found to be insufficient. My analysis makes clear that the potential evidence of fraud is a product of data heterogeneity. Moreover, I show that certain indications of potential fraud discussed by other researchers are, instead, examples of the electoral dynamics in the states, rather than products of manipulation. In particular, analyses of the vote counts of candidates with a poor performance in a particular region are more susceptible to positive results. Also, other irregularities, such as high voter turnout rates, are a product of unintentional errors in the counting process, and there is no evidence that they affected any of the candidates in a systematic way. In Venezuela, positive results of fraud are difficult to locate and are not supported by other findings. Meanwhile, the results in Bolivia confirm what both the media reports and the political opposition claim. Specifically, that there is evidence that manipulation occurred in two different ways: while the regions with strong MAS support had an unexpected rise in voter turnout, the opposition regions experienced an increase in the size of their voting rolls.

To expand this research further, an evaluation of the effects of the new electoral biometric system needs to be conducted in Bolivia. This system began in 2009 as an answer to the concerns by the opposition regarding the voting lists (Peñarada and Candia, 2010). It is expected that the new system will prevent irregular increases in the voting rolls; yet, as the last three chapters of the dissertation show, electoral criminals are able to adapt their technologies to fulfill their goal. An alternative study should focus on the unusual swing of support for López Obrador from 2006 to 2012 in those municipalities in Oaxaca with the "Usos and Costumbres" system for local elections. The findings are consistent with what Benton (2011) finds in elections in Oaxaca, where the alternation of power represents a signal of intra-elite struggles. The puzzle in these regions is contradictory effects of these rules for local democracies and policy outcomes. On the one hand, the lack of secret voting allows party brokers to have more control over the behavior of the voters. On the other hand, there is evidence of the positive effects of this system with respect to the level of public services (Díaz-Cayeros, Magaloni and Ruiz Euler, 2012). The findings of this chapter may suggest that voters in these communities behave as corporatist members, coordinating massive support for a candidate in exchange for requested favors (Huntington and Nelson, 1976).

Table 5.9: Summary of the Results for each Forensic Analysis in Mexico, Venezuela and Bolivia.

	Mexico, 2012	Venezuela, 2012	Bolivia, 2008
Turnout distribution	No significant results	No significant results	Bimodal turnout distributions in La Paz, Oruro, Potosí and Cochabamba
Regressing votes for a candidate on turnout	Negative coefficients for López Obrador in Campeche and Yucatán; both regressions have an R^2 below 0.02	Coefficients greater than 1 for Chávex in Cojedes, Distrito Capital, and Vargas; R^2 are equal or below 0.35	Coefficients outside the $[0,1]$ interval in Cochabamba, La Paz, Oruro, Potosí; R^2 values between 0.35 and 0.66
Last Digit	Significant deviations from the uniform distribution in Guerrero, Michoacán, Nuevo León, Tlaxcala, and Yucatán for the PRI; Guerrero and Hidalgo for the PAN	Significant deviation from the uniform distribution for Chávet's votes in Portuguesa	Significant deviation for the *Yes* alternative in Chuquiasca and Oruro; and for the in *No* alternative in Chuquiasca, Cochabamba, La Pax, Oruro, and Pando.
Last two digits	No significant results	No significant results	No significant results
Second digit	Significant results for the PRI in Aguascalientes, Coahuila, Guanajuato, Jalisco, State of Mexico, Morelos, San Luis Potosí, Sinaloa, Tabasco, and Veracruz; for the PRD in Aguascalientes, Colima, Mexico City, and State of Mexico; for the PAN in Chiapas, Colima, Mexico City, Guanajuato, Jalisco, Nuevo León, and Yucatán	Siginificant results for Capriles in Carabobo	Significant results for the *Yes* alternative in Tarija.
Flow of votes	Results comply with the hypothesis	The share of votes for *No* in 2009 received by Capriles in 2012 is equal to 105%	Results comply with the hypothesis
Outlier analysis	Errors are due to accidental errors; no significant bias to any candidate	No significant results	Significant results in Beni, Santa Cruz, and Pando

Chapter 6

Conclusion

This dissertation discusses the problems of detecting electoral manipulation in developing democracies. My argument explains why parties in competitive elections engage in corruption and demonstrates how spending resources on fraud is sometimes an inefficient activity. The empirical analysis provides a novel technique to identify electoral irregularities. In the course of this work, I evaluate the most common forensic tools and apply them in my analyses of elections in Mexico, Bolivia, and Venezuela.

In Chapter 2, I provide a theoretical argument to understand the allocation of resources for electoral manipulation. The model provides a framework to harmonize the disagreements in the literature about the goals and observability of fraud. Moreover, it suggests dividing the empirical cases into two groups. On the one hand, there are those cases where the decision to engage in fraud, and the amount of resources available, are exclusive to the incumbent party, providing opportunities for blatant and non-outcome driven fraud.On the other hand, there are those cases where the resources and opportunity for fraud are available to more than one group, thus increasing the possibility that the perpetrators will be exposed.

Chapter 3 provides a survey of the most common forensic tools. For each technique, I discussed its and highlighted the assumptions behind the test. My work builds on the existent literature of election forensics that use multiple tests to evaluate the integrity of elections (Mebane, 2009; Mebane and Kalinin, 2009; Levin et al., 2009; Jiménez, 2011; Klimek et al., 2012). In contrast to the existing literature, I provide a survey approach of precisely when and how each forensic technique should be utilized. To illustrate each of the tools, I present an exhaustive forensic analysis of the 2012 presidential elections in Mexico.

In Chapter 4, I propose a new identification strategy for electoral manipulation. This technique exploits a feature of the electoral system in certain countries: namely, within each electoral precinct, voters are assigned to polling stations according to their childhood surname. Consequently, the only difference between voters in contiguous polling stations should be their last names. This tool contributes to the growing literature using statistical tools to evaluate the quality of elections (Myagkov, Ordeshook and Shakin, 2009; Mebane, 2006; Beber and Scacco, 2008; Levin et al., 2009; Fukumoro and Horiuchi, 2011). For the particular analysis of the 2010 local elections in Mexico, this tool uncovers those "brown areas" (O'Donnell, 1998), where the interests of local power are in conflict with national-level regulations that still exist in the country.

Chapter 5 evaluate the national elections in Mexico, Bolivia, and Venezuela. The findings for Mexico and Venezuela show no strong evidence of electoral manipulation. Meanwhile, the results in Bolivia confirm what both the media reports and the political opposition claim. Specifically, while the regions with strong MAS support had an unexpected rise in voter turnout, the opposition regions experienced an increase in the size of their voting rolls.

This project represents the start of a research agenda that I will continue to pursue. The three potential avenues for exploration that stem from this dissertation are related to the methodology, evaluation, and implications of electoral manipulation. First, although the methodology described in Chapter 4 is a first approximation to understand electoral irregularities in Mexico, it is limited to finding irregularities only when political machines are incapable of manipulating all of the polling stations in a particular precinct. As such, this tool is a conservative way to detect irregularities. A possible further extension of this methodology is to compare voter turnout rates among precincts with similar covariates rather than contiguous polling stations. I expect that the new findings will not change the results, but rather they will refine the detection of irregularities and provide further evidence of fraud occurring on the election day.

Second, I seek to expand on the research of electoral integrity beyond the most recent history of Mexico. Electoral fraud appears as a given fact in the study of the political development in the country, yet the evidence to study the mechanisms and strategies of electoral manipulation is still inconclusive. I am particularly interested in solving questions about the effects of fraud in the 1988 Mexican presidential election (Castañeda, 2000, p.231-239) as well as local elections during the early 1990s (Molinar, 1991; Espíritu, 1984). This research will solve some elusive puzzles of Mexican political history. Moreover, the expansion of the analysis to other countries will yield more ways in which perpetrators choose among different alternatives to manipulate specific elections. The evidence here will test the hypothesis recently proposed in the literature of fraud as a “scare-off” effect aimed at the opposition (Magaloni, 2006; Little, 2011; Simpser, 2013).

Third, another relevant question in the literature on democratization is the gap between true electoral manipulation and the perceptions of this manipulation. The relevance of elections in democratic systems relies not only on the results they produce but also on the citizens’ reactions to those outcomes (Linz and Stepan, 1996; Przeworski, 1991). In other words, the perception of a fair procedure to select political representatives is essential to legitimate the outcomes (Nadeau and Blais, 1993, p. 553). Moreover, it seems that perceptions of electoral manipulation are very sensitive to negative events regardless of the quality of the electoral administration. The question, therefore, is whether these perceptions are accurate when evaluating electoral administration performance or, instead, are the result of factors unrelated to the electoral process.

While writing these lines, the Venezuelan opposition expects a full recount of the votes for the presidential election held April 14, 2013. The margin of victory between Hugo Chávez’s successor, Nicolás Maduro, and the opposition candidate, Henrique Capriles, is just 1.59%. Capriles claimed to have 3,200 irregularities on the election day and contested the official results of the election with a Supreme Court filling. This is an instance in which the tools reviewed in this dissertation could help distinguish whether fraud allegations are signs of a "sore loser" or are a legitimate reaction to a truly unfair election. I expect that this research will help both scholars and electoral administrators better uncover electoral manipulation. The research methodologies discussed above will raise the costs of illicit behavior because more resources, in terms of both money and time, will be needed on the part of the cheaters to mask fraudulent activities. At the same time, disgruntled political parties will be less likely to make unfounded accusations of fraud knowing that their allegations will soon be exposed as false.

Appendix A

Chapter 2

A.1 Proof for Proposition 1

Proof. Let $p_{1|\alpha_H}$ and $p_{2|\alpha_H}$ be the equilibrium probabilities for parties 1 and 2 to disrupt in the polling station given their effort levels at $t_1 = t_2$. Similarly, $p_{1|\alpha_L}$ and $p_{2|\alpha_L}$ be the equilibrium probabilities for parties 1 and 2 to disrupt in the polling station given their effort levels at $t_1 = t_2/\alpha_L; \alpha_L < 1$. Given these two different probability combinations, the governor chooses the α level that maximizes the probability of a victory for Party 1 in order to receive a reward αK. That is, she will opt for an asymmetry in the resources between the parties of magnitude α_L only when:

$$K[p_{1|\alpha_H} + (1 - p_{1|\alpha_H} - p_{2|\alpha_H})(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega})] <$$
$$\alpha_L K[p_{1|\alpha_L} + (1 - p_{1|\alpha_L} - p_{2|\alpha_L})(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega})]$$

Rearranging the terms, and substituting $1-p_{1|\alpha_i}-p_{2|\alpha_i}$ for $p_{S|\alpha_i}$ where $i \in 1,2$, we have:

$$p_{1|\alpha_H} + p_{S|\alpha_H}(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega}) < \alpha_L[p_{1|\alpha_L} + p_{S|\alpha_L}(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega})]$$

Which leads to

$$\frac{p_{1|\alpha_H} + p_{S|\alpha_H}(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega})}{p_{1|\alpha_L} + p_{S|\alpha_L}(\frac{1}{2} + \frac{\xi\omega\eta}{1-\omega})} < \alpha_L$$

Observe that while $p_{1|\alpha_H} < p_{1|\alpha_L}$, $p_{S|\alpha_H} > p_{S|\alpha_L}$. So the magnitude of these differences determine whether winning the election compensates the depreciation of the reward on K.

□

A.2 Proof for Comment 1

Proof. (i) To show that a more efficient technology expands the party's effort frontier, I derivate the reaction function of the party $i \in \{a,b\}$ with respect to its efficiency technology t_1. The reaction function $\hat{e}_{i\in\{a,b\}}$ changes with respect to t_1 as follows:

$$\frac{\partial \hat{e}_1}{\partial t_1} = \sqrt{\frac{R(2e_2+S)}{2t_1}}$$

Since the efficiency technology endowed to the parties is positive, the effort exerted by the Party increases as t_1 becomes larger.

(ii) To show how similar levels of efficient technologies shared by the parties led to a higher aggregated level of electoral corruption, I proceed as follows. The reaction function $\hat{e}_{i\in\{a,b\}}$ changes with respect to the effort that the opposite party exerts e_2 as:

$$\frac{\partial \hat{e}_1}{\partial e_2} = \frac{\sqrt{t_1 R}}{\sqrt{2e_2 + S}} - 1$$

This function is maximized setting $\frac{\delta \hat{e}_1}{\delta e_2} = 0$, which results in $t_1 = \frac{2e_1+S}{W}$. Given that both parties share the same efficiency technology $t_1 = t_2$, $\frac{2e_1+S}{R} = \frac{2e_2+S}{R}$, which leads to $t_1 = t_2$. To be sure that the function reaches a maximum at $t_1 = t_2$, it is enough to show that the second derivative of the function is negative or:

$$\frac{\partial^2 \hat{e}_1}{\partial e_2^2} = -\frac{1}{(2e_2 + S)^{\frac{3}{2}}} < 0$$

Therefore, the relative change of Party i's effort with respect to Party $-i$ effort reaches a maximum when $t_1 = t_2$. □

A.3 Proof of Proposition 2

Proof. First, suppose both parties engage in electoral corruption. I must show then that e_1^* and e_2^* are the effort levels for a Nash equilibrium. These quantities are the unique solution (e_1^*, e_2^*) to the first order conditions. To show that each player is maximizing against the other, I evaluate the expected utility for either player as follows:

$$\frac{\partial^2 u_1}{\partial e_1^2} = -R[\frac{2e_2(e_1+e_2+S)+S(\frac{1}{2}+\frac{\xi\omega\eta}{1-\omega})}{(e_1+e_2+S)^3}]$$

Given that parties only exert a positive effort when manipulating votes affect the probability of winning the election, $|\frac{\xi\omega\eta}{1-\omega}| < \frac{1}{2}$, $e_{i\in\{a,b\}}$, $\frac{\partial^2 u_{i\in\{a,b\}}}{\partial e^2_{i\in\{a,b\}}} < 0$, which is a sufficient condition for a pure strategy equilibrium.

Now, let $t_1 = (t_{i,1}, ..., t_{i,D})$ and $t_1^{'} = (t_{i,1}^{'}, ..., t_{i,D}^{'})$ be two vectors of efficiency technologies across districts such that $t_{i,1} > t_{i,1}^{'}$, $t_{i,1} = t_{-i,1}^{'}$, and $t_1 = t_{1,d}^{'}$, $t_1 = t_{2,d}^{'}$ for all $d \neq 1$. That is, the only difference between t_1 and $t_1^{'}$ is I take equation (2.45) as the utility functions of Party i and compare their utility as the expected number of votes less the total cost from the effort $u_1^* = \Pi_1 - C_1$, or:

$$u_1^* = R\sum_{d=1}^{D}\left[\frac{1}{2}+\omega_d\eta_d+(1-\omega_d)\frac{1}{2\xi}\frac{t_1-t_2}{t_1+t_2}\right]-\sum_{d=1}^{D}\left[\frac{2W_d t_{1,d}^2 t_2}{(t_1+t_2)^2}-\frac{S}{2t_1}\right]$$

Observe that Π_1 will be the same under t_1 as it is under $t_1^{'}$. As $t_1 = t_2$ and $t_{1,d}^{'} = t_{2,d}^{'}$, both parties expect the same number of votes regardless of their efficiency technology level. The difference is in the total cost exerted by the party given its effort level. In this case, C_1 becomes $\sum_{d=1}^{D}\left[W_d t_1 - \frac{S}{2t_1}\right]$, which is greater under t_1 than under $a_1^{'}$. The derivative of the cost function, $\frac{\delta C_1}{\delta t_1} = W - \frac{S}{2}$ will increase for all the values of t_1 when Assumption 1 holds, raising the costs while the effect on the expected number of votes remains only determined by the partisan votes.

□

A.4 Proof of Proposition 3

Proof. Let $p_{1|\alpha_H}$ and $p_{2|\alpha_H}$ be two vectors of equilibrium probabilities for parties 1 and 2 across precincts when $t_1 = t_2$. Similarly, $p_{1|\alpha_L}$ and $p_{2|\alpha_L}$ be the equilibrium probabilities vectors for parties 1 and 2 to disrupt in the precincts given their effort levels at $t_1 = t_2/\alpha_L; alpha_L < 1$. Given these two different probability combinations, the governor chooses the α level that maximizes the probability of a victory for Party 1 in order to receive a reward αK. That is, she will opt for an asymmetry in the resources between the parties of magnitude α_L only when:

$$K\sum_{d=1}^{D}\gamma_d[\frac{1}{2}+\omega_d\eta d+(1-\omega_d)\frac{1}{2\xi}(p_{1,d|\alpha_H}-p_{2,d|\alpha_H})] <$$
$$\alpha_L K\sum_{d=1}^{D}\gamma_d[\frac{1}{2}+\omega_d\eta d+(1-\omega_d)\frac{1}{2\xi}(p_{1,d|\alpha_L}-p_{2,d|\alpha_L})]$$

Rearranging the terms, the equation above leads to:

$$\frac{\sum_{d=1}^{D}\gamma_d(1-\omega_d)(p_{1,d|\alpha_H}-p_{2,d|\alpha_H})}{\sum_{d=1}^{D}\gamma_d(1-\omega_d)(p_{1,d|\alpha_L}-p_{2,d|\alpha_L})} < \alpha_L$$

In this case, the willingness of a governor to intervene in the election depends on the reaction function of both parties at different levels of α.

□

Appendix B

Chapter 4

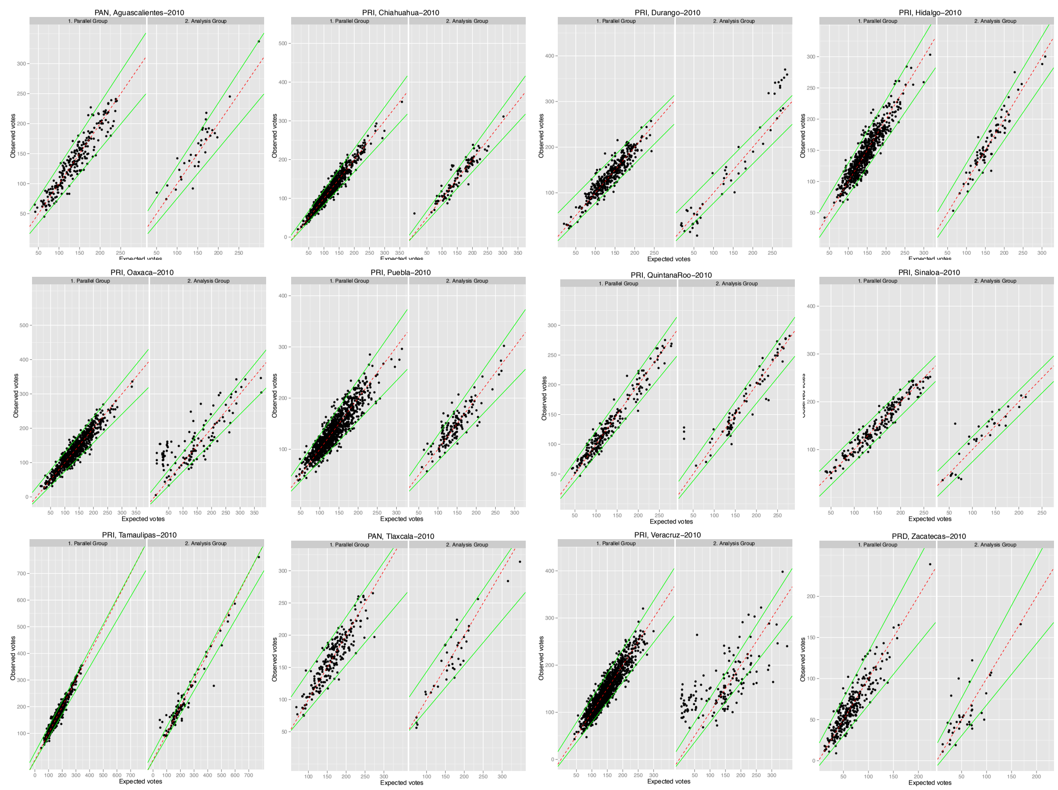


Figure B.1: Observed and expected votes for the incumbent party or coalition.

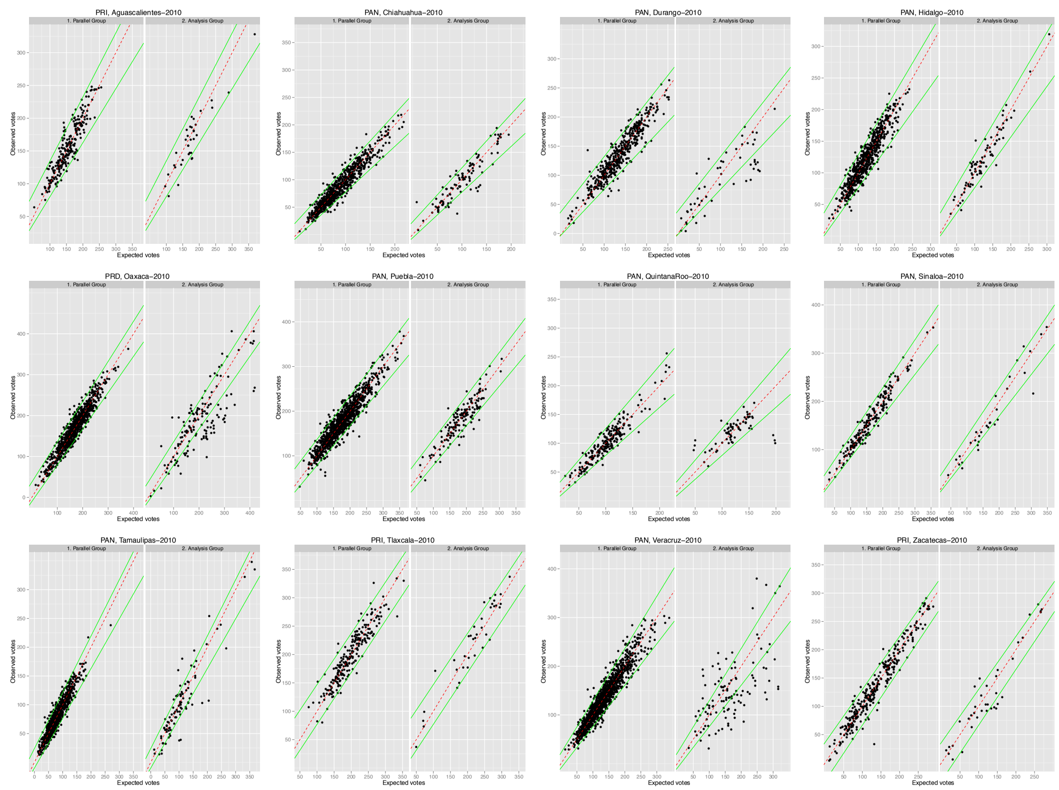


Figure B.2: Observed and expected votes for the challenger party or coalition.

Table B.1: Electoral results for Governor elections in Mexico, 2010.

State	Total of votes	Incumbent	Challenger	Others	Null and non-registered votes
Aguascalientes	429,308	PAN 42.47%	PRI+PVEM+PANAL 47.66%	PRD, PT 7.04%	2.83%
Chihuahua	1,081,825	PRI+PVEM+PANAL+PT 55.50%	PAN 39.14%	PRD 1.99%	3.37%
Durango	610,105	PRI 46.5%	PAN+PRD+CD+PT 44.6%	PT, PVEM, PD 5.6%	3.3%
Hidalgo	871,165	PRI+PVEM+PANAL 50.28%	PAN+PRD+PH+CD 45.23%		4.48%
Oaxaca	1,464,237	PRI+PVEM 41.9%	PAN+PRD+PT+CD 50.11%	PUP, PANAL 4.71%	3.24%
Puebla	2,203,868	PRI+PVEM 40.1%	PAN+PRD+CD+PANAL 50.4%	PT 5.6%	3.9%
Quintana Roo	351,752	PRI+PVEM+PANAL 52.42%	PRD+PT+CD 26.19%	PAN 15.43%	5.6%
Sinaloa	1,111,891	PRI+PVEM+PANAL 46.4%	PAN+PRD+CD 51.8%		1.8%
Tamaulipas	1,101,698	PRI+PVEM+PANAL 61.58%	PAN 30.81%	PRD, PT, CD 5.14%	2.42%
Tlaxcala	498,423	PRI+PVEM 46.47%	PAN+PANAL+PAC 38.86%	PRD+PT+CD, PS 11.39	3.27%
Veracruz	3,115,790	PRI+PVEM+PV 43.54%	PAN+PANAL 40.99%	PRD+PT+CD 12.90%	2.57%
Zacatecas	658,212	PRD+CD 23.22%	PRI+PVEM+PANAL 43.19%	PAN 30.54%	2.71%

Table B.2: Mean and standard errors for the T.V. viewership poll grouping respondents by their last name. July, 2011.

Variable	Group 1	Group 2
Age	41.3	40.68
	(0.64)	(0.58)
Female	0.7	0.71
	(0.02)	(0.02)
T.V. viewership	1.65	1.65
(1-4 scale)	(0.02)	(0.02)

Table B.3: Mean and standard errors for the T.V. viewership poll grouping respondents by their last name. July, 2011.

Variable	Group 1	Group 2	Group 3
Age	41.66	41	40.23
	(0.75)	(0.77)	(0.72)
Female	0.71	0.68	0.73
	(0.02)	(0.02)	(0.02)
T.V. viewership	1.63	1.67	1.64
(1-4 scale)	(0.03)	(0.03)	(0.03)

Table B.4: Mean and standard errors for the T.V. viewership poll grouping respondents by their last name. July, 2011.

Variable	Group 1	Group 2	Group 3	Group 4
Age	41.34	41.27	40.77	40.54
	(0.93)	(0.97)	(0.76)	(0.91)
Female	0.68	0.71	0.7	0.73
	(0.03)	(0.03)	(0.02)	(0.03)
T.V. viewership	1.64	1.66	1.66	1.64
(1-4 scale)	(0.03)	(0.03)	(0.03)	(0.03)

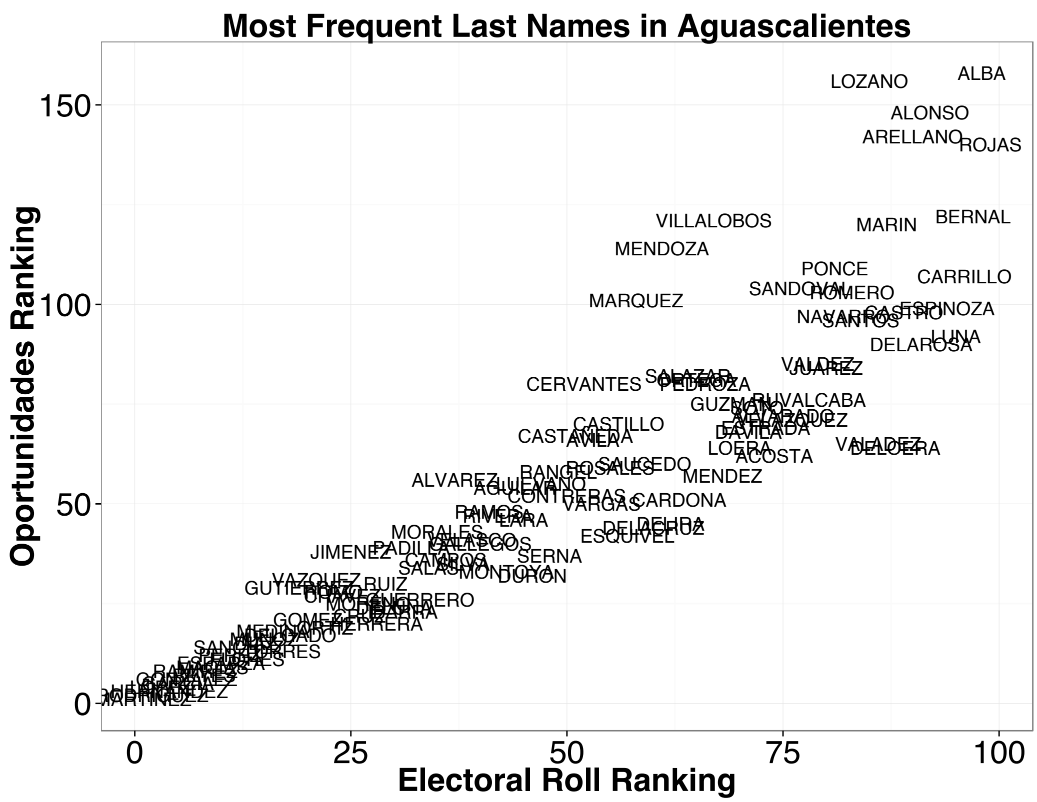


Figure B.3: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Aguascalientes.

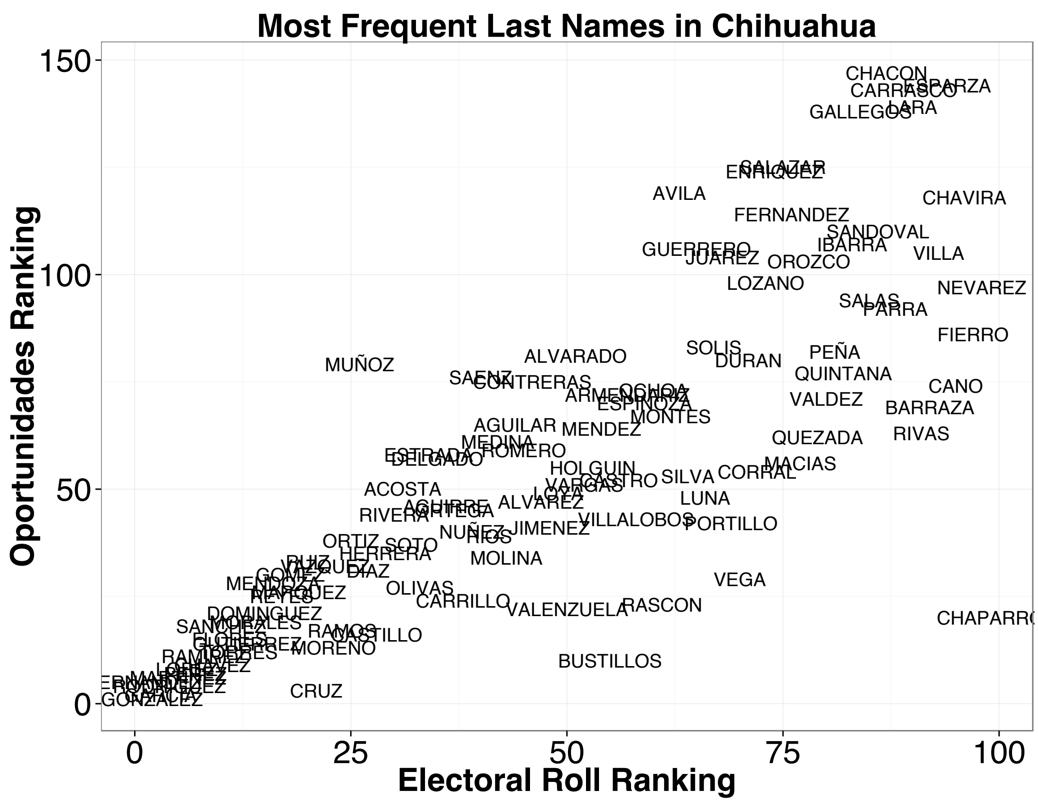


Figure B.4: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Chihuahua.

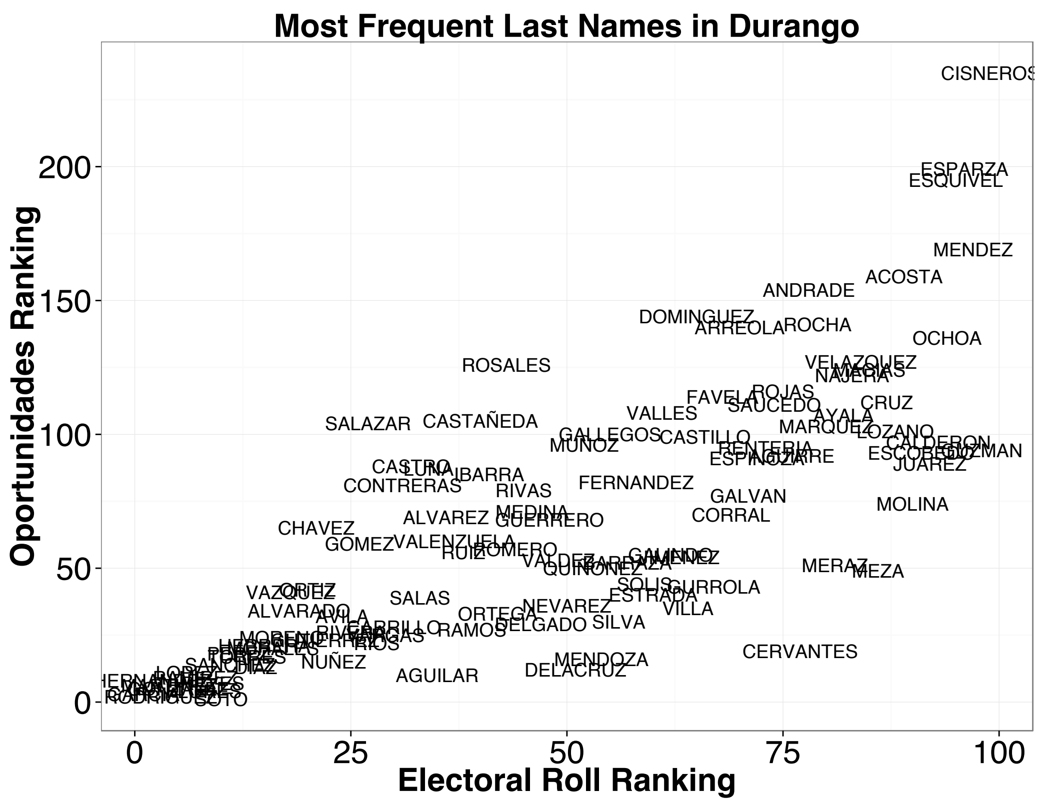


Figure B.5: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Durango.

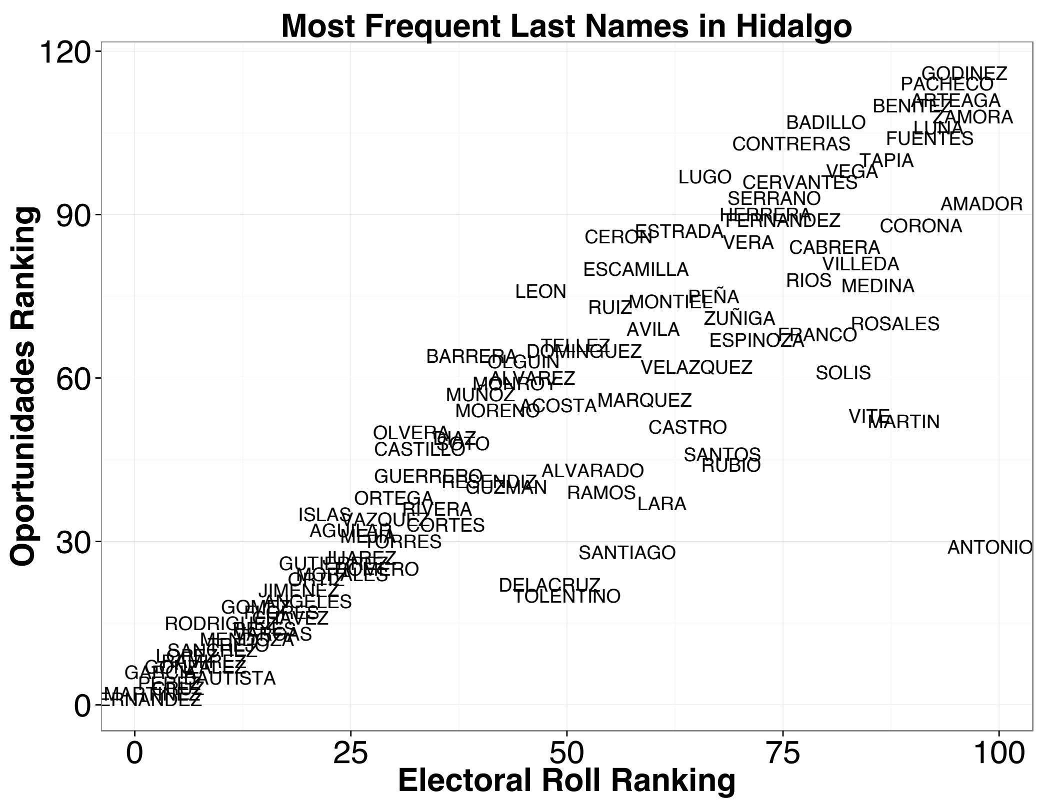


Figure B.6: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Hidalgo.

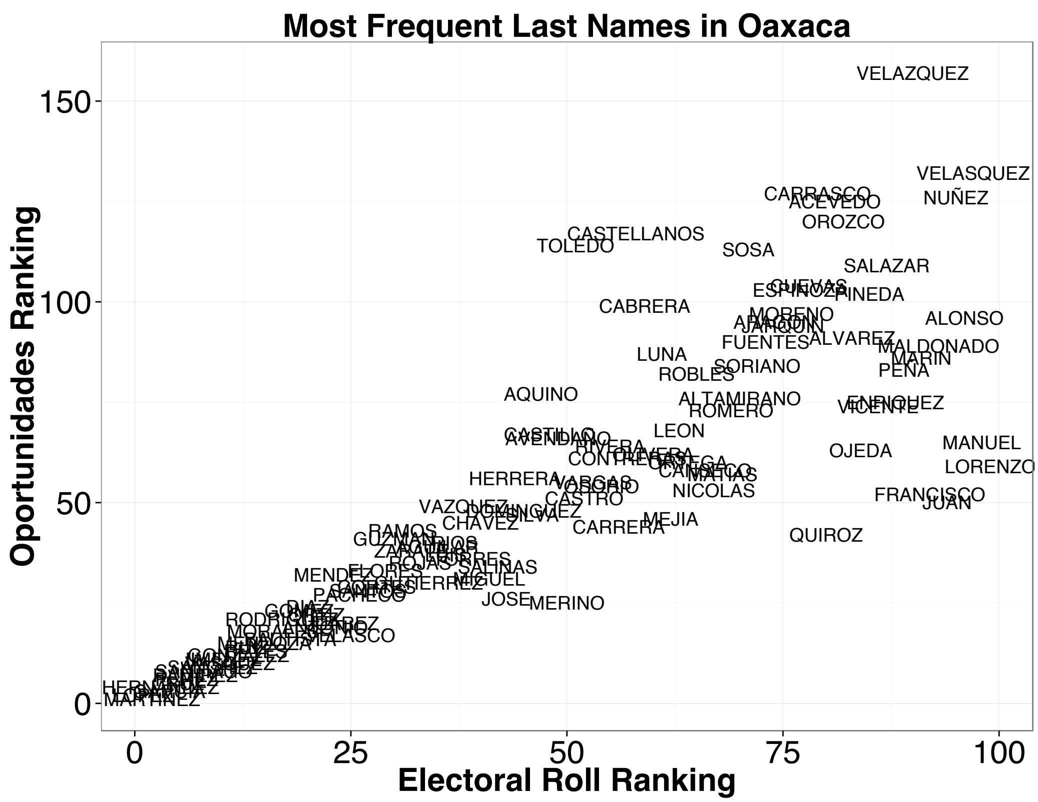


Figure B.7: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Oaxaca.

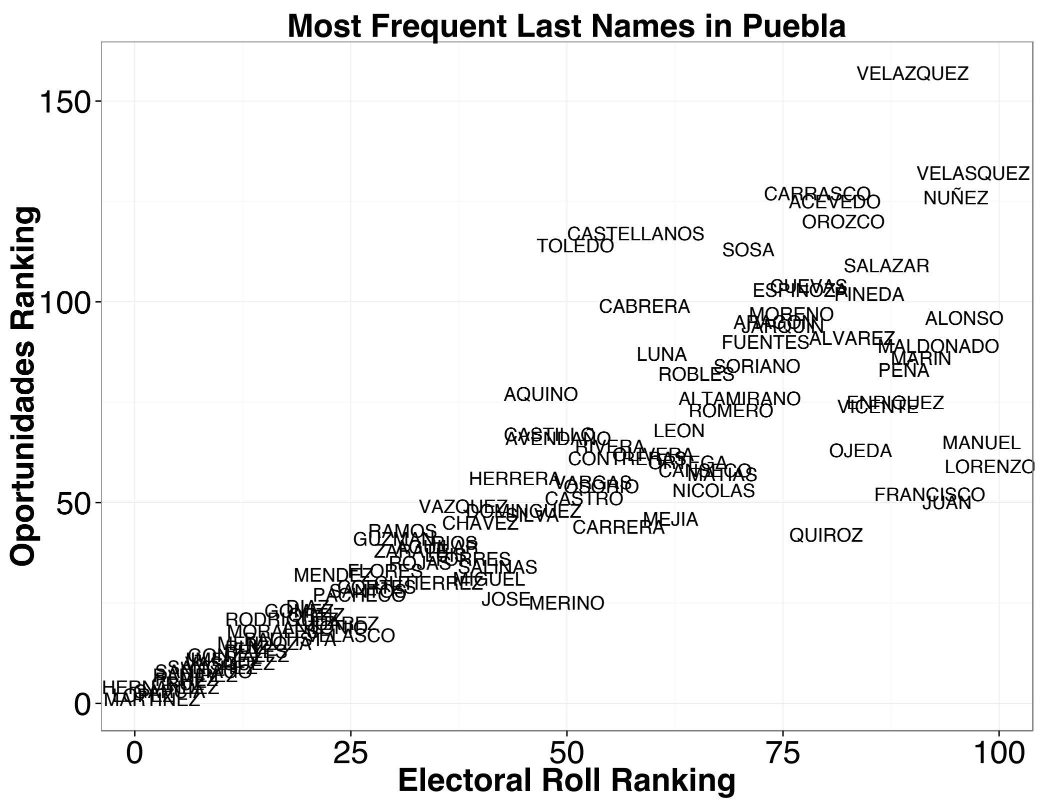


Figure B.8: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Puebla.

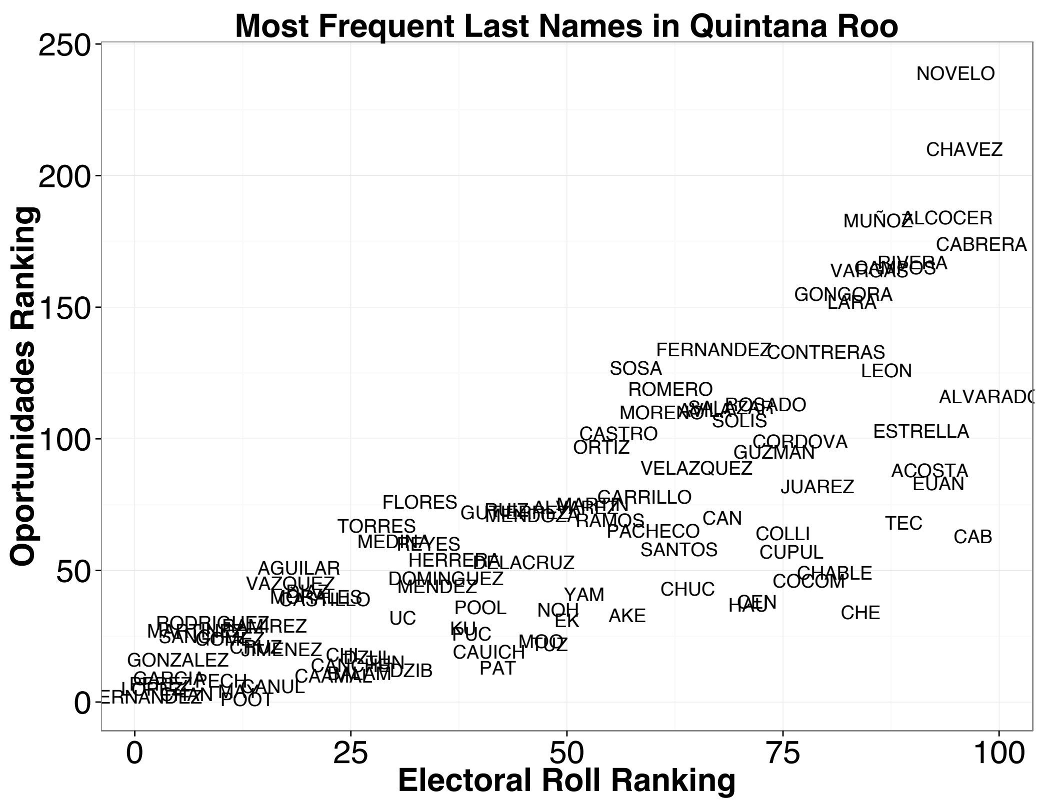


Figure B.9: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in QuintanaRoo.

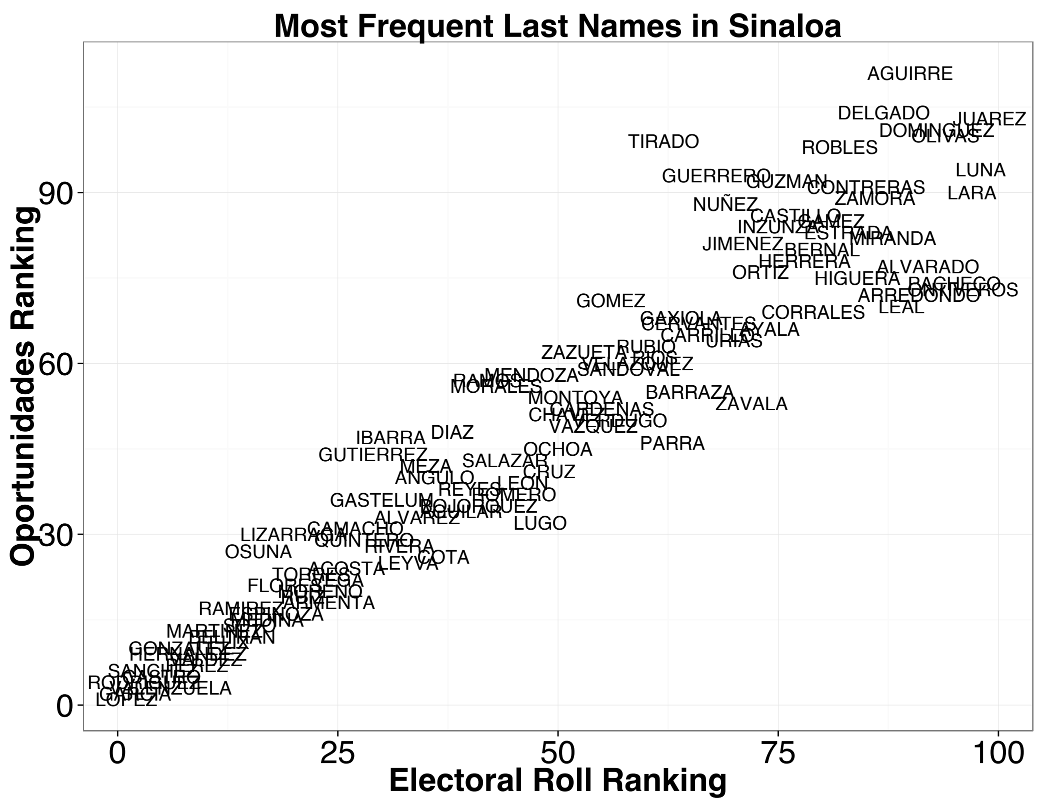


Figure B.10: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Sinaloa.

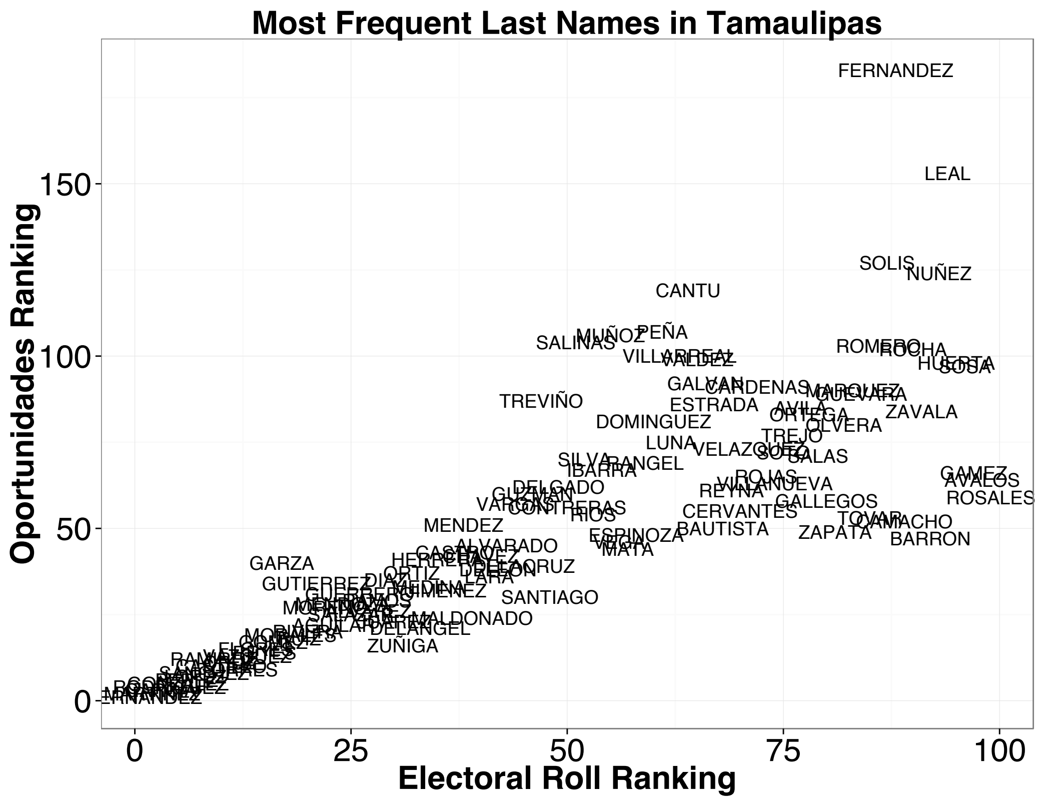


Figure B.11: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Tamaulipas.

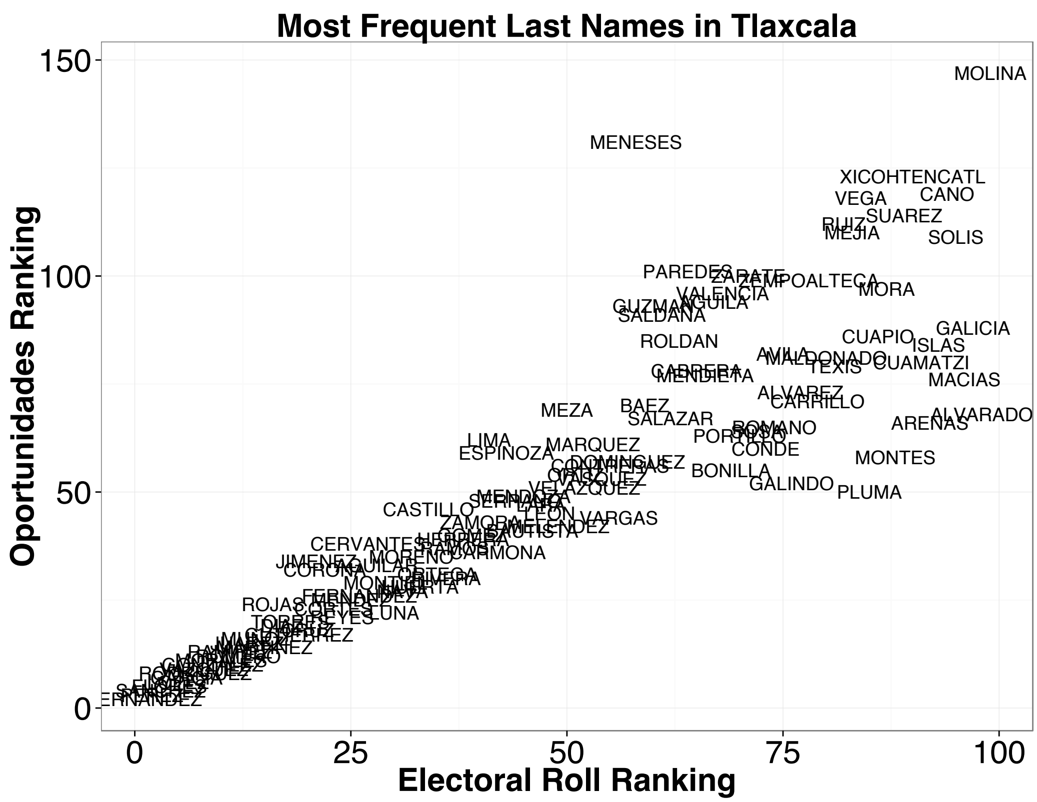


Figure B.12: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Tlaxcala.

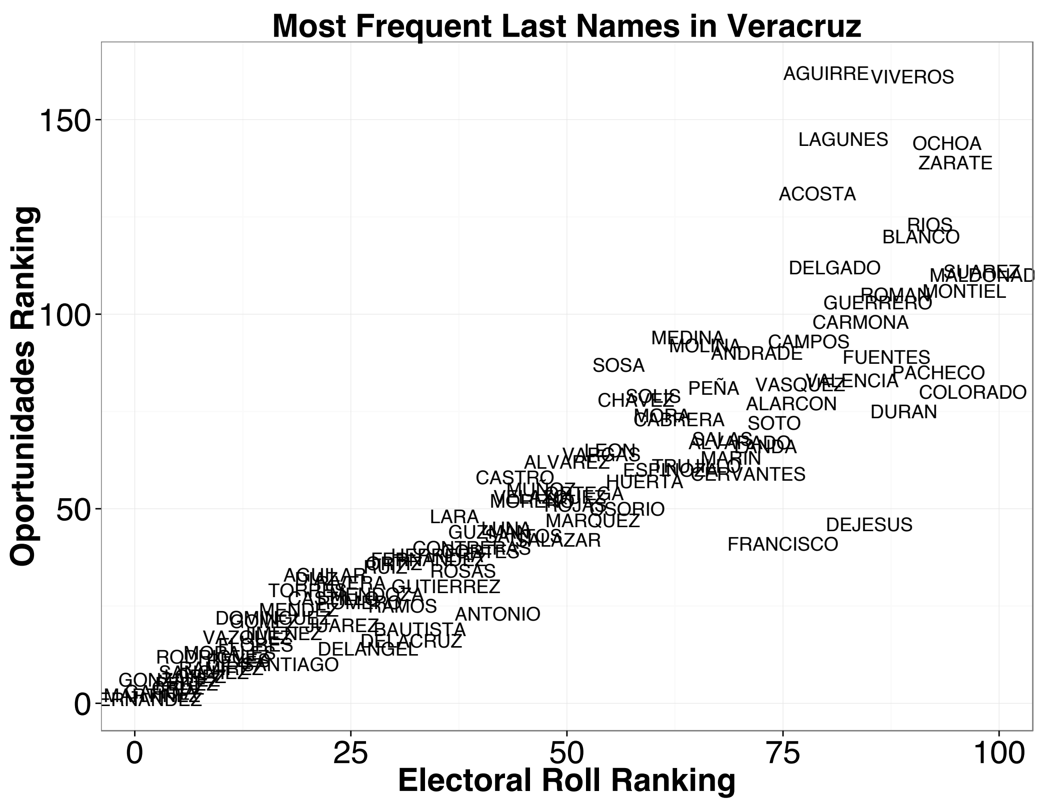


Figure B.13: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Veracruz.

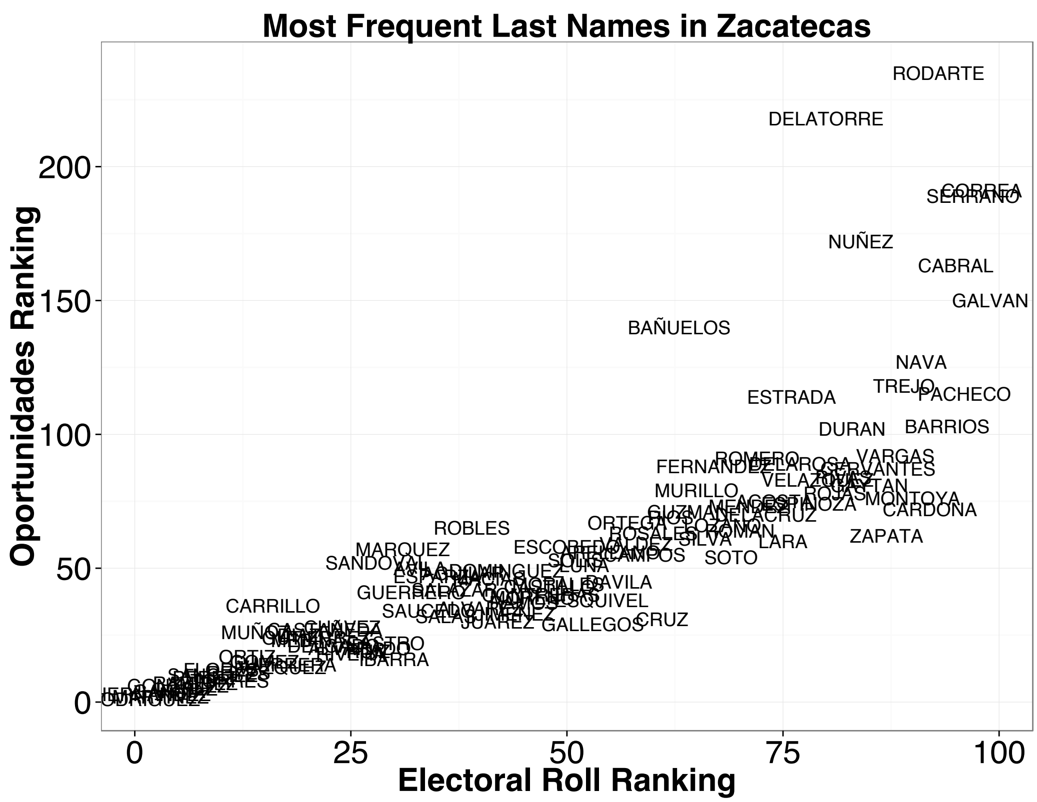


Figure B.14: Ranking comparison for the last names in the voting roll and the children who benefit from Oportunidades in Zacatecas.

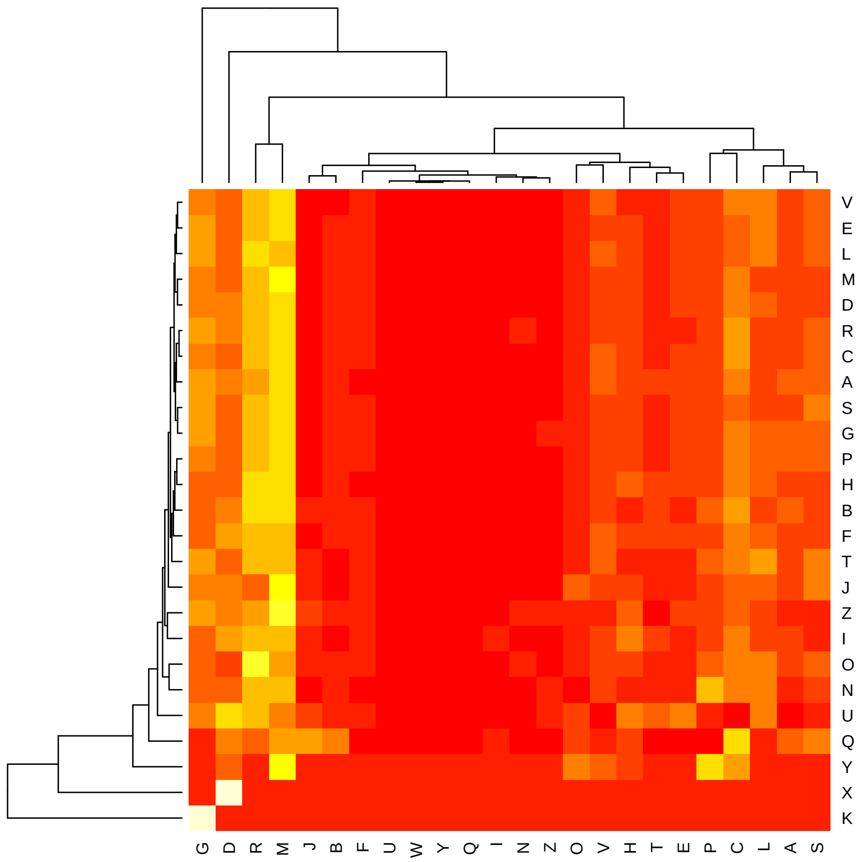


Figure B.15: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Aguascalientes, 2012.

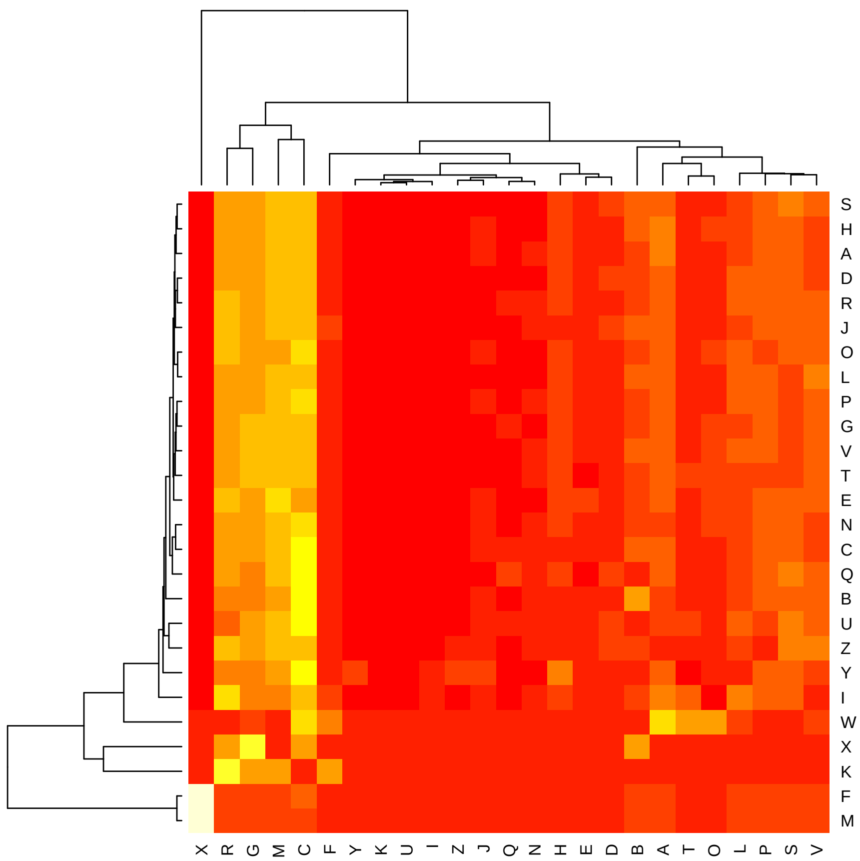


Figure B.16: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Chihuahua, 2012.

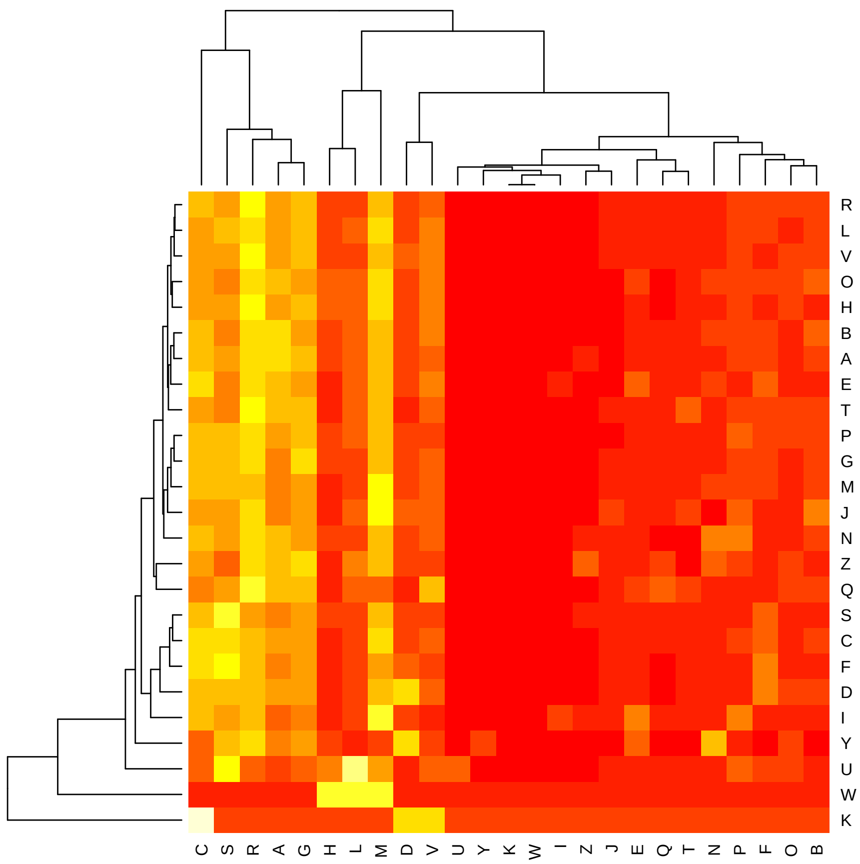


Figure B.17: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Durango 2012.

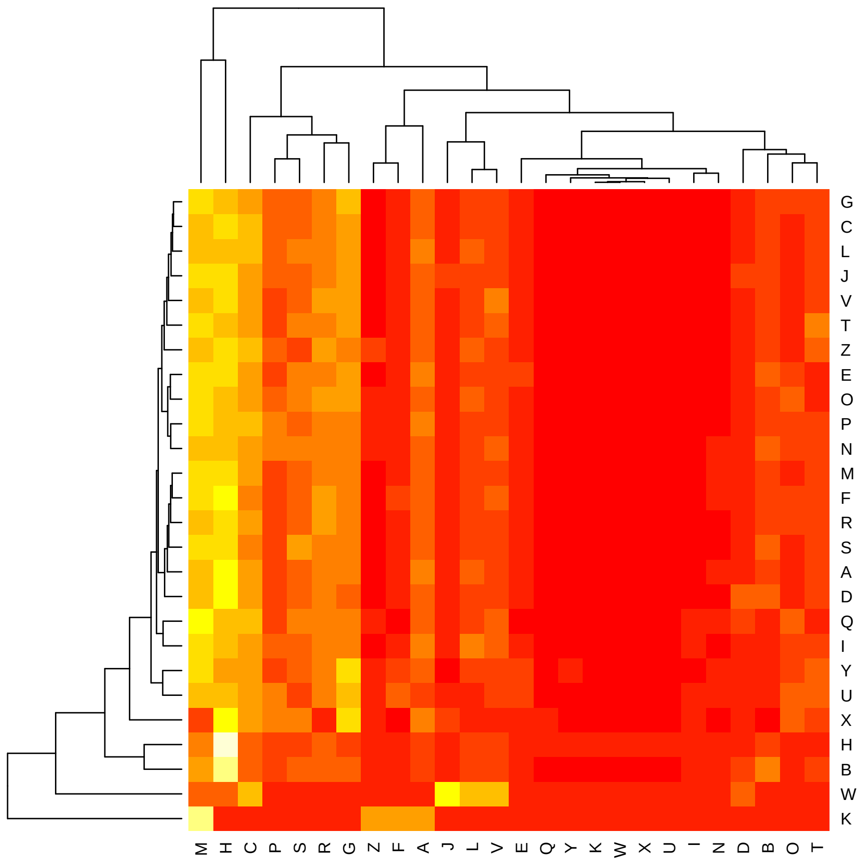


Figure B.18: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Hidalgo, 2012.

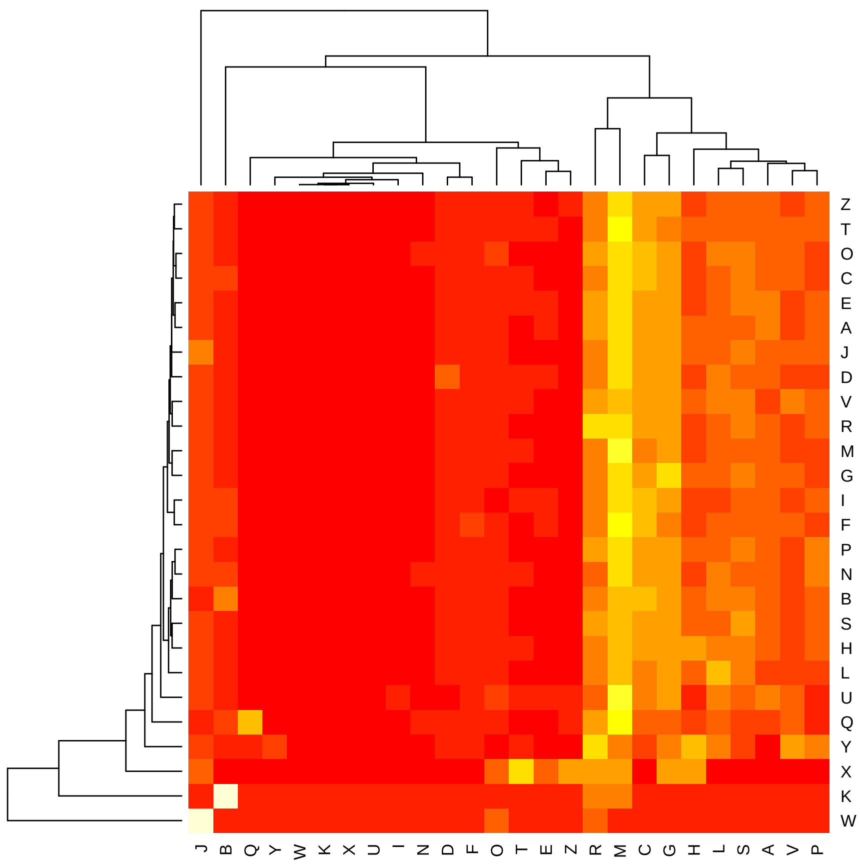


Figure B.19: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Oaxaca, 2012.

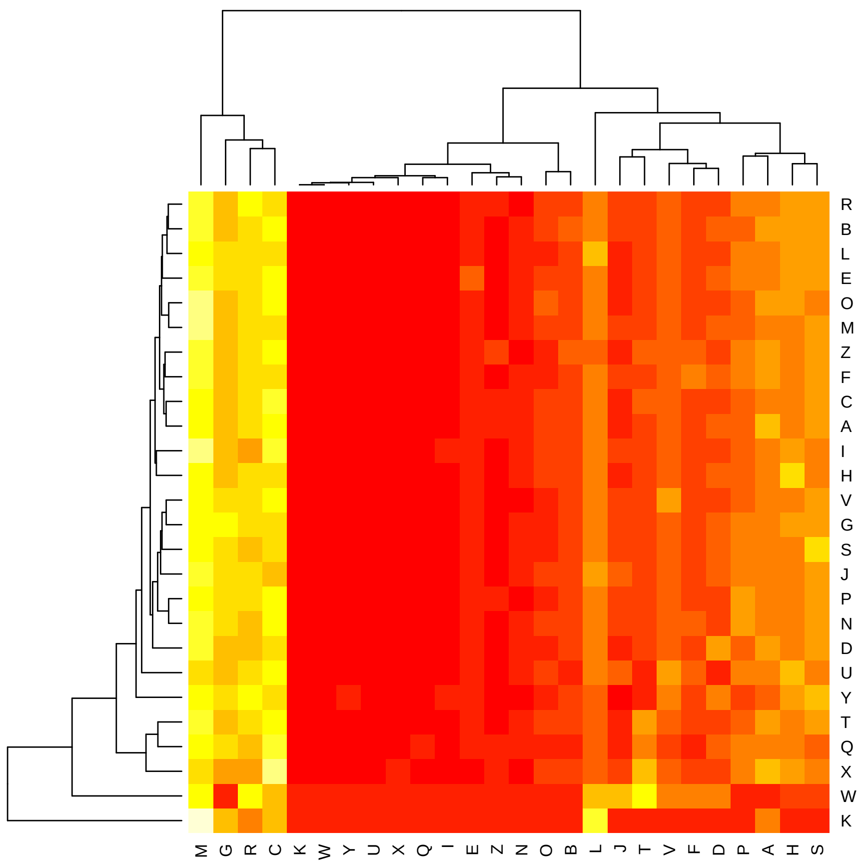


Figure B.20: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Puebla 2012.

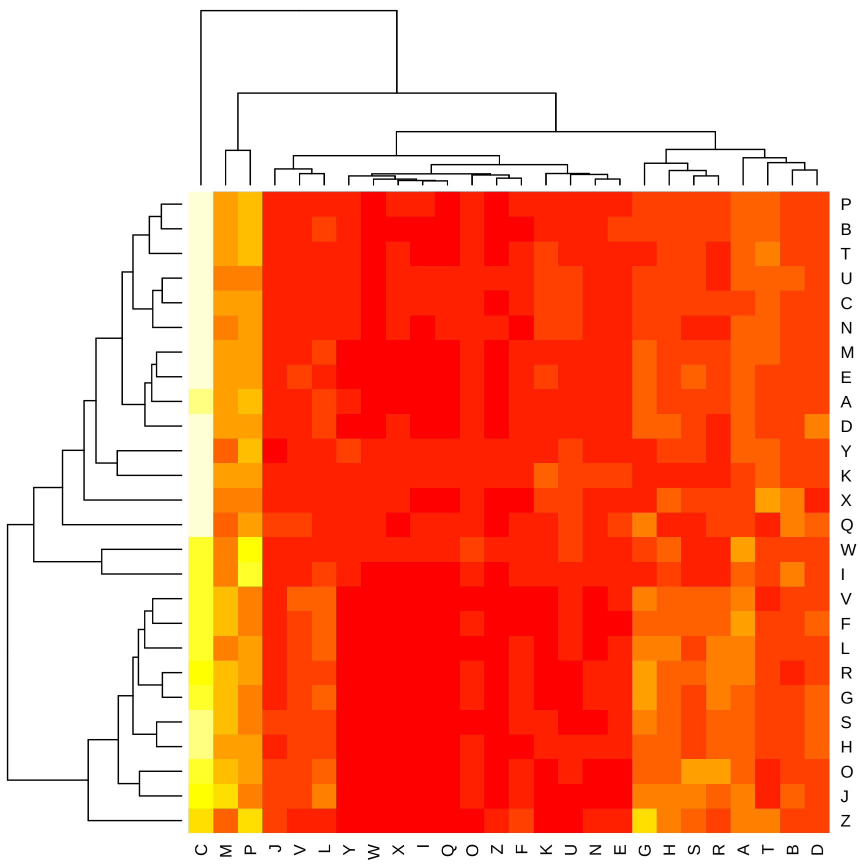


Figure B.21: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Quintana Roo, 2012.

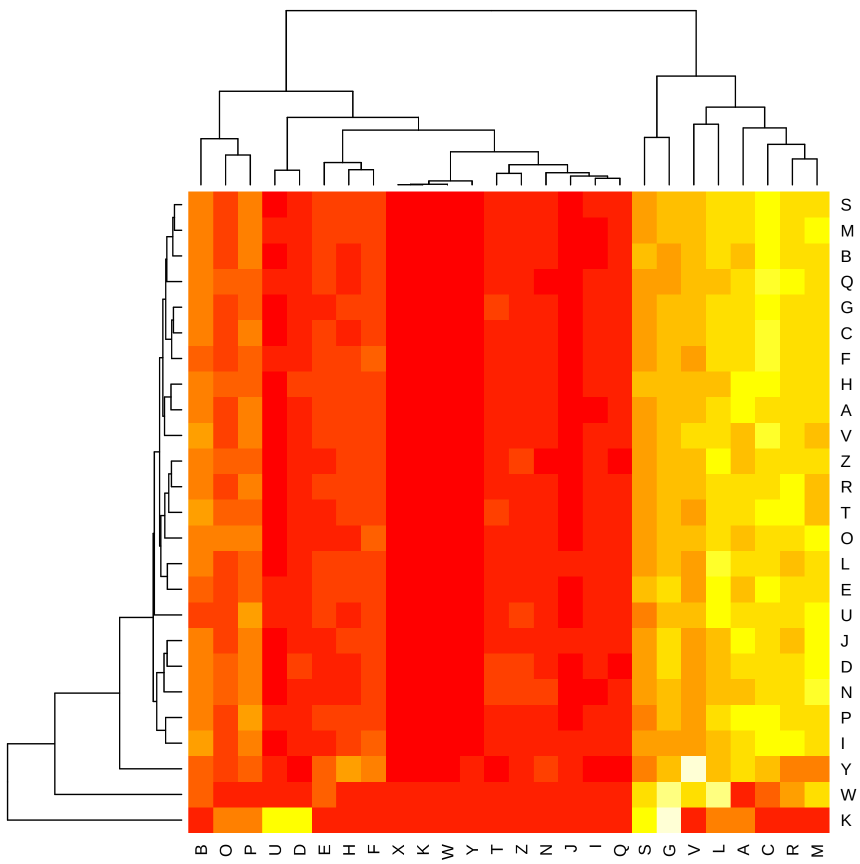


Figure B.22: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Sinaloa, 2012.

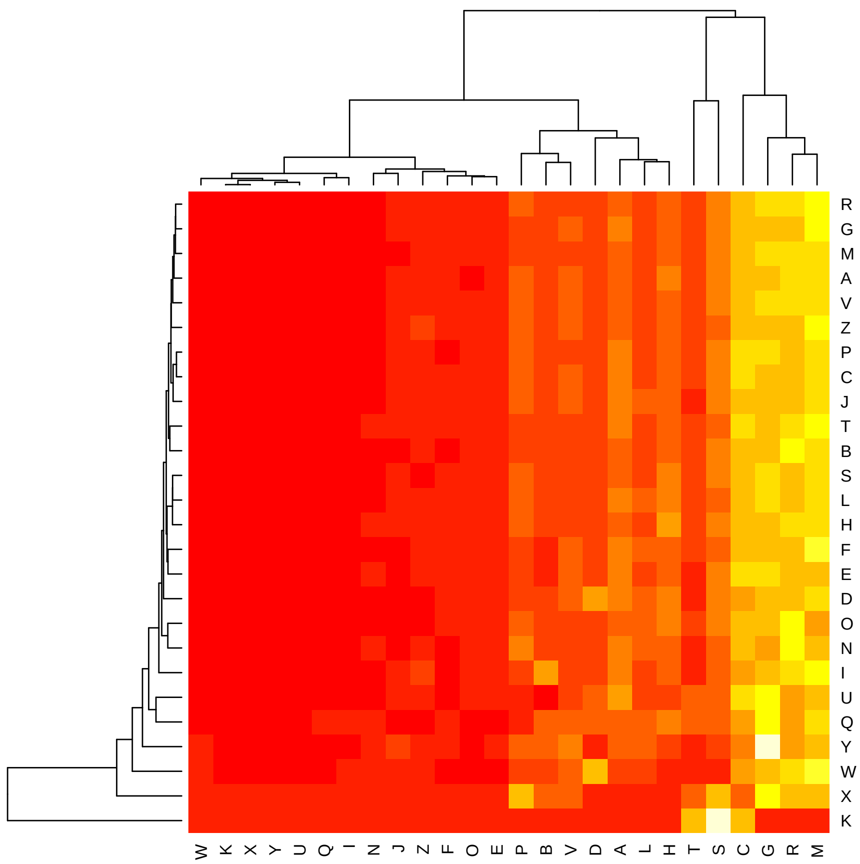


Figure B.23: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Tamaulipas, 2012.

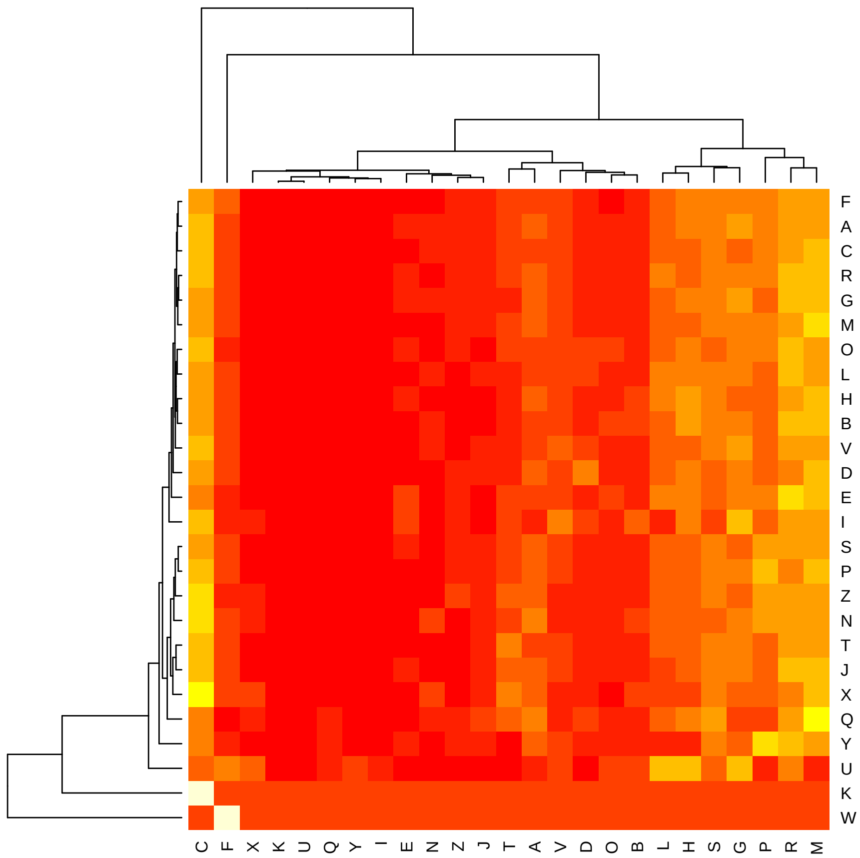


Figure B.24: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Tlaxcala, 2012.

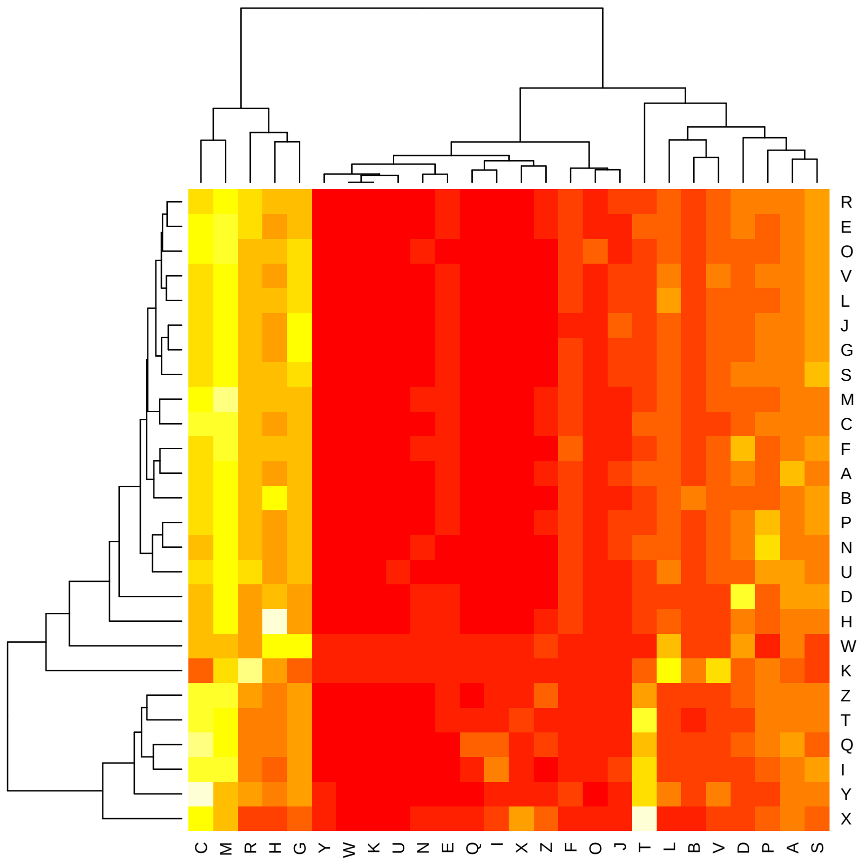


Figure B.25: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Veracruz, 2012.

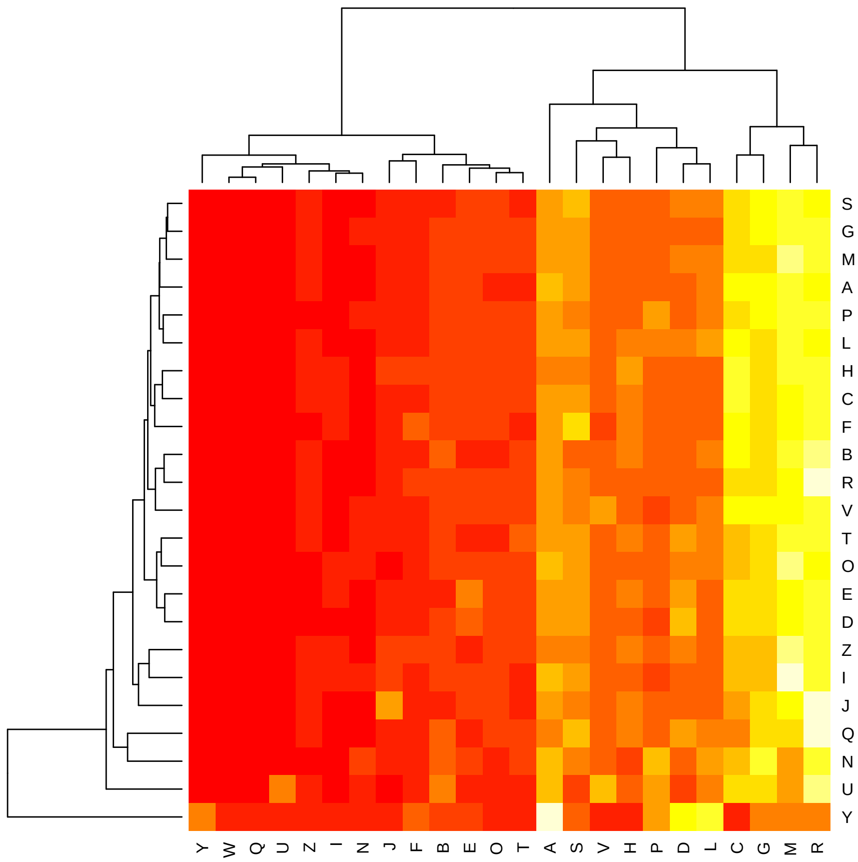


Figure B.26: Heat maps and endrograms for frequencies of the first letter in father's last name and mother's last name among children who benefit from Oportunidades. Zacatecas, 2012.

Table B.5: Correlation coefficient of the 100 most frequent last names among beneficiaries of Oportunidades and the electoral-roll.

State	Correlation coefficient
Aguascalientes	0.880
Chihuahua	0.796
Durango	0.760
Hidalgo	0.868
Oaxaca	0.835
Puebla	0.922
Quintana Roo	0.764
Sinaloa	0.944
Tamaulipas	0.815
Tlaxcala	0.878
Veracruz	0.894
Zacatecas	0.828

Table B.6: Frequency of "suspicious polling stations" by label

Number polling stations in the precinct	Precincts	Polling Stations	B	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c19
2	453	488	0.49	0.51																	
3	260	326	0.35	0.30	0.35																
4	85	115	0.37	0.20	0.20	0.23															
5	40	68	0.21	0.18	0.24	0.25	0.13														
6	20	34	0.24	0.21	0.18	0.18	0.09	0.12													
7	16	26	0.12	0.23	0.15	0.12	0.19	0.12	0.08												
8	7	16	0.19	0.31	0.06	0.13	0.13	0.13	0.00	0.06											
9	8	10	0.20	0.10	0.10	0.00	0.00	0.00	0.20	0.20	0.20										
10	8	20	0.15	0.00	0.10	0.05	0.10	0.05	0.15	0.15	0.10	0.15									
11	3	4	0.00	0.25	0.00	0.00	0.00	0.25	0.00	0.00	0.25	0.25	0.00								
12	5	24	0.08	0.13	0.04	0.08	0.08	0.08	0.04	0.04	0.13	0.13	0.50	0.08							
13	4	20	0.10	0.10	0.05	0.05	0.10	0.10	0.05	0.10	0.10	0.05	0.04	0.10	0.05						
14	1	3	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.33	0.00	0.33					
15	3	5	0.20	0.00	0.00	0.20	0.00	0.00	0.00	0.20	0.00	0.20	0.00	0.00	0.00	0.20	0.00				
16	2	6	0.00	0.17	0.00	0.00	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.17	0.17			
17	4	23	0.09	0.13	0.04	0.04	0.09	0.04	0.04	0.00	0.04	0.09	0.17	0.09	0.09	0.04	0.00	0.09	0.04		
18	2	17	0.00	0.06	0.06	0.00	0.06	0.06	0.06	0.06	0.06	0.06	0.09	0.06	0.06	0.06	0.06	0.06	0.06	0.06	
19	1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
20	2	12	0.17	0.00	0.00	0.08	0.00	0.17	0.08	0.17	0.00	0.00	1.00	0.00	0.17	0.00	0.00	0.08	0.00	0.00	0.00
21	1	1	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table B.7: Analysis of local elections in Mexico by other fraud detection algorithms. Incumbent parties in 2010.

	Second Digit (p-value)	Last Digit (p-value)	Turnout (β coefficient)
Aguascalientes	0.76	0.44	0.26
Chihuahua	0.02	0.38	0.58
Durango	0.48	0.77	0.40
Hidalgo	0.82	0.43	0.50
Oaxaca	0.20	0.43	0.62
Puebla	0.19	0.65	0.45
Quintana Roo	0.00	1.00	0.57
Sinaloa	0.02	0.90	0.29
Tamulipas	0.19	0.10	0.61
Tlaxcala	0.09	0.63	0.47
Veracruz	0.18	0.64	0.39
Zacatecas	0.02	0.45	0.48

Table B.8: Analysis of local elections in Mexico by other fraud detection algorithms. Incumbent parties in 2010.

	Second Digit (p-value)	Last Digit (p-value)	Turnout (β coefficient)
Aguascalientes	0.15	0.70	0.63
Chihuahua	0.03	0.76	0.40
Durango	0.37	0.41	0.54
Hidalgo	0.11	0.95	0.46
Oaxaca	0.69	0.68	0.35
Puebla	0.15	0.76	0.46
Quintana Roo	0.99	0.30	0.25
Sinaloa	0.02	0.14	0.70
Tamulipas	0.25	0.34	0.32
Tlaxcala	0.07	0.97	0.40
Veracruz	0.16	0.59	0.45
Zacatecas	0.40	0.02	0.36

Appendix C

Chapter 5

Table C.1: Regression Results from Turnout versus Votes for each Candidate. In those precincts where López Obrador had the plurality of votes in 2006.

	PRI		PRD		PAN	
	β_V	R^2	β_V	R^2	β_V	R^2
Aguascalientes	0.01	0.00	0.41	0.39	0.60	0.41
Baja California	0.22	0.23	0.35	0.38	0.41	0.49
Baja California Sur	0.43	0.28	0.28	0.17	0.23	0.13
Campeche	0.58	0.64	0.11	0.04	0.23	0.23
Chiapas	0.59	0.35	0.26	0.09	0.09	0.02
Chihuahua	0.36	0.27	0.26	0.24	0.35	0.37
Coahuila	0.11	0.02	0.28	0.23	0.59	0.41
Colima	0.20	0.12	0.41	0.40	0.39	0.39
Distrito Federal	0.22	0.15	0.11	0.01	0.65	0.38
Durango	0.29	0.12	0.26	0.19	0.42	0.33
Guanajuato	0.31	0.20	0.15	0.07	0.50	0.36
Guerrero	0.35	0.23	0.51	0.34	0.11	0.05
Hidalgo	0.34	0.18	0.44	0.25	0.18	0.08
Jalisco	0.08	0.01	0.20	0.08	0.69	0.38
Mexico	0.37	0.17	0.12	0.02	0.50	0.31
Michoacan	0.39	0.35	0.23	0.10	0.32	0.26
Morelos	0.31	0.17	0.49	0.31	0.18	0.06
Nayarit	0.37	0.26	0.37	0.21	0.25	0.15
Nuevo Leon	-0.12	0.02	0.14	0.06	0.99	0.44
Oaxaca	0.40	0.32	0.45	0.28	0.15	0.06
Puebla	0.20	0.07	0.38	0.18	0.40	0.19
Queretaro	0.09	0.01	0.20	0.12	0.72	0.45
Quintana Roo	0.48	0.40	0.29	0.30	0.20	0.10
San Luis Potosí	0.32	0.21	0.31	0.24	0.32	0.19
Sinaloa	0.35	0.24	0.35	0.29	0.28	0.17
Sonora	0.33	0.32	0.36	0.26	0.27	0.17
Tabasco	0.31	0.13	0.57	0.29	0.11	0.06
Tamaulipas	0.20	0.17	0.34	0.38	0.41	0.36
Tlaxcala	0.39	0.19	0.40	0.14	0.19	0.07
Veracruz	0.39	0.20	0.36	0.10	0.25	0.05
Yucatan	0.50	0.26	0.20	0.06	0.30	0.09
Zacatecas	0.22	0.11	0.27	0.17	0.49	0.35

Table C.2: Regression Results from Turnout versus Votes for each Candidate. In those precincts where López Obrador did not have the plurality of votes in 2006.

	PRI		PRD		PAN	
	β_V	R^2	β_V	R^2	β_V	R^2
Aguascalientes	0.23	0.09	0.19	0.11	0.54	0.34
Baja California	0.29	0.22	0.06	0.01	0.60	0.58
Baja California Sur	0.60	0.27	0.23	0.29	0.20	0.05
Campeche	0.55	0.49	0.06	0.01	0.35	0.26
Chiapas	0.67	0.30	0.28	0.09	0.06	0.01
Chihuahua	0.48	0.36	0.07	0.03	0.40	0.43
Coahuila	0.21	0.06	0.17	0.11	0.61	0.41
Colima	0.51	0.29	0.14	0.05	0.34	0.22
Distrito Federal	0.44	0.68	0.12	0.04	0.45	0.43
Durango	0.48	0.20	0.17	0.07	0.32	0.17
Guanajuato	0.28	0.12	0.15	0.12	0.54	0.34
Guerrero	0.37	0.22	0.43	0.29	0.09	0.03
Hidalgo	0.44	0.20	0.29	0.12	0.22	0.06
Jalisco	0.27	0.09	0.22	0.07	0.46	0.21
Mexico	0.56	0.25	0.14	0.03	0.30	0.10
Michoacan	0.54	0.40	0.17	0.08	0.28	0.13
Morelos	0.57	0.47	0.51	0.40	-0.14	0.05
Nayarit	0.55	0.30	0.14	0.04	0.30	0.11
Nuevo Leon	0.10	0.02	0.17	0.13	0.71	0.53
Oaxaca	0.47	0.32	0.35	0.19	0.18	0.06
Puebla	0.51	0.28	0.01	0.00	0.42	0.19
Queretaro	0.49	0.18	0.18	0.06	0.31	0.11
Quintana Roo	0.72	0.47	0.11	0.02	0.16	0.04
San Luis Potosí	0.48	0.25	0.21	0.09	0.25	0.10
Sinaloa	0.45	0.19	0.19	0.10	0.34	0.12
Sonora	0.45	0.49	0.10	0.07	0.37	0.42
Tabasco	0.40	0.17	0.47	0.16	0.15	0.04
Tamaulipas	0.56	0.37	0.06	0.01	0.38	0.25
Tlaxcala	0.60	0.41	0.12	0.04	0.23	0.12
Veracruz	0.52	0.26	0.05	0.01	0.42	0.14
Yucatan	0.56	0.41	-0.12	0.03	0.57	0.42
Zacatecas	0.46	0.25	0.08	0.03	0.45	0.24

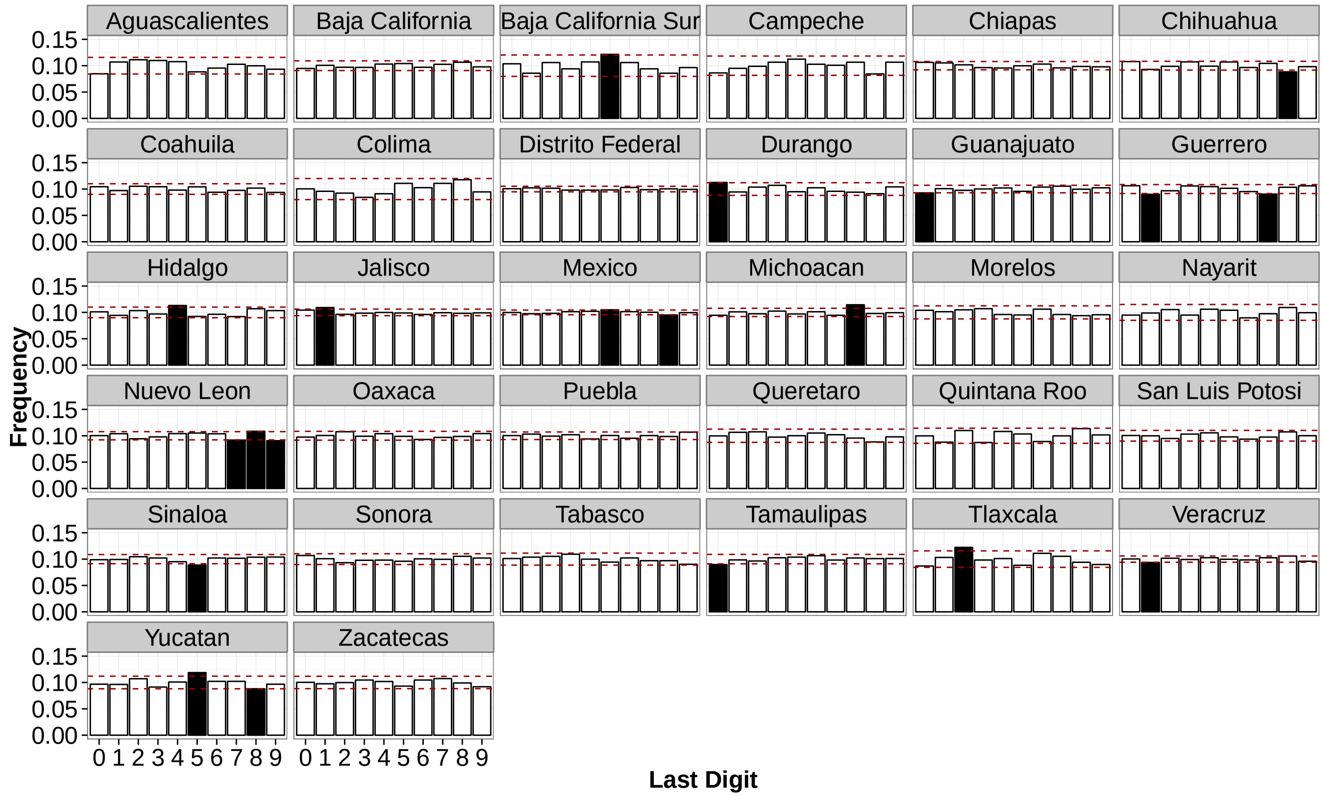


Figure C.1: Last digit distribution for Peña Nieto by state.

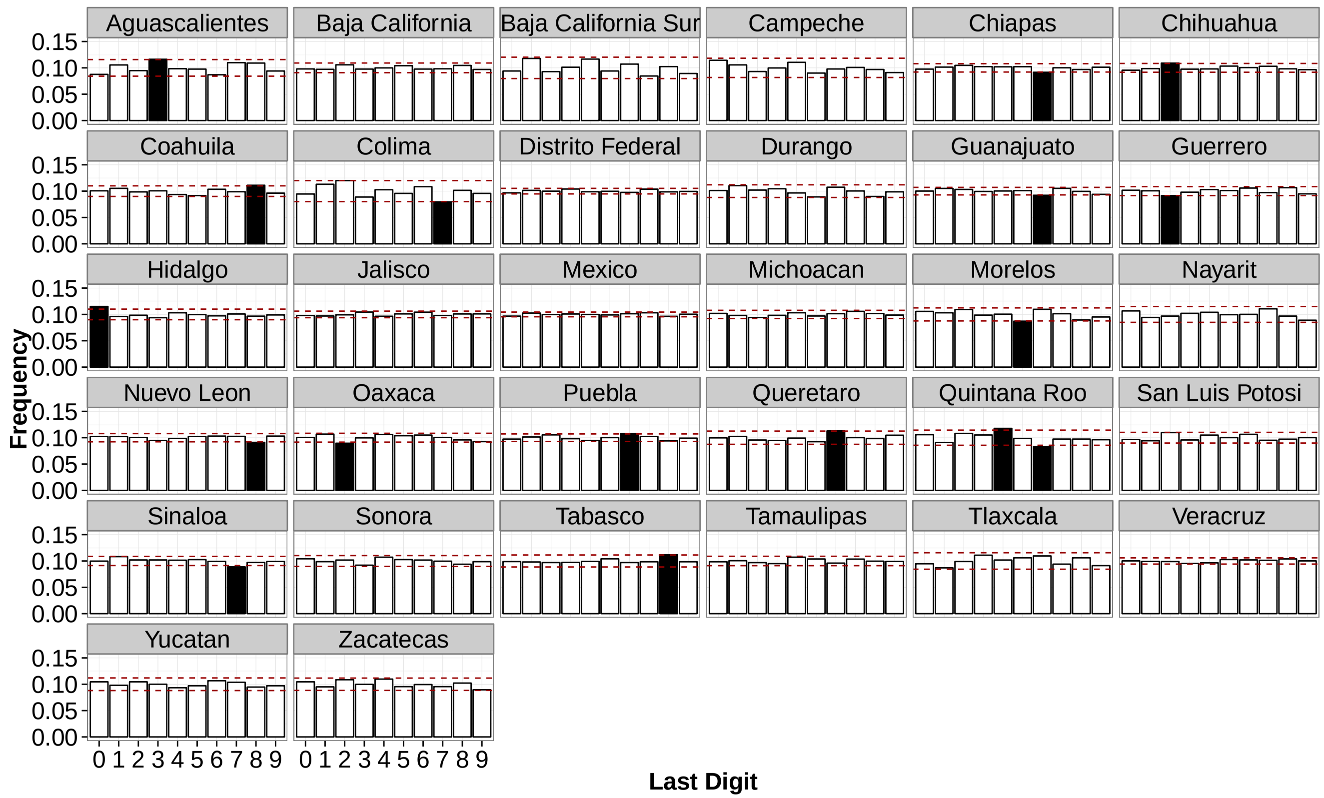


Figure C.2: Last digit distribution for López Obrador by state.

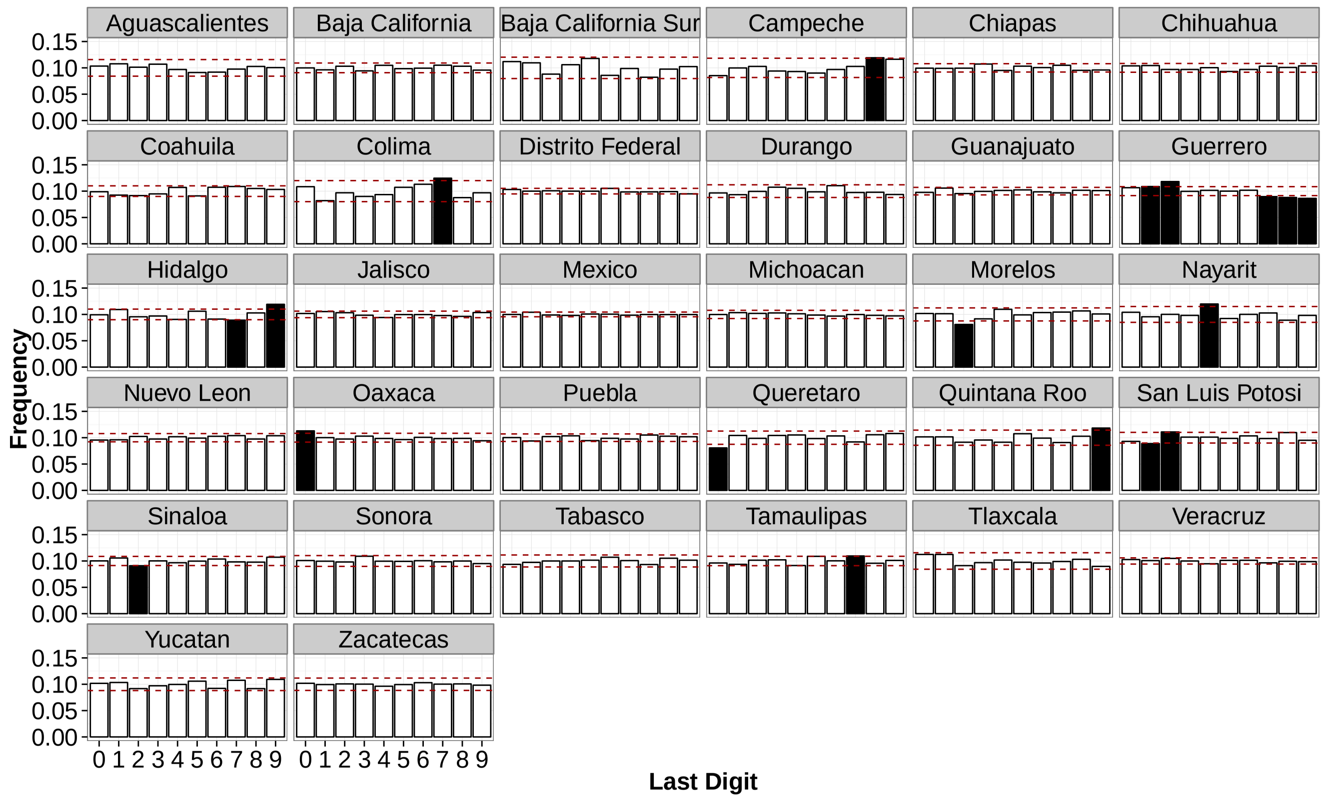


Figure C.3: Last digit distribution for Vazquez Mota by state.

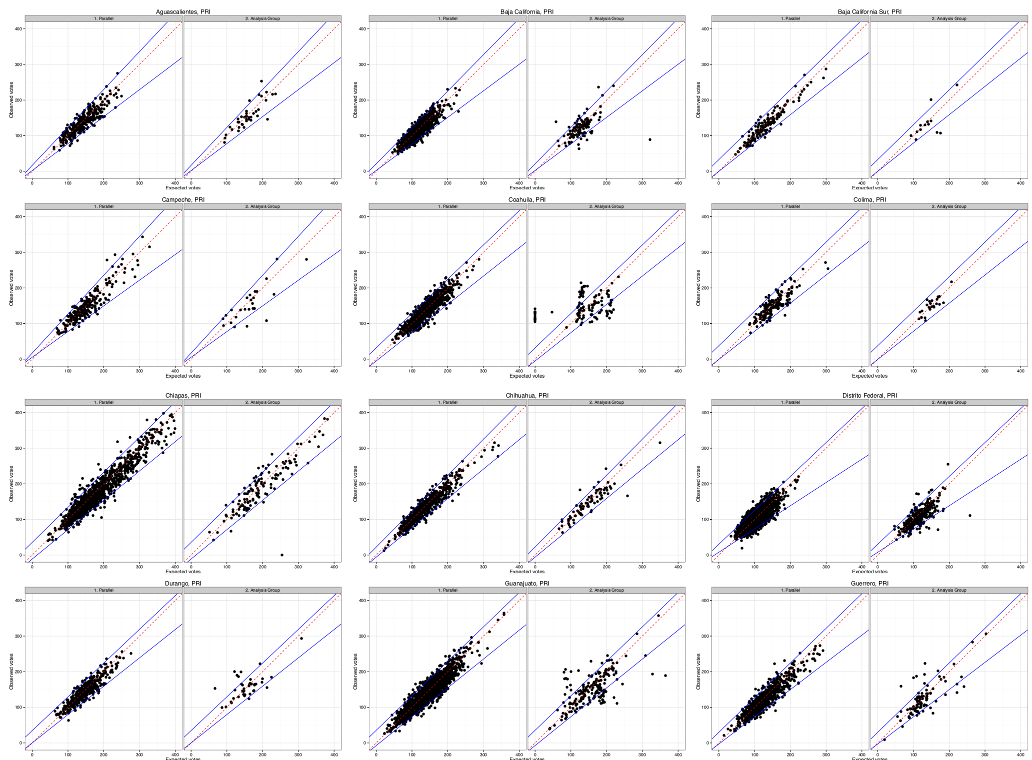


Figure C.4: Outlier analysis for Peña Nieto's votes. Mexico, 2012

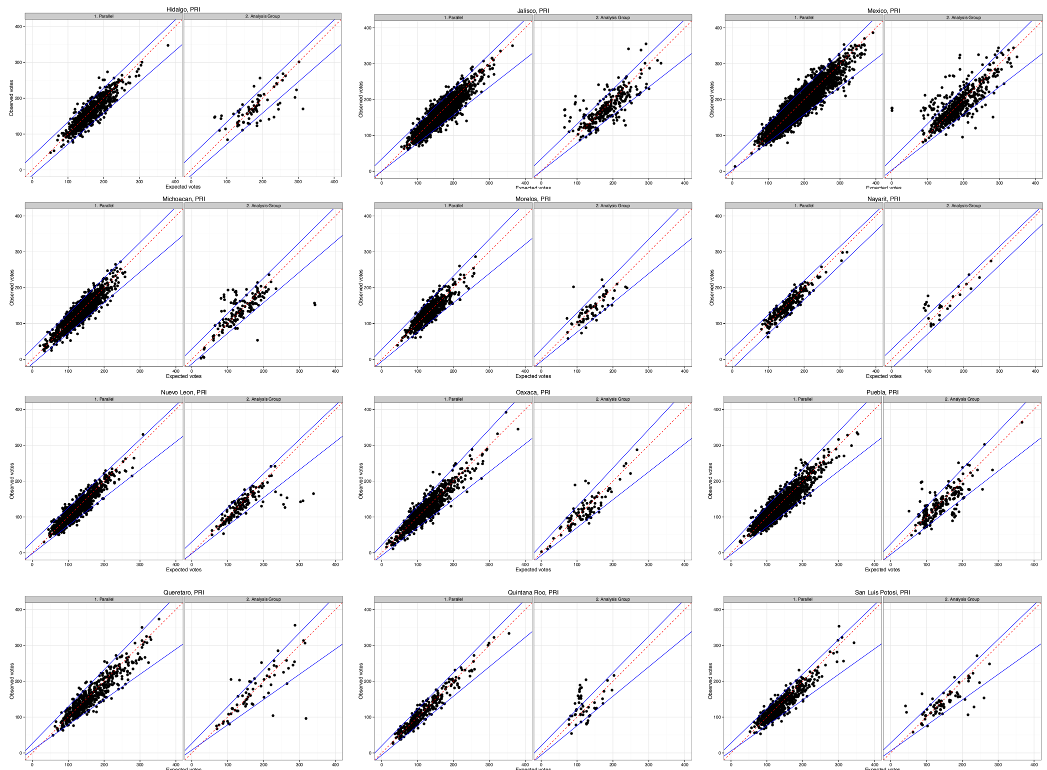


Figure C.5: Outlier analysis for Peña Nieto's votes. Mexico, 2012

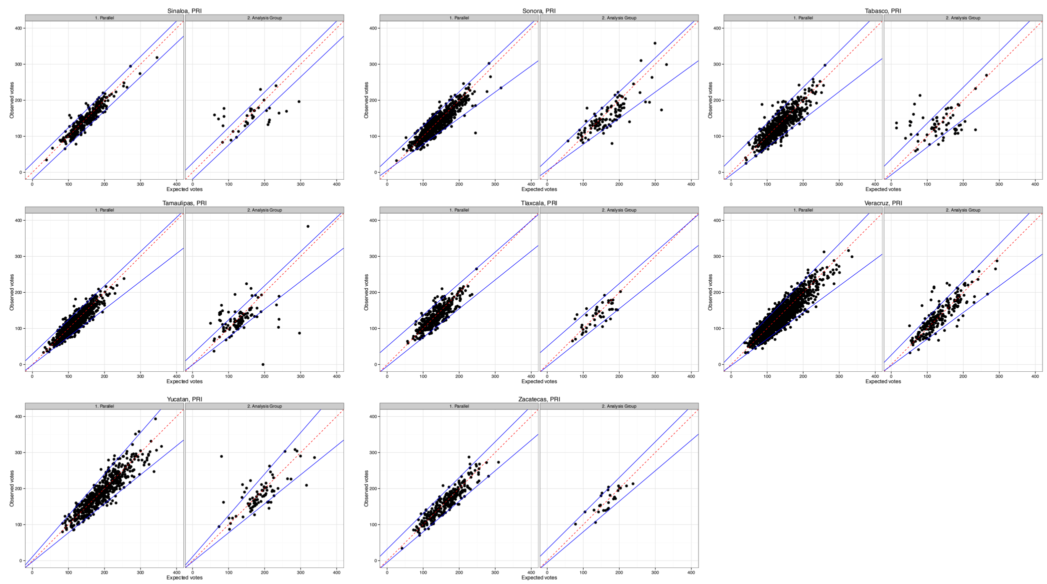


Figure C.6: Outlier analysis for Peña Nieto's votes. Mexico, 2012

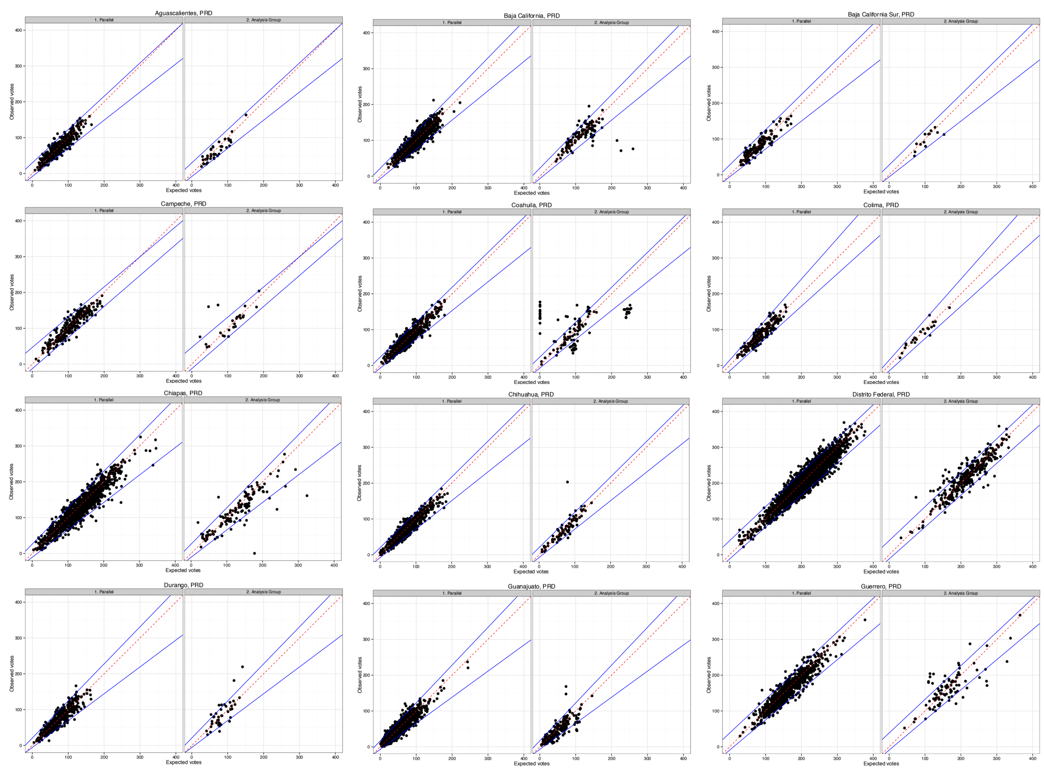


Figure C.7: Outlier analysis for López Obrador's votes. Mexico, 2012.

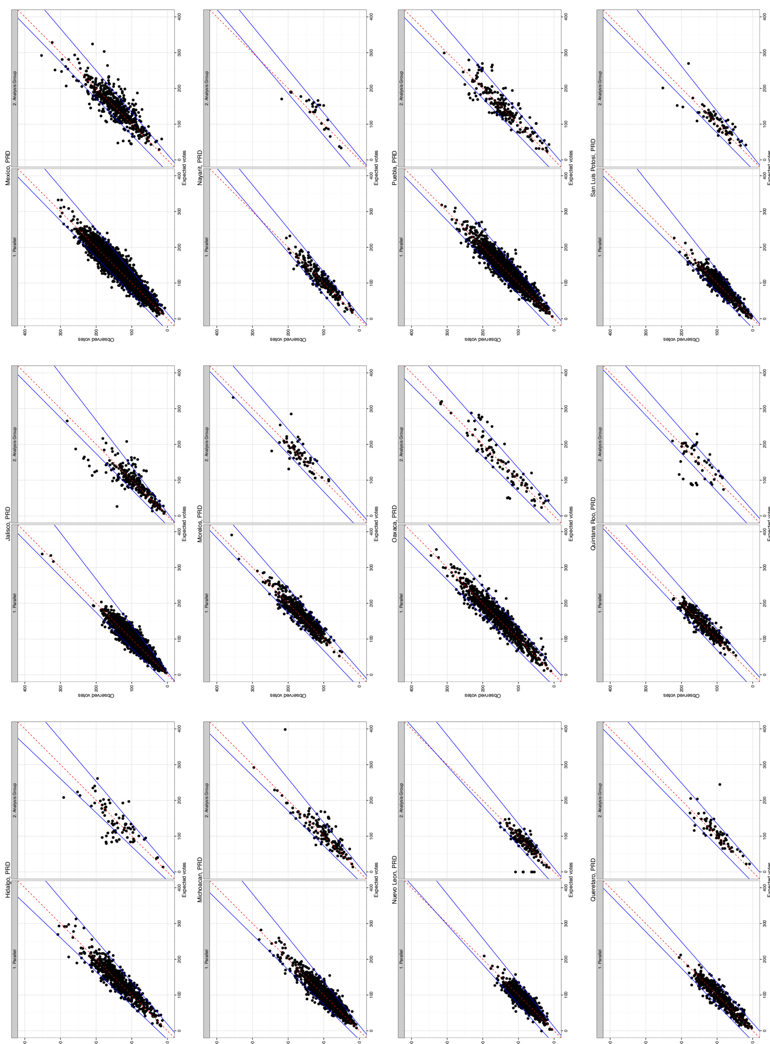


Figure C.8: Outlier analysis for López Obrador's votes. Mexico, 2012.

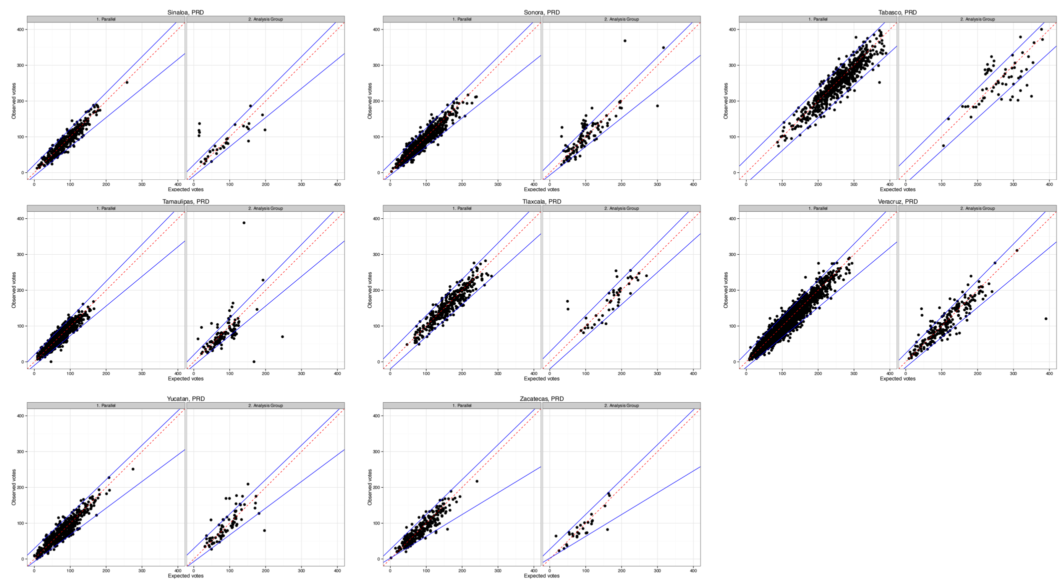


Figure C.9: Outlier analysis for López Obrador's votes. Mexico, 2012.

Table C.3: Regression Results from Turnout versus Votes for each Candidate. In those polling places where the proportion of votes for the *Yes* alternative was above the average in 2009. Venezuela, 2012.

	Chávez		Capriles	
	β_V	R^2	β_V	R^2
Amazonas	0.23	0.07	0.16	0.05
Anzoategui	0.66	0.24	0.31	0.06
Apure	0.65	0.30	0.31	0.10
Aragua	0.43	0.08	0.28	0.03
Barinas	0.78	0.23	0.17	0.01
Bolivar	0.45	0.19	0.37	0.15
Carabobo	0.51	0.15	0.34	0.07
Cojedes	1.15	0.34	-0.10	0.00
Delta Amacuro	0.44	0.21	0.50	0.29
Distrito Capital	1.24	0.35	-0.26	0.02
Falcon	0.67	0.21	0.17	0.02
Guarico	0.47	0.16	0.33	0.09
Lara	0.54	0.14	0.25	0.04
Merida	0.57	0.16	0.40	0.09
Miranda	0.72	0.21	0.16	0.01
Monagas	0.58	0.23	0.24	0.05
Nueva Esparta	0.59	0.29	0.39	0.16
Portuguesa	0.55	0.14	0.14	0.01
Sucre	0.55	0.26	0.38	0.15
Tachira	0.69	0.28	0.30	0.07
Trujillo	0.40	0.08	0.44	0.10
Vargas	1.12	0.32	-0.13	0.01
Yaracuy	0.71	0.19	0.08	0.00
Zulia	0.59	0.30	0.34	0.13

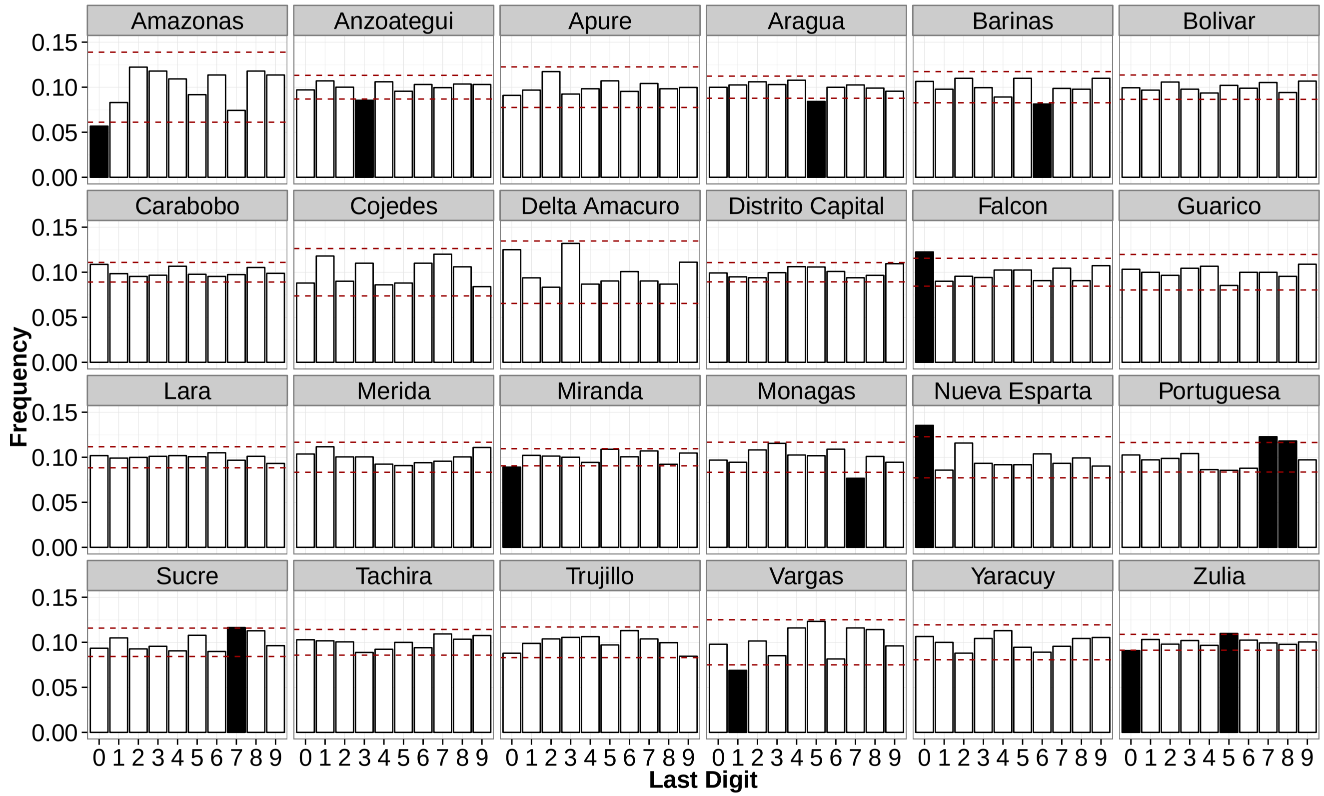


Figure C.10: Last digit distribution for Chávez by state.

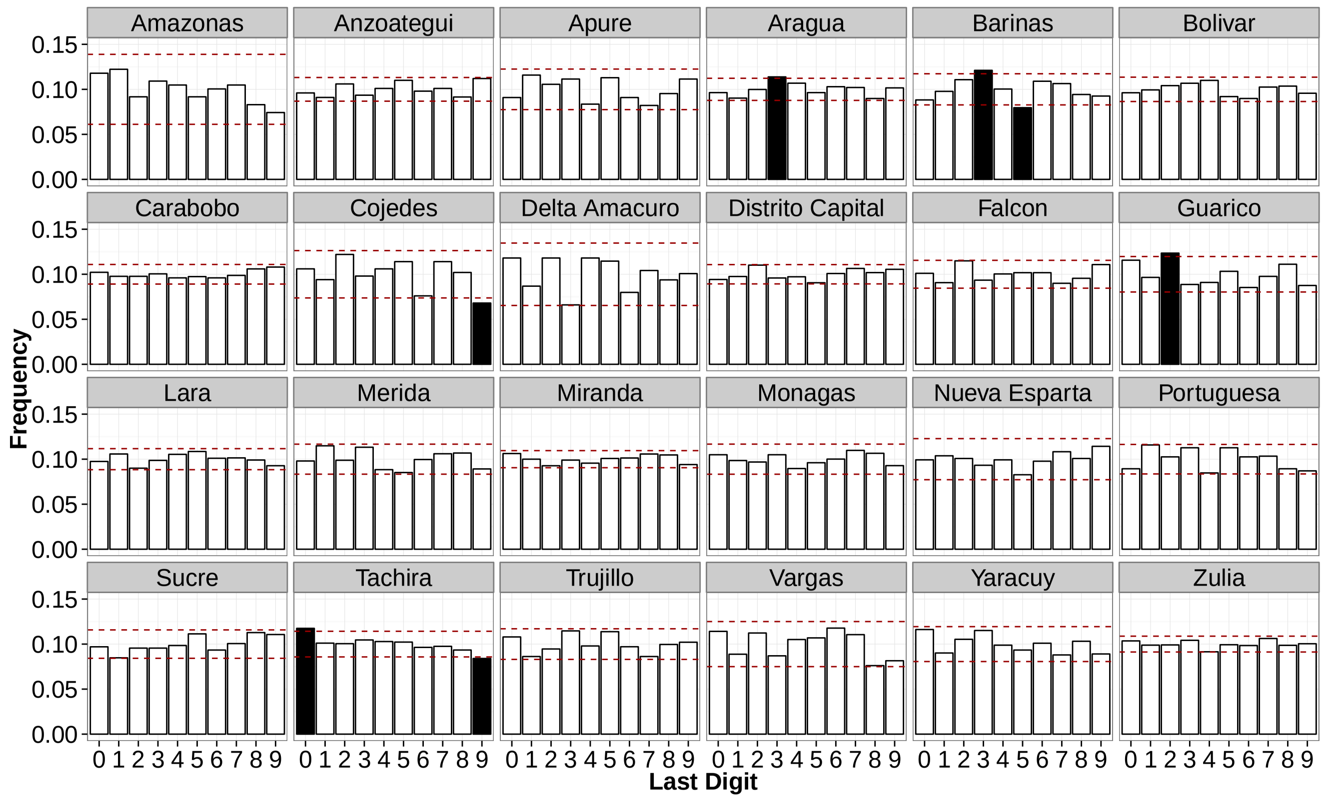


Figure C.11: Last digit distribution for Capriles by state.

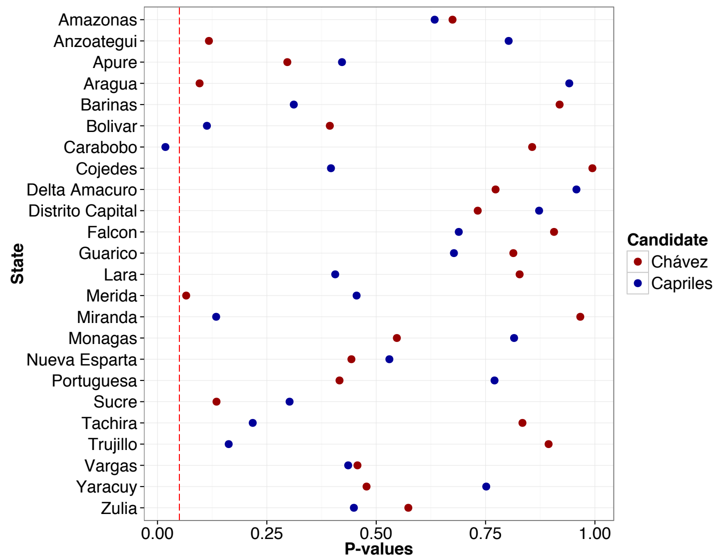


Figure C.12: P-values for chi-squared tests for the last digit distribution by candidate and state. Observations to the left of the red line are for those p-values below 0.05.

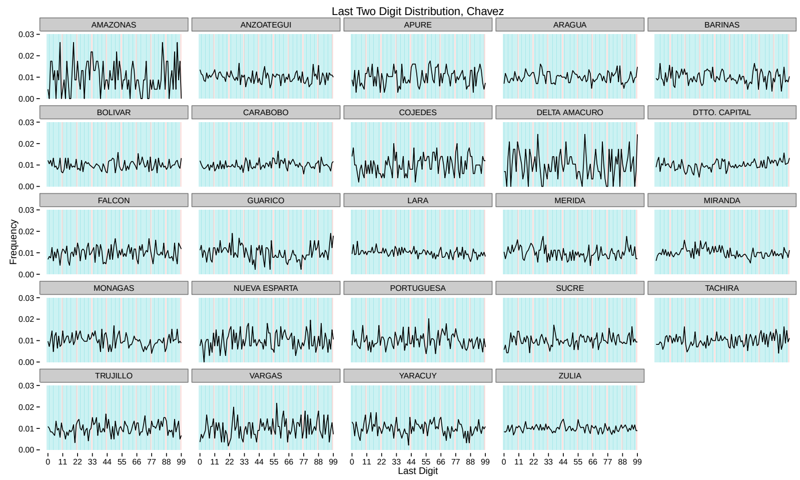


Figure C.13: Last two digits distribution for Chávez by state. Venezuela, 2012.

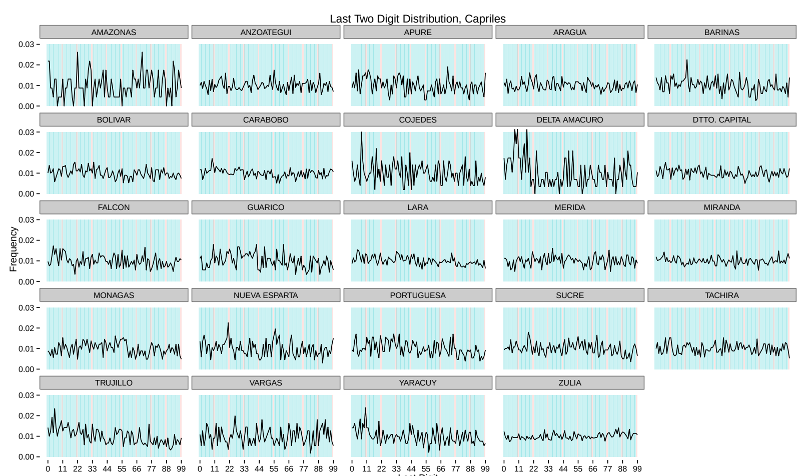


Figure C.14: Last two digits distribution for Capriles by state. Venezuela, 2012.

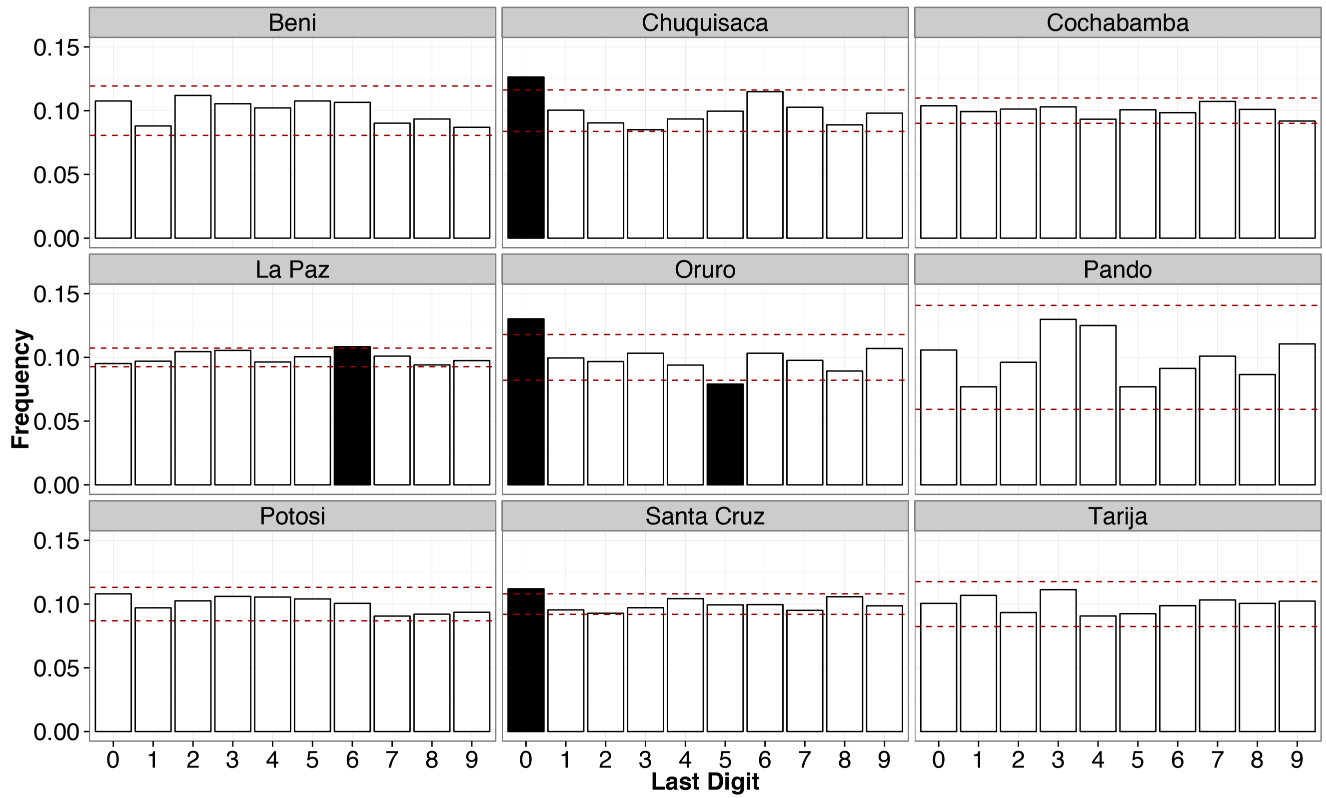


Figure C.15: Last digit distribution for Yes by department.

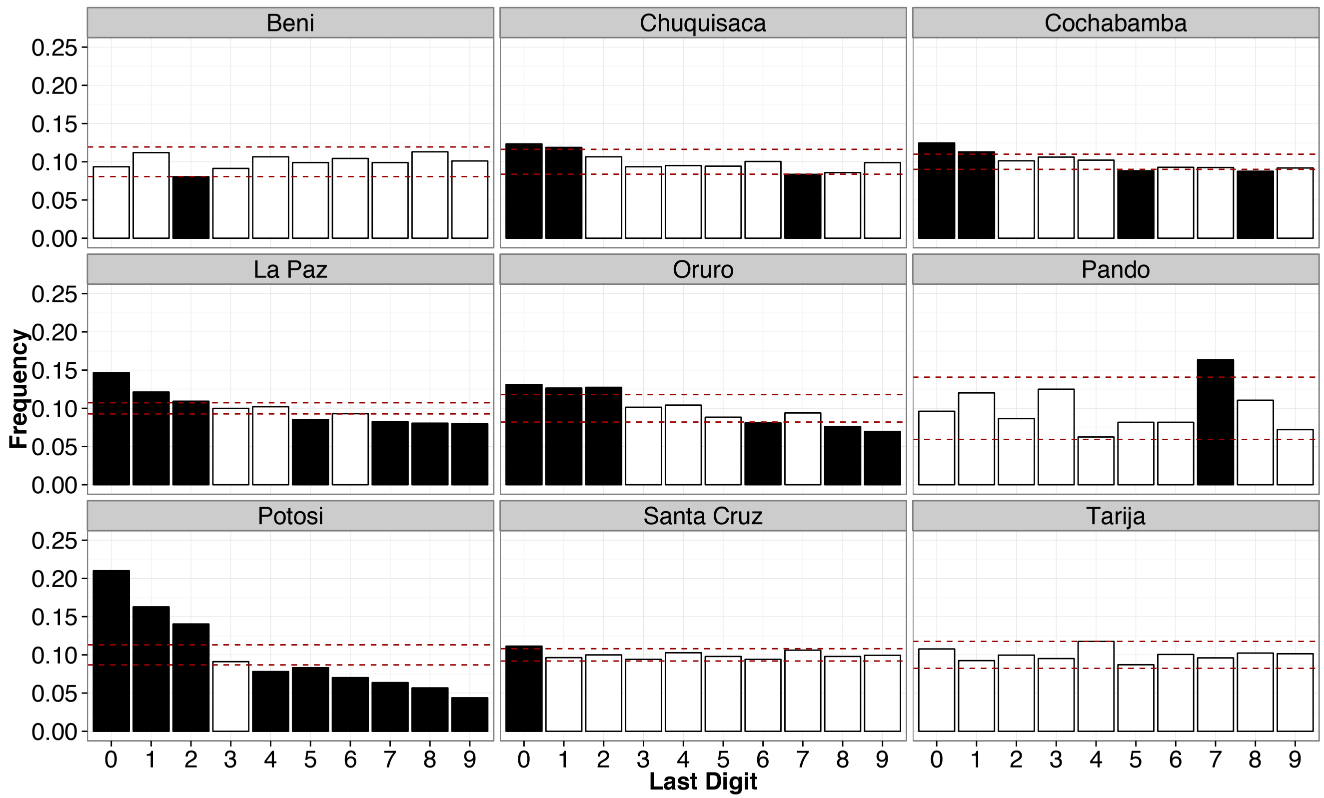


Figure C.16: Last digit distribution for No by department.

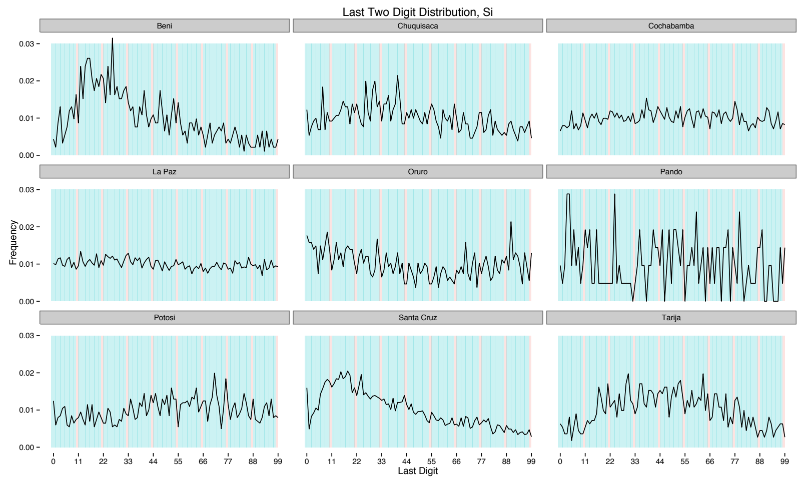


Figure C.17: Last two digits distribution for *Yes* by state. Bolivia, 2008.

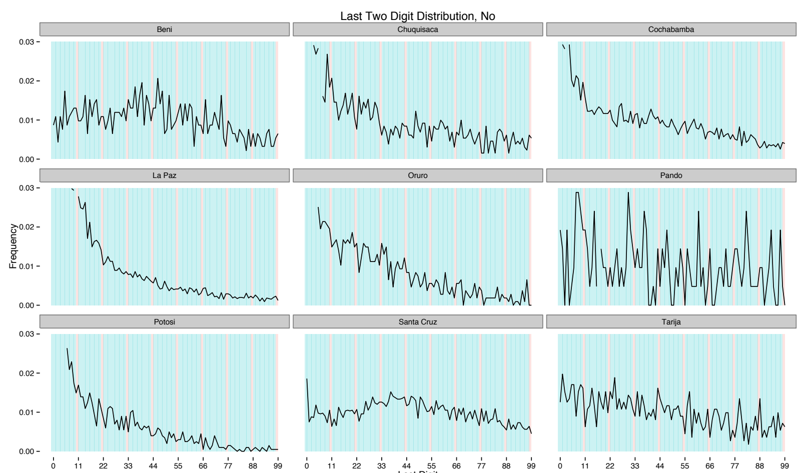


Figure C.18: Last two digits distribution for *No* by state. Bolivia, 2008.

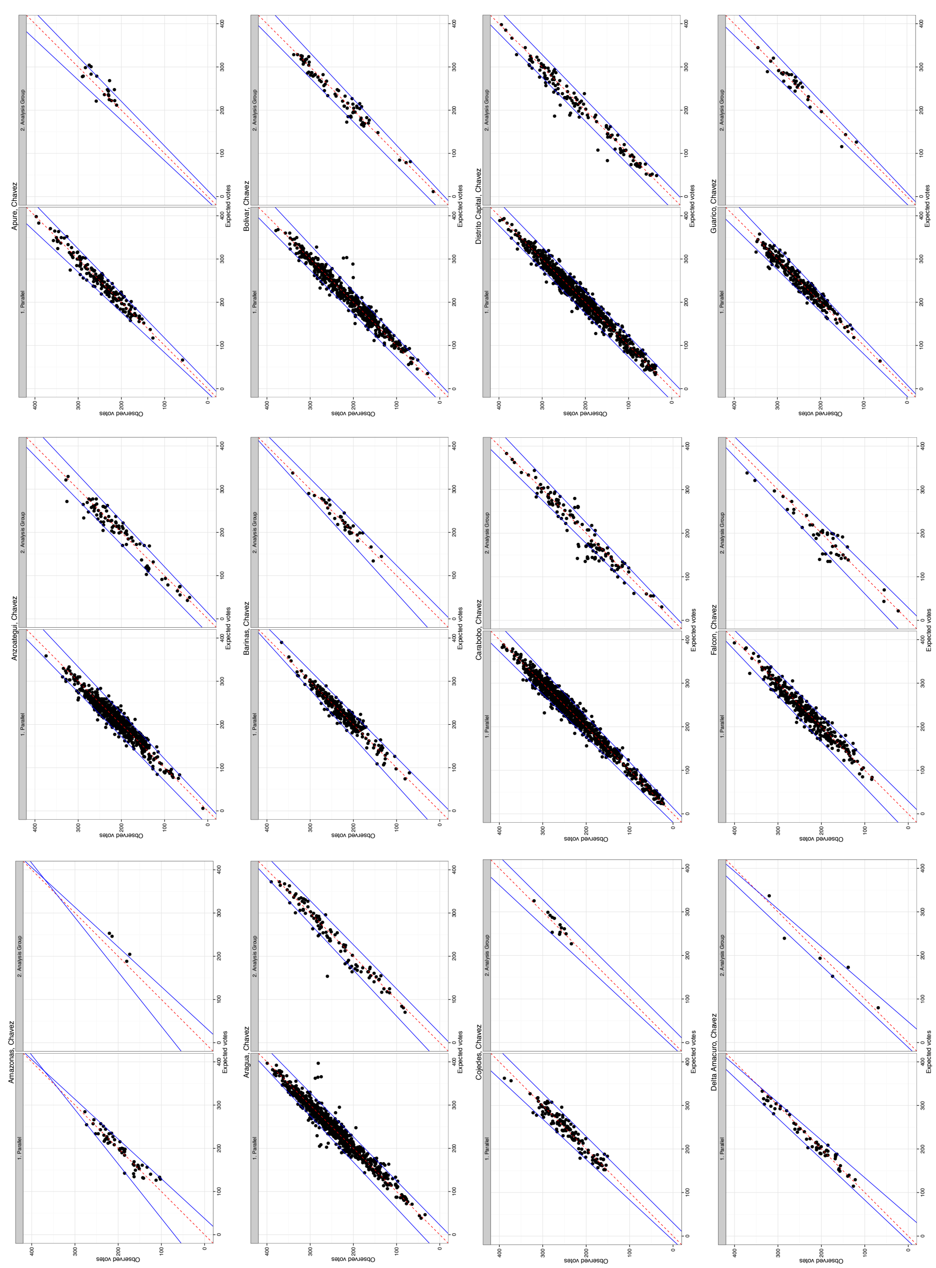


Figure C.19: Outlier analysis for Chávez's votes. Venezuela, 2012.

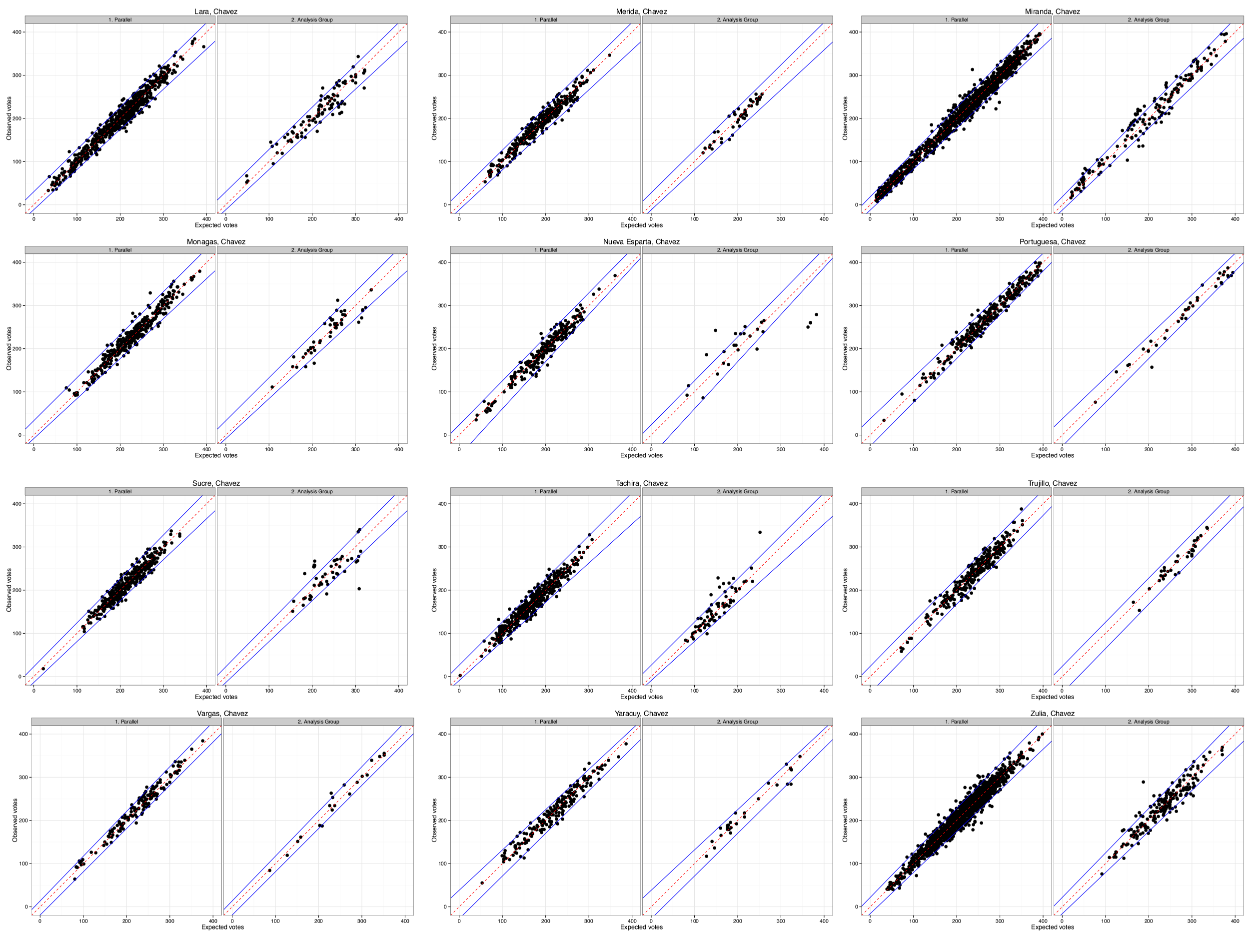


Figure C.20: Outlier analysis for Chávez's votes. Venezuela, 2012.

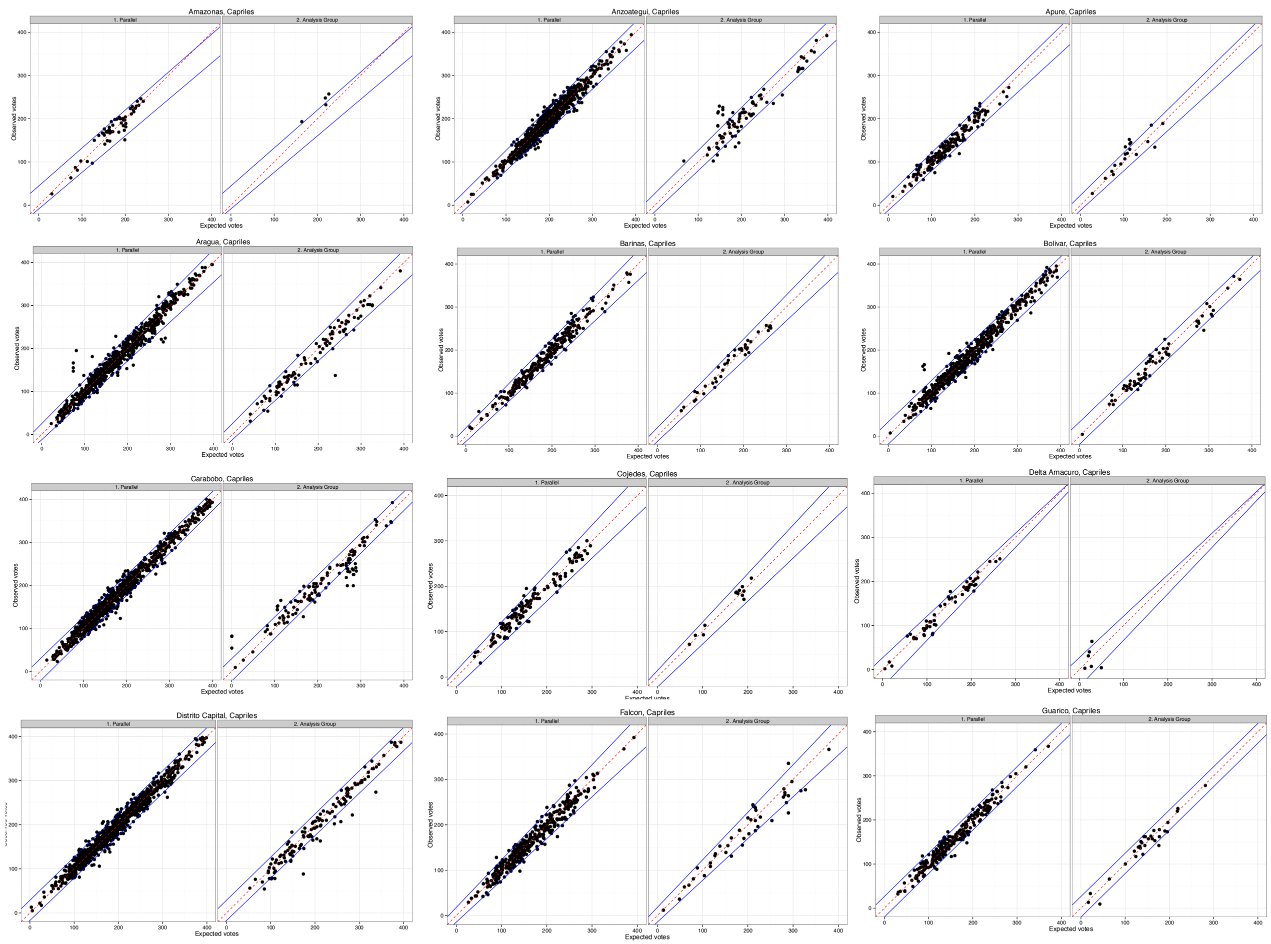


Figure C.21: Outlier analysis for Capriles's votes. Venezuela, 2012.

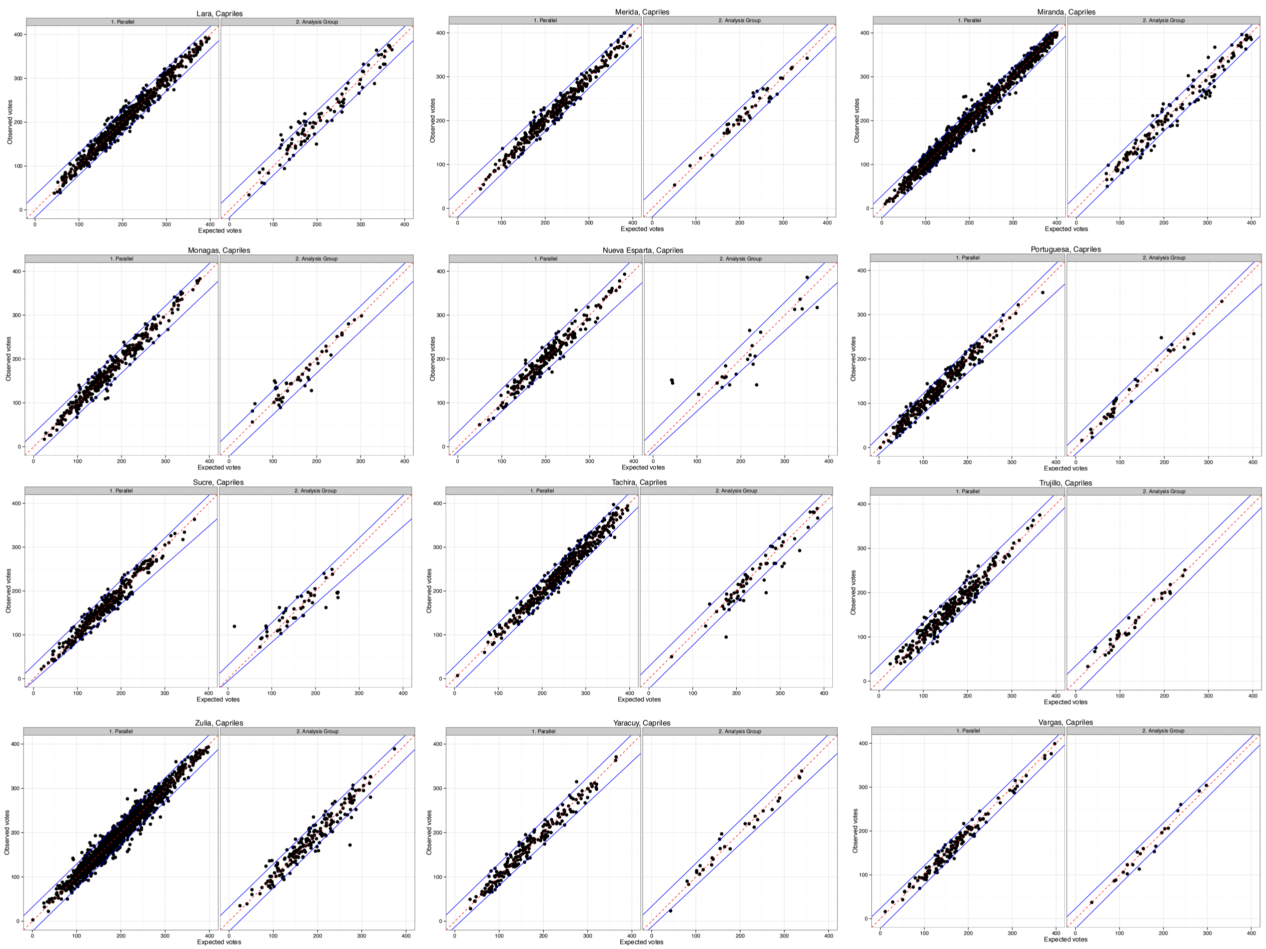


Figure C.22: Outlier analysis for Capriles's votes. Venezuela, 2012.

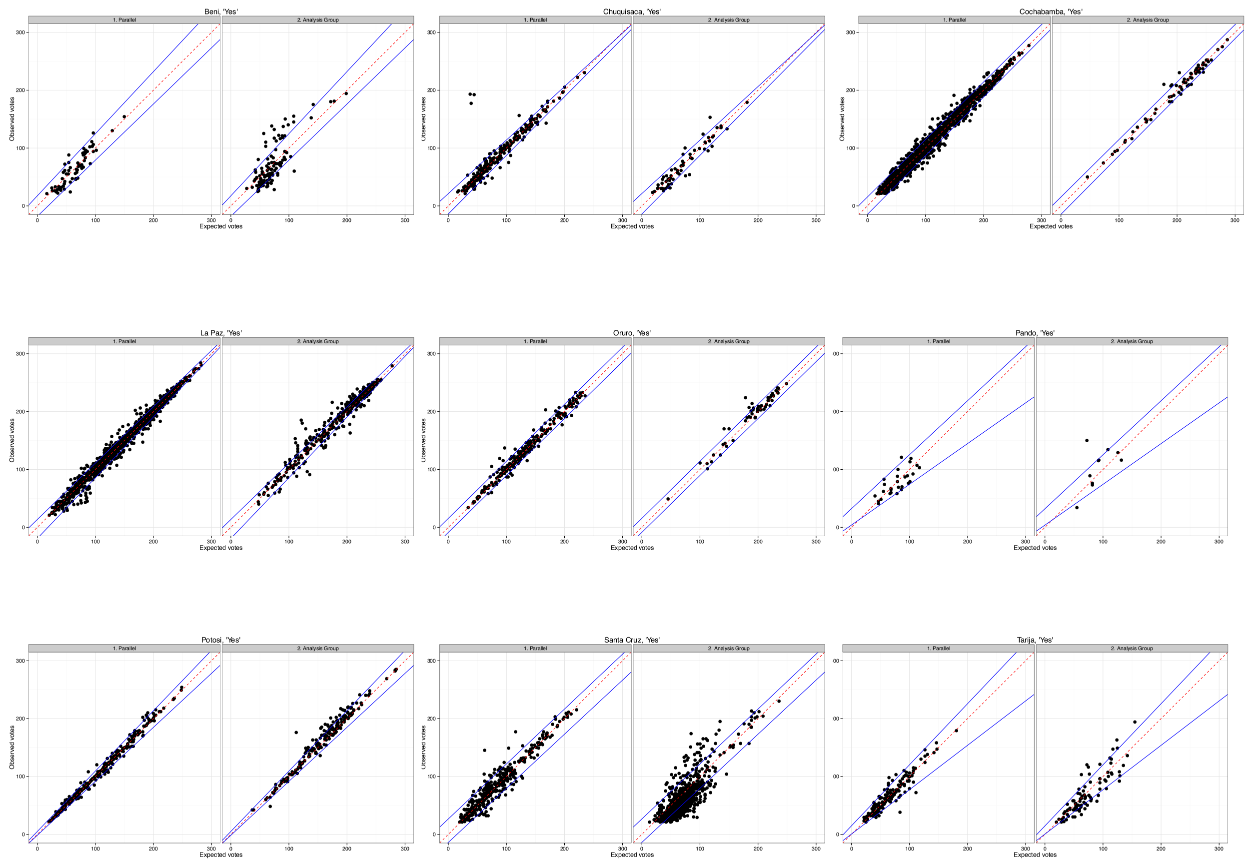


Figure C.23: Outlier analysis for the *Yes* votes. Bolivia, 2008.

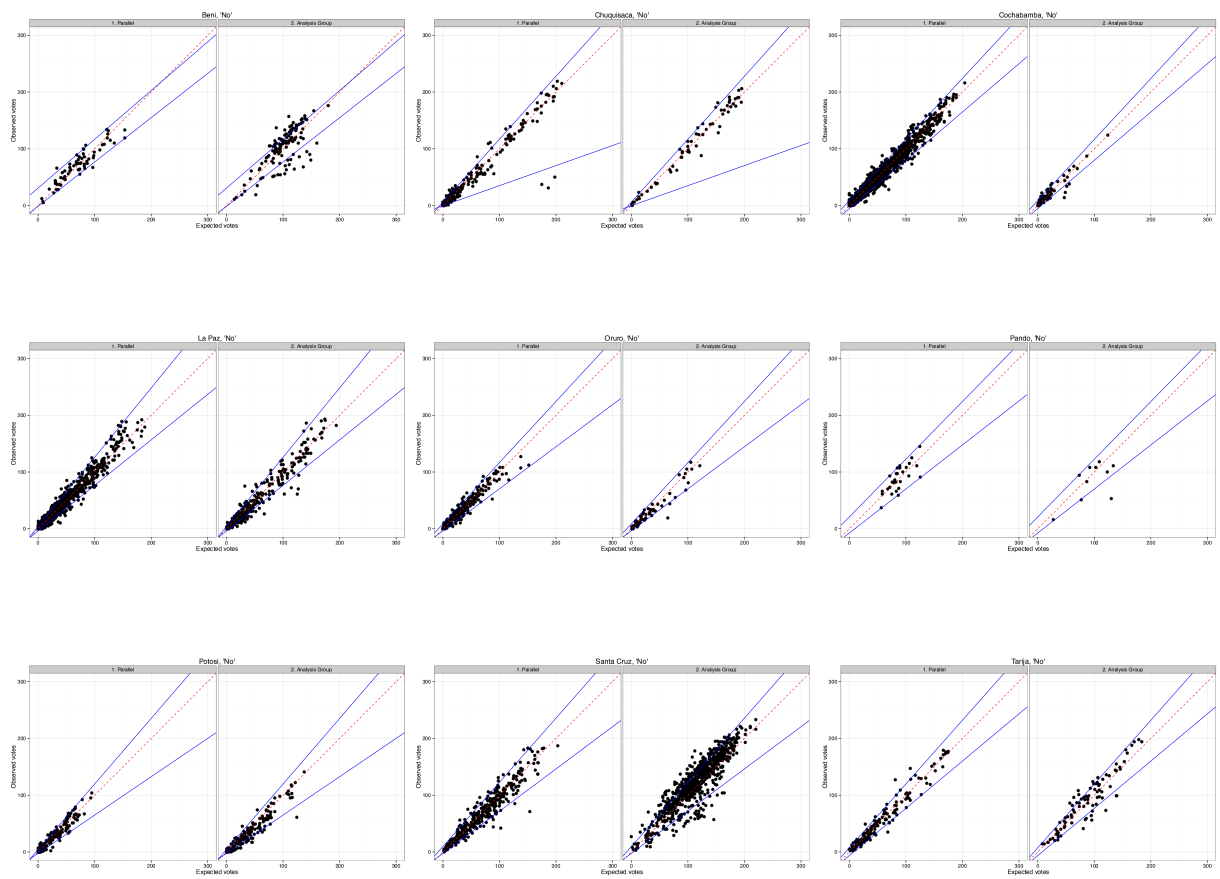


Figure C.24: Outlier analysis for the *No* votes. Bolivia, 2008.

Bibliography

Afshari, Ali and H. Graham Underwood. 2009. "The Green Wave." *Journal of Democracy* 20(4):6–10.

Alabama League of Municipalities. 2011. "Special Report." https://alalm.org/PDF

Alston, Lee J. and Andres A. Gallo. 2010. "Electoral fraud, the rise of Peron and demise of checks and balances in Argentina." *Explorations in Economic History* 47:179–197.

Alvarez, R. Michael and Jonathan N. Katz. 2008. The Case of the 2002 General Election. In *Election Fraud: Detecting and Deterring Electoral Manipulation*, ed. R. Michael Alvarez, Thad E. Hall and Susan D. Hyde. Washington D.C.: The Brookings Institution pp. 149–161.

Alvarez, R. Michael, Lonna Rae Atkenson and Thad E. Hall. 2013. *Evaluating Elections: A Handbook of Methods and Standards*. New York: Cambridge University Press.

Alvarez, R. Michael and Thad E. Hall. 2008. Measuring Perceptions of Election Threats: Survey Data from Voters and Elites. In *Election Fraud*, ed. R. Michael Alvarez, Thad E. Hall and Susan D. Hyde. Washington D.C.: Brooking Institution Press.

Aparicio, Javier. 2006. "Fraud or human error in Mexico's presidential election?" *Voices of Mexico* (77).

Aparicio, Javier. 2009. "Análisis estadístico de la elección presidencial de 2006 ¿fraude o errores aleatorios?" *Política y Gobierno* Volumen Tematico:225–243.

Arriagada, Genaro and José Woldenberg. 2012. "The Elections in Venezuela, October 7, 2012.".

Baland, Jean-Marie and James A. Robinson. 2008. "Land and Power: Theory and Evidence from Chile." *American Economic Review* 98(5):1737–1765.

Beber, Bernd and Alexandra Scacco. 2008. "What the Numbers Say: A Digit-Based Test for Election Fraud Using New Data from Nigeria." Working Paper.

Beber, Bernd and Alexandra Scacco. 2009. "The Devil Is in the Digits." *The Washington Post* .

Beber, Bernd and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud Using New Data from Nigeria." *Political Analysis* 20:211–234.

Benton, Allyson. 2011. "Subnational Regime Dynamics After national Authoritarian Breakdown: When Political Competition and Turnover Don't Indicate Democratization." Working Paper.

Benton, Allyson. 2013. "The Territorial Logic of Electoral Manipulation in an Electoral Authoritarian Regime: The Case of Mexico." Working Paper.

Besley, Timothy. 2006. *Principled Agents?* Cambridge University Press.

Borges, André. 2008. "State Government, Political Competition and Education Reform: Comparative Lessons from Brazil." *Bulletin of Latin American Research* 27(2):235–254.

Breunig, Christian and Achim Goerres. 2011. "Searching for Electoral Irregularities in an Established Democracy: Applying Benford's Law Tests to Bundestag Elections in Unified Germany." *Electoral Studies* 30(3):534–545.

Buttorf, Gail. 2008. "Detecting Fraud in America's Gilded Age." Working Paper.

Callen, Michael and James D. Long. 2012. "Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan." Working Paper.

Calvo, Ernesto and María Victoria Murillo. 2004. "Who Delivers? Partisan Clients in the Argentine Electoral Market." *American Journal of Political Science* 48(4):742–757.

Cantú, Francisco and Sebastián Saiegh. 2011. "Fraudulent Democracy? An Analysis of Argentina's Infamous Decade using Supervised Machine Learning." *Political Analysis* . Forthcoming.

Carothers, Thomas. 2002. "The End of the Transition Paradigm." *Journal of Democracy* 13(1).

Carriquiry, Alicia L. 2011. "Election Forensics and the 2004 Venezuelan Presidential Recall Referendum as a Case Study." *Statistical Science* 26(4):471–478.

Castañeda, Jorge G. 2000. *Perpetuating Power*. New York: The New Press.

Centellas, Miguel and Miguel A. Buitrago. 2009. "The Political Economy of Bolivia's New Regionalism: A Look at Electoral Patterns in Santa Cruz, Tarija, and Chuquisaca." Prepared for delivery at the Annual Meeting of the Midwest Political Science Association, Chicago.

Cho, Wendy K. Tam and Brian J. Gaines. 2007. "Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance." *The American Statistician* 61(3):218–223.

Ciofalo, Michele. 2009. "Entropy, Benford's first digit law, and the distribution of everything." Dipartamento di Ingenieria Nucleare, Universita degli Studi di Palermo, Italy.

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

Código Federal de Institutciones y Preocedimientos Electorales. 2012.

Collier, Paul and Pedro C. Vicente. 2008. "Votes and Violence: Evidence from a Field Experiment in Nigeria." CSAE WPS/2008-16.

Cornelius, Wayne A. 2002. "La eficacia de la compra y coacción del voto en las elecciones mexicanas de 2000." *Perfiles Latinoamericanos* (20):11–32.

Cornes, Richard and Roger Hartley. 2005. "Assymetric contests with general technologies." *Economic Theory* 26.

Corrales, Javier. 2010. The Repeating Revolution: Chávez New Politics and Old Economics. In *Leftists governments in Latin America: successess and shortcomings*, ed. Kurt Gerhard Weyland, Raúl Madrid and Wendy Hunter. New York: Cambridge University Press pp. 28–56.

Cox, Gary W. 2009. "Authoritarian elections and leadership succession, 1975-2000." Working Paper. Department of Political Science. University of California, San Diego.

Cox, Gary W. and Mathew D. McCubbins. 1986. "Electoral Politics as a Redistributive Game." *The Journal of Politics* 48(2):370–389.

Cox, Gary W. and Morgan Kousser. 1981. "Turnout and Rural Corruption: New York as a Test Case." *American Journal of Political Science* 25(4).

Craig, Ann L. and Wayne A. Cornelius. 1995. Houses Divided. Parties and Political Reform in Mexico. In *Building Democratic Institutions: Party Systems in Latin America*, ed. Scott Mainwaring and Timothy Scully. Stanford, California: Stanford University Press.

Crespo, José Antonio. 1996. *Votar en los estados*. Miguel Ángel Porrúa.

Crespo, José Antonio. 2006. *2006: hablan las actas*. Random House Mondadori.

Crespo, José Antonio. February 2, 2008. "El modelo del "fraude hormiga"." *Excelsior* p. 11.

Dasgupta, A. and K.O. Nti. 1998. "Designing an optimal contest." *European Journal of Political Economy* 14(4):587–604.

de la Madrid, Miguel. 2004. *Cambio de Rumbo*. Mexico, D.F.: Fondo de Cultura Económica.

Deckert, Joseph, Mikhail Myagkov and Peter C. Ordeshook. 2011. "Benford's Law and the Detection of Election Fraud." *Political Analysis* 19(3):245–268.

Delfino, Gustavo and Guillermo Salas. 2011. "Analysis of the 2004 Venezuela Referendum: The Official Results Versus the Petition Signatures." *Statistical Science* 26(4):479–501.

Diamond, Larry. 1999. *Developing Democracy: Toward Consolidation*. Baltimore: John Hopkins University Press.

Díaz-Cayeros, Alberto, Beatriz Magaloni and Alexander Ruiz Euler. 2012. "Traditional Governance, Citizen Engagement and Local Public Goods: Evidence from Mexico." Working Paper.

Dixit, Avinash and John Londregan. 1996. "The Determinants of Success of Special Interests in Redistributive Politics." *The Journal of Politics* 58(4):1132–1155.

Domínguez, Jorge I. and James A. McCann. 1996. *Democratizing Mexico*. Baltimore: The John Hopkins University Press.

Donno, Daniela and Nasos Roussias. 2012. "Cheating Pays: The Effect of Electoral Manipulation on Party Systems." *Comparative Political Studies* 45(5).

Drake, Paul W. 2009. *Between tyrany and anarchy: a history of democracy in Latin America, 1800-2006.* Stanford, California: Stanford University Press.

Dresser, Denise. 2003. Mexico: From PRI Predominance to Divided Democracy. In *Constructing democratic governance in Latin America*, ed. Jorge I. Domínguez and Michael Shifter. The John Hopkins University Press.

Eisenstadt, Todd A. 2004. *Courting democracy in Mexico: party strategies and electoral institutions*. Cambridge University Press.

Eisenstadt, Todd A. 2007. "Usos y Costumbres and Postelectoral Conflicts in Oaxaca, Mexico, 1995-2004: An Empirical and Normative Assessment." *Latin American Research Review* 42(1):52–77.

Eisenstadt, Todd A. and Alejandro Poiré. 2006. "Explaining the Credibility Gap in Mexico's 2006 Presidential Election, Despite Strong (Albeit Perfectable) Electoral Institutions." Working Paper.

El País. October 7, 2012. "La Hora de Capriles." p. p. 4.

El Siglo de Durango. July 15, 2010. "Una prueba más del fraude electoral fue encontrada en Gómez Palacio." p. 6.

El Universal. August 13, 2012. ""Poco más de la mitad aprueba elección".".

Electoral Commission of Pakistan. 2002. *Election Laws (Volume 1)*. Pakistan: .

Espíritu, Jorge Orlando. 1984. "Evaluación de las elecciones locales durante 1983." *Nueva antropología* VII(25):100–124.

Fearon, James D. 2011. "Self-Enforcing Democracy." *The Quarterly Journal of Economics* (126):1661–1708.

Fewster, R. M. 2009. "A Simple Explanation of Benford's Law." *The American Statistician* 63(1):26–32.

Figueros, Mark. 1985. "An Assessment of Overvoting in Jamaica." *Social and Economic Studies* 34(3):71–106.

Fukumoro, Kentaro and Yusaku Horiuchi. 2011. "Making Outsiders' Votes Count: Detecting Electoral Fraud through a Natural Experiment." *American Political Science Review* 105(3):583–603.

Gabinete de Comunicación Estratégica. 2010. "Reporte Postelectoral 2010." `http://www.gabinetece.com.mx/images/uploads/documents/documentos/n_estudios/reporte_060810.pdf`.

Gallardo, Jorge Alberto López. 2009. *2006 ¿Fraude Electoral?* Chihuahua: Doble Hélice.

Gallardo, Jorge Alberto López. 2012. *2012 ¿Fraude Electoral?* Universidad de Guadalajara.

Gandhi, Jennifer and Adam Przeworski. 2009. "Holding onto Power by Any Means? The Origins of Competitive Elections." Working Paper.

Gel'Man, Vladimir and Cameron Ross. 2010. *The Politics of Sub-National Authoritarianism in Russia*. Ashgate Publishing.

Gervasoni, Carlos. 2008. "Measuring Variance in Subnational Regimes: Results from an Expert-Based Operationalization of Democracy in the Argentine Provinces." *Journal of Politics in Latin America* 2(2):13–52.

Gibson, Edward. 2005. "Boundary Control: Subnational Authoritarianism in Democratic Countries." *World Politics* (58):101–32.

Gibson, Edward L. 2010. "Politics of the Periphery: An Introduction to Sub-national Authoritarianism and Democratization in Latin America." *Journal of Politics in Latin America* (2):3–12.

Gillingham, Paul. 2012. Mexican Elections, 1910-1994: Voters, Violence, and Veto Power. In *The Oxford Handbook of Mexican Politics*, ed. Roderic A. Camp. Oxford University Press pp. 53–72.

Gonzalez-Ocantos, Ezequiel, Chand Kiewiet de Jonge, Carlos Melendez, Javier Osorio and David W. Nickerson. 2012. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science* 56(1):202–217.

Gray Molina, George. 2010. The Challenge of Progressive Change under Evo Morales. In *Leftists governments in Latin America: successess and shortcomings*, ed. Kurt Gerhard Weyland, Raúl Madrid and Wendy Hunter. New York: Cambridge University Press pp. 57–76.

Harberes, Imke and Matthew Ingram. 2012. "Democratic Institutions Beyond the Nation State: Measuring Institutional Dissimilarity in Federal Countries." Working Paper.

Hartlyn, Jonathan and Jennifer McCoy. 2006. Observer Paradoxes: How to Assess Electoral Manipulation. In *Electoral Authoritarianism*, ed. Andreas Schedler. Lynne Rienner Publishers.

Hausmann, Ricardo and Roberto Rigobón. 2011. "In Search of the Black Swan: Analysis of the Statistical Evidence of Electoral Fraud In Search of the Black Swan: Analysis of the Statistical Evidence of Electoral Fraud in Venezuela." *Statistical Science* 26(4):543–563.

Hidalgo, Fernando Daniel. 2011. "Digital Democratization: Suffrage Expansion and the Decline of Political Machines in Brazil." Working Paper.

Hiskey, Jonathan T. and Shaun Bowler. 2005. "Local Context and Democratization in Mexico." *American Journal of Political Science* 49(1):57–71.

Hunter, Paul R. and Michael A. Gaston. 1988. "Numerical Index of the Discriminatory Ability of Typing Systems: an Application of Simpson's Index of Diversity." *Journal of Clinical Microbiology* 26(11):2465–2466.

Huntington, Samuel and Joan Nelson. 1976. *No Easy Choice: Political Participation in Developing Democracies*. New York: Cambridge University Press.

Hyde, Susan D. 2007. "The Observer Effect in International Politics: Evidence from a Natural Experiment." *World Politics* 50(1).

Hyde, Susan D. 2011. *The Pseudo-Democrat's Dilemma: Why Election Observation Became and International Norm*. Cornell University Press.

Hyde, Susan D. and Nikolay Marinov. 2012. "Which Elections Can Be Lost?" *Political Analysis* 49(4):503–516.

Ichino, Nahomi and Matthias Schundeln. 2012. "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana." *Journal of Politics* 74(1):292–307.

Instituto Federal Electoral. 2010. Análisis de las Actas de Escrutinio y Cómputo de la Elección de Diputados Federales 2009. Technical report.

Janvresse, Élise and Thierry de la Rue. 2004. "From Uniform Distributions to Benford's Law." *Journal of Applied Probability* 41:1203–1210.

Jia, Hao, Stergios Skaperdas and Samarth Vaidya. Forthcoming. "Contest Functions: Theoretical Foundations and Issues in Estimation." *International Journal of Industrial Organization* .

Jiménez, Raúl. 2011. "Forensic Analysis of the Venezuelan Recall Referendum." *Statistical Science* 26(4):564–583.

Jockers, Heinz, Dirk Kohnert and Paul Nugert. 2009. The successful Ghana election of 2008: A convenient myth? Etnicity in Ghana's elections revisited. GIGA working paper 109 German Institute of Global and Area Studies.

Kalinin, Kirill and Walter R. Mebane. 2011. "Understanding electoral frauds through evolution of Russian federalism: From 'Bargaining Loyalty' to 'Signaling Loyalty.'." Paper prepared for the 2011 Annual Meeting of the Midwest Political Science Association, Chicago, IL, March 31-April 2.

Kelley, Judith. 2012. *Monitoring Democracy*. Princeton University Press.

King, Ronald F. 2001. "Couting the Votes: South Carolina's Stolen Election of 1876." *Journal of Interdisciplinary History* 32(2):169–191.

Klimek, Peter, Yuri Yegorov, Rudolf Hanel and Stefan Thurner. 2012. Statistical detection of systematic election irregularities. In *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 10.1073.

Koenker, Roger and Kevin F. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives* 15(4):143–156.

Konrad, Kai A. 2009. *Strategy and Dynamics in Contests*. Oxford University Press.

Kozumi, Hideo and Genya Kobayashi. 2009. Gibbs Sampling Methods for Bayesian Quantile Regression. Technical report Graduate School of Business Administration, Kobe University.

Kuo, Didi, Jan Teorell and Daniel Ziblatt. 2011. "Election Fraud and Contested Congressional Elections: An Analysis of the United States, 1840-1940." Paper presented at the Annual Meeting of the American Political Science Association.

La Jornada. August 16, 2010. "Ratifican triunfo del PRI en Durango." p. 31.

La Razón. January 6, 2009*a*. "El padrón se disparó en las regiones cocaleras del país.".

La Razón. January 7, 2009*b*. "Insólito, en agosto hubo 289 mesas con 100% de votantes.".

Langston, Joy. 2012. *Party Change. Mexico's PRI, from Hegemony to Democracy, 1980 to 2006*. chapter 5. Book manuscript.

Larreguy, Horacio. 2012. "Monitoring Political Brokers: Evidence from Clientelistic Networks in Mexico." Working Paper.

Lawson, Chappell. 2007. "How Did We Get Here? Mexican Democracy after the 2006 Elections." *Political Science and Politics* .

Leemis, Lawrence M., Bruce Schmeiser and Diane L. Evans. 2000. "Survival Distributions Satisfying Benford's Law." *The American Statistician* 54(4):236–241.

Lehoucq, Fabrice. 2003. "Electoral Fraud: Causes, Types, and Consequences." *Annual Review of Political Science* 6:233–256.

Lehoucq, Fabrice. 2010. Bolivia: From Neoliberal Democracy to Plebiscitarian, Multi-Ethnic Politics. In *Latin American Politics and Development*, ed. Howard Wiarda and Harvey Kline. 7th ed. Boulder, Colorado: Westview Press pp. 324–362.

Lehoucq, Fabrice E. 2002. "Can Parties Police Themselves? Electoral Governance and Democratization." *International Political Science Review / Revue internationale de science politique* 23(1):29–46.

Lehoucq, Fabrice and Iván Molina. 2002. *Stuffing the Ballot Box*. Cambridge University Press.

Levin, Inés, Gabe A. Cohn, Peter C. Ordeshook and R. Michael Alvarez. 2009. Detecting Voter Fraud in an Electronic Voting Context: An Analysis of the Unlimited Reelection Vote in Venezuela. In *Proceedings of the 2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE '09)*.

Levitsky, Steven and James Loxton. 2012. "Populism and Competitive Authoritarianism in the Andes." Working Paper.

Levitsky, Steven and Lucan A. Way. 2010. *Competitive Authoritarianism*. New York: Cambridge University Press.

Levy, Santiago. 2006. *Progress Against Poverty*. Washington D.C.: The Brookings Institution.

Linz, Juan J. and Alfred Stepan. 1996. *Problems of Democratic Transition and Consolidation*. Baltimore: The John Hopkins University Press.

Little, Andrew T. 2011. "Rational Expectations and Electoral Fraud." Presented at the 2011 Midwest Science Association Meeting.

Lohr, Sharon L. 1999. *Sampling: Design and Analysis*. Boston, MA: Brooks/Cole.

Long, James D., Karuti Kanyinga and David Ndii. 2010. Was it Rigged? A Forensic Analysis of Vote Returns in Kenya's 2007 Election. In *Democratic Gains and Gaps: A Study of the 2007 Kenyan General Elections*, ed. Society for International Development. Nairobi, Kenya: Society for International Development.

Los Angeles Times. December 27, 2002. "Allegations of Fraud Precede Kenyan Vote.".

Los Angeles Times. February 01, 2010. "Mexico's PAN and PRD. Love letters in the making?".
URL:
http://articles.latimes.com/2010/feb/01/world/la-fg-mexico-alliance2-2010feb02

Madrid, Raúl. 2011. Bolivia: Origins and Policies of the Movimiento al Socialismo. In *The Resurgence of the Latin American Left*, ed. Steven Levitsky and Kenneth M. Roberts. Baltimore: The John Hopkins University Press pp. 239–259.

Magaloni, Beatriz. 2006. *Voting for Autocracy: Hegemonic Party Survival and Its Demise in Mexico*. Cambridge University Press.

Magaloni, Beatriz. 2007. "The Game of Electoral Fraud and the Ousting of Authoritarian Rule." *American Journal of Political Science* 54:751–765. Working Paper.

Mares, Isabela and Boliang Zhu. 2012. "Economic Concentration, Rural Inequality and Electoral Fraud: Re-Examining the Case of Imperial Germany." Working Paper.

Martin, Denis. 1978. The 1975 Tanzanian Elections: the Disturbing Six Per Cent. In *Elections without choice*, ed. Hermer Guy and Richard Rose. The Macmillan Press LTD.

Martín, Isbelia. 2011. "2004 Venezuelan Presidential Recall Referendum (2004 PRR): A Statistical Analysis from the Point of View of Electronic Voting Data Transmissions." *Statistical Science* 26(4):528–542.

Mayorga, René Antonio. 2009. Sociedad civil y Estado bajo un populismo plebscitario y autoritario. In *La "Nueva Izquierda" en América Latina.*, ed. Cynthia J. Arnson, Ariel C. Armony, Catalina Smulovitz, Gastón Chillier, Enrique Peruzzotti and Giselle Cohen. Washington DC: Woodrow Wilson Press Center pp. 106–114.

McElduff, Fiona, Pablo Mateos, Angie Wade and Mario Cortina. 2008. "What's in a name? The frequency of geographic distributions of UK surnames." *Significance* 5(4):189–192.

Mebane, Walter R. 2006. "Election Foresincs: Vote Counts and Benford's Law." 2006 Summer Meeting of the Political Methodology Society, UC-Davis.

Mebane, Walter R. 2007. "Statistics for Digits." 2007 Summer Meeting of the Political Methodology Society.

Mebane, Walter R. 2008*a*. "Election Foresincs: Outlier and Digit Tests in America and Russia." Working Paper.

Mebane, Walter R. 2008*b*. Elections Foresincs: The Second-Digit Bendford Law's Test and Recent American Presidential Elections. In *Election Fraud*, ed. R. Michael Alvarez, Thad E. Hall and Susan D. Hyde. The Brookings Institutions.

Mebane, Walter R. 2009. "Note on the presidential election in Iran, June 2009." Unpublished.

Mebane, Walter R. 2010. "Election Fraud or Strategic Voting? Can Second-digit Test Tell the Difference?" Prepared for presentation at the 2010 Summer Meeting of the Political Methodology Society, University of Iowa.

Mebane, Walter R. and Jasjeet S. Sekhon. 2004. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science* 48(2):392–411.

Mebane, Walter R. and Kirill Kalinin. 2009. "Comparative Election Fraud Detection." Prepared for presentation at the 2009 Annual Meeting of the American Political Science Association, Toronto, Canada.

Mendizabal, Yuritzi and Alejandro Moreno. 2010. La confianza electoral: el IFE y los partidos políticos. In *La confianza en las instituciones. México en perspectiva comparada*, ed. Alejandro Moreno. Mexico City: Centro de Estudios Sociales y de Opinión Pública pp. 227–248.

Mochán, Luis. 2011. "Elecciones presidenciales, México 2006.".
URL: *http://em.fis.unam.mx/ mochan/elecciones/*

Mochán, Luis. 2012. "Elecciones presidenciales, México 2012.".
URL: *http://em.fis.unam.mx/elecciones2012/*

Molinar Horcasitas, Juan. 1987. "Regreso a Chihuahua." *Nexos* 10(111):21–32.

Molinar, Juan. 1991. *El Tiempo de la Legitimidad*. Mexico, D.F.: Cal y Arena.

Moreno, Alejandro. 2012. Who Is the Mexican Voter? In *The Oxford Handbook of Mexican Politics*, ed. Roderic A. Camp. Oxford University Press pp. 571–595.

Muñoz-Pogossian, Betilde. 2008. *Electoral rules and the transformation of Bolivian politics: the rise of Evo Morales*. New York: Palgrave Macmillan.

Myagkov, Mikhail and Peter C. Ordeshook. 2010. "Ukraine 2010: Were Tymoshenko's Cries of Fraud Anything More than Smoke?" Caltech/MIT Voting Technology Project. Working Paper No. 10.

Myagkov, Mikhail, Peter C. Ordeshook and Dimitri Shakin. 2005. "Fraud or Fairytales: Russia and Ukraine's Electoral Experience." *Post-Soviet Affairs* 21(2):91–131.

Myagkov, Mikhail, Peter C. Ordeshook and Dimitri Shakin. 2009. *The foresincs of election fraud: Russia and Ukraine*. Cambridge University Press.

Nadeau, Richard and André Blais. 1993. "Accepting the Election Outcome: The Effect of Participation on Losers' Consent." *British Journal of Political Science* 23(4):553–563.

Navia, Patricio. 2000. "Fair Play at Voting Precincts: A Comparison of Mexican and Chilean Elections." *Democratization* 7.1:183–189.

Noticias24. 2012. "En video: "Aquí no hubo fraude, lo digo clarito", aseguró Capriles ante rumores.".
URL: *http://www.noticias24.com/venezuela/noticia/130475/en-video-aqui-no-hubo-fraude-lo-digo-clarito-asegura-capriles-ante-rumores/*

Nti, Kofi O. 1999. "Rent-seeking with asymmetric valuations." *Public Choice* (98):415–430.

Ochoa-Reza, Enrique. 2004. Multiple Arenas of Struggle: Federalism and Mexico's Transition to Democracy. In *Federalism and Democracy in Latin America*, ed. Edward L. Gibson. John Hopkins University Press.

O'Donnell, Guillermo. 1998. "Polyarchies and the (Un)Rule of Law in Latin America." Paper presented at the Meeting of the Latin American Studies Association, Chicago.

O'Donnell, Guillermo. 1999. *Counterpoint: Selected Essays on Authoritarianism and Democratization*. Notre Dame, Indiana: University of Notre Dame Press.

Organization of American States. August 11, 2008. "Comunicado de Prensa Final.".

Pastor, Robert. 2006. "A Full Recount to Achieve Closure: Comment of the Eisenstadt and Poire Working Paper." Unpublished manuscript.

Pe Lero, Cecilia. 2010. "Election Forensics - The Effect of Socioeconomic Characteristics on Voting Behavior in the Philippines." Institute for Popular Democracy Report.

Peñarada, Raúl and Oswaldo Candia. 2010. *Padrón Biométrico, una proeza de los Bolivianos.* La Paz: Fundación Boliviana para la Capacitación Democrática y la Investigación.

Pericchi, Luis R. and David Torres. 2011. "Quick Anomaly Detection by the Newcom-Benford Law, with Applications to Electoral Process Data from the USA, Puerto Rico and Venezuela." *Statistical Science* 23(4):502–515.

Persson, Torsten and Guido Tabellini. 2000. *Political Economics*. The MIT Press.

Peschard, Jacqueline. 2010. Federal and Local Electoral Institutions: From a National to a Fragmented System. In *Mexico's Democratic Challenges: Politics, Government, and Society*, ed. Andrew Selee and Jacqueline Peschard. Stanford, California: Woodrow Wilson Press Center chapter 4.

Poiré, Alejandro and Luis Estrada. 2006. "Allegations of Fraud in Mexico's 2006 Presidential Election." Paper presented at the 102th meeting of the American Political Science Association.

Powell, Lawrence N. 1989. "Correcting for Fraud: A Quantitative Reassessment of the Mississippi Ratification Election of 1868." *The Journal of Southern History* 55(4):633–658.
URL: *http://www.jstor.org/stable/2209043*

Prado, Raquel and Bruno Sansó. 2011. "The 2004 Venezuelan Presidential Recall Referendum: Discrepancies Between Two Exit Polls and Official Results." *Statistical Science* 26(4):517–526.

Prats, Juan José Rodríguez. August 8, 2010. "El Trife no puede convalidar Durango, Hidalgo y Veracruz: correligionarios o cómplices?" *Siempre!* 57(2982):38.

Preston, Julia and Samuel Dillon. 2004. *Opening Mexico: The Making of a Democracy*. New York: Farrar Straus and Giroux.

Przeworski, Adam. 1991. *Democracy and the Market*. Cambridge University Press.

Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub and Fernando Limongi. 2000. *Democracy and Development*. New York: Cambridge University Press.

Raphael, Ricardo. 2007. *Los Socios de Elba Esther*. Mexico City: Planeta.

Rebolledo, Juan. 2012. Voting with the Enemy: A Theory of Democratic Support for Subnational Authoritarians. PhD thesis Yale University.

Romero Ballivián, Salvador. 2003. *Geografía Electoral de Bolivia*. La Paz: Fundación Boliviana para la Capacitación Democrática y la Investigación.

Roukema, Boudewijn F. 2013. "A first-digit anomaly in the 2009 Iranian presidential elections." *Journal of Applied Statistics* . Forthcoming.

Schedler, Andreas. 2002*a*. "Citizens resent clientelism: on the moral economy of vote trading in Mexico." Presented at the 2002 Annual Meeting of the American Political Science Association. Boston.

Schedler, Andreas. 2002*b*. "The Menu of Manipulation." *Journal of Democracy* 13(2):36–50.

Schedler, Andreas. 2002*c*. "The Nested Game of Democratization by Elections." *International Political Science Review* 23(1):103–122.

Schedler, Andreas. 2006. The logic of Electoral Authoritarianism. In *Electoral Authoritarianism*, ed. Andreas Schedler. London: Lynne Rienner Publishers.

Sidel, John. 2004. Bossism and democracy in the Philippines, Thailand, and Indonesia: towards an alternative framework for the study of 'local strongmen'. In *Policising democracy: the new local politics of democratisation.*, ed. John Harriss, Kristin Stokke and Olle Tornsquist. Basingstone, UK: Palgrave Macmillan pp. 51–74.

Simpser, Alberto. 2005. Making votes not count: Strtegic incentives for electoral corruption. PhD thesis Stanford University.

Simpser, Alberto. 2012*a*. "Does Electoral Manipulation Discourage Voter Turnout? Evidence from Mexico." *The Journal of Politics* 74(3):782–795.

Simpser, Alberto. 2012*b*. *Why Governments and Parties Manipulate Elections: Theory, Practice, and Implications*. Cambridge University Press.

Simpser, Alberto. 2013. *Why Governments and Parties Manipulate Elections: Theory, Practice, and Implications*. Cambridge University Press.

Skidmore, Thomas E., Peter H. Smith and James N. Green. 2010. *Modern Latin America*. New York: Oxford University Press.

Smith, Peter H. 2011. *Democracy in Latin America*. 2 ed. Oxford University Press.

Snyder, James M. 1989. "Election Goals and the Allocation of Campaign Resources." *Econometrica* 57(3):637–660.

Snyder, James M. 1990. "Resource Allocation in Multiparty Elections." *American Journal of Political Science* 34(1):59–73.

Sobyanin, Alexander, E. Gelman and O. Kaiunov. 1994. "The Political Climate of Russia's Regions: Voters and Deputies, 1991-3." *The Soviet and Post-Soviet Review* 21:63–84.

Stokes, Susan. 2005. "Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina." *American Political Science Review* 99(03):315–325.

Szidarovszky, Ferenc and Koji Okuguchi. 1997. "On the Existence and Uniqueness of Pure Nash Equilibrium in Rent-Seeking Games." *Games and Economic Behavior* 18(1):135–149.

The Carter Center. October 2012. "Study Mission to the October 7, 2012, Presidential Election in Venezuela.".

The Miami Herald. September 11, 2012. "Capriles closing in on Chávez in Venezuela presidential race Capriles closing in on Chavez in Venezuela presidential race.".

The New York Times. July 4, 2010. "Once-Dominant Party in Mexico Seems Resurgent." p. p. A7.

The New York Times. July 6, 2012. "Mexican Election Officials Confirm Presidential Victory." p. p. A5.

Tofias, Michael. 2007. "Of Colonels and Generals: Understanding Asymmetry." Prepared for presentation at the 2007 Midwest Political Science Annual Meeting.

Tucker, Joshua. 2007. "Enough! Electoral Fraud, Collective Action Problems, and Post-Communist Colored Revolutions." *Perspectives on Politics* 5(3).

Tullock, Gordon. 1980. Efficient rent seeking. In *Toward a theory of the rent-seeking society*, ed. James M. Buchanan, Robert D. Tollison and Gordon Tullock. Texas AM University Press.

Vicente, Pedro C. and Leonard Wantchekon. 2009. "Clientelism and vote buying: lessons from field experiments in African elections." *Oxford Review of Economic Policy* 25(2):292–305.

Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr, Michael Herron and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Journal of Political Science* 95(4).

Worden, Scott. 2010. "Afghanistan: An Election Gone Awry." *Journal of Democracy* 21:11–25.

Ziblatt, Daniel. 2009. "Shaping Democratic Practice and the Causes of Electoral Fraud: The Case of Nineteenth-Century Germany." *American Political Science Review* 103(1):1–21.