# Gone or Hidden? Identifying Electoral Irregularities in Mexican Local Elections

Francisco Cantú

**Abstract**

This paper argues that, despite efforts to dismantle the structural artifacts of electoral fraud after the fall of the Institutionalized Revolutionary Party (PRI) in the 2000 elections, the liberalization of the political system in Mexico has not eliminated political machines' fraudulent behavior. Instead, such practices have been concealed from the public eye. To detect these hidden electoral irregularities, I exploit a feature of the Mexican electoral system: 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. 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. I illustrate this methodology by evaluating the 2010 gubernatorial elections held in twelve states. My findings validate the view that electoral fraud has not disappeared in Mexico but has merely changed in nature.

# 1 Introduction

The defeat of the Institutionalized Revolutionary Party (PRI, according to its Spanish initials) in the 2000 presidential election marked a watershed in Mexican politics. After the PRI left power and lost control over the organization and certification of elections, expectations arose that electoral fraud would become an obsolete practice. Were those expectations fulfilled?

In this paper, I argue that fraud has not disappeared from Mexico, having changed in nature instead. Under the rule of the PRI, fraud was a highly visible practice performed in a centralized manner (Simpser, 2008; Magaloni, 2006). In contrast, contemporary fraud is an invisible practice performed in a decentralized manner: it is often carried out by local brokers, who tamper with the electoral process at the polling station in exchange for political favors from the local candidates.

Given the hidden nature of these fraudulent practices, locating substantive evidence of their existence is problematic. To overcome this challenge, I propose an empirical way to detect irregularities at the polling-station level. My identification strategy exploits a feature of the Mexican 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 use suspicious differences in turnout levels across contiguous polling stations to identify fraudulent practices. I illustrate this methodology by evaluating the 2010 gubernatorial elections held in twelve states.

The study's findings support the view that fraud has not disappeared. In particular, candidates in the states of Durango, Oaxaca, and Veracruz benefitted from electoral irregularities during the 2010 gubernatorial elections. In fact, some of the polling stations that

are flagged as suspicious according to the method proposed in this paper were notorious for violent disruptions or the use of duplicated paper ballots. The findings are further validated with the help of other sources, such as reports of electoral irregularities and post-electoral surveys.

This paper makes two contributions. First, the findings contribute empirical evidence to recent literature focusing on the challenges of democratization at the federal level (Gibson, 2008; Gervasoni, 2008; Rebolledo, 2010), and confirm the existence of regions in Mexico that actively obstruct the completion of the country's democratic transition (Cornelius, 2000). The second contribution is empirical: the proposed approach contributes to a growing literature that uses statistical tools to evaluate the quality of the elections (Myagkov, Ordeshook and Shakin, 2009; Mebane, 2006; Beber and Scacco, 2008; Levin et al., 2009). My method, however, identifies a specific mechanism in which fraud is observed, avoiding the "lack of theory" problem identified by Deckert, Myagkov and Ordeshook (2011).

This paper is organized as follows. Section 2 describes the opportunities for fraud in Mexico resulting from two different institutional contexts. Section 3 describes the foundations of the empirical model and explains the methodology I propose for measuring electoral fraud. Section 4 discusses the results and their validity. Finally, Section 5 presents the conclusions and recommendations for future research.

# 2 New and old fraud

## 2.1 Cheating under a hegemonic party regime

Elections in Mexico, similar to other countries in Latin America, have been held regularly throughout its national history.[1] However, these elections were rarely legitimate processes for selecting public officials (Drake, 2009). With the emergence of the PRI in 1929, elections became a ceremonial process of power rotation and distribution among different elites.[2] Fraud was a selective practice at the beginning of the PRI's one-party regime, but it became a common activity as the political competition intensified in the 1980s.[3] Corrupt officials not only abused the direct connection between the government and the election organizers, but also took advantage of a weak opposition that lacked the means to monitor or prevent irregularities.

Under the party's hegemonic rule, electoral fraud was a visible, centralized, and monopolized activity. As Magaloni (2006) and Simpser (2005) argue, the reasons that the PRI committed fraud extended beyond electoral victory. Perpetrators made their illegal activities visible in order to discourage potential competitors from participating in future elections. In addition, electoral manipulation was ultimately validated, if not performed, by the Ministry of Interior or the President. Such was the case in the 1988 presidential election, when the early uncertainty regarding the winning candidate prompted President de la Madrid to manipulate the official vote returns (de la Madrid, 2004, p. 814-

[1] In 1847, Mexico, El Salvador, and Greece were the only three countries in the world with a broad male suffrage (Przeworski, 2010, p.20).

[2] President Plutarco Elías Calles founded the National Revolutionary Party (PNR, according to its Spanish initials) in 1929 as an institution to mediate power conflicts among local bosses and mass organizations. By 1938, President Lázaro Cárdenas modified the internal party structure to a corporatist scheme and renamed it the Party of the Mexican Revolution (PRM, according to its Spanish initials). Finally, President Miguel Alemán reorganized and gave it its present name in 1946 (Craig and Cornelius, 1995, p. 252-253).

[3] See Lujambio (2001) for a detailed description of how the Partido Acción Nacional (PAN, according to its Spanish initials) was affected by fraud throughout the party's history of opposition, beginning in the 1940s.

824). The significant role of the government in perpetrating fraud reduced elections to a mere instrument for opposition parties and the federal government to trade political positions (Eisenstadt, 2003). This occurrence was clearly illustrated in the outcome of protests against the evident fraudulent activities in the 1987 gubernatorial elections in Chihuahua. After the protests, bargaining between civil society, the opposition, and the PRI was held in the offices of the Minister of Internal Affairs in Mexico City (Preston and Dillon, 2004, p. 138-147). Finally, as mentioned above, the ruling party monopolized opportunities for cheating during elections. Due to the inequality of resources, the weakness of the judicial system, and the intervention of the federal government, fraud was commonly and exclusively committed by and for the PRI.

For nearly 40 years, the PRI ruled through noncompetitive elections while providing incentives to the opposition parties to legitimize its power (Haber et al., 2008, p.125). The first step in the institutional development of the electoral system began after 1976, when there was only one single candidate for the presidential elections. The urge for legitimization through multiparty elections was the impulse for a series of marginal electoral reforms.[4] Electoral reforms peaked when the Federal Electoral Institute (IFE), the agency created to organize the elections, received full autonomy in 1996, causing the government to lose its *de jure* ability to overtly control the electoral process. With this institutional change, political parties became institutionally unable to carry out electoral malpractice (Magaloni, 2006, p.36-38).

## 2.2 Cheating under a competitive system

As has been the case in countries such as Costa Rica, Chile, and Uruguay, once parties were denied control over organizing and certifying election results, the threat of blatant

[4]For a detailed description of each electoral reform see Ochoa-Reza (2004) and Craig and Cornelius (1995).

manipulation significantly decreased (Lehoucq, 2002). Since the mid-1990s, elections in Mexico have been impartially organized, and the process has been opened to all participants. Moreover, the IFE is the only agency in the country with the authority to issue voter identification cards, which are required for participation in an elections at any level. Because this institution maintains the number of registered voters according to geographical location, it is also responsible for establishing the electoral precincts (or *sección electoral*), in which citizens are assigned to vote. Furthermore, the Supreme Court holds full constitutional control over electoral laws, being allowed to manipulate state electoral codes to prevent unconstitutional activities (Lujambio, 2000). With the creation of institutions to monitor and penalize electoral misconducts, the likelihood of fraudulent activities has been significantly decreased. The electoral process can be monitored by not only the political parties but also any citizen, reducing the opportunities for fraud.[5]

However, the quality of local elections varies by state.[6] The Constitution provides the authorities of each sub-national unit with the right to organize the elections for governors, local legislators, and municipal councils, although they need to fulfill minimum legitimacy (Mex. Const., Art CXVI §4).[7] States exercise their sovereignty to freely interpret the Constitution. While some states have changed their electoral codes in response to the increasing competitiveness and political pressure of the opposition, incumbents in

[5]However, post-electoral protests have been used a political tool for losers even when evidence of electoral irregularities does not exist (Eisenstadt and Poiré, 2006; Poiré and Estrada, 2006; Eisenstadt, 2004). The challenge then is to distinguish electoral irregularities from the losing contender's false protests.

[6]Mexico is a federal system composed of thirty-one states and a Federal District. Governors serve a tenure of six years, while local legislators, mayors, and municipal councils are elected every three years. Consecutive re-election is not allowed for any of these positions. When local elections occur in the same year as federal elections, both are held on the first Sunday of July. Otherwise, each state is free to determine its own election day. In 2010, parties agreed to hold the same election day in the twelve states where a new governor was to be elected.

[7]Peschard (2010, p. 74) summarizes the content of the eight principles as follows: "(1) universal suffrage and free, secret, and direct vote, (2) an autonomous administrative authority, (3) equitable public funding for the parties, (4) limits to private funding and campaign expenses, (5) equitable access of parties to media, (6) party accountability, including control over party income and expenses, (7) regulated electoral justice, and (8) a typology of electoral infractions and their corresponding penalties."

other states have modified rules to prevent electoral defeat (Peschard, 2010).

Consequently, the change in the political system has not eliminated the occurrence of electoral irregularities. Evidence of vote buying is common in the modern literature on Mexican politics (Cornelius, 2002; Aparicio, 2002; Gibson, 2005), and the state governors' interference in local electoral institutions was revealed during the 2010 elections.[8]

Under the new institutional framework, and compared to previous forms of electoral manipulation, contemporary fraud has three characteristics:

1. Current electoral manipulation has a unique instrumental incentive: winning elections. Due to the presence of partisan, independent, and international observers in the polling stations and the creation of judicial bodies capable of ruling on electoral issues, electoral misconduct is committed only when the opportunity arises.

2. Current electoral irregularities are decentralized. The relative distance of political parties from the electoral administration reduces the possibility of manipulating vote returns in a centralized way; therefore, misconduct is perpetrated by party brokers at the polling-station level.

3. Electoral manipulation is now a plausible activity for any political party. Traditionally, fraud was associated with the incumbent candidate and the ruling party, -i.e., the PRI. Currently, however, any political machine is able to commit fraud during elections as major irregularities can be made at the polling-station level.

In brief, election interference has developed from a centralized and monopolized practice by the ruling party to a decentralized and open practice by any party.

[8] See, for example, http://www.eluniversal.com.mx/notas/689196.html and http://www.milenio.com/cdb/doc/noticias2011/7c954b2a681dd8c89db2c26caf4d5a34 .

## 2.3 The micro-foundations of fraud

Considering the current institutional conditions, fraud is most likely a hidden activity, plausible at the polling-station level, available to any political party, and attractive only when elections appear close *ex-ante*. The fraud machinery has developed from a tool exclusively controlled by the central committee of a party or a public office into a descentralized activity, in which local brokers can manipulate the ballot boxes in their own regions in exchange for resources or political favors from candidates who will benefit from such activity.[9]

In particular, a strategy known as *fraude hormiga* may be prevalent.[10] 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 local vote counts by an an amount unlikely to be noticed but considerable enough to be decisive in the aggregate count. Mexico's electoral code prescribes the nullification of vote tallies in 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 competitors. This limitation of the electoral legislation creates an opportunity for *fraude hormiga*, which becomes more attractive in close elections (Crespo, February 2, 2008.).[11]

[9] A similar line of thought is presented by Szwarcberg (2008) in her analysis of electoral mobilization in Argentina. In brief, she stated that the pace of the development of a broker's political career within the party is determined by the number of votes she can mobilize. Moreover, parties rely on local brokers to commit fraud, as these brokers better understand a particular region's context. Díaz-Cayeros, Estévez and Magaloni (2007) provide similar arguments for employing brokers to mobilize voters in Mexico.

[10] The closest translation of this term in English for this term would be "leaking fraud."

[11] This strategy may be concomitant to "buying" poll workers and replacing them with partisan agents. Political parties typically offer either positive incentives (e.g., money and private goods) or negative incen-

For perpetrators, the opportunity to commit fraud depends on (1) the presence of electoral observers, (2) the opportunity to "buy" poll workers, and (3) the opportunity for perpetrators to commit fraud at a particular moment. Anticipating this behavior, electoral observers strategically position their members in those precincts with "red flags," i.e., areas where they expect fraud given the precincts' characteristics. If they cannot place one observer in each polling station, they attempt to position at least one observer in each of these precincts.[12] Therefore, fraud is the result of opportunities that, when existent, appear under stochastic conditions.

The allocation of both electoral observers and perpetrators of fraud resembles a Colonel Blotto game in which both sets of players allocate resources to achieve their respective goals (i.e., either the illicit increase in the number of votes for a particular candidate or the inhibition of that practice), and no *ex-ante* optimal strategy exists for either player (Roberson, 2006). In this model, the opposition members and observers allocate their resources to those "red flag" regions where they expect to observe fraud, while perpetrators may find opportunities to commit fraud in the "safe zones," or regions that rival parties and electoral observers leave unattended. Because neither player possesses a dominant strategy, the game outcome is determined by the resource asymmetry (Tofias, 2007).

This analysis, therefore, focuses on the opportunities that the political machines seize to manipulate some, but not all, of the polling stations in the precinct. I assume that this decision depends on neither the voter characteristics at the polling stations nor the expected vote returns for the candidates.

---

tives (e.g., violent threats) in exchange for their absence on the election day (The New York Times, July 4, 2010). Placing 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.

[12]Personal interviews with local organizations of electoral observers. Puebla, Veracruz, Hidalgo, and Oaxaca, June 2010.

# 3 Empirical Analysis

## 3.1 Identification Strategy

Precincts are the smallest subunits of the electoral district. They group voters in units of 50 to 1,500 (COFIPE, 155).[13] Within each precinct, there must be one polling station for every 750 voters. Due to demographic changes, the number of voters in a precinct can be greater than 1,500 after a precinct has been drawn; 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.), creating an additional *casilla contigua* for each additional group of 750 in the precinct.[14]

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 name (COFIPE, 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 to provide similar transit conditions for all voters (COFIPE, Art. 239).

As an illustration, consider the electoral geography of the state of Oaxaca, which is shown in Figure 1. The lines define the boundaries of the 2,694 precincts in the state, which are delineated by geographical and demographic characteristics. There are 4,452 polling stations throughout the state, as the number of polling stations in each precinct

[13]The electoral district (*distrito*) constitutes the territory of the single member districts' (SMD) deputies. These districts follow municipal boundaries, such that a district that includes more than one municipality must include the entire territory of those municipalities. At the federal level, there are 300 electoral districts, each corresponding to a deputy elected by a plurality of the vote.

[14]In each polling station, a three-stage procedure is used to randomly select poll workers. First, the local electoral organization randomly selects a month of the year. Citizens whose birthdays fall in that month are identified as potential poll workers. From this subset of registered voters, local officials select at least 10% of the names by lottery. Selected citizens are invited to attend a brief workshop, and those who are more skillful, i.e, those "with higher education (COFIPE, Art. 240, 1.d)," are chosen to participate in the final lottery to identify which individuals will be given assignments on the election day.

depends on its demographic characteristics. Consider precinct 152, which is marked in red. At the time of the 2010 local elections, it had 1,416 registered voters assigned according to alphabetical order to two different polling stations, each of which permitted 708 voters on its voting rolls. Suppose that there is a household in the precinct with two registered voters, withe last names of Abasolo and Zurita. Although they share the same address, voter Abasolo is assigned to a different polling station than the one assigned to voter Zurita. In short, the only condition for voters' assignment to a particular precinct is their home address: within each precinct, the assignment of voters to polling stations depends only on their last name.[15]

Because voters in each precinct are distributed among polling stations according to the first letter of their last name, and last names are not correlated with any relevant characteristic related to voting behavior, each polling station appears to be an unbiased sample of the precinct's voters. Relaxing the assumption that sorting the voters by alphabetical order does not create other types of covariates implies that the alphabetical assignment of voters should be sufficient to group voters according to their political preferences (i.e., there is no overlap among the members of different social groups in the precinct). To illustrate the implausibility of this situation, consider a precinct inhabited by two families, **C**orona and **T**alavera, each of whom has preference for a different political candidate. To ensure that each family is assigned to a different polling station and that there is no overlapping of members from the different families, it has to be the case that all members of each family have a very similar last name.[16] This condition means, for example, that

[15]There are two other types of polling stations that have additional requirements above the *básica* or *contigua*. "Special polling stations," or *casillas especiales*, receive the paper ballots of voters who are temporarily located outside of their assigned precinct (COFIPE, Art. 270). Similarly, "extraordinary polling stations," or *casillas extraordinarias*, are designed for those precincts whose sociocultural or geographic conditions make it difficult for all voters to commute to the same place (COFIPE, Art. 239). These polling stations represent fewer than 1% of the observations and are not taken into account in this analysis.

[16]In Mexico, both women and men keep their original last name after marriage.

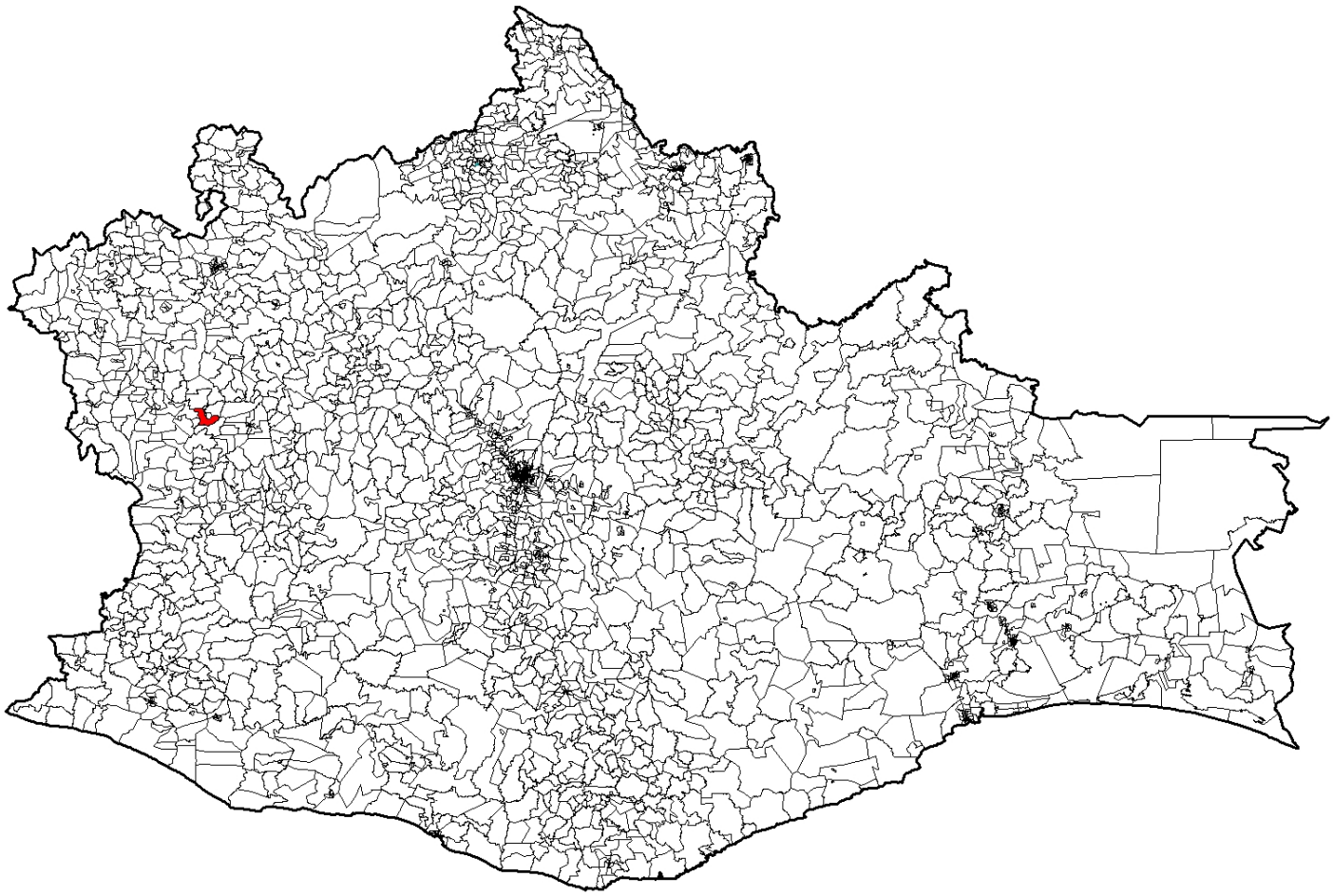

Figure 1: Electoral precincts in Oaxaca.

members of the Corona family cannot marry someone whose last name begins with letters O to Z. Otherwise, voters from the Corona family will vote in the same polling station where members of the Ochoa family are assigned.

Because the irregular activities described above must be clandestine, the challenge is to identify fraud within the vote tallies. I identify potential polling stations with irregularities by comparing the turnout and the electoral preferences of each unit with those of the other polling stations within the precinct. The units are the polling stations on the election day, and the "red flag" indicating that a polling station is suspicious is a higher turnout rate than other polling stations in the precinct. The potential finding of fraud-

ulent activity among these suspicious units is the illicit transfer of additional votes to a particular candidate. In contrast, the potential finding for the non-suspicious units is the absence of change in the vote distribution among the candidates. In consequence, I aim to measure the change in the vote returns for the candidates given differences in turnout rates.[17]

Previous studies on electoral irregularities in Mexico have focused on detecting irregularities during the process (Crespo, 2006; Mebane, 2006; Instituto Federal Electoral, 2010) and assessing whether these inconsistencies affect the electoral outcome (Aparicio, 2006, 2009; Poiré and Estrada, 2006). This paper's proposed methodology has two advantages over the previous analyses. First, the algorithm I use is able to distinguish between random and systematic effects at the polling station level. While Aparicio (2009), Mebane (2006), and Poiré and Estrada (2006) group all of their observations without distinguishing accidental from fraudulent acts, the methodology used in this paper can identify

[17]The proposed identification strategy relies on two important assumptions. The first assumption is the single-unit treatment value assumption, or SUTVA (Rubin, 1986), which refers to the individualistic treatment assignment of the unit. This assumption implies the stability of both the treatment exposures of the other units and the potential values of the treatment. The former condition is sustained given that the potential outcomes of each unit remain the same regardless of the type of treatment that other units receive. This arrangement implies that the number and distribution of the votes at a polling station are independent of the treatment assignment in other polling stations.

SUTVA also implies the stability of potential values, which must be relaxed given the lack of information to distinguish *a priori* the mechanism to assign the treatment and its effect on the vote returns. Specifically, because any political machine is able to manipulate elections, the hidden identity of the perpetrator hinders the estimation of how irregularities affect the votes for a specific party. Relaxing this assumption, however, does not pose a problem as long as each of the hypothetical manipulations is estimated on a probability distribution (Greiner and Rubin, 2011, p. 779).

The second assumption is ignorability, which implies that the potential outcomes are independent of the treatment assignment condition on other covariates. This assumption means that the choice of the treatment within the precinct does not depend on the specific characteristics of the units. For the purpose of this analysis, reliance on the ignorability assumption is based on two characteristics. First, because voters in the precinct are grouped according to alphabetical order, it seems unlikely that political behavior is correlated with voters' last names. In other words, the covariates of the polling stations in a precinct remain unaltered after grouping voters according to alphabetical order. Second, the political machine's decision is based on the probability of its activities remaining undetected. Despite the lack of data to identify the presence of observers in a polling station or the replacement of poll workers on election day, neither of these variables affects citizens' vote. Therefore, due to the independence of electoral irregularities from covariates that may affect the distribution of votes, the use of the ignorability assumption for analysis is plausible.

which errors may not be considered deliberate fraud at each polling station. This feature reduces the possibility of confusing cases of deliberate manipulation at the polling station with cases classified as accidental errors. Second, given the high variance of electoral behavior at the precinct level (Navia, 2000), grouping polling stations according to the electoral results of the district or state is vulnerable to ecological fallacy. In contrast, this paper does not categorize the observations according to the aggregate result at the district level, making the estimates to be independent of the winning candidate above the precinct level.

## 3.2 Methodology

The challenge in assessing electoral irregularities is distinguishing fraud from two other situations. First, the analysis should be able to distinguish fraud from simple human error. Inaccuracies are expected in massive elections because they involve vast mobilization and coordination in a short time span (Mozaffar and Schedler, 2002). Therefore, it is necessary to distinguish random errors from systematic ones. Second, the analysis should "distinguish electoral fraud from normal politics" (Mebane, 2010, p. 2). In other words, the empirical tools that detect electoral irregularities must be able to clarify whether the vote returns are deliberate manipulations or the consequence of expected political phenomena such as strategic voting or voter mobilization (Cox, 1997).[18]

To identify and estimate electoral irregularities in Mexico, this study undertakes the following steps: (1) collection of electoral data, (2) classification of the data, and (3) estimation of differences in electoral support for a given candidate to assess the existence of irregularities in the vote returns. Figure 2 summarizes and graphically represents the

[18]For an example of this problem during the analysis of the 2009 Iranian elections see http://thelede.blogs.nytimes.com/2009/06/13/landslide-or-fraud-the-debate-online-over-irans-election-results/

procedure for the analysis of each election.

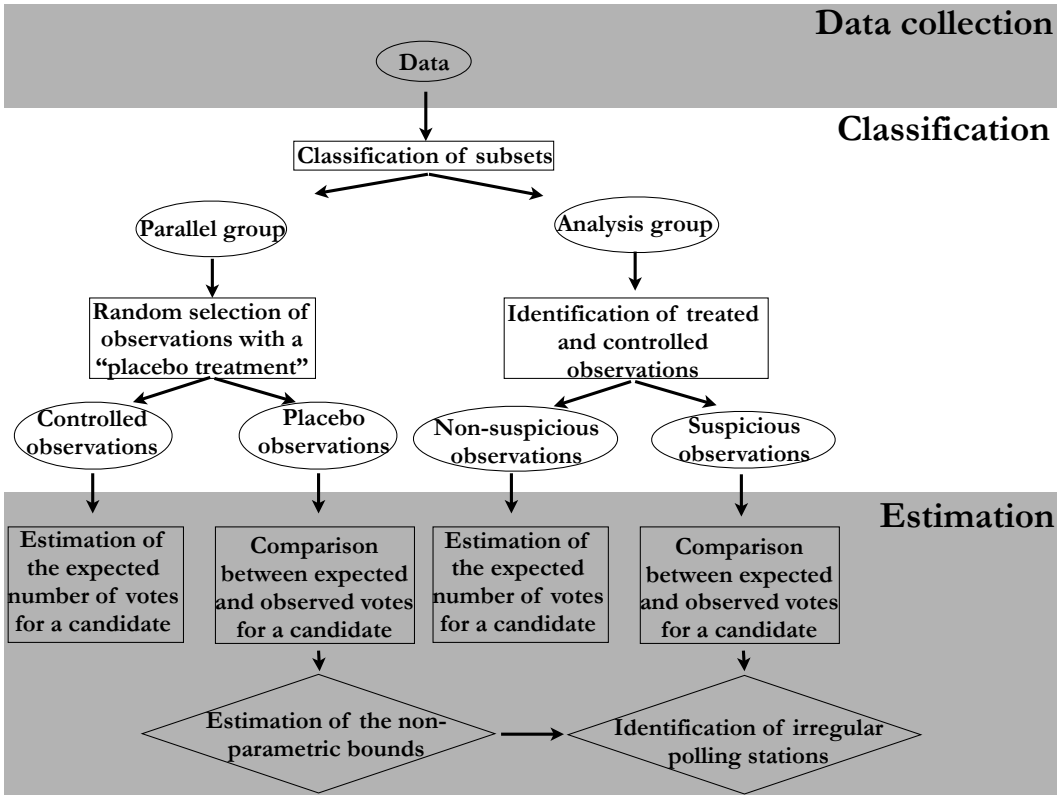


Figure 2: Diagram of the procedure for detecting irregular observations.

Every step of the process was repeated for each of the examined elections (twelve gubernatorial elections in Mexico in 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 in the polling station that are neither spoiled nor for a nonregistered candidate), the number of registered voters, and the number of votes for each candidate.

### 3.2.1 Data sources

**3.2.1.1 Case selection** When the PRI lost the presidency in 2000, many of its members sought consolation in their remaining bastions of political power: the sub-national gov-

ernments. Ten years later, the PRI had positioned itself to control not only the regions still controlled by the old guard *priistas*, but also the states run by any other political parties. Given that voters in twelve out of the thirty-two states in the country elected their governor, the gubernatorial elections in 2010 represented a decisive phase in defining the political strength of each party in anticipation of the 2012 presidential election.

In many of the states, the election pitted the PRI candidate against an unusual alliance among most of the 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 for the PRI. This strategy was controversial, even among the members of the coalition parties,[19] but represented a pragmatic way to respond to the common practices of the local political machines.

The overall results of the election were mixed for the contenders. While the coalition won in Oaxaca, Sinaloa, and Puebla, the PRI retook Aguascalientes, Tlaxcala, and Zacatecas, the first two of which were formerly governed by the PAN while the last had a governor from the PRD. Although both PAN and PRD celebrated seized three of the PRI's historical bastions, they contested the elections in Durango, Hidalgo, and Veracruz, claiming fraud. The Federal Electoral Court did not find substantive evidence to support their claim that the elections were rigged, certifying the legitimacy of the outcomes.[20]

Table D in the Appendix summarizes the results of the gubernatorial elections in 2010. It lists the members of each electoral coalition and the aggregated results of the election. It also describes the electoral data for each election, specifying the number of available observations and their sources.

[19]The PRI's leader in the Senate, Manlio Fabio Beltrones, commented that "alliances between enemies that dont 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 leading to fraud." (Los Angeles Times, February 01, 2010)

[20]See Prats (August 8, 2010).

**3.2.1.2 Data collection** For eight out of the twelve elections, information was obtained through the websites of the local electoral institutions that had already published the official results. In the case of Tamaulipas and Oaxaca, the data were acquired via a formal request to the local electoral institution. For Hidalgo and Veracruz, the official results at the polling-station level were not yet available at the time of this study. To overcome the lack of data, I used the information downloaded three days after the election through the websites that each of the local electoral organizations used to publish the preliminary vote counts. This system is known as the Program for the Preliminary Electoral Results (PREP), and it has the task of reducing the uncertainty of the electoral results before the official results are announced. The PREP is updated in real time and is typically disabled once the electoral outcome is publicly available and is recognized by the candidates. The local electoral authorities publish the information in the PREP after they receive the official form completed by the poll workers in each polling station.[21]

The vote returns for each polling station were matched with the number of registered voters in each polling station, a number deduced by dividing the number of registered voters in the precinct by the number of polling stations open on election day.[22] The data were provided by a formal information request to the IFE, which is in charge of providing voter identification to all citizens in the country.[23]

This study only considers observations in the database that are free of inconsistencies identified by the electoral authorities.[24] I also exclude both extraordinary and special

[21] Because the PREP stops publishing the results once the political actors recognize the electoral outcome, the information is not available for all polling stations. The time at which ballot boxes are received by local electoral organizations varies according to the geographical location of the polling station. Thus, data from isolated polling stations may not be available, minimizing the data by which electoral irregularities may be detected.

[22] As an illustration, and recalling the example in Figure 1, precinct number 152 in Oaxaca has two polling station for its 1,416 registered voters. Thus, there are 708 voters for every polling station.

[23] Prior to the 2010 local elections, this information was last updated in July 2009.

[24] 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

polling stations.[25]

### 3.2.2 Classification

One of the basic assumptions of this study is that within each electoral precinct, electoral behavior is unrelated to the assignment of voters to the polling stations. Hence, I do not expect to find significant differences in the turnout levels among the polling stations of the same precinct. The challenge is, therefore, to distinguish significant differences due to random factors. To analyze how different the observations within precincts are, I consider the general dispersion of the observations in each precinct.

In every precinct $X$, the differences in turnout between a polling station $i$ and the rest of the polling stations $j \neq i$ in the precinct was measured. The turnout is estimated by dividing 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:

$$d_X = |Turnout_i - Turnout_j|; \quad i, j \in X, i \neq j$$

That is, the absolute difference is calculated between the turnout of polling station $i$ and the turnout of polling station $j$ in precinct $X$. Therefore, a precinct is considered irregular whenever the differences in turnout across precincts are greater than the quantile $d = Q(.95)$ in $D$.[26] In other words, observations $i$ in which $F^{-1}(d_i) < 0.05$ are considered.

As an illustration of the methodology, consider an election between two candidates,

form, or the submission of the form without the official envelope. These observations represent less than 0.5% of the total observations.

[25]While the *casillas especiales* allow voters from outside the precinct to vote, the *casillas extraordinarias* are designed for voters living in communities distant from the regular polling stations but in an insufficient number to constitute a precinct.

[26]Consider $D$ a random variable with a probability function $F(d) = Prob(D \geq d)$. Following (Parzen, July 2003, p.2), I assign $\tau = .95$ so that $\tau(d)$ represents the proportion of the population whose values are less than or equal to $d$. Formally, define $d^* = Q(\tau)$ and suppose that $\tau$ is in the range of $F$; there exists a value $d$ such that $\tau = F(d)$.

María and Juan, in a state with 100 precincts and 1,500 registered voters in every precinct (i.e., with two polling stations per precinct). Consider the vote returns for the first four precincts, which are shown in Table 1.

| Precinct | Polling station i.d. | Registered voters | María (% in parenthesis) | Juan (% in parenthesis) | Total (turnout in parenthesis) |
|---|---|---|---|---|---|
| 1 | 1a | 750 | 314 (65%) | 168 (35%) | 482 (64%) |
| 1 | 1b | 750 | 317 (63%) | 178 (36%) | 495 (66%) |
| 2 | 2a | 750 | 500 (66%) | 250 (33%) | 750 (100%) |
| 2 | 2b | 750 | 500 (66%) | 249 (33%) | 749 (100%) |
| 3 | 3a | 750 | 450 (73%) | 168 (27%) | 618 (82%) |
| 3 | 3b | 750 | 310 (62%) | 190 (38%) | 500 (66%) |
| 4 | 4a | 750 | 314 (45%) | 376 (54%) | 690 (92%) |
| 4 | 4b | 750 | 249 (46%) | 301 (54%) | 550 (73%) |

Table 1: Vote returns for the first four precincts in a simulated example.

To calculate the differences in voter turnout in the first precinct, the absolute value of the difference between the turnout observed in $1a$, which is $482/750 = 0.64$, and the turnout observed in $1b$, which is $495/750 = 0.66$, is used. As a result, $d_1 = |0.64 - 0.66| = 0.02$. The same procedure is used for each precinct in the state, and the value of the percentile $95$ in $d^* = 0.058$ is estimated. Only the precincts whose differences between polling stations are greater than $d^*$ are considered in this analysis. This case holds for precincts $3$ and $4$, where the differences in turnout, $d_3 = |(618/750) - (500/750)| = 0.160$ and $d_4 = |(690/750) - (550/750)| = 0.186$, are greater than the critical value $d^*$.

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 $d_i > Q(0.95)$; this subset is labeled the *analysis group*.

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 $d^* = Q(0.95)$. In contrast, the polling stations in the analysis group that cannot be classified as treated are considered to be the non-suspicious observations.

In the hypothetical example described above, while precincts $1$ and $2$ are assigned to the parallel group, precincts $3$ and $4$ are part of the analysis group. In the latter group, polling stations $3a$ and $4a$ are classified as treated observations because their turnout rates are higher than those of the other observations in their precinct.

### 3.2.3 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 counting the votes should not affect the differences in the estimated proportion of votes for each candidate.

If this condition holds, the expected number of votes for a candidate $A$ in a suspicious polling station, $s$, of precinct $X$ is

$$E(\text{Votes for candidate A})_s = \frac{(\text{Votes for candidate A})_n}{(\text{Effective votes})_n}(\text{Effective votes})_s; \quad n, s \in X$$

Regardless of the differences in the total number of votes between the polling stations (i.e., $Turnout_s > Turnout_n$) the proportion of votes in a suspicious polling station for any given party should be similar to the number observed for the non-suspicious polling station, $n$, in the same precinct.[27]

Recalling the hypothetical example described in Table 1, the expected number of votes for candidate María in polling station $3b$ is $E(\text{Votes for Maria})_s = \frac{310}{500} * 618 = 383.16$, while the expected vote tally polling station $4a$ is: $E(\text{Votes for Maria})_s = \frac{249}{550} * 690 = 312.38$. Comparing the number of expected votes for candidate María with the observed number of votes in these polling stations, a difference of more than $67$ votes in polling station $3b$ and less than two votes in $4a$ may be observed.

To estimate the expected random error for the estimated number of votes, I consider the dispersion of votes for a party within the parallel group. Because the observations lack significant differences in the total number of votes within the precincts, I can evaluate the differences in the proportion of votes for each party when no irregularities are observed. To estimate these differences, I randomly assign a placebo label of suspicious polling station to serve as a reference to estimate the expected number of votes for each polling station in the precinct. To assign this "placebo treatment," each observation in the subset is treated as a Bernoulli trial with two possible outcomes and a probability $p = 0.5$ of being a treated/controlled observation in the precinct is established.

The left plot in Figure 3 shows the difference between the expected and observed num-

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

$$E(\text{Votes for candidate A})_s = \frac{\sum_{n=1}^{N}(\text{Votes for candidate A})_n}{\sum_{n=1}^{N}(\text{Effective votes})_n} * \text{Effective votes})_s; \quad n, s \in X$$

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

bers of votes for María in the parallel group. This plot only displays 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 $1a$ and $2b$ are randomly considered suspicious, and each station's expected number of votes for María is estimated using the vote returns in polling stations $2a$ and $1b$.

A perfect correlation between expected and observed votes would imply that all observations fall over the red line; i.e., the proportion of votes for any given candidate is the same across all polling stations in the same precinct. Because the measurement is subject to random events that may affect the number of votes in any polling station, I estimate the continuous quantiles that include 95% of the observations. In Figure 3, the green lines in the plot show the Bayesian quantile regression for quantiles 0.025 and 0.975.[28] The area between these two lines denotes the variance between the expected and observed votes 95 out of 100 times.

Analyzing the accuracy of the estimated votes in the parallel group provides an opportunity to observe how much dispersion is expected in the number of votes for each party in any given precinct. I estimate non-parametric bounds that include at least 95% of the observations. For this, the Bayesian quantile regressions for $\tau = .025$ and $\tau = .975$ are estimated.

[28]Following (Chen, 2005, p.3), the linear conditional function purpose for a given quantile, $\tau$, $Q(\tau|X = x) = x'\beta(\tau)$ can be estimated by solving the following:

$$\hat{\beta}(\tau) = argmin_{\beta \in \mathcal{R}} \sum_{i=1}^{n} \rho_\tau (y_i - x_i^{'}\beta)$$

Yo and Moyeed (2001, p. 439) implemented the MCMC method to get the quantiles via the Bayesian framework. The posterior distribution of $\beta$, $\pi(\beta|y)$ is given by

$$\pi(\beta|y) \propto L(y|\beta)p(\beta),$$

where $p(\beta)$ is the prior distribution of $\beta$ and the likelihood function $L(y|\beta)$ is

$$L(y|\beta) = p^n(1-p)^n exp\ \left\{-\Sigma_i \rho_p (y_i - x_i^{'}\beta)\right\}$$

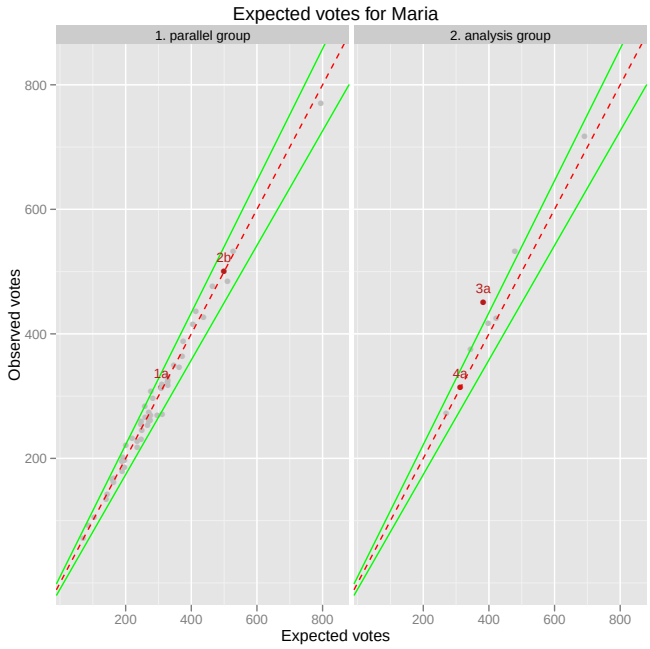


Figure 3: Estimation of expected votes for a simulated example.

The final step in classifying irregular observations is to use the non-parametric bounds from the parallel group to analyze the treated polling stations that belong to the analysis group. If turnout differences can be explained by non-intentional 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 in the 95% confidence interval of the expected vote estimation). Otherwise, it is plausible that the irregularities observed in the polling station at the turnout levels disproportionally affect a particular political party.

The right panel of Figure 3 shows the difference between the expected and observed numbers of votes in the analysis group. I use the same non-parametric bounds from the left plot to consider the variation when no differences in the total number of valid votes were observed. The observed number of votes for María in polling station $4a$ falls within the green lines; that is, despite the difference in the turnout rates between the polling

stations of precinct number $4$, the number of votes for the candidates are still similar. In contrast, polling station $3a$ displays not only a greater number of valid votes compared with other observations in the precinct but also a higher proportion of votes for candidate María. This difference falls outside the bounds and would be considered a suspicious observation.

# 4 Results

Figures A and B in the appendix depict the graphical analyses for the incumbent party and the most significant challenger in each of the elections examined in this paper. For each of these elections, there are two panels. The left panel displays the differences between the expected and observed votes for the observations in the parallel group that were assigned a random "placebo treatment," while the right panel shows the differences between expected and observed votes for the treated observations in the analysis group. The bounds in both panels are the Bayesian quantile regressions for $\tau = 0.025$ and $\tau = 0.975$, covering approximately 95% of the observations in the parallel group. In other words, to identify the irregular observations in the analysis group, I use the bounds produced in the panel where no irregularities are expected.

To make the graphs more comprehensible, I discuss three of the cases represented in the following figures. First, consider Figure 4, 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, for the case of Oaxaca illustrated on Figure 5, the proportion of observations above the upper bound is evidently larger in the analy-

sis group than in the parallel group. These observations not only indicate a proportion of the total number of observations but also represent the group with the highest difference between the expected and observed number of votes. Third, the graphical analysis for Durango, illustrated in Figure 6, represents an very interesting case. Note that most of the suspicious observations 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). In addition, the cluster of observations on the top-left side of the graph is unusual pattern in comparison with the rest of the graphical analyses.

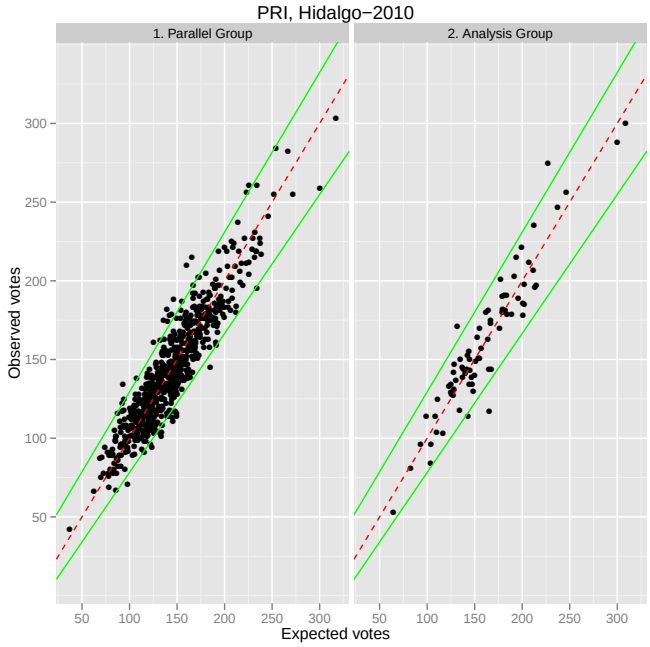


Figure 4: Observed and expected votes for the incumbent coalition in Hidalgo, 2010.

The results identify specific events that occurred on election day. Consider the analysis of the election in Durango, where most of the detected irregular observations occurred in the city of Gómez Palacio. On the day of the election, a group of armed people disrupted the electoral process in precinct 447's polling stations, representing most of the observa-

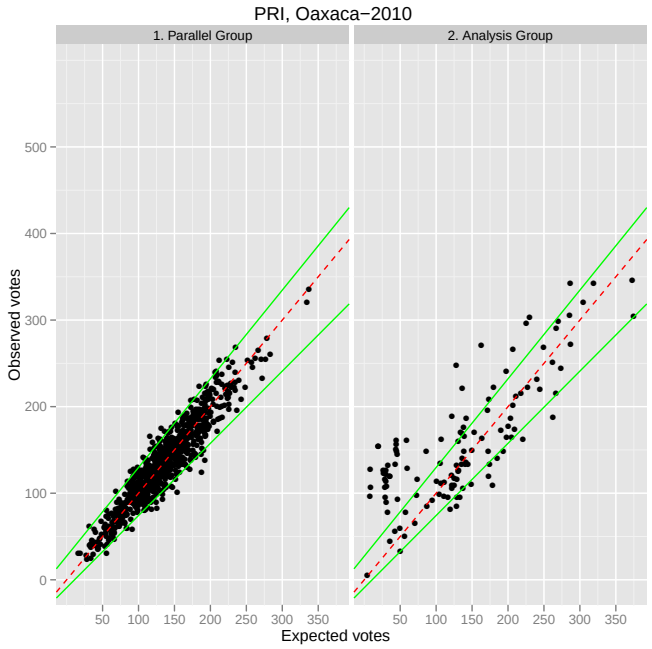


Figure 5: Estimation of expected votes for the incumbent coalition in Oaxaca, 2010.

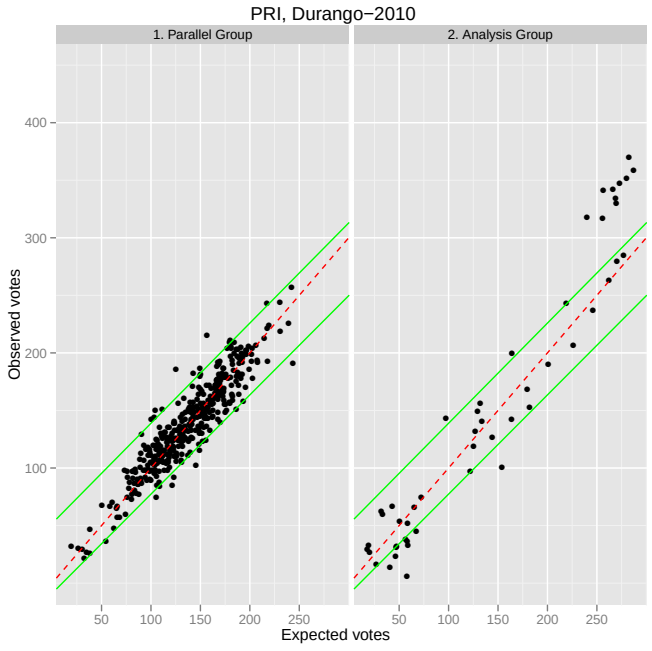


Figure 6: Estimation of expected votes for the incumbent coalition in Durango, 2010.

tions in the unusual cluster discussed above. This event resulted in two dead policemen and caused voters and poll workers to flee the polling stations. Despite the evidence, the local electoral institute counted the votes for the eighteen polling stations located in the affected precinct (El Siglo de Durango, July 15, 2010). Another case is the election in Oaxaca, where the algorithm detects irregularities in the municipality of Tututepec. Several days prior to election day, 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.[29]

Table 2 summarizes the information in Figures A and B in the Appendix, accounting for the number of observations with an unexpected number of votes benefitting a particular candidate. Because the identification strategy locates the additional number of votes that contribute to one party's tally, this paper only considers observations where the observed number of votes is greater than the expected number of votes or observations above the bound of $\tau = 0.975$.[30] As depicted in Figure A, the number of observations outside the bounds in the analysis group is higher than that those observed in the parallel group in most cases.

The next step in the analysis is to determine whether these observations affect the vote returns of any candidate. Because the proportion of observations outside the bounds in the parallel group is approximately 5%, the out-of-bound observations may be the result of random errors. As a consequence, a similar proportion of these cases is expected in the analysis group. If the evidence suggests that the proportion of observations outside the bounds is higher than expected (i.e., the observed proportion in the parallel group),

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

[30] Table A in the Appendix shows the table and the analysis of the observations below the bound for $\tau = 0.025$. Although irrelevant for this paper, it is worth mentioning that the results in most cases are reciprocal to the analysis for $\tau = 0.975$. That is, the number of observations above the bounds for the expected votes of a party is very similar to the observations below the bounds for the opposite party in the election.

| | Incumbent | | Challenger | |
|---|---|---|---|---|
| State | Parallel group | Analysis group | Parallel group | Analysis group |
| Aguascalientes | 6 (239) | 4 (45) | 7 (239) | 0 (45) |
| Chihuahua | 21 (794) | 9 (102) | 20 (794) | 7 (102) |
| Durango | 10 (404) | 14 (51) | 10 (404) | 2 (51) |
| Hidalgo | 16 (668) | 2 (80) | 17 (668) | 1 (80) |
| Oaxaca | 22 (915) | 48 (140) | 23 (915) | 19 (140) |
| Puebla | 29 (1116) | 41 (224) | 29 (1116) | 10 (224) |
| Quintana Roo | 6 (207) | 16 (54) | 6 (207) | 13 (54) |
| Sinaloa | 7 (198) | 4 (43) | 5 (198) | 7 (43) |
| Tamaulipas | 21 (859) | 17 (94) | 22 (859) | 9 (94) |
| Tlaxcala | 6 (227) | 1 (45) | 7 (227) | 3 (45) |
| Veracruz | 31 (1249) | 63 (152) | 31 (1249) | 25 (152) |
| Zacatecas | 8 (300) | 5 (37) | 9 (300) | 4 (37) |

Table 2: Observations above the Bayesian quantile regression for $\tau = 0.975$. Each number in the parentheses is the total number of observations in each sample.

then a bias for a particular candidate in the polling stations of the analysis group can be inferred. In short, we must assess the way in which the evidence either confirms or rejects two rival models of the observations outside the bounds. The observed data ($y$) are either the product of a systematic bias ($M_1$) or the result of a random error ($M_0$). Therefore, assuming that either model is true, the posterior probability for $M_0$ is estimated. I use the Bayes theorem to calculate the posterior probability for each model as follows:

$$Pr(M_k|y) = \frac{p(y|M_k)Pr(M_k)}{p(y|M_0)Pr(M_0) + p(y|M_1)Pr(M_1)}; k = 0, 1$$

The information provided in Table 2 is sufficient to estimate the posterior probability for every each case. To illustrate this point, let us examine the case of Aguascalientes, whose quantities of interests are displayed in the first several rows of the table. The probability of obtaining observations above the Bayesian quantile regression for $\tau = 0.975$ in the parallel group is $\frac{6}{239} = 0.025$, while the number of observations in the parallel group represents a proportion of $\frac{239}{226+45} = 0.841$ of the total observation in the sample. Similarly, the proportion of observations above the upper bound in the analysis group is $\frac{4}{45} = 0.088$, while the observations in the analysis group represent $\frac{45}{226+45} = 0.159$ of the observations in the sample. Consequently, the posterior probabilities for both models are

$$Pr(M_0|y) = \frac{(0.025) * (0.841)}{(0.025) * (0.841) + (0.088) * (0.159)} = 0.600$$

and

$$Pr(M_1|y) = \frac{(0.088) * (0.159)}{(0.025) * (0.841) + (0.088) * (0.159)} = 0.4000$$

With the above quantities, the posterior odds in favor of $M_1$ are estimated:

$$\frac{Pr(M_1|y)}{Pr(M_0|y)} = \frac{P(y|M_1)}{P(y|M_0)} * \frac{Pr(M_1)}{Pr(M_0)}$$

The above implies that the posterior odds are a product of the prior odds using Bayes factor, which is:

$$B_{10} = \frac{P(y|M_1)}{P(y|M_0)}$$

I use the Bayes factor as a summary of the evidence provided by the data as opposed to the prior supposition. In particular, I assess whether the frequency of observations out-

side the bounds is due to systematic differences in voter turnout. Jeffreys (1961) suggests using the Bayes factor as evidence against the null hypothesis, which is the unexpected number of votes for a candidate as a result of a random error. The interpretation of this factor proposes that a value of $B_{10} \geq 10$ provides "strong evidence" against the null hypothesis.[31]

Table 3 indicates the value of the Bayes factor that was derived from the division of marginal likelihoods for each model, $P(y|M_k)$. To illustrate how to calculate the estimation, consider the election in Hidalgo. The probability of identifying observations above the bound for $\tau = 0.975$ in the parallel group is $P(y|M_0) = \frac{16}{668} = 0.024$, while the proportion of observations above the bound in the analysis group is $P(y|M_1) = \frac{2}{80} = 0.025$. Consequently, the ratio of the posterior odds of $M_1$ to its prior odds is $B_{10} = \frac{0.025}{0.024} = 1.043$. In this case, the value of $B_{10}$ suggests that the evidence against $M_0$ is "not worth more than a bare mention."(Jeffreys, 1961, p. 471).

The Bayes factor finds strong evidence for the alternative hypotheses, $M_1$, in Durango, Oaxaca, Veracruz, and Quintana Roo. For the first three states mentioned, there is strong evidence of systematic bias for incumbent candidates, suggesting a systematic bias in the irregular observations to the benefit of one candidate. There is no reason to assume that random error would benefit the incumbent party, but the distribution of the political capital among the candidates makes it plausible that non-accidental irregularities were triggered by the manipulation of the vote in favor of candidates supported by the local government.

Bayes factors are sensitive to the priors, as they may substantially influence the results

[31]Jeffreys (1961) assesses that a value of $10^{-1} > B_{01} > 10^{-\frac{3}{2}}$ provides "strong evidence" against the null hypothesis. I take the inverse of that value just to clarify the interpretation of the coefficient. Theoption to consider only those values greater than 10 follows the discussion that Goodman (2005) highlights at the threshold for "substantive evidence," $B_{10} \geq 3.2$ is a very low standard to assess the evidence for any of the hypotheses.

| | Incumbent | | | Challenger | | |
|---|---|---|---|---|---|---|
| State | $P(y\|M_1)$ | $P(y\|M_0)$ | $B_{10} = \frac{P(y\|M_1)}{P(y\|M_0)}$ | $P(y\|M_1)$ | $P(y\|M_0)$ | $B_{10} = \frac{P(y\|M_1)}{P(y\|M_0)}$ |
| Aguascalientes | 0.089 | 0.025 | 3.541 | 0.000 | 0.029 | 0.000 |
| Chihuahua | 0.088 | 0.026 | 3.336 | 0.069 | 0.025 | 2.725 |
| Durango | 0.275 | 0.025 | **11.090** | 0.039 | 0.025 | 1.584 |
| Hidalgo | 0.025 | 0.024 | 1.044 | 0.013 | 0.025 | 0.491 |
| Oaxaca | 0.343 | 0.024 | **14.260** | 0.136 | 0.025 | 5.399 |
| Puebla | 0.183 | 0.026 | 7.044 | 0.045 | 0.026 | 1.718 |
| Quintana Roo | 0.296 | 0.029 | **10.222** | 0.241 | 0.029 | 8.306 |
| Sinaloa | 0.093 | 0.035 | 2.631 | 0.163 | 0.025 | 6.447 |
| Tamaulipas | 0.181 | 0.024 | 7.398 | 0.096 | 0.026 | 3.738 |
| Tlaxcala | 0.022 | 0.026 | 0.841 | 0.067 | 0.031 | 2.162 |
| Veracruz | 0.414 | 0.025 | **16.699** | 0.164 | 0.025 | 6.627 |
| Zacatecas | 0.135 | 0.027 | 5.068 | 0.108 | 0.030 | 3.604 |

Table 3: Marginal likelihoods and Bayes factors for observations above the Bayesian quantile regression for $\tau = 0.975$.

(Kass, 1992). Furthermore, the amount of data available for the analysis determines how sensitive the Bayes factor is to the prior distribution (O'Hagan, 1995, p. 112). To measure the robustness of the results, I depict the asymptotic of the Bayes factor via bootstrapping. In particular, I take random subsamples of varying sizes from both the training and the analysis sets and follow the procedure described above to produce the Bayes factor. For every election and coalition, I repeat the procedure 10,000 times and count the number of times that $B10 < 10$. The results are shown in Table B in the Appendix, but the only case in which the result is sensitive to the sample size is in Quintana Roo. In consequence, the available data for that state are insufficient to test both hypotheses. In sum, the analysis finds evidence of a systematic bias among the vote returns for the incumbent party in three states: Durango, Oaxaca, and Veracruz.

To test the face validity of the results, I compare the estimations with other sources of information related to the legitimacy of each election. First, I consider the citizens' evaluation of the legality of the election using the only post-electoral poll on the state level in 2010 (GCE, 2010). Two months after the elections, citizens in each state were asked the

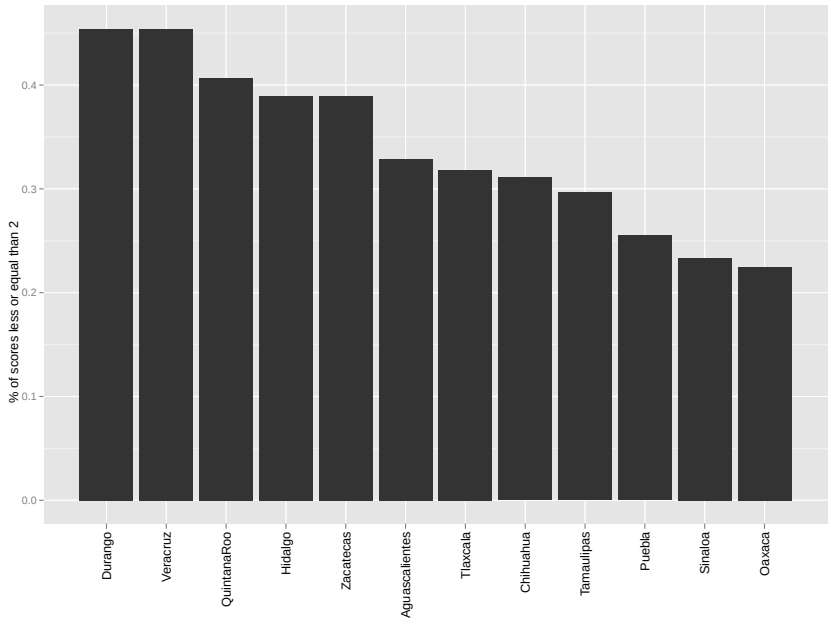


Figure 7: Public opinion: % of the respondents qualifying the legality of the election with a score equal to or below the rating of 2 in a 1-4 scale). Gabinete de Comunicación Estratégica, 2010.

following: "On a scale from 1 to 4, where 1 is 'not legal at all' and 4 means 'totally legal,' how legal do you think the election in your state last July 4th was?" Figure 7 displays the percentage of people in each state who responded with answers equal to or below the rating of 2. Veracruz and Durango have the highest proportion of responses questioning the legality of their elections. In contrast, citizens in Oaxaca appear to be more confident about the legality of their elections; however, this result could be a result of the survey's timing. The opposition's victory has influenced citizens' opinions regarding the quality of the electoral process. Forty-five days prior to the election, 48% of the population predicted fraudulent activities in the gubernatorial election, and only 31% believed that no fraud would take place (Moreno and Mendizabal, May 20, 2010, p.10).

The second method to validate the results uses a number of reported irregularities submitted to the Special Prosecutor's Office for Electoral Crimes (FEPADE), which is an

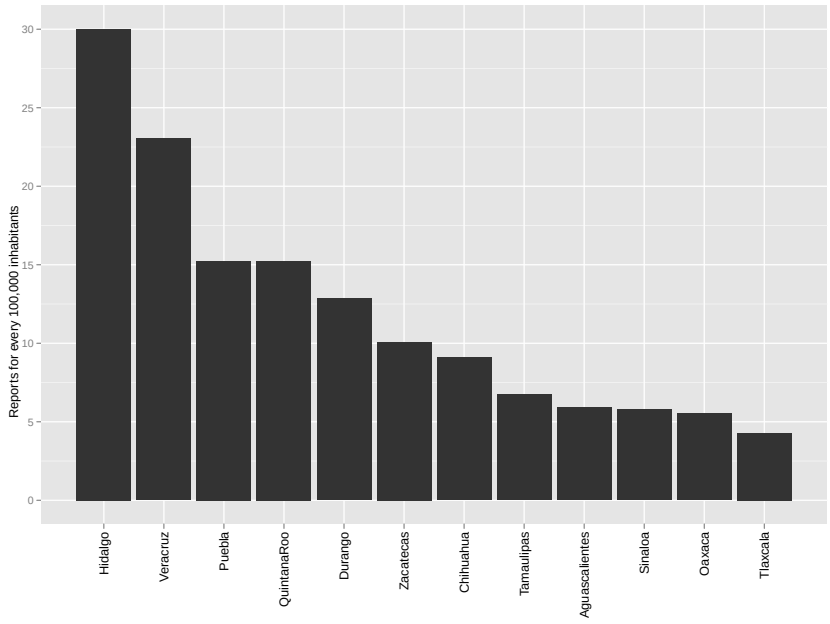


Figure 8: Reports of electoral irregularities reports for every 100,000 inhabitants. FEPADE, 2010.

independent branch of the Office of the Attorney General (PGR) that prosecutes electoral crimes. Because the Mexican Penal Code considers fraudulent activities during elections to be federal crimes, FEPADE receives all of the allegations of irregularities concerning with electoral procedures. Figure 8 depicts the rate of reports received by FEPADE in 2010 in the states that are part of this study. I normalized the number of reports according to the population of each state, obtaining the number of reports for every 100,000 state residents. It turned out that Hidalgo, Veracruz, and Durango have the highest rate of reported irregularities, while Oaxaca retains its place among the group of states with the lowest rates of anomalies reported.

# 5 Conclusion

Przeworski et al. (2000, p. 26) describes the behavior and ambitions of politicians in general, characterizing them as "just PRIstas by nature." In the particular case of Mexico, this paper's findings reveal that this statement is indeed true. In particular, the liberalization of the political system in Mexico has not eliminated the fraudulent behavior of political machines, having made such illegal practices more covert.

The results provide evidence of the new way in which political machines attempt to interfere in elections. Once visible fraud came to be prosecuted, perpetrators have attempted to distort the voting returns by an unobservable yet effective means. The findings demonstrate evidence of vote-rigging in three of the twelve states that held gubernatorial elections in 2010. Specifically, electoral irregularities favored the PRI in Durango, Oaxaca, and Veracruz. The proposed methodology also provides quantitative evidence to suggest which candidate benefitted from violent disruptions, duplicated paper ballots, and missing ballots.

Sub-national democracy is the new challenge for Mexico and other federal systems in consolidating their democracies. During the late 1990s, Cornelius (1999) signaled that the concomitant decentralization and democratization of Mexico may result in the configuration of an archipelago of autocracies within the country, where governors and local elites would be the ultimate obstacles to full democratic consolidation. Ten years later, it is evident that despite the democratization of the political system at the federal level, sub-national units remain vulnerable to authoritarian practices. Although the integrity of the elections is not a sufficient condition for a democratic system, the findings of this paper encourage scholars and policy-makers to examine the ultimate connection between citizens and politicians: the ballot box.

# Appendix

| | Incumbent | | | Challenger | | |
|---|---|---|---|---|---|---|
| State | $P(y\|M_1)$ | $P(y\|M_0)$ | $B_{10} = \frac{P(y\|M_1)}{P(y\|M_0)}$ | $P(y\|M_1)$ | $P(y\|M_0)$ | $B_{10} = \frac{P(y\|M_1)}{P(y\|M_0)}$ |
| Aguascalientes | 0.022 | 0.029 | 0.759 | 0.089 | 0.021 | 4.249 |
| Chihuahua | 0.039 | 0.025 | 1.557 | 0.108 | 0.025 | 4.281 |
| Durango | 0.176 | 0.030 | 5.941 | 0.294 | 0.025 | 11.882 |
| Hidalgo | 0.025 | 0.024 | 1.044 | 0.038 | 0.025 | 1.474 |
| Oaxaca | 0.100 | 0.024 | 4.159 | 0.357 | 0.024 | 14.854 |
| Puebla | 0.049 | 0.022 | 2.283 | 0.125 | 0.027 | 4.650 |
| Quintana Roo | 0.204 | 0.029 | 7.028 | 0.278 | 0.029 | 9.583 |
| Sinaloa | 0.163 | 0.030 | 5.372 | 0.163 | 0.035 | 4.605 |
| Tamaulipas | 0.074 | 0.024 | 3.046 | 0.170 | 0.024 | 6.963 |
| Tlaxcala | 0.178 | 0.018 | 10.089 | 0.022 | 0.026 | 0.841 |
| Veracruz | 0.164 | 0.027 | 6.042 | 0.414 | 0.023 | 17.851 |
| Zacatecas | 0.135 | 0.027 | 5.068 | 0.162 | 0.023 | 6.950 |

Table A: Marginal likelihoods and Bayes factors for observations above the Bayesian quantile regression for $\tau = 0.025$.

| | | | Size of the subsample | | |
|---|---|---|---|---|---|
| State | Coalition | Bayes factor | $\alpha = 0.99$ | $\alpha = 0.975$ | $\alpha = 0.95$ |
| Durango | Incumbent | 11.090 | 0.000 | 0.000 | 0.116 |
| Oaxaca | Incumbent | 14.260 | 0.000 | 0.000 | 0.000 |
| Quintana Roo | Incumbent | 10.222 | 0.295 | 0.312 | 0.303 |
| Veracruz | Incumbent | 16.699 | 0.000 | 0.000 | 0.000 |

Table B: Results for the bootstrapping test after 50,000 iterations. Parallel and the analysis sets were bootstrapped with samples of size $\alpha$. The table shows the proportion of iterations in which, $B_{10} = \frac{P(y|M_1)}{P(y|M_0)} < 10$.

| State | Number of polling stations | Source |
|---|---|---|
| Aguascalientes | 1334 | http://www.eleccionesenmexico.org.mx/ |
| Chihuahua | 5003 | http://www.ieechihuahua.org.mx/ |
| Durango | 2403 | http://www.ieedgo.org.mx |
| Hidalgo | 3316 | PREP |
| Oaxaca | 4761 | Formal request to the local electoral institute |
| Puebla | 6174 | http://www.ieepuebla.org.mx/ |
| Quintana Roo | 1526 | http://www.eleccionesenmexico.org.mx/ |
| Sinaloa | 4492 | http://www.cee-sinaloa.org.mx/ |
| Tamaulipas | 4285 | Formal request to the local electoral institute |
| Tlaxcala | 1246 | http://www.ietlax.org.mx/ |
| Veracruz | 9826 (8214 available) | PREP |
| Zacatecas | 2553 | http://www.ieez.org.mx/ |

Table C: Elections analyzed in the paper.

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

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

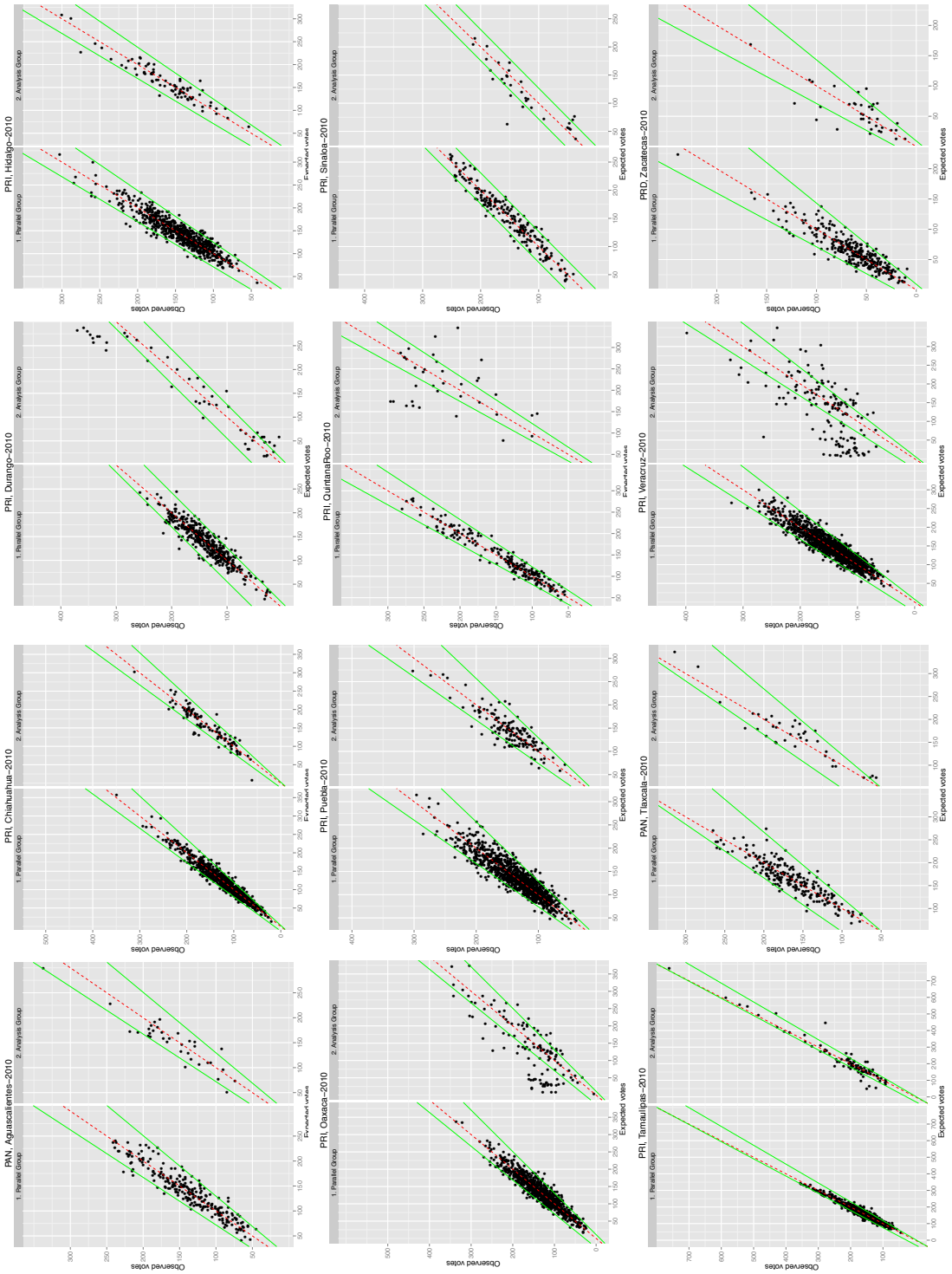


Figure A: Observed and expected votes for the incumbent party or coalition.

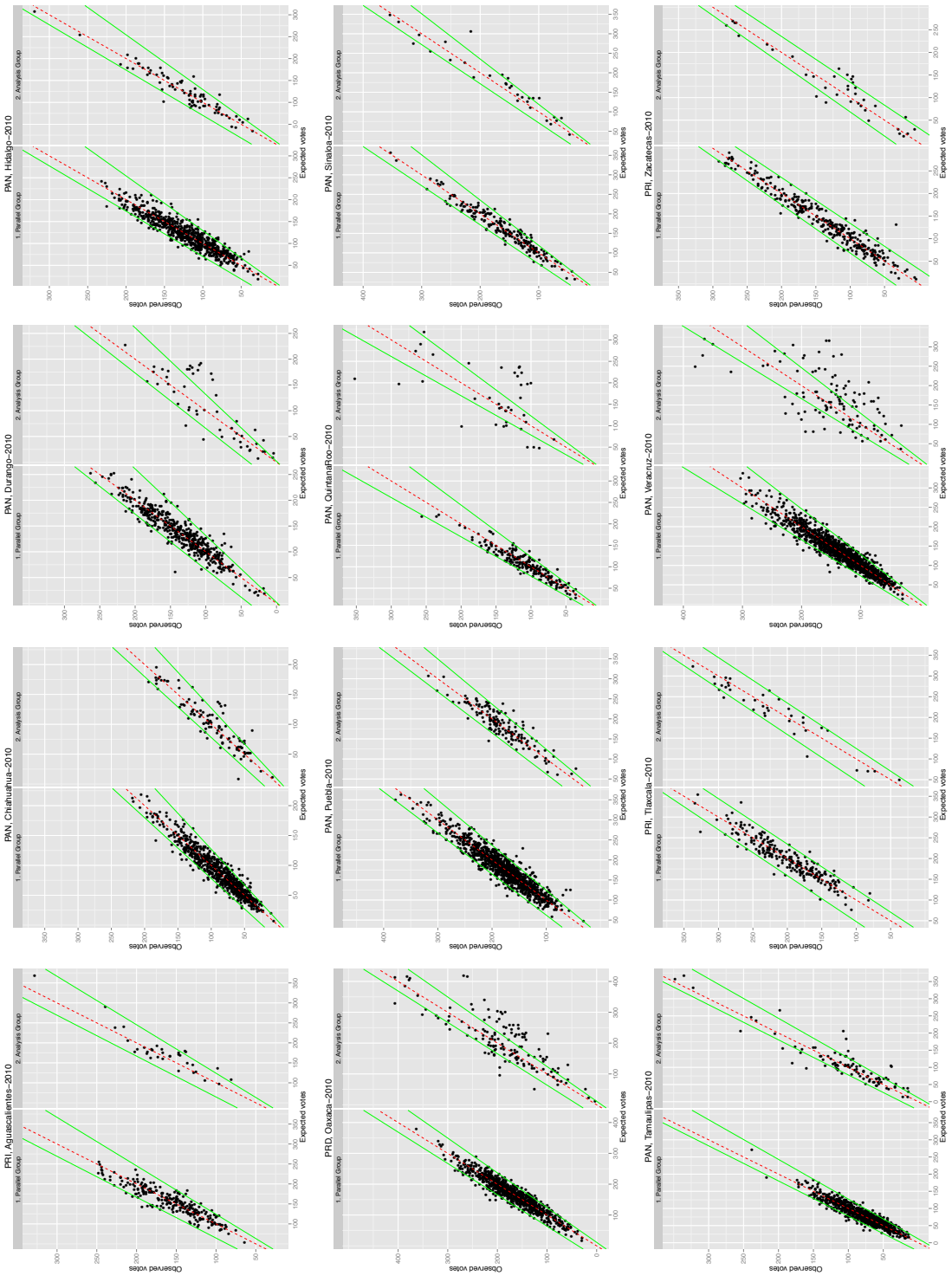


Figure B: Observed and expected votes for the challenger party or coalition.

# References


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.

Aparicio, Ricardo. 2002. "La Magnitud de la Manipulación del Voto en las Elecciones Federales del Año 2000." *Perfiles Latinoamericanos* (20):79–99.

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

Chen, Colin. 2005. "An Introduction to Quantile Regression and the QUANTREG Procedure." Proceedings of the THirtieth Annual SAS Users Group.

Cornelius, Wayne A. 1999. Subnational Politics and Democratization: Tensions between Center and Periphery in the Mexican Political System. In *Subnational Politics and Democratization in Mexico*, ed. Wayne A. Cornelius, Todd A. Eisenstadt and Jane Hindley. Center for U.S.-Mexican Studies.

Cornelius, Wayne A. 2000. "Blind spots in democratization: Sub-national politics as a constraint on Mexico's transition." *Democratization* 7(3):117–132.

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

Cox, Gary W. 1997. *Making Votes Count*. Cambridge University Press.

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. 2006. *2006: hablan las actas*. Random House Mondadori.

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

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.

Díaz-Cayeros, Albero, Federico Estévez and Beatriz Magaloni. 2007. "The Core Voter Model: Evidence From Mexico." Leitner Working Paper 2007-12, Yale University.

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

Eisenstadt, Todd A. 2003. "Thinking outside the (Ballot) Box: Informal Electoral Institutions and Mexico's Political Opening." *Latin American Politics and Society* 45(1):25–54.
**URL:** *http://www.jstor.org/stable/3177062*

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

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

GCE. 2010. "Reporte Postelectoral 2010." Survey.
**URL:** *http://www.gabinetece.com.mx/images/uploads/documents/documentos/*$n_e studios/reporte_0 60810.pdf$

Gervasoni, Carlos. 2008. "Conceptualizing and Measuring Subnational Regimes: An Expert-based Operationalization of Democracy in the Argentine Provinces." Presented at the 2008 Annual Meeting of the American Political Science Association. Boston, MA.

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

Gibson, Edward L. 2008. "Subnational Authoritarianism and Territorial Politics: Charting the Theoretical Landscape." Paper presented at the annual meeting of the APSA 2008 Annual Meeting.

Goodman, Steven N. 2005. "Introduction to Bayesian methods I: measuring the strength of evidence." *Clinical Trials* 2:282–290.

Greiner, D. James and Donald B. Rubin. 2011. "Causal Effects of Perceived Immutable Characteristics." *The Review of Economics and Statistics* 93(3):775–785.

Haber, Stephen, Hebert S. Klein, Noel Maurer and Kevin J. Middlebrook. 2008. *Mexico Since 1980*. New York: Cambridge University Press.

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

Jeffreys, Harold. 1961. *Theory of Probability*. 3: Oxford University Press.

Kass, Robert E. 1992. "Bayes Factors in Practice." *Journal of the Royal Statistical Society. Series D (The Statistician)* 42(5):551–560.

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.

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)*.

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*

Lujambio, Alonso. 2000. *El Poder Compartido*. Mexico, D.F.: Editoreal Oceano.

Lujambio, Alonso. 2001. Democratization through Federalism? The National Action Party Strategy, 1939-2000. In *Party Politics and the Struggle for Democracy in Mexico*, ed. Kevin J. Middlebrook. La Jolla: Center for U.S.-Mexican Studies, University of California, San Diego pp. 47–94.

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

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. 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.

Moreno, Alejandro and Yuritzi Mendizabal. May 20, 2010. "Encuesta: Oaxaca." *Reforma* .

Mozaffar, Shaheen and Andreas Schedler. 2002. "THe Comparative Study of Electoral Governance–Introduction." *International Political Science Review* 23(5).

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

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

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'Hagan, Anthony. 1995. "Fractional Bayes Factors for Model Comparison." *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):99–138.

Parzen, Emanuel. July 2003. Quantile Probability and Statistical Data Modeling. Technical report Texas AM University.

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.

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. 2010. *Democracy and the Limits of Self-Government*. New York: Cambridge University Press.

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

Rebolledo, Juan. 2010. "Resisting Democracy: Explaining the Existance and Presistence of Subnational Authoritarian Practices." Unpublished manuscript.

Roberson, Brian. 2006. "The Colonel Blotto game." *Economic Theory* 29:1–24.

Rubin, Donald B. 1986. "Which Ifs Have Causal Answers (Comment of 'Statistics and Causal Inference' by Paul W. Holland." *Journal of the American Statistical Association* 81:377–400.

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

Simpser, Alberto. 2008. "Cheating Big: On the Logic of Electoral Corruption in Developing Countries." Working Paper.

Szwarcberg, Mariela. 2008. "Counting Heads and Votes: Authoritarian and Democratic Strategies of Electoral Mobilization in Argentina." Prepared for delivery at the Comparative Politics Workshop, Yale University.

The New York Times. July 4, 2010. "Once-Dominant Party in Mexico Seems Resurgent.".

Tofias, Michael. 2007. "Of Colonels and Generals: Understanding Asymmetry." Prepared for presentation at the 2007 Midwest Political Science Annual Meeting.

Yo, Keming and Rana A. Moyeed. 2001. "Bayesian quantile regression." *Statistics and Probability Letters* 54:437–447.