RIGHT ON TIME: AN ELECTORAL AUDIT FOR THE PUBLICATION OF VOTE RESULTS

July 11, 2019

Abstract

The publication of electoral results in real time is a common practice in contemporary democracies. However, delays in the reporting of electoral outcomes often stir up skepticism and suspicion in the vote-counting process. This issue urges us to construct a systematic test to distinguish delays attributable to manipulation to those resulting from a limited administrative capacity. This paper proposes a method to assess the potential sorting of the electoral results given the moment at which polling stations publish their vote totals. To do so, we model the time span for a polling station to report its electoral results, to identify those observations whose reported times are poorly explained by the model, and to assess a potential bias in the candidates' vote trends. We illustrate this method by analyzing the 2006 Presidential Election in Mexico, a contest that aroused suspicion from opposition parties and public opinion alike regarding how the electoral results were reported. The results suggest that polling stations' time logs mostly respond to their specific geographic, logistic, and sociodemographic features. Moreover, those observations that took longer than expected to report their returns had no systematic effect on the electoral outcome. The proposed method can be used as an additional post-election audit to help officials and party representatives evaluate the integrity of an election.

1 Introduction

Patience is a virtue scarce in politics and absent on election night. The interlude between the closing of polls and the announcement of results evokes both anxiety and distress, leading candidates and voters to doubt the integrity of the election after any unexpected delay or oddity in the process. In Honduras, for example, the sudden interruption of the results during the 2017 presidential election prompted fraud allegations from the main opposition candidate and his supporters.[1] Likewise, during the 2017 Senate Primary elections in Argentina, the vote margin between the two main candidates in the Buenos Aires district progressively narrowed throughout the election night, fueling suspicion from political parties and voters.[2] Similar cases in Florida, Ecuador, and Perú illustrate the challenges election officials face to announce the results in a prompt and efficient manner.[3] In fact, concerns about undue delays in electoral results are present in about one in every five elections worldwide.[4]

Yet, despite the anger and frustration of protesters, the speed and appearance of the results are not necessarily signs of electoral manipulation. Instead, they often reflect the resources that are readily available to election officials. Protests did little to accelerate the two-week vote count during the 2012 local elections in Arizona, where an unexpected number of provisional ballots surpassed authorities' capacity to verify and tabulate each individual vote.[5] Similarly, despite

[1] *The Economist*, "Honduras's disputed election provokes a crisis," November 30, 2017 (https://www.economist.com/news/americas/21731881-loser-unlikely-accept-result-hondurass-disputed-election-provokes-crisis).

[2] Clarín, "Elecciones PASO 2017: Cristina Kirchner denunciará la "trampa electoral" del Gobierno y apuntará a todos los votos peronistas." August 14, 2017. https://www.clarin.com/politica/elecciones-paso-2017-cristina-kirchner-denunciara-trampa-electoral-gobierno-apuntara-votos-peronistas_0_SJghMNJ_Z.html

[3] For the case of Florida in 2012 see *The New York Times*, "Vote Count Confirms Obama Win in Florida." November 10, 2012. (http://www.nytimes.com/2012/11/11/us/politics/florida-to-address-delays-as-it-confirms-obama-victory.html) and *National Public Radio*, "Four Days Later, Florida Declares For Obama." November 10, 2012. (http://www.npr.org/sections/thetwo-way/2012/11/10/164859656/florida-finishes-counting-obama-wins). For the anxiety during the vote count in the 2016 Peru's presidential election see: *RPP Noticias*. "ONPE: Resultados al 100% estarán listos a más tardar el sábado." June 5, 2016. (http://rpp.pe/politica/elecciones/resultados-al-100-estaran-listos-a-mas-tardar-el-proximo-sabado-onpe-noticia-968684) and *El País*. "El recuento apuntala la victoria por la mínima de Kuczynski en Perú". June 7, 2016. (http://internacional.elpais.com/internacional/2016/06/05/america/1465153235_953203.html). For the account of the vote count during the 2017 presidential election in Ecuador see: *El País*. "El candidato de Correa encabeza el escrutinio con dudas sobre la segunda vuelta". February 20, 2017. (http://internacional.elpais.com/internacional/2017/02/19/america/1487517563_758291.html)

[4] The Expert Survey of Perceptions of Electoral Integrity (Norris, Martínez i Coma and Frank, 2016) shows that 19 percent of the elections analyzed between 2012 and 2016 presented an average response lower than three in a scale from 1 to 5, where 1 is "Strongly disagree" and 5 is "Strongly agree" to the statement: "The results were announced without undue delay."

[5] See *The New York Times,* "Races in Arizona Still Hang in the Balance." November 9,

the riots and violent demonstrations during the 2016 presidential election in Haiti, authorities took nine days to overcome logistical problems after Hurricane Matthew forced them to manually handle all of the country's ballots.[6] Some of the existing evidence for the recent incidents in Honduras and Argentina suggests that the delays and observed trends of the reported results can be explained by the logistics of the respective electoral administration (European External Action Service, 2017; Calvo et al., 2017; Antenucci, Mascioto and Page, 2017). However, if election officials cannot show that those delays do *not* distort electoral outcomes, and voters are unable to verify such claims, then election results are condemned to skepticism and mistrust.

This paper proposes a method to disentangle potential irregularities from mere logistical issues for the time in which polling stations report their results. We illustrate our methodology by analyzing the 2006 presidential election in Mexico, an event associated with a biased management of the vote-counting process. The allegations in this case centered on the Preliminary Electoral Results Program (PREP), a system that publicizes electoral results in real time as soon as polls close. This system is held responsible for sorting the publication of the vote returns in such a way as to favor a specific candidate, giving him a narrow yet consistent lead throughout the count. The concerns regarding how electoral authorities publish the results, plus the availability of data to test these claims, give us the opportunity to illustrate our methodology. Therefore, a substantive contribution of this paper is to evaluate the fraud claims for one of the most contested electoral results in the history of Mexico.

This analysis focuses on a particular type of fraud: the potential sorting of results by electoral authorities at the moment of publicizing the vote count. We test for this possibility in two ways. First, we check for a potential sorting of the announced results to put one of the candidates leading the race early in the count. We do so by comparing the observed time trends with a set of simulated vote counts that randomize the order in which the polling stations were reported by electoral authorities. Second, we model the variance for the timestamps in which the results of the polling stations were reported. After assessing the precision of our estimation, we identify those polling stations whose reporting times systematically differ from the rest of the observations and assess their effects on the vote trends.

Our approach builds upon recent studies that look for outlying observations in the electoral results (Wand et al., 2001; Alvarez and Katz, 2008; Jiménez and Hidalgo, 2014). We distinguish

2012. (http://www.nytimes.com/2012/11/10/us/politics/arizona-races-still-hang-in-the-balance-over-uncounted-votes.html); *Los Angeles Times*, "Arizona ballots finally counted – and Latinos ask, Why so long?" November 21, 2012. (http://articles.latimes.com/2012/nov/21/nation/la-na-nn-arizona-latinos-voting-20121121); and *Tucson Sentinel* "Why is Arizona still counting votes?" November 21, 2012. (http://www.tucsonsentinel.com/local/report/112012_az_vote_count/why-arizona-still-counting-votes/).

[6] *BBC*, "Haiti starts counting votes in long-delayed election." November 21, 2016. (http://www.bbc.com/news/world-latin-america-38042585); *Reuters*, "Haiti police clash with demonstrators ahead of election results." November 22, 2016. (https://www.reuters.com/article/us-haiti-election/haiti-police-clash-with-demonstrators-ahead-of-election-results-idUSKBN13I05K).

from these works by identifying outliers by looking not at their vote totals, but rather at the timestamps for when the results were reported. This is an important departure because we can assess whether potential oddities in the electoral results correlate with a time delay in the vote-counting process for a particular polling station. This post-election audit can also be a useful tool for election officials and party representatives to evaluate concerns about the order in which the results were sorted by identifying the geographic location and operational ways in which anomalies occur (Alvarez, Atkenson and Hall, 2013; Mebane, 2015).

We proceed as follows. Section 2 briefly discusses the 2006 presidential election in Mexico and the concerns about how the results were reported. Section 3 assesses the likelihood of finding the observed vote trends in such tight elections. Section 4 models the timestamps for the polling stations, identifies the potential outlying observations, searches for control units, and assesses the existence of a potential bias in the results. Finally, Section 5 discusses the implications of this work when it comes to other elections.

2 Contextual Background: Mexico, 2006

The July 6, 2006 election remains the closest presidential race in Mexican history. Pre-electoral polls anticipated a technical tie between the National Action Party's (PAN) Felipe Calderón and the Democratic Revolutionary Party's (PRD) Andres Manuel López Obrador. The closeness of the race prompted both Calderón and López Obrador to make condemning statements against each other and to publicly anticipate irregularities that could occur during election day (Bruhn and Greene, 2007; Lawson, 2007). Despite this adverse environment, candidates and voters expected the Federal Electoral Institute (IFE), the national body in charge of administering federal elections, to announce the results on election night.[7]

A few hours after the polls closed, however, the IFE's president announced that the preliminary vote count was "too close to call," and pleaded with candidates to abstain from announcing any result. Calderón and López Obrador ignored the petition, proclaiming themselves winners and polarizing citizens' perceptions of the integrity of the election. The preliminary results gave Calderón a 1.04% lead over López Obrador, a margin that shrank to 0.56% in the official vote count – a difference of about 250,000 votes. López Obrador refused to accept the results, alleging several irregularities in the electoral process. These fraud claims, together with Calderón's resistance to a ballot recount, produced the most critical post-electoral crisis in the democratic era of the country (Tello, 2007; Tuckman, 2012).

Most of the fraud allegations centered on the vote trends published by the Preliminary Electoral Results Program (PREP), a system designed to provide the real-time results of all polling booths in the country.[8] The PREP collects information from the country's 300 districts, where elec-

[7] After 2013, the Institute changed its name to the National Electoral Institute (INE).

[8] This system is similar to those in other countries with manual scrutiny and a system for results transmission, such as Chile, Colombia, and Costa Rica. For more information

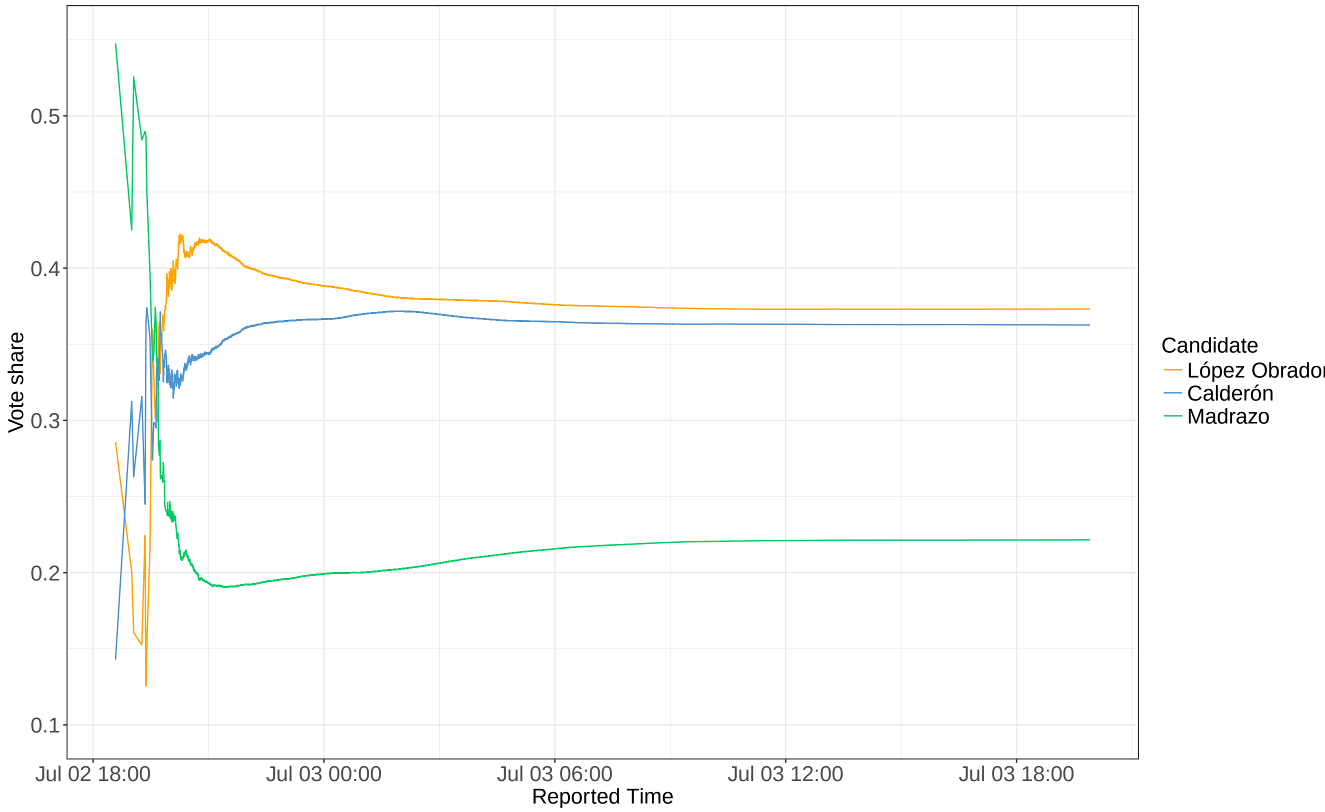


Figure 1: Presidential candidates' cumulative vote shares at each time the results of the polling stations were reported. Mexico, 2006.

tion officials upload the results of every polling station as soon as poll workers count the ballots, record the results on a tally sheet, and deliver the electoral materials to their respective election district offices. Therefore, the sequence through which the PREP publishes the results should denote the time it took poll workers to count the votes and commute to the district office, as well as the capacity of IFE officials to receive the electoral materials and upload the results.

Figure 1 depicts the progression of the results from the 117,287 tallies reported by the PREP in 2006. The system started publicizing the results of the first vote tallies at 6 PM CST, the closing time for most of the polling stations. While they cross each other early in the vote count, the candidates' vote shares level off after only 26 polling stations reported, showing a narrow and consistent advantage to Calderón throughout the rest of the count. Overall, the average difference between Calderón and López Obrador was, after midnight on July 3, 1.3% and it never got narrower than 0.8%.

The trends shown during the preliminary vote count were interpreted by López Obrador and his supporters as evidence of the PREP's bias in favor of the PAN's candidate (Eisenstadt, 2007).

see http://www2.ine.mx/archivos3/portal/historico/contenido/Centro_internacional_capacitacion_investigacion_electoral/idiomas/img/Use_Technology_2014.pdf.

In the words of López Obrador, "[The IFE] must explain why the first 25% of the reported vote tallies gave Calderón an advantage of 10 points, which then started to shorten until 2:30 am to increase then again in an inexplicable way."[9] A group of scholars and political observers backed up this skepticism and suggested that the PREP used an algorithm to sort the results and both give Calderón an early lead and to deflate the votes for López Obrador (de Icaza-Herrera, 2006; Rochín, 2006; Mochán, 2006; López Gallardo, 2009).[10]

Defending the integrity of the process, election officials refuted deliberate bias within the PREP. They vindicated the observed vote trends as by-products of the polling stations' proximity to the 300 district councils in the country, mostly located in cities and sizable towns. Efforts to debunk bias in the PREP were backed by a different group of scholars, who showed that the reporting times correlated with the distances between the precincts and the district councils (Hernández, Vallejo and Vergara, 2006; Pliego Carrasco, 2007), making urban areas to report their results, on average, two and a half hours faster than those from the countryside (Aparicio, 2006, 2009).

Promoters and detractors of the impartiality in the preliminary vote count did not find common ground, and the discussion about the PREP remains ongoing. Concerns about the PREP resurfaced six years later, when a few opposition parties were wary of the publicized results of the presidential election.[11] The skepticism also prevailed during the 2016 and 2017 local elections, wherein gubernatorial candidates continuously disregarded the preliminary results and proclaimed themselves winners of the election.[12]

In summary, the concerns about the partiality of the 2006 preliminary vote count are based on two arguments. First, the narrow margin between the top two candidates seems at odds with the fleeting volatility of their vote shares during most of the overall vote count. Second, Calderón's early, steady lead suggests the possibility that electoral authorities sorted the results to immediately report positive results for the PAN's candidate while deflating López Obrador's vote returns for those tallies reported later on election night.

Unfortunately, the existent approaches to empirically support or reject these concerns remain inconclusive. On the one hand, the conjecture that authorities sorted the results assumes that the publication of the vote returns should follow a random generating process in which every polling station has the same probability of being reported at a given moment. This assumption, however,

[9] "AMLO: el PREP, manipulado y con infinidad de inconsistencias," *La Jornada*, July 4, 2006. p.2

[10] See also Villoro, Juan "La división," *Reforma*, July 7, 2006, and Hernández, Julio. "Astillero," *La Jornada*, July 7, 2006.

[11] See, for example, Sonnleitner, Willibald, "¿Inconsistencias o irregularidades? Los votos y la ley," *Letras Libres*, August 2012. (http://www.letraslibres.com/revista/dossier/inconsistencias-o-irregularidades).

[12] *La Jornada*, "Fallan los PREP; todos los partidos se dicen ganadores," June 6, 2016 (http://www.jornada.unam.mx/ultimas/2016/06/06/fallan-los-prep-todos-los-partidos-se-dicen-ganadores.1); *The Huffington Post* "Las inconsistencias del PREP en el Edomex que dejan dudas en 241 mil votos a favor de Del Mazo," June 7, 2017 (http://www.huffingtonpost.com.mx/2017/06/07/las-inconsistencias-del-prep-en-el-edomex-que-dejan-dudas-en-241_a_22129855/).

ignores the existence of any contextual variables that may affect the time for the results to be publicized, such as the resources in each polling station to count the ballots and fill out the tallies or the proximity of the polling stations to their district councils. On the other hand, rebuttals to the allegations of fraud highlight the correlation between a polling station's timestamp and its location in either a rural or urban area. However, under the assumption of careful manipulation, this relationship overlooks any potential strategic behavior of election officials, in which case they could modify a few observations without disrupting the general trend.

Our analysis assesses the claims about the sorting of the electoral results in two ways. We first evaluate the feasibility of observing such a small number of lead changes given the tight margin between the two candidates' vote shares. Our second approach identifies those polling stations whose results were reported either earlier or later than expected, given the stations' contextual characteristics. We model the expected time at which each polling station should have publicized its results, identified the potential outliers, and evaluated their vote returns.

Before presenting our analysis, we must emphasize that our methodology focuses on how electoral authorities reported the vote tallies. Since most of the claims made during the 2006 Mexican election focused on the method by which election officials uploaded the data across the 300 district councils, the analyses below leave aside any other irregularity that could occur at the polling station, such as ballot stuffing or vote buying. Our findings complement existing research into allegations of irregularities at other stages of the election (Poiré and Estrada, 2006; Crespo, 2006; Aparicio, 2009).

3 Sorting the Results

The first argument in favor of the suspicion of the reported results is that, despite the narrow vote margin between the top two candidates, their vote shares intersect just three times in the entire vote count, with the last crossing occurring after only 26 tallies had been reported. This complaint sustains that the candidates' vote trends in a hard-fought election would have to cross multiple times along the count. In the words of López Obrador himself, "[I]f they had counted the votes as they arrived, at some point the lines would have had to cross over."[13]

The assumption that such a narrow vote margin should produce multiple crossings between the candidates' vote shares rests on two premises: (1) vote shares are very similar across polling stations, and (2) all polling stations have the same probability of being reported at any given time. While Section 4 evaluates both premises in the specific context of the 2006 election, we first examine whether the observed trends displayed in Figure 1 can be reproduced while holding both premises. Our goal, then, is to see how these two assumptions may diverge from what is expected by the law of the large numbers, in which a sequence of random variables converges to their respective probability as the number of trials increase (DeGroot, 1984, p. 229-231). In this case, the

[13] López Obrador, Andrés Manuel. *0.56%*. DVD. Directed by Lorenzo Hagerman. Mexico City: IMCINE, 2010.

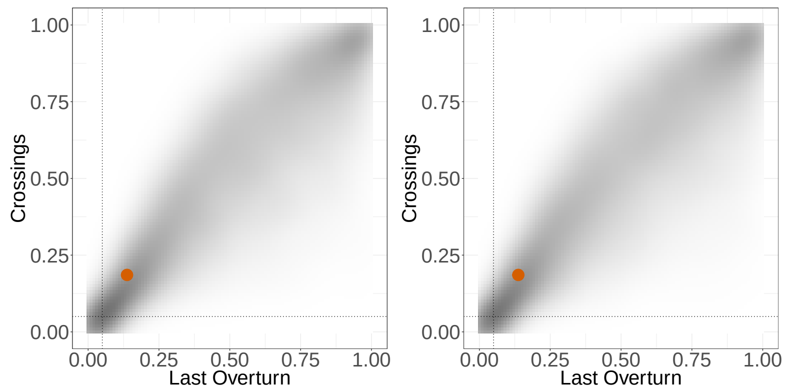
progressive count of the more than 110,000 tallies should gradually approach the candidates' final vote shares, decreasing the probability of an intersection between the vote trends later in the vote count.

We use both premises to simulate the intersection of the candidates' vote shares in two ways. Our first test uses a set of 10,000 simulated elections between two candidates and model their electoral returns based on Calderón and López Obrador's vote distributions. Each iteration considers a contest in 117,287 polling stations, the number of tallies reported by the PREP, while the vote returns in each of them represent a random draw from a gamma distribution, $Gamma(\alpha, \theta)$. The properties of the gamma distribution allows us to model the votes in a given polling station as a positive, non-bounded number. Moreover, it allows us to simulate the vote counts assuming independence across observations, which implies that the amount of votes in a polling station does not affect the amount of votes in other polling stations (DeGroot, 1984, p. 288-289). We use the shape (α) and scale (θ) parameters for each candidate by estimating $E[X] = \alpha\theta$ and $Var(X) = \alpha\theta^2$ given the moments of the empirical distributions for Calderón ($\bar{x} = 119.435$ and $\sigma = 72.707$) and López Obrador ($\bar{x} = 116.069$ and $\sigma = 71.853$). In this case, our resulting parameters are $\alpha_{Calderon} = 2.698$ and $\theta_{Calderon} = 44.261$ for candidate A and $\alpha_{LopezObrador} = 2.609$ and $\theta_{LopezObrador} = 44.481$ for candidate B. The second test performs 10,000 permutations for the sequence of the actual results reported by the PREP. The goal of each test is to randomize the order in which the tallies are aggregated to the final outcome and to trace the frequencies of *crossings*, or the times in which the vote shares of the candidates intersect each other, and *last overturn*, the number of reported tallies before the leading candidate gets an unsurpassable vote share.

Both exercises demonstrate that any volatility in the candidates' vote trends is likely to vanish very early in the race. In nine out of every ten simulations, the leading candidate takes the final lead of the vote count with less than 2% of the tallies counted.[14] Similarly, the 80% confidence intervals for the number of *crossings* are $[1, 25]$ for the simulated elections and $[0, 30]$ for the permutations of the PREP results. As Figures A and B in the Appendix show, the most frequent values for both parameters are lower than what is observed in the 2006 election. In fact, in about 10% of the iterations for both tests, a candidate uninterruptedly lead the race since the first reported polling station.

We summarize the results of both tests in Figure 2. The plots show the p-values of *crossings* and *last overturn* for each iteration of the simulated vote counts and the permuted PREP results. Each p-value represents the proportion of iterations with similar or lower value in the entire set. To facilitate their visualization, we normalize the parameter values by estimating their one-tailed p-values, or the probability of observing a similar or lower value in the entire set of iterations. Observations to the left or below the dashed lines identify those iterations with at least one of its parameters with a value lower than at least 95% of all observations. The empirical p-values for the 2006 election are represented by the solid dot, suggesting that its three turnovers and the definite lead at the 26th polling station is far from infrequent. In fact, 13% of the iterations in both

[14] The value of the 90 percentile for *last overturn* is 2448 in both approaches



(a) Simulated vote counts (b) Permutations of the PREP results

Figure 2: Heat map plots for the p-values of *crossings* and *last overturn*

approaches presented less than three crossings, and 18% of them had its last overturn within the first 25 polling stations in the count.

In sum, despite the fleeting volatility of the vote trends, both the number of intersections between the candidates' vote shares and the PAN candidate's early lead are not as unexpected as critics claim. The section below goes a step farther and assesses whether these results were sorted in a manner to give one of the candidates the early lead.

4 Identifying Out-of-Sync Timestamps

The previous section demonstrates that the vote shares in the 2006 Mexican election follow an expected trend even when we assume that all observations have the same probability to be reported at a given time. This finding goes against the claim of a potential sorting of the results. We now verify the possibility that the vote trends were clinically altered by a subset of polling stations to either inflate the early vote returns towards the winning candidate or to delay those observations favoring the runner-up.

We do so by evaluating the timestamps of the polling stations in three steps. We first model the duration for a vote tally to be reported in the system. After checking for the robustness of our estimations, we identify those polling stations whose timestamps significantly depart from the rest of the observations. Having identified those outliers, we evaluate how the vote trends would look by either excluding these observations from the analysis or imputing them with different timestamps.

4.1 Duration model

We estimate the time it took electoral authorities to publicize the results of the polling stations by using a Cox proportional hazard model. Our dependent variable is the duration, in minutes, for the results of a given polling station to be published, starting at 6:00 PM, CST, on July 2, 2006, the time at which polls closed and the ballot-counting started. The average polling station duration is about six and a half hours ($Mean = 401$ minutes, $SD = 187$ minutes). Our goal is to explore the time variance after considering the sociodemographic and logistic characteristics of the polling stations.

We account first for the distance between the polling stations and the district councils. *DrivingTime* measures the driving time between each polling station and its respective district council. To build this variable, we matched the geographic coordinates of the polling stations and the 300 district councils, using the Python module `ArcPy` to estimate the driving times considering path distance, average route traffic, and road conditions.[15] Our expectation is that the logged times are positively correlated with the travel time for poll workers to deliver the results to the electoral authorities.

We also consider the idiosyncratic characteristics of the precincts that may affect the speed of the vote count. Based on the existent literature on polling work (Hall, Monson and Patterson, 2008; Alvarez, Atkenson and Hall, 2013; Burden and Milyo, 2015), we expect that the efficiency of the vote count depends on the poll workers' ability to count ballots, sum the vote totals, and fill out the vote tallies. We proxy for these skills in two ways. First, we account for the educational, economic, and demographic characteristics of the precincts. This battery includes rates for *illiteracy*, *high school* completion, and adults who are at least *65 years old*, as well as the household shares of access to *electricity* and *computers*. Since poll workers in Mexico are randomly selected from all the registered voters in the precinct, we use these variables as proxies for the characteristics of the group of poll workers in the precinct on Election Day.[16] All of these data come from the 2005 national census conducted by the National Institute of Statistics and Geography (INEGI).[17]

Second, we include the ratio of poll workers in the polling station who did not receive adequate training (*no training*). The electoral code prescribes that those citizens selected as poll workers attend a training session in the weeks leading to the election. However, if at least three selected

[15]The coordinates for the polling stations were obtained by an information request to the IFE. The coordinates of the district councils were hand coded based on the information available at: `http://www2.ine.mx/archivos3/portal/historico/contenido/informacion_juntas_locales_y_distritales/`.

[16]The Mexican electoral organization required a lottery system for selecting and training poll workers in every precinct (Eisenstadt, 2004, p. 46). Drafted citizens are contacted for receiving the training, and the IFE selects those most qualified to serve as poll workers among those who accepted the invitation and attended the training.

[17]`http://www.ine.mx/archivos3/portal/historico/contenido/interiores/Detalle_geografia_electoral_y_cartografia_transparencia-id-04a9d8bd4ac04210VgnVCM1000000c68000aRCRD/`

	N	Mean	S.D.	Min	Max
Duration time	110,139	401.60	186.94	60.28	1554.00
Log Driving time	110,139	3.39	3.96	-4.64	6.77
65 years old population	110,139	0.06	0.04	0	1
Illiteracy share	110,139	0.08	0.06	0	1
High School share	110,139	0.11	0.05	0	0.75
Computer share	110,139	0.20	0.17	0	1
No Spanish share	110,139	0.003	0.03	0	0.62
Registered Voters	110,139	0.76	0.15	0.07	1
Agents PAN	110,139	0.81	0.39	0	1
Agents PRD	110,139	0.799	0.40	0	1
Agents PRI	110,139	0.887	0.317	0	1
Absent Poll Workers	110,139	0.18	0.2	0	1

Table 1: Summary Statistics

officials do not show up at the polling station on election day, then they are replaced by the first citizens in line to vote. These extemporized poll workers have not received adequate training to be a poll worker, so we expect that their performance may affect the overall speed of the vote count of the polling station.

Finally, we include two logistical issues that may also decelerate the vote count. We expect slower vote counts in polling stations with a greater number of voters. Therefore, we include the proportion of *registered voters* relative to the maximum number of registered voters (750) allowed in each of the polling stations. We also consider the presence of party agents in the polling station. The task of these representatives is to witness the vote count process and prevent any irregularity. At the same time, the presence of these party agents may delay the process, since poll workers pay more attention to the the vote-counting process.

Table 1 displays the summary statistics of our data. With the exception of driving time, which is log-transformed, all our dependent variables are rescaled to represent the share of the population in the precinct with a given attribute. See that, for example, the mean proportion of non-Spanish speakers in a polling station is 0.3%.

The model specification we employ is

$$h_{ij}(t)|(X_{ij}\beta, v_j) = h_0(t)v_j e^{(X_{ij}\beta)}; \quad E(v_j) = 1, \text{ and } var(v_j) = \theta \tag{1}$$

where the hazard rate of publishing the results for polling station i in district j at time t is a function of a common baseline hazard ratio, $h_0(t)$, and the slopes, β, of covariates X_{ij}. To account for unobserved heterogeneity across districts, such as time zone, unexpected traffic conditions, or the efficiency of election officials in the district councils, we include the vector $v_j = e^{(w_j\theta')}$, where w_j are the district-level frailties with variance θ (Box-Steffensmeier and Jones, 2004). Note that as θ converges to 0, v_j converges to 1. Thus, district-level heterogeneity approaches null, resulting in a model that simplifies to the standard Cox proportional hazards model.

Variable	Coef.
Absent Poll Workers	-0.061^{**}
	(0.020)
Registered Voters	0.0075
	(0.0243)
log(Driving Distance)	-0.175^{***}
	(0.0043)
PAN Observers	14.7544^{***}
	(0.1362)
PRD Observers	8.3793^{***}
	(0.1261)
PRI Observers	23.9628^{***}
	(0.1838)
65-Year-Olds	114.6258^{***}
	(1.2901)
High School	207.6447^{***}
	(1.3329)
Computer	44.1353^{***}
	(0.4896)
No Spanish	-99.8115^{***}
	(3.1297)
Illiteracy	227.0505^{***}
	(1.0154)
PAN Observers * log(Minutes)	-2.4492^{***}
	(0.0223)
PRD Observers * log(Minutes)	-1.3835^{***}
	(0.0208)
PRI Observers * log(Minutes)	-3.9264^{***}
	(0.0294)
65-Year-Olds * log(Minutes)	-19.3735^{***}
	(0.2218)
High School * log(Minutes)	-35.0485^{***}
	(0.2318)
Computer * log(Minutes)	-7.4468^{***}
	(0.0854)
No Spanish * log(Minutes)	16.0395^{***}
	(0.5039)
Illiteracy * log(Minutes)	-37.7975^{***}
	(0.1715)
θ	1.024
l-likelihood	-912221.1

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Cox model on the reporting time for the vote counts in every polling station. Mexico, 2006.

Model models nonlinear effects by using splines for *Absent Poll Workers*, *Registered Voters*, and *Driving Time*, and corrects for violations of proportionality assumption by adding time interactions for the violating covariates.

We first check for violations of the proportionality assumption by obtaining the model's Schoenfeld residuals. These residuals represent the difference between the observed and expected values of the covariates at different times (Box-Steffensmeier and Jones, 2004, p. 121). We then use a Therneau-Grambsch non-proportionality test to check for any correlation between the residuals and the reporting times (Grambsch and Therneau, 1994; Box-Steffensmeier and Jones, 2004; Keele, 2010).[18] The results of this diagnostic test suggest that, with the exception of those variables capturing the share of the population over 65 years of age and with a high school graduation, all of the covariates correlate with the polling stations' reporting times.

A correlation uncovered by the Therneau-Grambsch test may indicate an incorrectly specified model, rather than a violation of the proportionality assumption (Keele, 2010). To distinguish both issues, we use splines to test for nonlinear effects of the specified covariates. These splines fit numerous, piecewise polynomials to the nonlinear effect of each covariate, resulting in a more accurate modeling (Schoenberg, 1946; Keele, 2010). A Wald test comparing the splined model to the baseline model shows that the nonlinear functional form fits the data better. A second check of the Therneau-Grambsch test reveals that the Schoenfeld residuals for the proportion of registered voters, the proportion of absent poll workers, and the driving distance are no longer significantly correlated with reporting times, suggesting that these variables do not violate the proportionality assumption.

We correct for the non-proportionality of the remaining variables by interacting them with our function of time (Box-Steffensmeier and Jones, 2004; Keele, 2010).[19] The parameter estimates for our corrected model are shown in Table 2.

Ordinarily, we can interpret the effect of each variable by estimating its hazard ratio, $\exp(\beta)$ for positive coefficients or $(\exp(\beta))^{-1}$ for negative coefficients. However, given how time interacts with many of our variables, we must account for the coefficients of the interaction and its constituent variables.[20] Since variables that use spline corrections are not easy to interpret, we discuss them in greater detail below.

We begin by calculating the hazard ratios for the time-interacted variables. In the case of binary variables, we can simply set each $x_i = 1$ to compare the change in the hazard ratio of that variable relative to a baseline model without that variable. For example, we calculate the hazard ratio of the presence of PAN observers (relative to no PAN observers) at the average value of the logged reporting time (5.883) as follows: $\exp(14.7544 - 2.446 * 5.883) = \exp(0.3646) = 1.40$. In other

[18] We use the *cox.zph* function in the `Survival` package in `R` (Therneau, 2015). See Table F in Appendix for the test's output.

[19] Following Jin and Boehmke (2016), we verified that the time variable used in our interactions is measured using the same units as the time variable used to determine our model's duration time (i.e. both the interaction variable and the variable for duration time are measured in minutes).

[20] In other words, if β_1 represents the coefficient of a variable, x_i, and β_2 represents the coefficient of the interaction of that same variable with a function of time (in our case, logged minutes), to find the hazard ratio we calculate $\exp(x_i(\beta_1 + \beta_2 * min))$, where *min* is the time (in the reporting process) at which we wish to calculate the hazard ratio.

words, at the mean reporting time in the sample, having a PAN observer produces a hazard ratio of just 1.40. We repeat this procedure by using the minimum and mean values of reporting time for each of the binary time-interacted variables reporting the results in Table A in the Appendix.

To facilitate interpretation, we illustrate the estimated hazard rations of each variable in Figure 3.[21] Each plot takes as the baseline the minimum value of x_i. Figures 3(d) to 3(f) show the changes of the hazard ratio for the binary covariates over the entire reporting period (as opposed to just calculating the hazard ratio at the reporting time's minimum and mean values). Meanwhile, Figures 3(g) to 3(k) display changes to the hazard ratio over the reporting period, when each time-interacted covariate is set to its first quartile, median, mean, and third quartile values, respectively.[22] As expected, the hazard ratios indicate that each characteristic initially delays the reporting of results, but the effect diminishes over time. For example, in Figure 3(g) we see that the greater the proportion of 65-year-olds, the greater the hazard ratio; however, at about the 250-minute mark, the hazard ratio for all levels of 65-year-olds converges to 1 (i.e. no effect). A similar pattern emerges in Figures 3(h), 3(i), and 3(k).

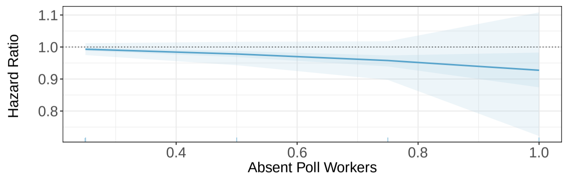
The only exception in the analysis is for the estimation of non-Spanish speakers (Figure 3(j)). The positive slope in this figure suggests that the positive correlation between the share of non-Spanish speakers in the precinct and the delay in the publication of the results increases as the time passes. One possible explanation for this result lies in the high number of observations that are equal to zero (11% of total observations) or that take on very small values (only roughly 1% of observations are greater than 0.05).

Due to the nature of splined variables, their coefficients are not as easily interpreted, nor are their hazard ratios easily calculated (Gandrud, 2015). We therefore show the hazard ratio of each splined variable—*Absent Poll Workers*, *Registered Voters*, and *Driving Distances*—in Figures 3(a), 3(b), and 3(c), respectively. The x-axis of each figure shows the range of values for each covariate, with their respective hazard-ratios plotted on the y-axis.[23] In the cases of absent poll workers and driving distance, higher values of each variable results in increasing delays in the reporting time, though initially neither variable has an effect on reporting times (as shown by the confidence intervals intersecting 1). Oddly, however, as the ratio of registered voters increases, the reporting time initially decreases, eventually leveling off at around the 1.25 mark. One possible explanation for this result may be its highly left-skewed distribution, where only 4% of observations take on a value of lower than 0.50.

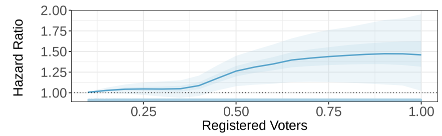
[21]Figures were generated using the *simGG* function in the `simPH` package in `R` (Gandrud, 2015). For each (non-splined) figure, only the middle 95% probability interval of simulations were kept (see Gandrud (2015) for a more in-depth discussion).

[22]With the exception of non-Spanish speakers, whose first, second, and third quartiles are less than 0.05, which we combine with the minimum value of zero for easier interpretation.

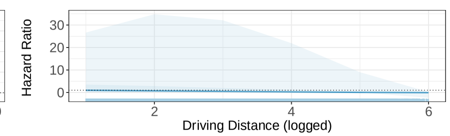
[23]We use cubic smoothing splines to produce a continuous line. The wider interval represents the minimum and maximum simulated hazard ratios of each value, and the narrower, darker interval represents the middle 50% of the simulated values.



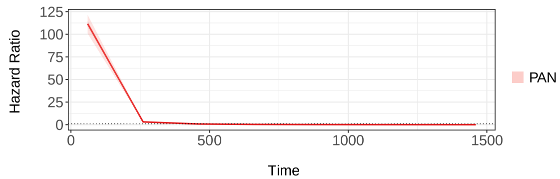
(a) Absent Poll Workers



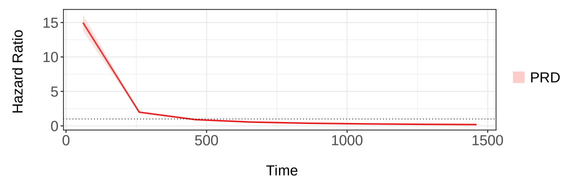
(b) Registered Voters



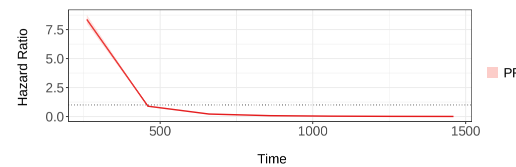
(c) Driving Distance



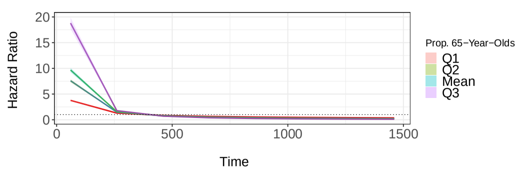
(d) PAN Observers



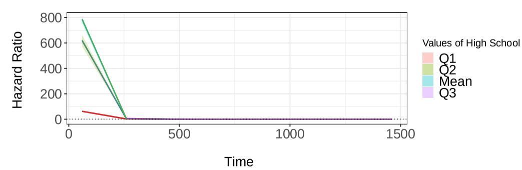
(e) PRD Observers



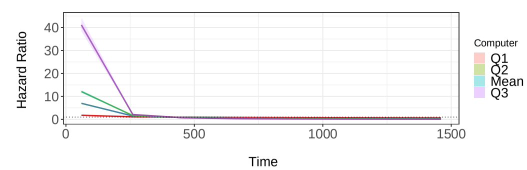
(f) PRI Observers



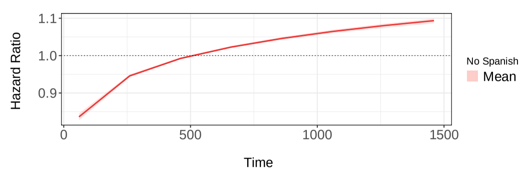
(g) Proportion 65-Year-Olds



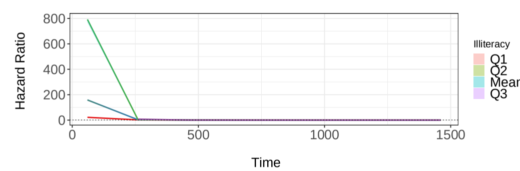
(h) High School



(i) Computer



(j) No Spanish



(k) Illiteracy

Figure 3: Relative Hazard Plots for Model Coefficients, Table 2

Note: Each figure above displays the hazard ratio of each covariate on the outcome. The dotted horizontal line at 1 represents no effect. For the splined variables (Figures 3(a) to 3(c)), the x-axis represents different values of the variable, the wider interval represents the minimum and maximum simulated hazard ratios of those values, and the narrower, darker interval represents the middle 50% of the simulated values. The remaining figures display the time (in minutes) on the x-axis, and shows the simulated hazard ratios for different values of each covariate relative to its minimum value. Figures were generated using the `simGG` function in the `simPH` package in `R` (Gandrud, 2015)

4.2 Identification of outlying units

After modeling the time duration for each vote tally, we use diagnostic methods to identify those units with reporting times so poorly explained by our model that they may arouse suspicions about the mechanism through which they were generated (Hawkins, 1980). For this goal, we estimate the Martingale residuals, or the difference between the observed duration time for each polling station and its conditionally expected time given the fitted model (Therneau and Grambsch, 2000, p. 80-82). The Martingale residual ($M_i(t)$) for each observation i and time t is denoted by:

$$M_i(t) = \delta_{ij}(t) - h_{ij}(t|X_{ij}\beta, v_j); \quad E(M_i) = 0, \text{ and } cov(M_i, M_j) = 0 \tag{2}$$

where $\delta_{ij}(t)$ and $h_{ij}(t|X_{ij}\beta, v_j)$ are the observed and predicted duration times for observation i in district j, respectively. These residuals are non-symmetrically distributed over the $(-\infty, 1)$ domain. To facilitate the identification of outlying observations, we transform the residuals into a normal-shaped distribution centered on 0 (Box-Steffensmeier and Jones, 2004, p. 123). The resulting transformation produces the deviance residuals (D_i) (Therneau and Grambsch, 2000, p. 83-84) and are denoted by:

$$D_i = sign(M_i(t))\sqrt{-2[M_i(t) + \delta log(\delta_i - M_i(t))]} \tag{3}$$

Figure 4 shows the deviance residual of each observation against its reported time. The plot reveals a fairly symmetrical distribution of the residuals around 0, and the flatness of the loess line suggests that the time-corrected model is correctly specified.

We identify as outliers those observations with deviance residuals outside the $[-2, 2]$ interval (Christensen, 1997).[24] In other words, observations with larger deviance residuals than 2 denote those polling stations with a reporting time significantly shorter than what estimated by the model. Similarly, the "falling snowflakes" below the -2 horizontal line represent those reporting times underestimated by our model. As Figure 4 shows, these outliers are spread along the horizontal axis, suggesting that their unexplained time component is not an issue exclusive to those results reported almost a day after polls closed.

The analysis found 5239 outliers, representing less than 5% of the polling stations in the election. This set of observations is comprised by 4263 and 976 polling stations whose reporting times were overestimated and underestimated by the model, respectively. As Figure D in the Appendix shows, these observations are distributed across the country and did not follow a geographic pattern.

A concern of using diagnostic plots to identify outliers is the potential bias that these observations may have on the estimates of our survival model. In other words, since we assume that all observations are initially derived from the same data-generating process, if the estimations of our

[24] The Appendix shows that the results hold when we take a less conservative approach that identifies outlying observations as those falling outside the $[-3, 3]$ range.

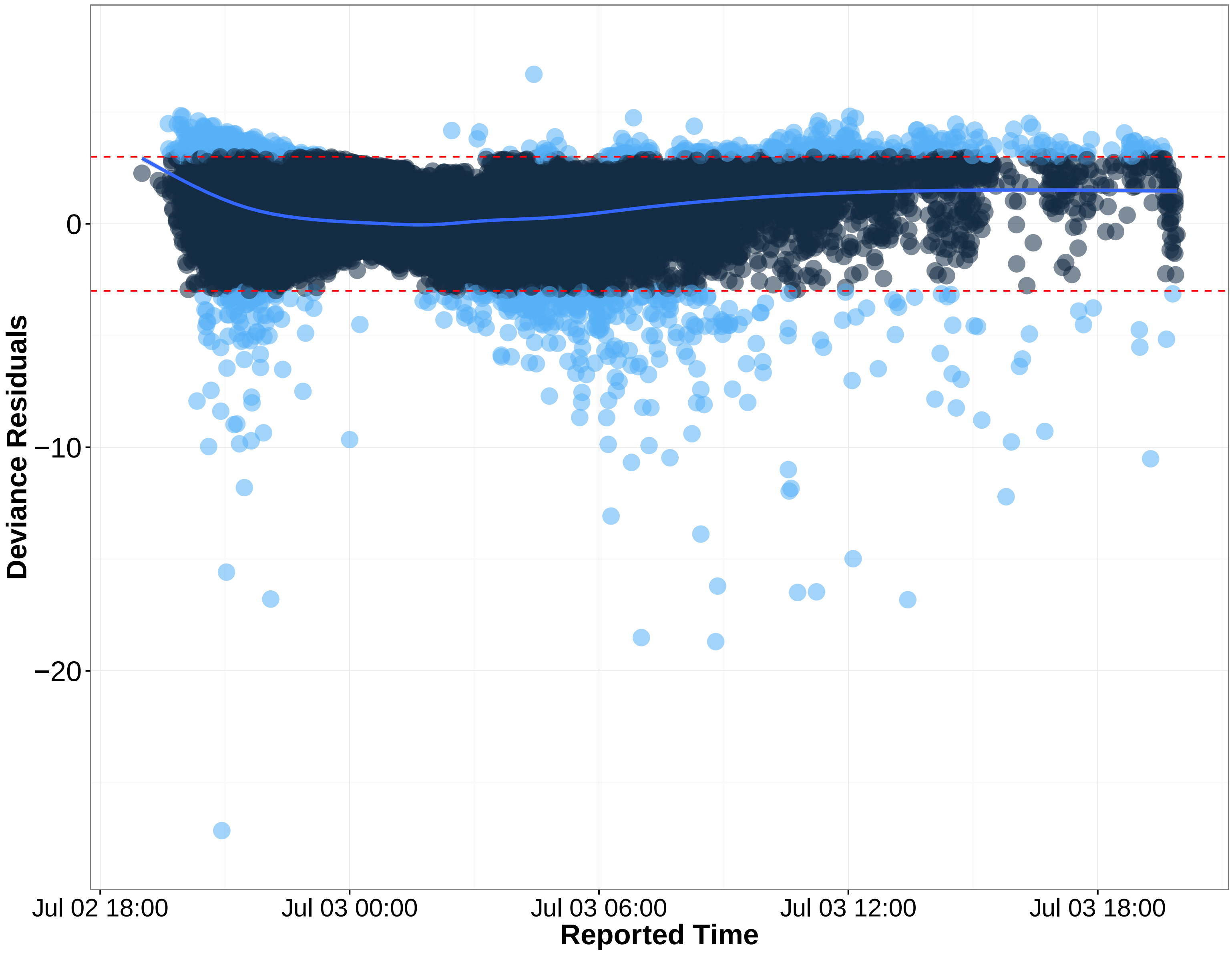


Figure 4: Deviance residuals and time in which each observation was reported by the PREP. Mexico, 2006.
Notes: The scatter plot show the relationship between the deviance residual of each observation and its correspondent time that reported the results. We denote outlying observations in light blue and define them as those with a deviance residual outside the $[-2, 2]$ range. The plot displays the results for the model that corrects for non-proportionality by including splines for the number of registered voters, the proportion of absent poll workers, and the driving distance. Each of the remaining variables is interacted with the duration time. Locally-weighted regression lines are shown in blue.

model are sensitive to outliers, there would be a disproportionate rate of false negatives or masked observations that should be declared outliers. We therefore assess bias in our model estimations in three ways. First, we rerun the model, excluding *all* of the observations identified as outliers and compare the results with those reported in Table 2. As the Appendix shows (Table H), the estimates are almost identical to our original model, providing evidence that the outlying polling stations do not bias our estimates. Second, we also compare our original model with the resultant estimates of bootstrapping the outlier observations. As Figure W in the Appendix shows, the estimates derived from our original model perform fairly well compared with the bootstrapped estimates. Finally, Section 6.5 in the Appendix shows that the results of our analysis hold only when considering those outlying units also identified by a *k-means* cluster analysis.

4.3 Searching for Control Units

After identifying the outlying observations, we look for units with equivalent characteristics but with a timestamp accurately estimated by the model.

We look for control units using two alternative procedures. First, we exploit a particularity of the Mexican electoral code, which assigns voters of a given precinct to polling stations according to a voter's surname. In this case, surnames are a poor predictor of voter behavior in Mexico, so sorting precinct voters alphabetically is orthogonal to electoral outcomes (Cantú, 2014). Moreover, since the polling stations of a given precinct often share the same location (Larreguy, Marshall and Querubin, 2016), this procedure compares observations with the same driving time to the district offices.

Comparing adjacent polling stations ensures that outlying observations share characteristics similar to their potential comparison units. However, this method suffers from two shortcomings. First, we can compare only those outlying observations in which not all the units in the precincts are identified as outliers. For the 2006 election, this method considers the vote shares of 4156 out of the 5239 outlying observations in the sample. Second, and crucial to extend this approach beyond the Mexican case, the comparison of adjacent polling stations can be applied only to elections with a quasi-random assignment of voters, restricting the application of the proposed method to just a handful of cases.[25]

To address these issues, we look for potential control units by using a sparse optimal matching algorithm (Keele, Pimentel and Yoon, 2015). This method simultaneously matches treated units to as many control units as possible while ensuring a fine balance between and among groups, allowing for an exact balance among nominal covariates and the shortest covariance distance between treated units and control ones (Pimentel, 2016). Using this approach, we take as potential control units those polling stations with a deviance residual inside the $[-2, 2]$ range, including

[25]For example, Argentina, where precinct voters are also assigned to polling stations depending on voter's surname (Casas, Diaz and Trindade, 2013). Similarly, voters in Venezuela are assigned to polling station depending on the last two-digits of their voter's ID number (Hausmann and Rigobón, 2011).

only those polling stations sharing very similar characteristics with the outliers.

The matching procedure works as follows. We first limit the potential matches to observations within the electoral district of each treated unit. We then estimate pairwise Euclidean distances between treated and potential control units, ruling out those matches with a difference in the propensity scores larger than a fifth of a standard deviation of the estimated propensity score using the full sample. Next, we coarsen our continuous covariates into categorical variables, allowing us to refine the covariance balance into two levels. Finally, we define the the final set of control units by adopting fine-balance constraints across our set of covariates.[26] This approach leaves us with 5133 treated and matched units. Table B in the Appendix shows that both approaches significantly improve the balance over the unmatched comparison.

4.4 Evaluating the effects

The last step in the analysis looks for anomalies of the vote results after considering the times in which they were reported. Our goal is to evaluate the evidence in light of the alleged *sorting* of electoral results in a fashion that gave the PAN's candidate an early lead. It is worth reiterating that our goal is *not* to assess whether there was a distortion of the vote totals. Existing analyses on the 2006 election find no evidence of manipulation in the vote counts, and the errors on the tallies show no consistent bias toward any of the candidates (Crespo, 2006; Poiré and Estrada, 2006; Aparicio, 2009). Confirming this observation, the Appendix presents the results of an exercise comparing the vote shares of the outlying observations with their control units.

We search for any slant in the results publicized by the electoral authorities by comparing the vote trends illustrated in Figure 1 with the "corrected" sequence of vote totals. This correction consisted of imputing the reporting times of the outlying observations with those from their control units. Our expectation is that, if the results were strategically sorted to publicize a certain set of results ahead of time, then we should find significant discrepancies between and among the distributions.

The most straightforward way to find such discrepancies is to visually compare the observed and imputed vote trends, as shown in the left plot of Figure 5. This plot displays the vote trends for every candidate during the first nine hours of the vote count, where there exists the larger discrepancies between the vote trends before their convergence.[27] The comparison suggests that the imputed distributions digress from the reported ones only before 10:30 p.m., and that the outlying observations behave in a similar way for the top two candidates. In particular, the imputed vote counts for both Calderón and López Obrador are consistently above the observed trend, a difference that contrasts with the direction of the difference for Madrazo. These differences suggest that

[26] The first level includes driving time, the number of registered voters, the number of absent poll workers and the presence of party representatives. The second level contains the rates of the population in the precinct that are illiterate, have a high school degree, are 65 years or older, or do not speak Spanish.

[27] The comparison for the entire vote count is in the Appendix.

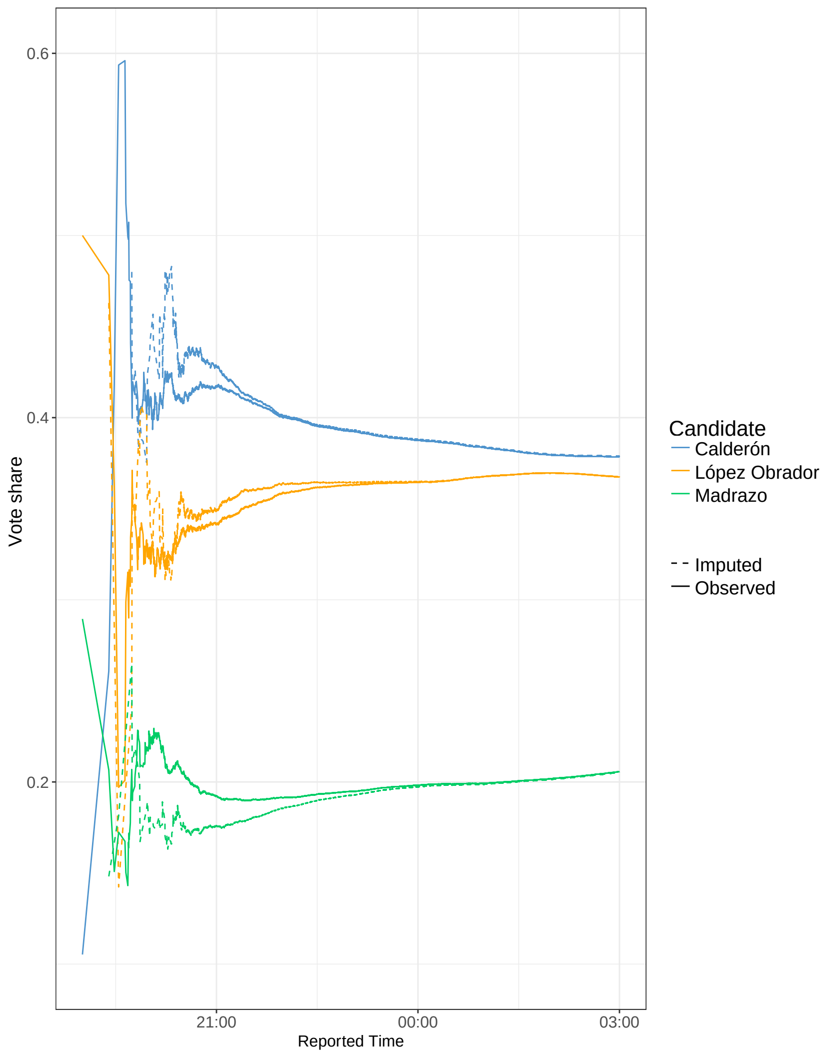


Figure 5: Observed and corrected vote-shares at different moments of the vote count. Mexico, 2006.

most of the observations that reported the results later than expected have larger vote shares for the third place candidate, deflating the vote shares for the winner and front-runner candidate to a similar magnitude. This first test falls short of documenting a potential sorting of the results for or against either of the top two candidates.

In sum, the method described above posits a way to identify suspicious observations given the speed at in which a polling station reported the results. In the case of the 2006 presidential election in Mexico, we find that polling stations reporting their results with an undue delay represent less than 5% of all observations. Moreover, and against the prevalent skepticism regarding the preliminary vote count in the country, we did not find significant biases in the vote shares of these outlying units. Our model shows, rather, that the reporting time of every polling station corresponds mostly to its geographic and logistic context and that the late results show no pattern in particular.

5 Conclusion

To build public confidence in elections, authorities must demonstrate competence in three competing goals: administrative efficiency, political neutrality, and transparency (Mozaffar and Schedler, 2002). The publication of the electoral results in real-time is a way of making the vote-counting process transparent, since it provides details for when and how the votes are aggregated. Nevertheless, the gains in transparency come at the cost of spotlighting administrative inefficiencies and raising suspicions about the impartiality of the process. This dilemma is particularly significant in developing democracies, where accidental errors or inefficiencies during the process are likely to be interpreted as signals of fraud (Pastor, 1999).

With the goal of distinguishing electoral irregularities from other factors concurrent to the election, we propose a methodology to identify potential irregularities at the polling-station level and assess the claims of manipulation during the vote count. This methodology estimates the time during which every polling station should deliver its results given its geographic, logistic, and socioeconomic characteristics. Moreover, to evaluate potential bias in the electoral results, our methodology identifies those observations predicted poorly by the model. By comparing polling stations that are similar in all characteristics except reporting time, we can assess whether locations with delayed reporting of the results systematically swung the vote trends in favor of one candidate. Our proposed methodology is useful for election officials and parties who seek to identify irregularities at this stage of the electoral process, assessing whether any discrepancies in reporting times significantly influenced the final outcome.

We illustrated our methodology using the case of the 2006 presidential elections in Mexico, where fraud allegations focused how electoral authorities reported the results. The findings show that, contrary to the conventional wisdom about the election and the PREP, those observations with underestimated report times do not show a particular bias towards either candidate in their vote trends. This analysis serves not only as a way to verify the allegations of fraud regarding the PREP in this election, but also as a practical post-electoral audit for subsequent elections. As the Appendix shows, the method can be applied to local elections to assess any potential tendency of results to favor a particular candidate.

We also hope this analysis can be a helpful tool for elections outside Mexico. The methodology can be applied to cases such as the recent elections in Argentina and Honduras, helping electoral authorities and observers flag a set of polling stations on which they should focus their attention. Nevertheless, the generalizability of this method requires a surprisingly absent input in many elections worldwide: the time logs for the reported results. With a very few exceptions, including the parliamentary elections in Croatia or the presidential elections in Ukraine, publicly available results with the time at which every electoral result was published appears to be the exception, rather than the rule, in contemporary elections. A potential explanation is the gradual adoption of alternative technologies to aggregate and report the votes in a more efficient way. But given that the speed of reporting results is a necessary requirement for electoral integrity and public trust, this approach could serve as an additional check for electoral transparency when anxiety and suspicion can harm the perception of electoral integrity.

References

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.

Antenucci, Pedro, Juan Matías Mascioto and María Page. 2017. "PASO 2017 en la provincia de Buenos Aires: el recuento provisorio explicado." *Revista SAAP* 11(2):341–364.

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.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling*. Cambridge University Press.

Bruhn, Kathleen and Kenneth F. Greene. 2007. "Elite Polarization Meets Mass Moderation in Mexico's 2006 Elections." *PS: Political Science and Politics* (1):33–38.

Burden, Barry C. and Jeffrey Milyo. 2015. "The Quantities and Qualities of Poll Workers." *Election Law Journal* (14):38–46.

Calvo, Ernesto, Juan Dodyk, Marcelo Escobar, Tomás Olego and Juan Pablo Pilorget. 2017. 1 Reporte: Evaluación de la Carga de Votos en las PASO 2017, Categoría Senador Nacional, Provincia de Buenos Aires. Technical report.

Cantú, Francisco. 2014. "Identifying Electoral Irregularities in Mexican Local Elections." *American Journal of Political Science* 58(4):936–951.

Casas, Agustin, Guillemo Diaz and Andre Trindade. 2013. "Effect of Non-Neutral Observers on Electoral Outcomes." Working Paper.

Christensen, Ronald. 1997. *Log-Linear Models and Logistic Regression*. New York: Springer-Verlag.

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

de Icaza-Herrera, Miguel. 2006. Signos Inequivocos de Manipulacion en el PREP. Working paper UNAM. Working Paper.

DeGroot, Morris H. 1984. *Probability and Statistics*. 2nd ed. Reading, MA: Addison-Wesley.

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

Eisenstadt, Todd A. 2007. "The Origins and Rationality of the "Legal versus Legitimate" Dichotomy Invoked in Mexico's 2006 Post-Electoral Conflict." *PS: Political Science and Politics* 40(1):39–43.

European External Action Service. 2017. EU EOM Honduras 2017: Statement following the declaration of provisional results of the 2017 general elections. Press release European Union.

Gandrud, Christopher. 2015. "simPH: An R Package for Illustrating Estimates from Cox Proportional Hazard Models Including for Interactive and Nonlinear Effects." *Journal of Statistical Software* 65(3):1–20.
URL: *http://www.jstatsoft.org/v65/i03/*

Gibson, Edward L. 2013. *Boundary Control: Subnational Authoritarianism in Federal Democracies*. New York: Cambridge University Press.

Giraudy, Agustina. 2015. *Democrats and Autocrats: Patwways of Subnational Undemocratic Regime Continuity within Democratic Countries*. Oxford University Press.

Grambsch, Patricia M. and Terry M. Therneau. 1994. "Proportional hazards tests and diagnostics based on weighted residuals." *Biometrika* 81(3):515–526.

Hall, Thad E., J. Quin Monson and Kelly D. Patterson. 2008. Poll Workers' Job Satisfaction and Confidence. In *Democracy in the States: Experiments in Election Reform*, ed. Bruce Cain, Todd Donovan and Caroline Tolbert. Washington: Brooking Institution Press.

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

Hawkins, Douglas M. 1980. *Identification of Outliers*. Springer.

Hernández, Rubén, Adrián Vallejo and Alejandro Vergara. 2006. Análisis de la probabilidad de cruces entre las curvas de votación acumulada del PAN y la CBT en el PREP de la Elección de Presidente de los Estados Unidos Mexicanos de 2006. In *Desempeño del PREP en las pasadas elecciones del 2 de julio del 2006*, ed. Instituto Federal Electoral.

Jiménez, Raúl and Manuel Hidalgo. 2014. "Forensic Analysis of Venezuelan Elections during the Chávez Presidency." *PLoS ONE* 9(6).

Jin, Shuai and Frederick J. Boehmke. 2016. "Proper Specification of Non-proportional Hazards Corrections in Duration Models." International Methods Colloquium Talk on March 25th, 2016.

Keele, Luke. 2010. "Proportionally Difficult: Testing for Nonproportional Hazards in Cox Models." *Political Analysis* (18):189–205.

Keele, Luke, Sam Pimentel and Frank Yoon. 2015. "Variable Ratio Matching with Fine Balance in a Study of the Peer Health Exchange." *Statistics in Medicine* 34(30):4070–4082.

Larreguy, Horacio, John Marshall and Pablo Querubin. 2016. "Parties, Brokers and Voter Mobilization: How Turnout Buying Depends Upon the Party's Capacity to Monitor Brokers." *American Political Science Review* 10(1):160–179.

Lawson, Chappell. 2007. "How Did We Get Here? Mexican Democracy after the 2006 Elections." *Political Science and Politics* 40(1):45–48.

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

Mebane, Walter R. 2015. "Election Forensics Toolkit." DRG Center Working Paper.

Mochán, W. Luis. 2006. Incertidumbre y errores en las elecciones de julio del 2006. Working paper Centro de Ciencias Físicas, UNAM.

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

Norris, Pippa, Ferran Martínez i Coma and Richard W. Frank. 2016. "The Expert Survey of Perceptions of Electoral Integrity, PEI 4.0." www.electoralintegrityproject.com.

Pastor, Robert A. 1999. "The Role of Electoral Administration in Democratic Transitions: Implications for Policy and Research." *Democratization* 6(4):1–27.

Pimentel, Samuel D. 2016. "Large, Sparse Optimal Matching with R package rcbalance." *Observational Studies* 2:4–23.

Pliego Carrasco, Fernando. 2007. *El Mito del Fraude Electoral en México*. Mexico City: Editorial Pax.

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.

Rochín, Victor Romero. 2006. Una Opinión sobre el PREP y el Conteo Distrital. Working paper UNAM.

Schoenberg, Isaac Jacob. 1946. "Contributions to the problem of approximation of equidistant data by analytic functions. Part A. On the problem of smoothing or graduation. A first class of analytic approximation formulae." *Quarterly of Applied Mathematics* 4(1):45–99.

Tello, Carlos. 2007. *2 de Julio: La Crónica Minuto a Minuto Del Día Más Importante de Nuestra Historia Contemporánea*. Mexico City: Editorial Planeta Mexicana.

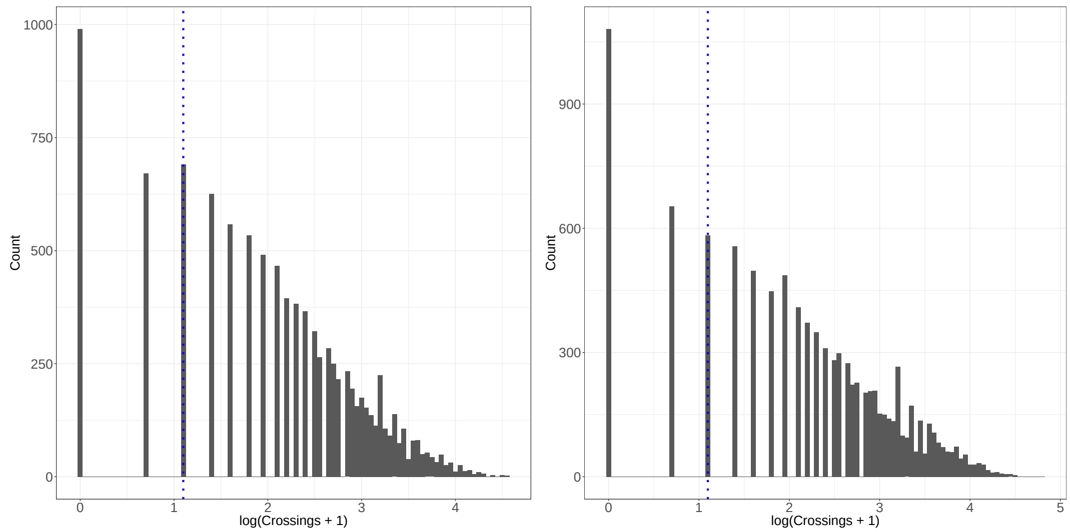
Therneau, Terry M. 2015. *A Package for Survival Analysis in S*.
URL: *http://CRAN.R-project.org/package=survival*

Therneau, Terry M. and Patricia M. Grambsch. 2000. *Modeling Survival Data*. New York: Springer.

Tuckman, Jo. 2012. *Mexico: Democracy Interrupted*. New Haven: Yale University Press.

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 Political Science Review* 95(4).

6 Appendix

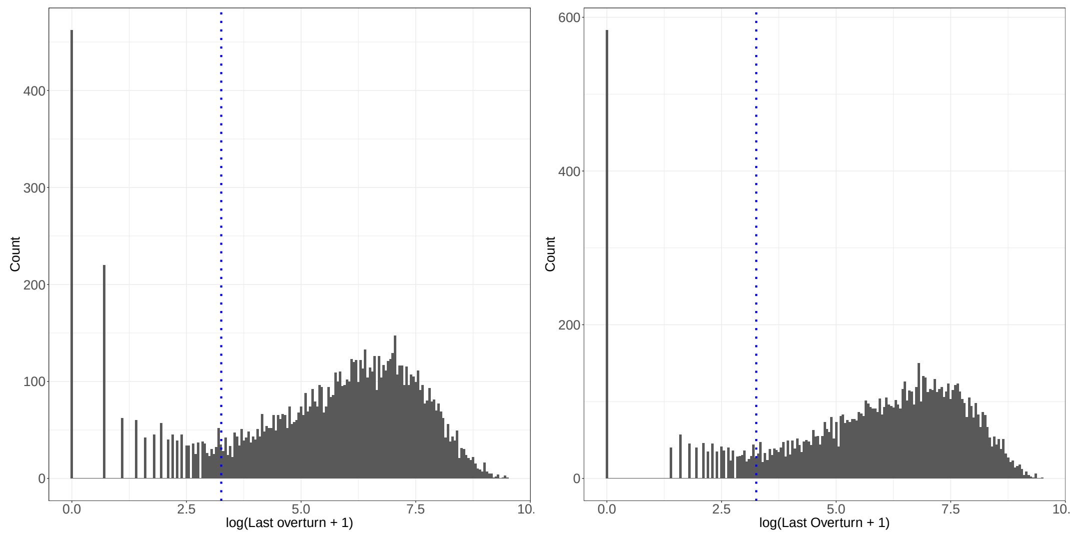


(a) Simulations

(b) Permutations

Figure A: Histogram for the values of *Crossings*

Notes: The plots show the frequency for the number of times in which the leading candidate overturns the vote aggregates of the runner-up in the simulated elections (Plot 1(a)) and permuted results (Plot 1(b)) described in Section 3.1. The dashed line represents the observed number of *crossings* in the candidates' vote shares reported by the PREP. We log-transform the number of overturns per iteration to facilitate its visualization.



(a) Simulations

(b) Permutations

Figure B: Histogram for the values of *Last overturns*

Notes: The plots show the frequency for the number of reported tallies before the leading candidate gets an unsurpassable vote share in the simulated elections (Plot 2(a)) and permuted results (Plot **??**) described in Section 3.1. The dashed line represents the observed *last overturn* in the candidates' vote shares reported by the PREP. We log-transform the number of last overturns overturns per iteration to facilitate its visualization.

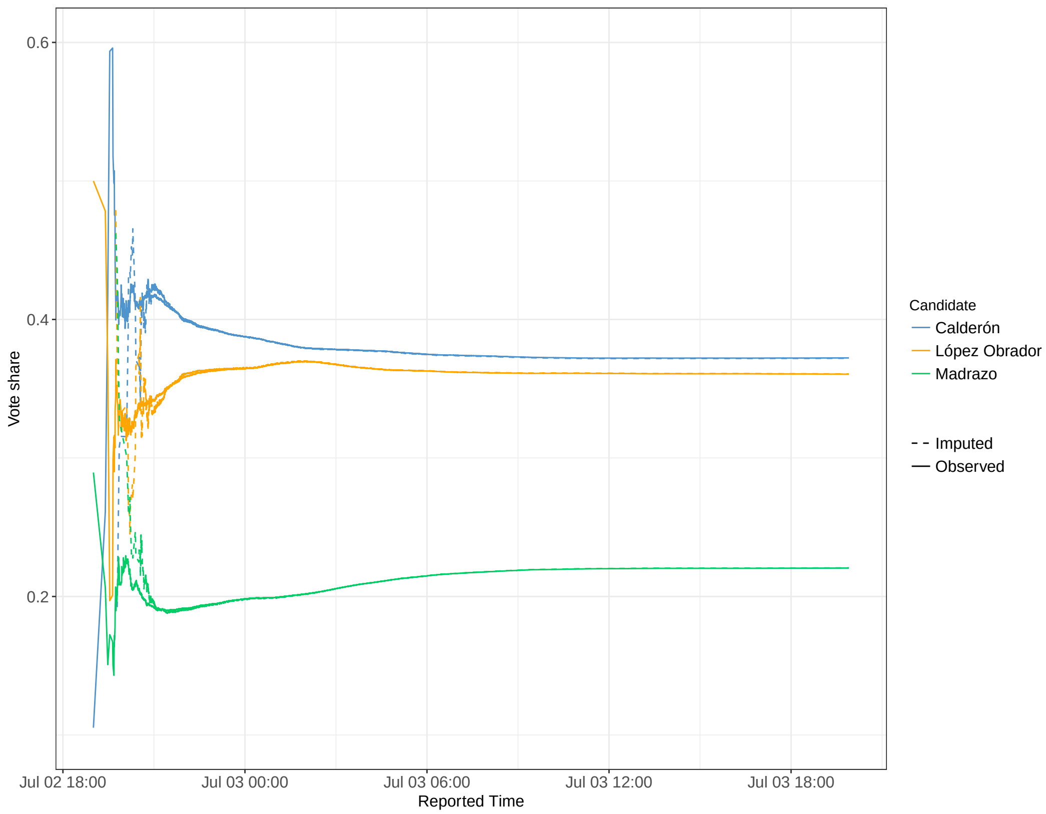


Figure C: Presidential candidates' vote shares by the time the results of the polling stations were reported. Mexico, 2006.

Notes: The figure shows the vote shares for the candidates given the reported time of the preliminary results. The dashed lines show the vote shares for the candidates after adjusting the reporting times of the outlying units.

Variable	**Hazard Ratio (Minimum)**	**Hazard Ratio (Mean)**
PAN Observers	111.6232	1.4131
PRD Observers	15.0043	1.2715
PRI Observers	2613.607	2.3721

Table A: Hazard ratios of Binary Variables with Time Interactions.
Note: Columns display hazard ratios at the minimum and mean (logged) reporting times, respectively. For each covariate, a hazard ratio for $x_i = 1$ is compared to a baseline of $x_i = 0$.

Variable	**Unbalanced**			**Balanced**		
	Treated	Control	Standarized Difference	Treated	Control	Standarized Difference
Log Driving Time	1.68	2.41	0.548	1.71	1.71	0.002
Registered Voters	0.70	0.76	0.421	0.71	0.71	0.011
65 years old	0.07	0.06	0.180	0.07	0.07	0.0003
Illiteracy	0.07	0.08	0.250	0.07	0.07	0.005
High School	0.12	0.11	0.220	0.12	0.12	0.003
No Spanish	0.001	0.003	0.078	0.001	0.001	0.008
Computer	0.24	0.20	0.240	0.24	0.24	0.019
Absentees	0.70	0.72	0.033	0.69	0.69	0.003

Table B: Absolute standardized differences for the covariates before and after sparse optimal matching.

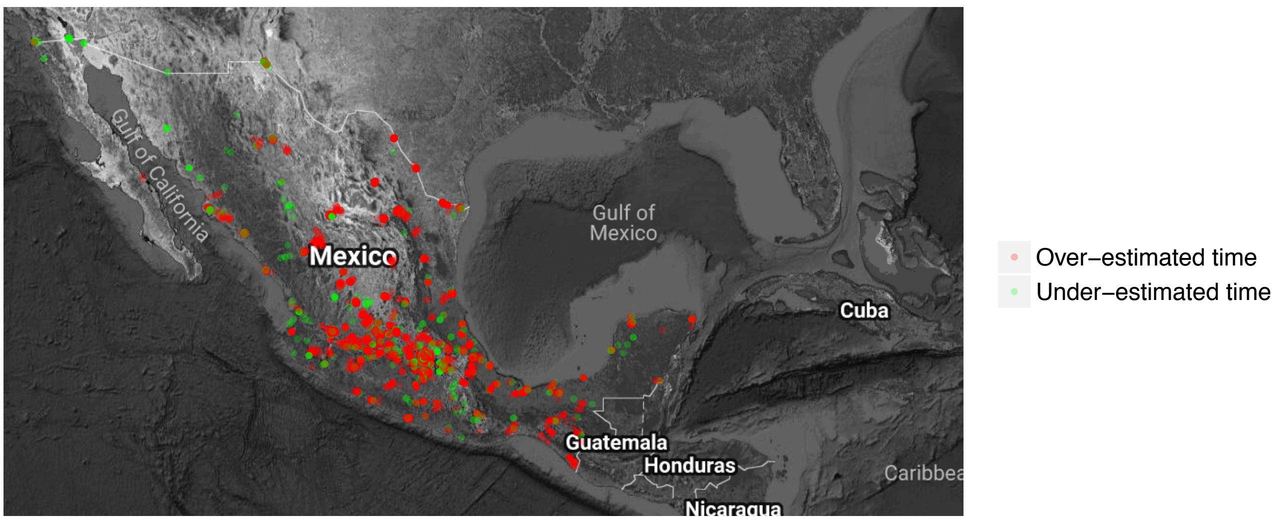
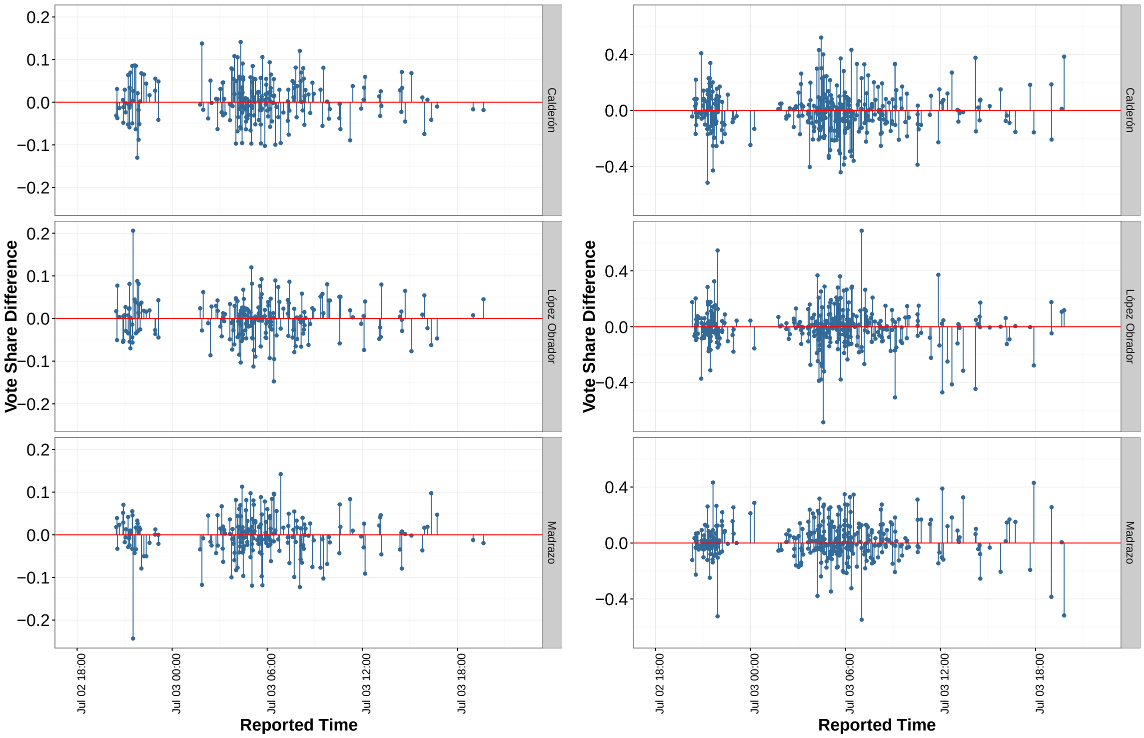


Figure D: Geographic Location of the Outliers Identified on Section 3.2.



(a) Adjacent Polling Stations. (b) Sparse Optimal Matching.

Figure E: Difference in Vote Shares for each outlier by candidate and comparison method

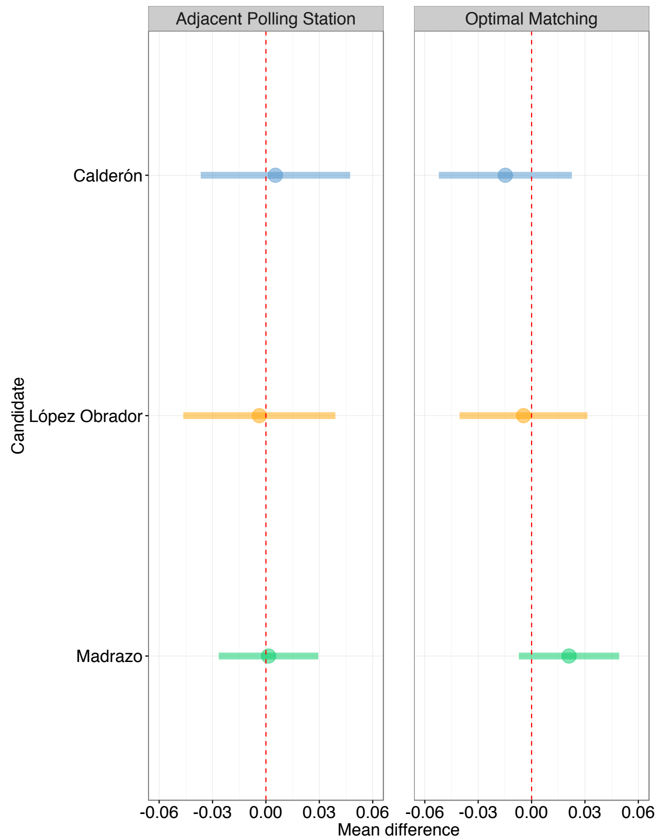
Notes: The plots show the difference in the vote shares for each candidate between those observations identified as outliers and their matched polling stations. Dots above (below) the red line depict higher (lower) vote shares of the outliers when compared to their matched observations.

6.1 Analysis of the Vote Shares

This test looks for a systematic bias in the reported results of the outlying observations. We do so by comparing the vote shares for each candidate between our outlying observations and their corresponding control units. This approach allows us to identify potential trends on the vote results over time and assess the overall effect on the electoral result.

The comparisons are illustrated in Figure E, which shows the differences between the vote shares of the three main candidates for each outlying observation and its correspondent control unit. We separate the figure by each of the matching approaches referenced above. Dots above the horizontal line indicate a positive difference for the candidate in the outlying observations compared with their control units. By contrast, a value below the horizontal line denotes a lower candidate's vote share for the outlying observation compared with that reported by its control units. The plots show no observable trend between the reported time of the outlying observation and the differences between the candidates' vote shares. Overall, the discrepancies for each candidate cancel each other out, failing to back the argument of a potential sorting of the results in favor of or against a particular candidate.

To estimate the overall effect of the outlying observations, we compare their mean vote share for each candidate with those reported by the control units. The results are summarized in Figure F, which shows the estimated difference of means and 95% confidence intervals of the vote shares between the outliers and their respective counterfactual sets when comparing adjacent polling stations and using sparse optimal matching, respectively. In both cases, the differences are never statistically different from 0, suggesting that the results reported later than expected had no effect on the reported trend for each candidate.



Notes: The dots and horizontal lines show the mean differences and 95% confidence intervals for comparing the vote shares of the outliers to the matched observations.

Figure F: Difference in Means for the Candidates' Vote Shares.

In sum, the method described above suggests a way to identify suspicious observations given the time in which a polling station reported the results. In the case of the 2006 presidential election in Mexico, we find that polling stations reporting their results with an undue delay represent less than 0.5% of all observations. Moreover, and against the conventional wisdom regarding the preliminary vote count in the country, we did not find significant biases in the vote shares of these outlying units. Our model shows that the reporting time of every polling station mostly corresponds to its geographic and logistic context, and that the late results show no particular pattern.

As an extension of the proposed methodology, the next subsection shows the analysis of similar cases for the 2016 local elections in two Mexican provinces. In both elections, the losing candidates claimed fraud and blamed the electoral authorities for multiple irregularities during the preliminary vote count. Using data for those elections, we find that the (un)timely reporting of the results had no effect on the overall trend for each of the candidates in their respective elections.

6.2 2016 Local Elections

We use the proposed methodology to assess the fraud allegations on the preliminary counts of two gubernatorial elections in Oaxaca and Veracruz on June 5, 2016. Both states are often identified among those states with the lowest levels of sub-national democracy (Gibson, 2013; Giraudy, 2015). In the case of Oaxaca, the PRI's candidate won the election with 32% of the votes, followed by the PAN and PRD coalition and the recently formed National Regeneration Movement (MORENA) with 25% and 23% of the vote, respectively. This result gave back control of the state to the PRI after losing the election six years earlier. Meanwhile, the PAN's candidate won the governorship in Veracruz with 34% of the vote, followed by the candidates of the PRI and MORENA with 30% and 27%, respectively. This result ended the PRI's 87-year rule of the state.

In both elections, losing candidates claimed fraud and blamed the electoral authorities for multiple irregularities during the preliminary vote count. In Oaxaca, after the PREP showed a lead for the PRI candidate, opposition parties alleged that the PREP was biased, suggesting a potential manipulation of the results to match them with the pre-electoral polls.[28] Similarly, the candidate of MORENA in Veracruz, Cuitláhuac García, accused the PREP of being "statistically manipulated," and argued that the electoral authorities inflated the preliminary results with more than 100,000 fake votes.[29]

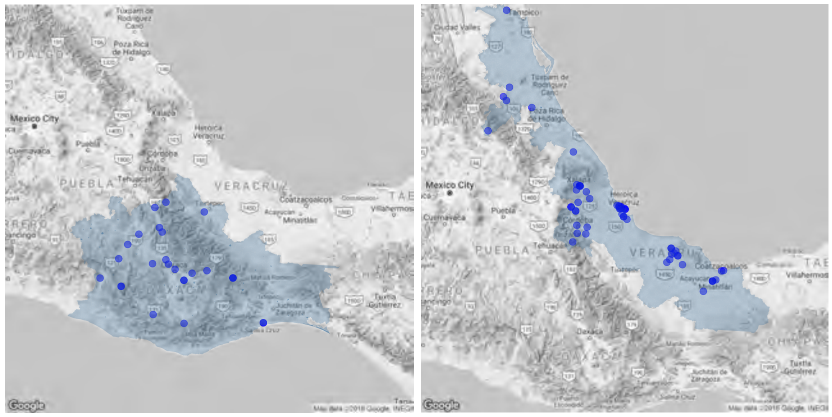
To assess the validity of these claims we proceeded in a similar way to the analysis described in the previous section. First, we obtained the coordinates of the local district offices in each state and estimated their driving distance to the polling stations from which they should receive the results. Then, using sociodemographic data from the 2010 census, we estimate a duration model for the time in which the results were reported in the PREP.[30] Unlike the analysis for the 2006 election, we lack information on the presence of party representatives and poll workers. For Oaxaca, we also lack information on the number of registered voters.

Following the procedures described above, we estimate the survival models and correct for non-proportionality. The results of the survival models for Oaxaca and Veracruz are shown in the Appendix in Tables C and D, respectively. We then obtain the Deviance residuals of the final model for each election and identify as outliers those observations with a residual value below

[28] See, for example, "Oaxaca: CNTE reclama fraude electoral en virtual triunfo de Alejandro Murat" (http://www.noticiasmvs.com/#!/noticias/oaxaca-cnte-reclama-fraude-electoral-en-virtual-triunfo-de-alejandro-murat-462), "Eleccion en Oaxaca se dirimira en tribunales, advierten Morena y PAN-PRD" (http://www.proceso.com.mx/443172/eleccion-en-oaxaca-se-dirimira-en-tribunales-advierten-morena-pan-prd), "Piden realizar una auditoría al PREP" (http://www.eluniversal.com.mx/articulo/estados/2016/06/9/piden-realizar-una-auditoria-al-prep), "Partidos preocupados ante lentitud del PREP en Oaxaca" (http://www.elfinanciero.com.mx/nacional/partidos-preocupados-ante-lentitud-del-prep-en-oaxaca.html)

[29] See "Protestas en Tlaxcala; recuento en Veracruz"(http://www.jornada.unam.mx/2016/06/09/politica/006n1pol), "OPLE exhorta a Morena a no desinformar con fraude"(http://e-veracruz.mx/nota/2016-06-09/elecciones/ople-exhorta-morena-no-desinformar-con-fraude), "PREP en Veracruz está manipulado, acusa Morena"(http://www.milenio.com/estados/Cuitlahuac_Garcia-Morena-Veracruz-Elecciones-2016_0_751125074.html), "Cierra PREP en Veracruz; Morena y PRI denuncian inconsistencias e irregularidades" (http://www.proceso.com.mx/443464/cierra-prep-en-veracruz-morena-pri-denuncian-inconsistencias-e-irregularidades)

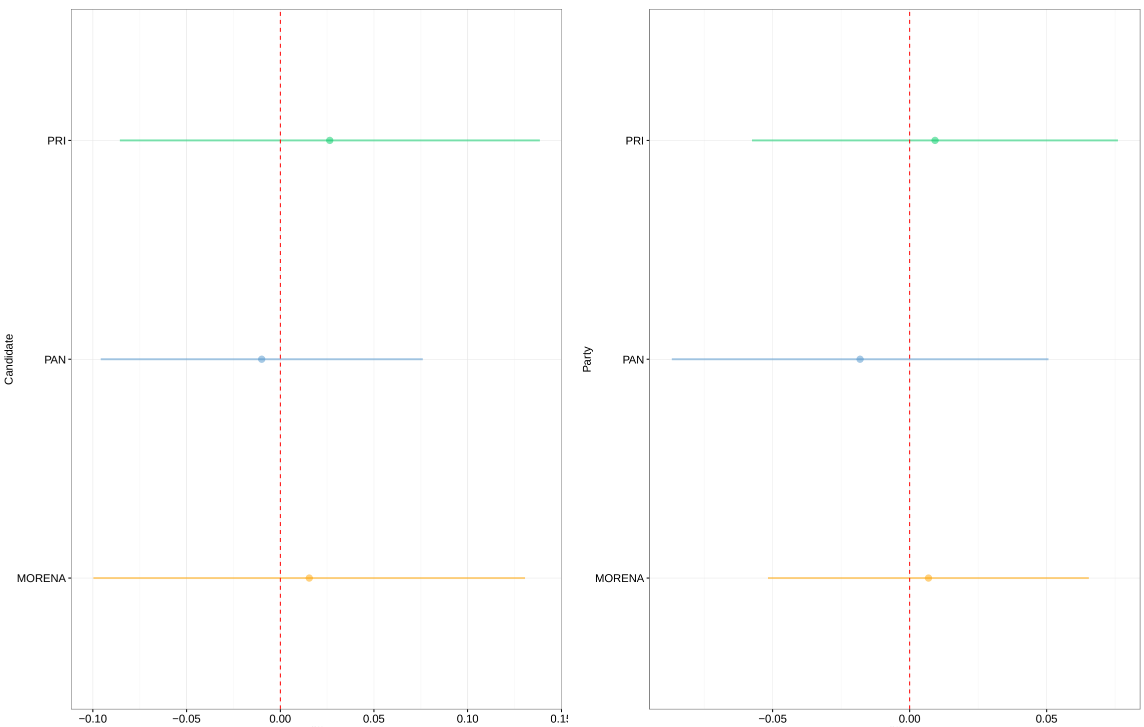
[30] http://gaia.inegi.org.mx/geoelectoral/viewer.html



(a) Oaxaca (b) Veracruz

Figure G: Geographic Location of the Outliers

Notes: The plots in each map show the geographic location of those polling station which reporting time was underestimated by the model.



(a) Oaxaca (b) Veracruz

Figure H: Difference in Means for the Candidates' Vote Shares

Notes: The dots and horizontal lines show the mean differences and 95% confidence intervals comparing the vote shares of the outliers to the matched observations.

-3; the geographic locations of these outliers are shown in Figure G. Finally, we create potential counterfactuals to estimate the potential effect of the unexpected survival time of these outliers on the electoral result. The complete results are in the Appendix.

Figure H shows the overall effects of the outliers on vote shares in Oaxaca and Veracruz, respectively. The difference in means for each candidate's vote share are plotted in both figures, along with 95% confidence intervals. As in the case of the 2006 Election, the model once again shows no statistical difference in outcomes when comparing the outliers to their respective counterfactuals, suggesting that the (un)timely reporting of the results had no effect on the overall trend for each of the candidates in their respective elections.

Variable	Coef.
log(Driving Distance)	-0.233
	(0.0173)
65-Year-Olds	0.0784
	(0.4024)
Illiteracy	330.8147
	(4.5249)
Computer	199.5805
	(3.1975)
No Spanish	-142.741
	(4.0912)
High School	5.0218
	(0.3341)
Illiteracy * log(Minutes)	-51.1548
	(0.7053)
Computer * log(Minutes)	-32.4932
	(0.5333)
No Spanish * log(Minutes)	22.0642
	(0.6213)
θ	0.958
I-likelihood	-25980.2

Table C: Cox model on the reporting time for the vote counts in every polling station, Oaxaca.

Variable	Coef.
log(Registered Voters)	0.0373
	(0.0522)
log(Driving Distance)	-0.2031
	(0.0151)
65-Year-Olds	0.4642
	(0.3974)
Illiteracy	251.0217
	(2.5864)
Computer	126.0903
	(1.3068)
No Spanish	-157.4623
	(7.568)
High School	0.9299
	(0.229)
Illiteracy * log(Minutes)	-38.3146
	(0.4041)
Computer * log(Minutes)	-19.6523
	(0.2084)
No Spanish* log(Minutes)	24.0149
	(1.1028)
θ	0.965
l-likelihood	-54040.2

Table D: Cox model on the reporting time for the vote counts in every polling station, Veracruz.

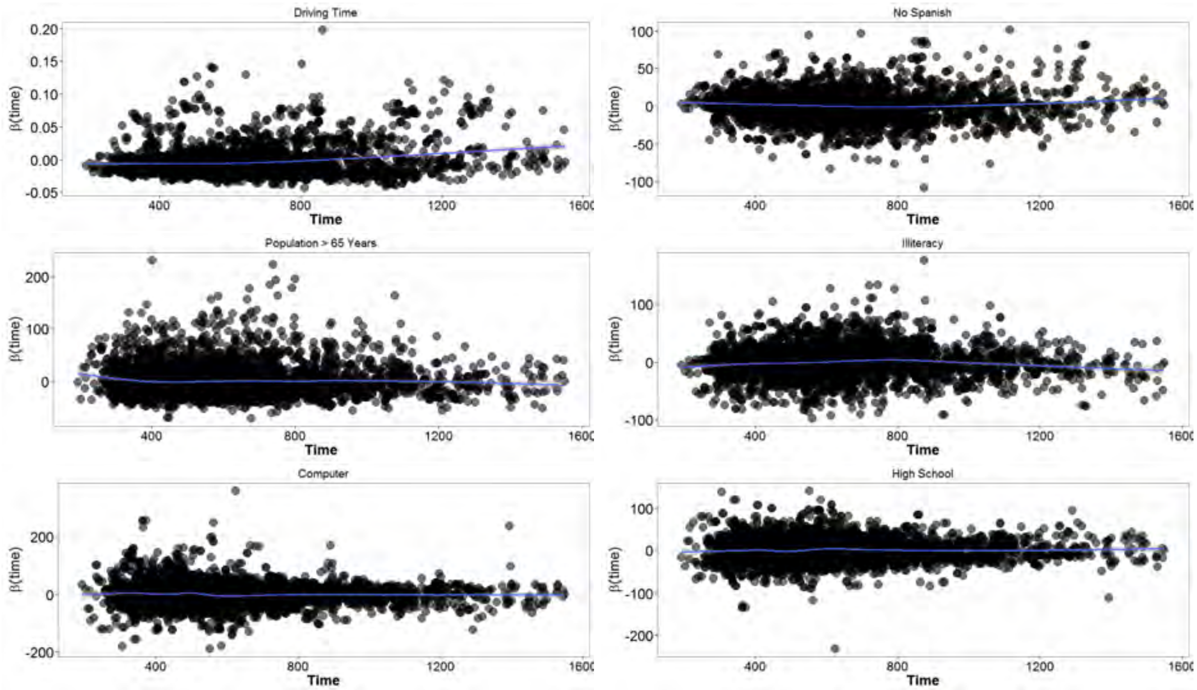


Figure I: Schoenfeld Residuals Showing Violations of Proportionality Assumption (Oaxaca Elections, Table C)

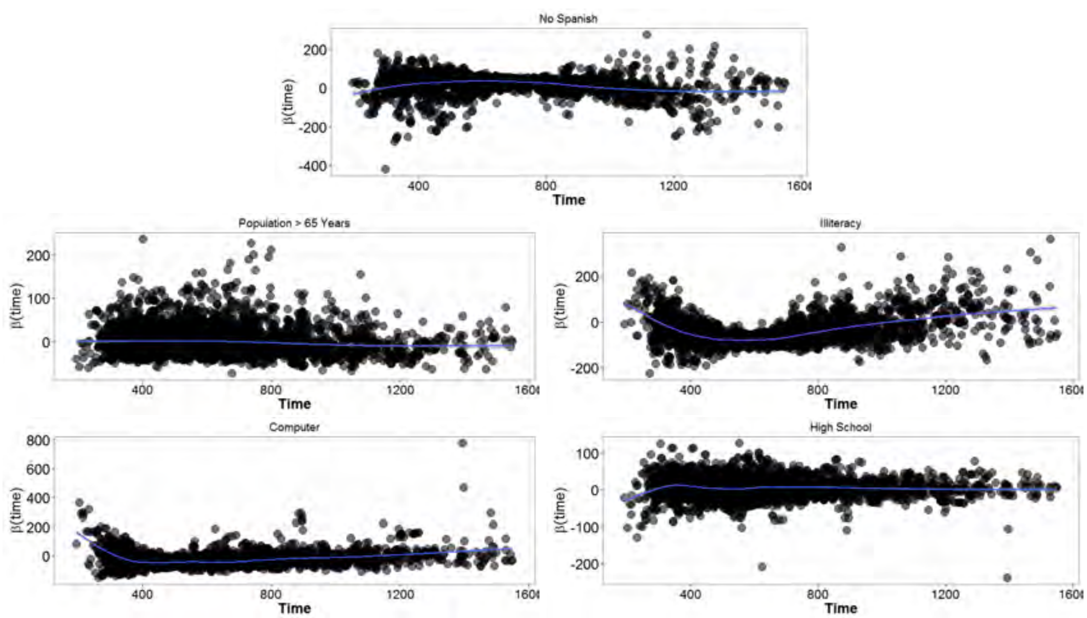


Figure J: Schoenfeld Residuals for Corrected Model (Oaxaca Elections)

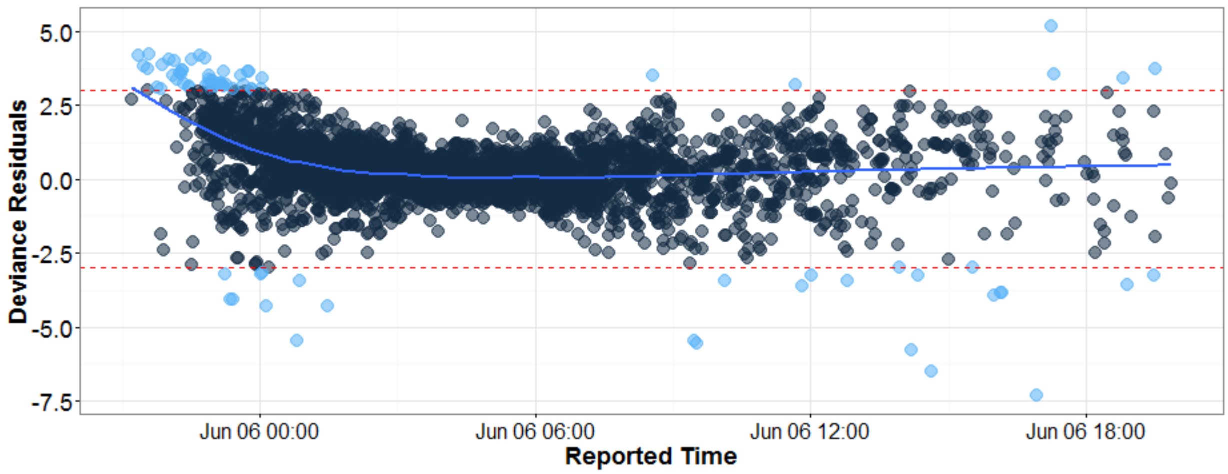


Figure K: Deviance residuals and time in which each observation was reported by the PREP. Oaxaca.

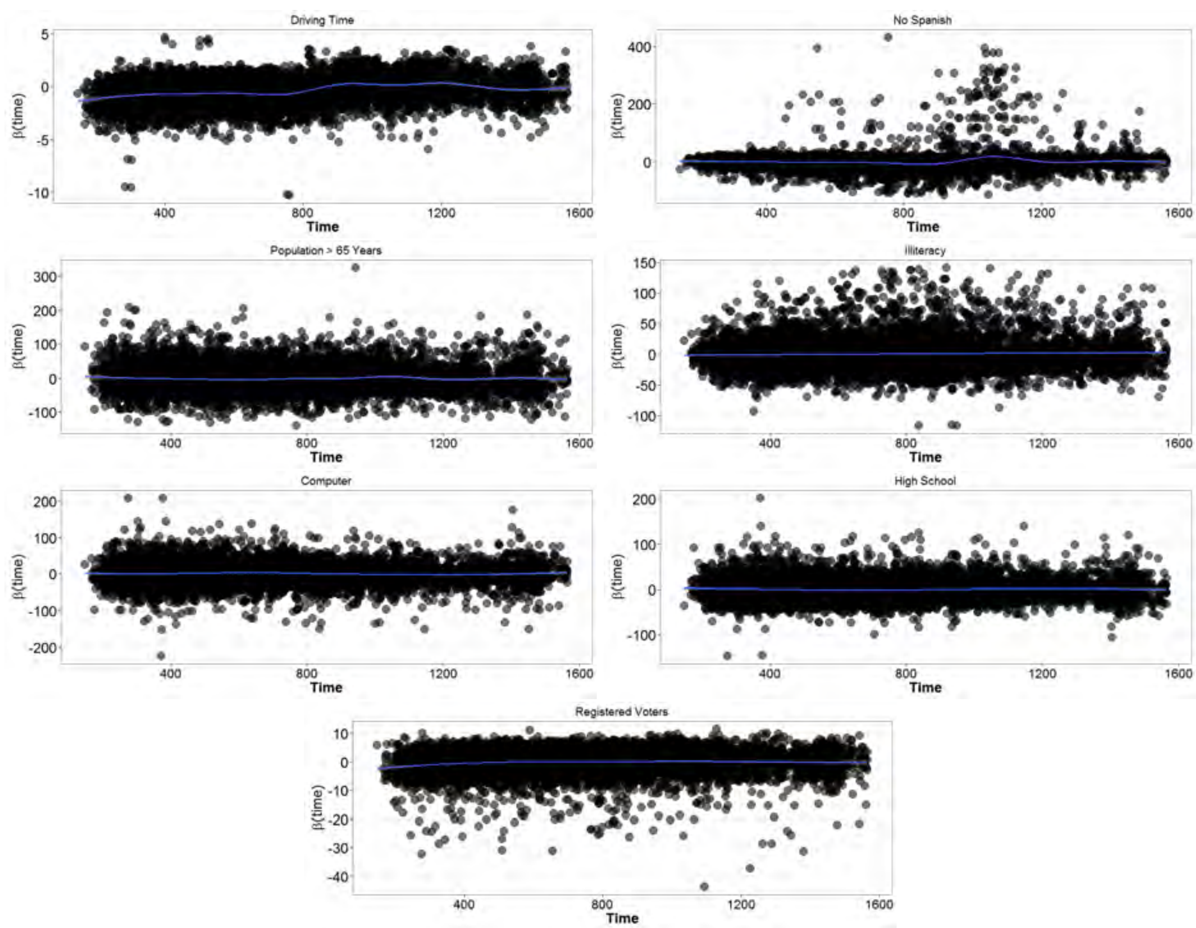


Figure L: Schoenfeld Residuals Showing Violations of Proportionality Assumption (Veracruz Elections, Table D)

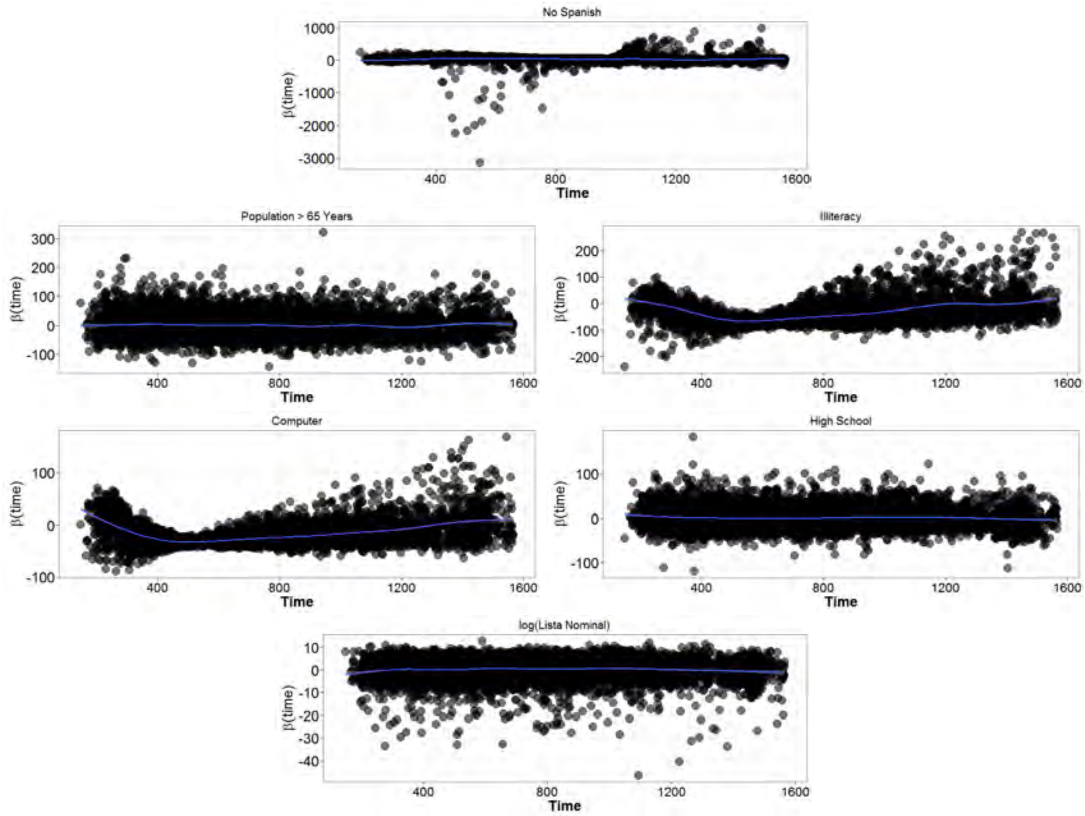


Figure M: Schoenfeld Residuals for Corrected Model (Veracruz Elections)

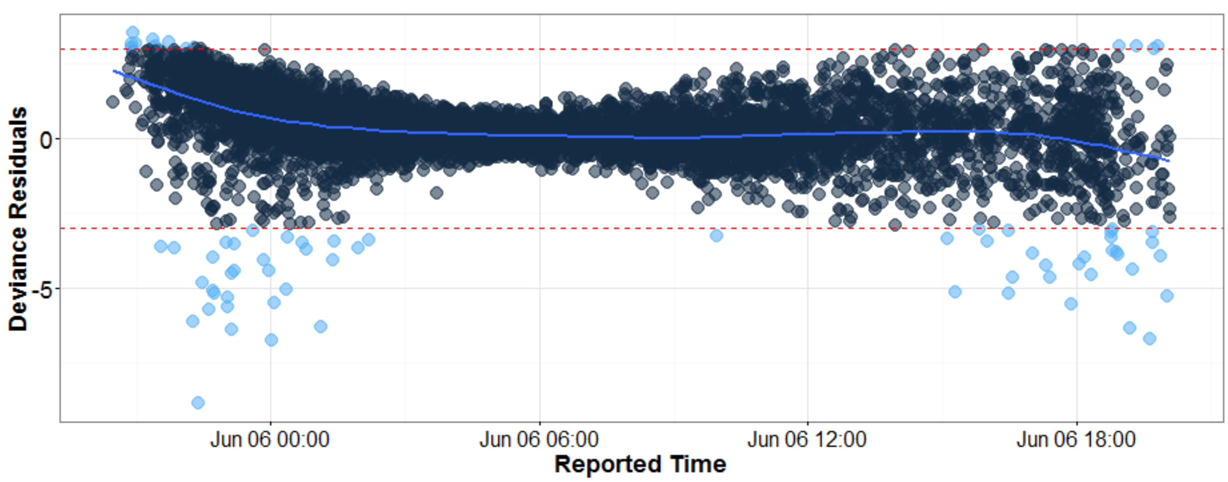


Figure N: Deviance residuals and time in which each observation was reported by the PREP. Veracruz.

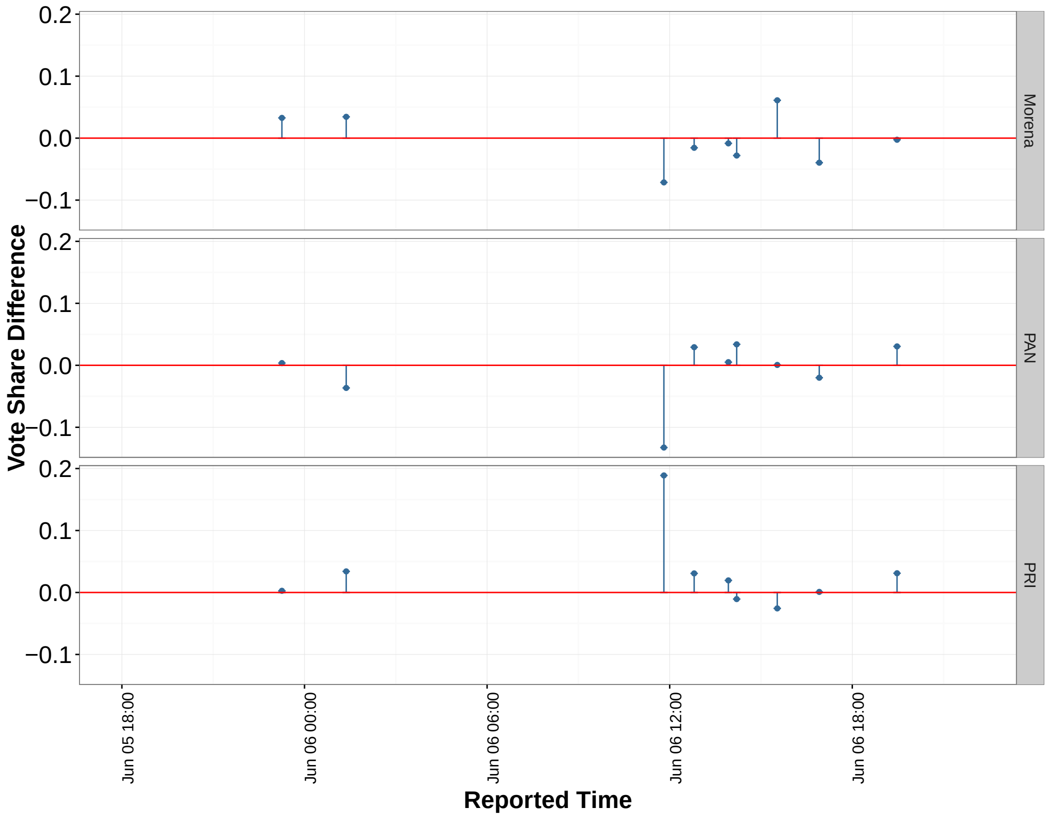


Figure O: Residuals Oaxaca (Under-Neighbors)

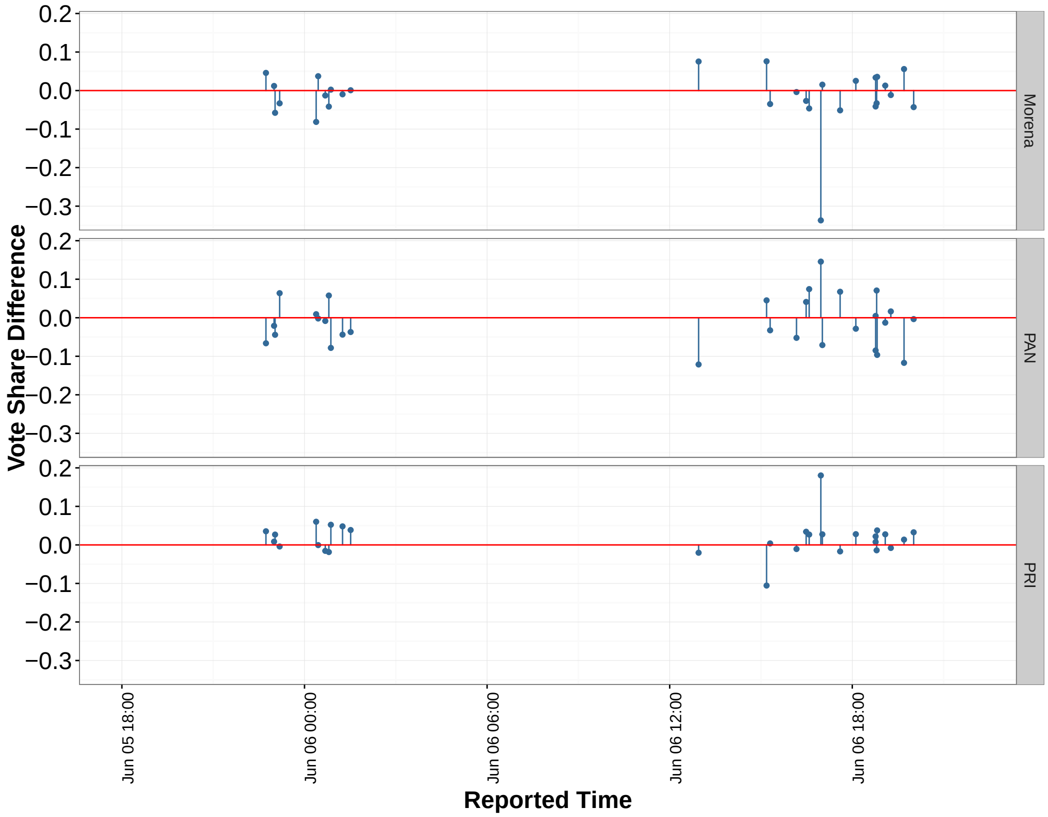


Figure P: Residuals Veracruz (Under-Neighbors)

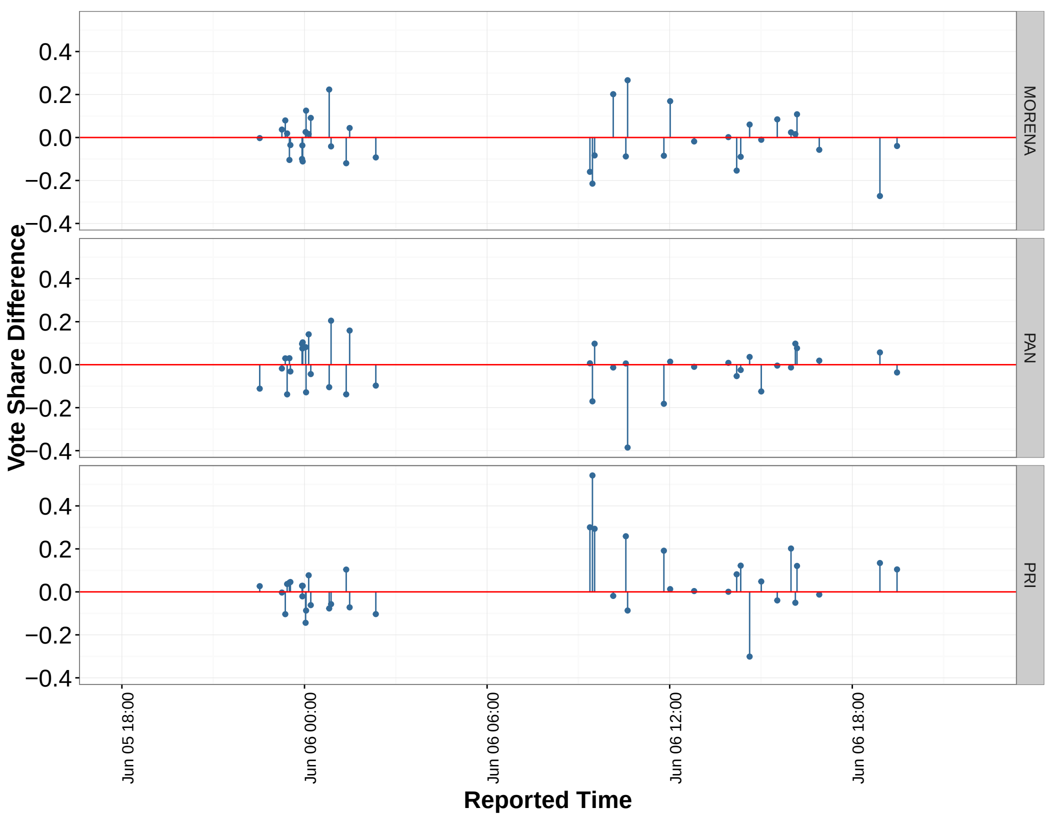


Figure Q: Residuals Oaxaca (Under-Optimal)

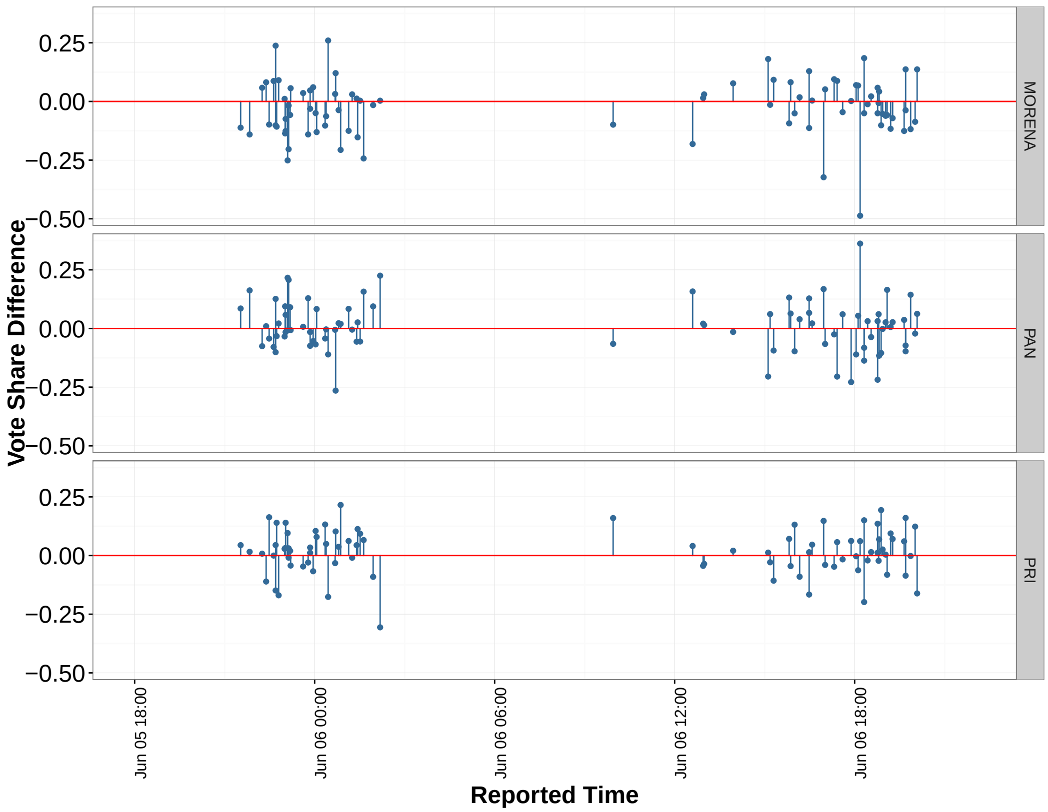


Figure R: Residuals Veracruz (Under-Optimal)

6.3 Alternative Specifications

Variable	Coef.
Absent Poll Workers	-0.1924^{***}
	(0.0167)
PAN Observers	-0.0425^{***}
	(0.0091)
PRD Observers	-0.0251^{**}
	(0.0086)
PRI Observers	-0.0192
	(0.0115)
Registered Voters	-0.9616^{***}
	(0.0230)
log(Driving Distance)	-0.6475^{***}
	(0.0033)
65-Year-Olds	-1.4157^{***}
	(0.1072)
High School	-0.0688
	(0.1355)
Computer	-0.0720^{*}
	(0.0337)
No Spanish	-0.5429^{**}
	(0.1772)
Illiteracy	-1.3082^{***}
	(0.1196)
θ	0.988
I-likelihood	-1069670.4

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table E: Baseline Cox model on the reporting time for the vote counts in every polling station. Mexico, 2006.

The baseline model contains the same base covariates as the model in Table 2, but does not correct for violations of non-proportionality.

Table E shows the results of the Cox Proportional Hazard Model prior to correcting for violations of the proportionality assumption. Overall, the direction of each covariate's effect is as expected, and most variables attain statistical significance. However, as detailed above, because these variables violate the proportionality assumption, this model is incorrectly specified, and, therefore, inaccurate. This is further supported by comparing the plot of the Deviance Residuals of the baseline model (Figure S) to the plot of the Deviance Residuals of the corrected model (Figure 4). In the former, the distribution of residuals is far from being symmetrically distributed around 0, tending instead towards the negative values for those units with the longest duration rates. This suggests that the model progressively underpredicts observations with the passage of time, as demonstrated by the larger residuals in the lower right quadrant of the figure. In contrast, the residuals for the time-corrected model show a more symmetric distribution of the residuals around 0, and the flatness of the loess line suggests that the corrected model overperforms the baseline model. Moreover, as seen in Table F and Figure U, the Schoenfeld Residuals further support the presence of violations of the proportionality assumption in the baseline model.

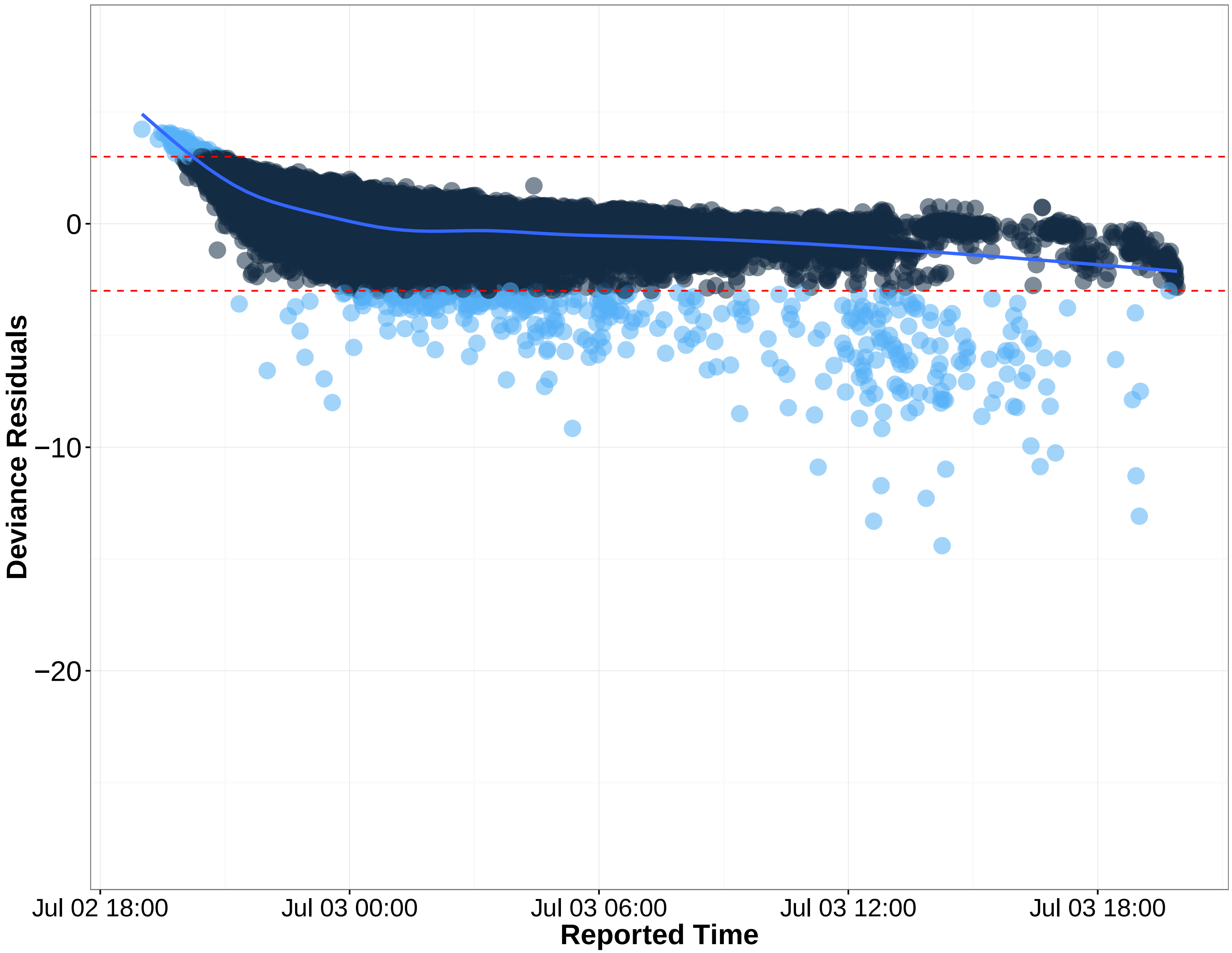


Figure S: Deviance residuals and time in which each observation was reported by the PREP. Mexico, 2006. Baseline Model.
Note: The scatter plot show the relationship between the deviance residual of each observation and its correspondent time that reported the results. We denote outlying observations in light blue and define them as those with a deviance residual outside the $[-3, 3]$ range. The plot displays the results of the baseline model that does not correct for violations of non-proportionality. Locally-weighted regression lines are shown in blue.

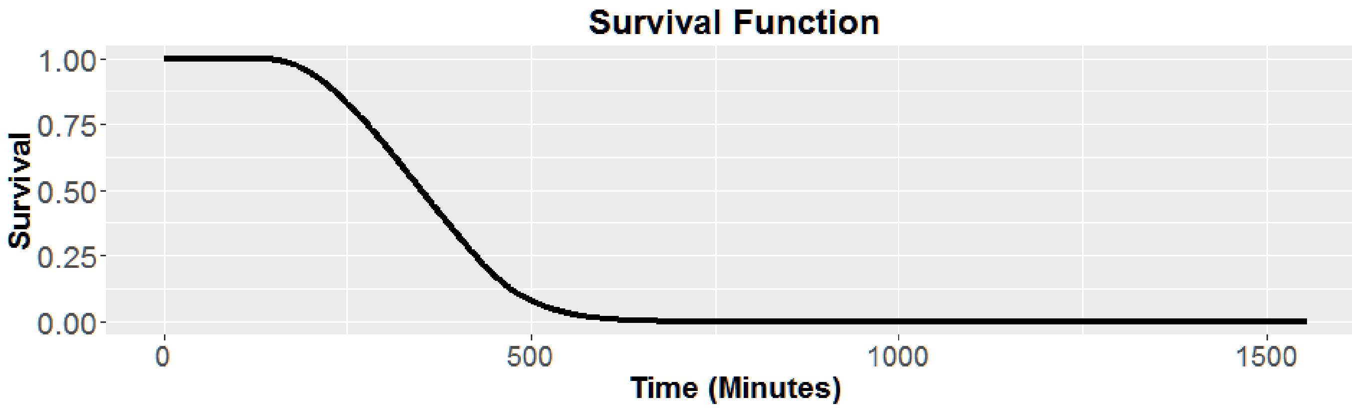


Figure T: Survival Function of Baseline Cox Model
Note: The baseline model contains the same base covariates as the model in Table 2, but does not correct for violations of non-proportionality.

Variable	ρ	χ^2	p-value
log(Registered Voters)	0.073096	42.6000	0.0000
log(Driving Distance)	0.263091	627.0000	0.0000
65-Year-Olds	0.010798	1.1900	0.2760
Illiteracy	0.04525	22.0000	0.0000
Computer	-0.01912	3.9300	0.0475
No Spanish	0.035912	13.0000	0.0003
High School	-0.000504	0.0026	0.9600
GLOBAL		1420.0000	0.0000

Table F: Schoenfeld residuals for initial, baseline model. Mexico, 2006.

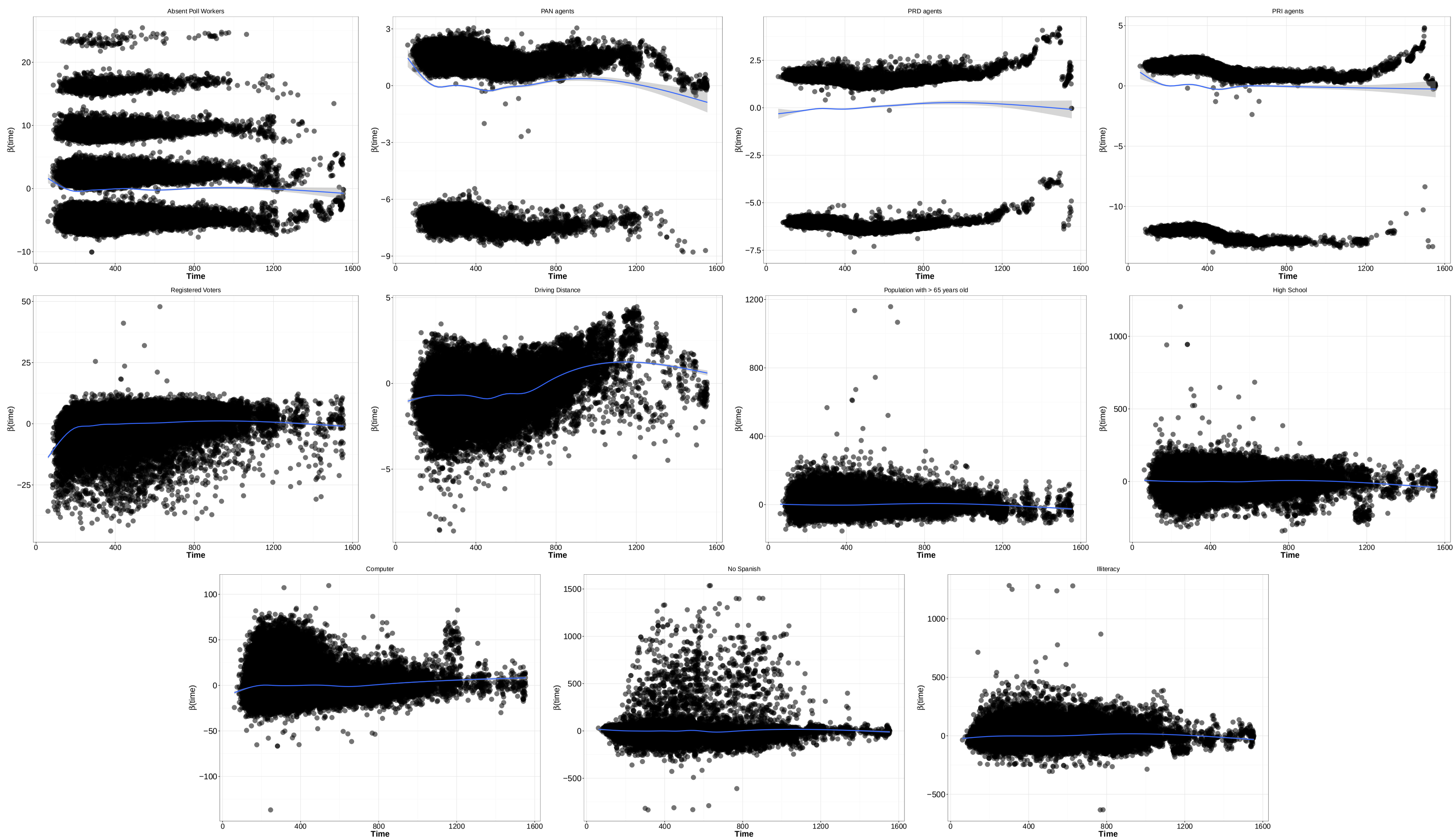


Figure U: Schoenfeld Residuals for Baseline Model, Table E

As an additional check, we rerun the hazard model with the same specification as our main model (Table 2), but with the right-censored observations included. Results of this model are shown in Table G, and a plot of this model's Deviance Residuals is displayed in Figure V. The right-censored model is not directly comparable to our main model due to the inability to correct for violations of proportionality using time interactions. This is because, by definition, right-censored observations are those still present after the reporting period is over; that is, these observations do not have a measure of time with which to create the interactions. However, the Deviance Residuals display a pattern similar to that of the uncorrected, baseline model: we observe a larger clustering of residuals in the right, bottom quadrant. Compared to the plot of the Deviance Residuals of our main model, this suggests that the main model outperforms the right-censored model, and is therefore more suitable for the analyses conducted above.

Variable	Coef.
Absent Poll Workers	-0.216^{***}
	(0.0198)
Registered Voters	-1.06
	$(0.0234)^{***}$
log(Driving Distance)	-0.688^{***}
	(0.0036)
PAN Observers	-0.0262^{**}
	(0.0091)
PRD Observers	-0.0184^{*}
	(0.0086)
PRI Observers	-0.0104
	(0.0115)
65-Year-Olds	0.3364^{**}
	(0.1074)
High School	-0.1162
	(0.1361)
Computer	0.0974^{**}
	(0.0336)
No Spanish	-0.9883^{***}
	(0.1803)
Illiteracy	-0.6900^{***}
	(0.1201)
θ	1.02
l-likelihood	-912231.5

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table G: Cox model on the reporting time for the vote counts in every polling station. Mexico, 2006.

Same model specification as Table 2, but with right censoring information included. Note that, due to the inclusion of right-censored information, time interactions are no longer possible.

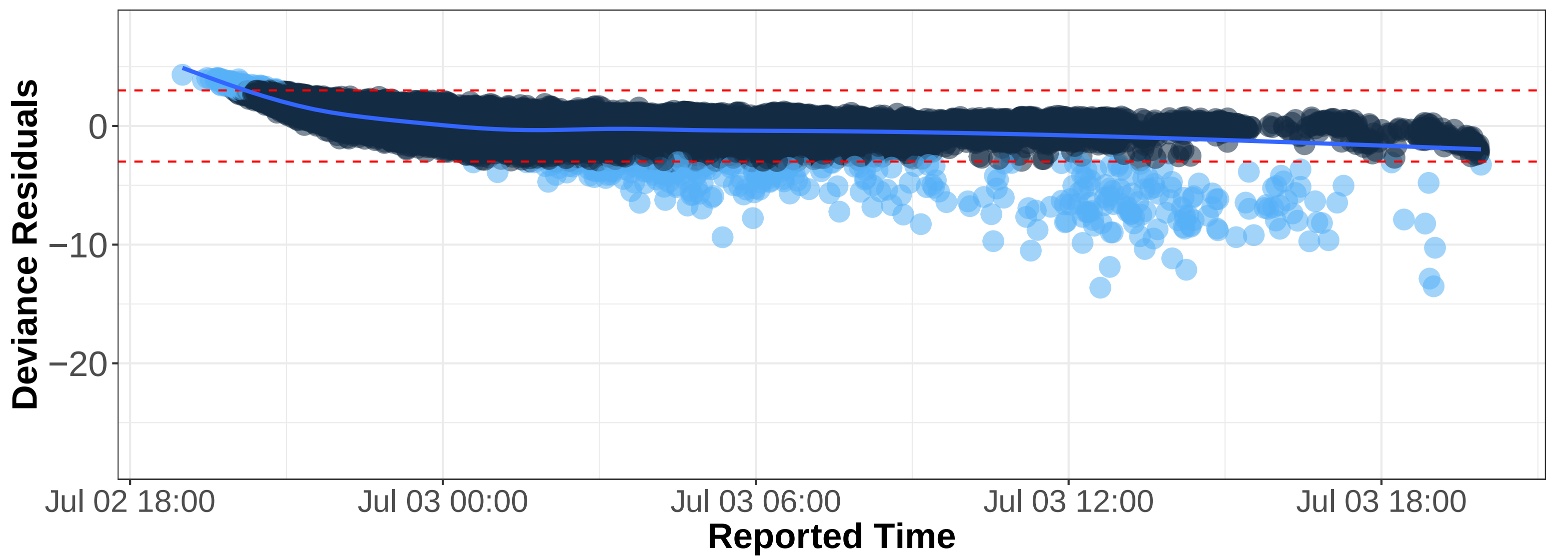


Figure V: Deviance residuals and time in which each observation was reported by the PREP. Mexico, 2006. Censored Model.
Note: The scatter plot show the relationship between the deviance residual of each observation and its correspondent time that reported the results. We denote outlying observations in light blue and define them as those with a deviance residual outside the $[-3, 3]$ range. The plot displays the results of from a model with the same specifications as the one found in Table 2, but includes the right censoring information with the survival time. Locally-weighted regression lines are shown in blue.

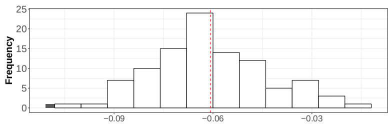
6.4 Robustness Checks

To further verify that the outliers identified in the above analyses are not effecting an undue bias on our model estimates, we conduct two additional robustness checks. The first of these is shown in Table H. Here, Model (1) is the model used for the above analyses, whereas Model (2) excludes *all* outlying observations. A comparison of the two models reveals that no major differences are observed, providing evidence that the outlying polling stations are not having a disproportionate effect on our model estimates. For the second robustness check, we take the following steps. First, we drop outlying observations with 0.20 probability. Then, we rerun the model and store the new model's estimates. We repeat this process one hundred times, each time storing the new coefficient estimates. Finally, we plot a histogram displaying the distribution of the coefficients derived from each of the one hundred simulations. Results can be seen in Figures 23(a) - 23(s). In each figure, the red, vertical, dotted line represents our model's coefficient estimate. As we can see, with the exception of illiteracy (and its interaction term), the model estimates are fairly centered around the mean of the distributions, providing further evidence that the outliers are not biasing our model estimates.

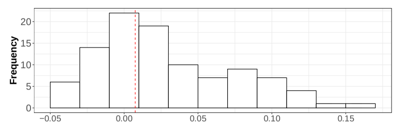
	(1)	(2)
Variable	Coef.	
Absent Poll Workers	-0.061^{**}	-0.0917^{***}
	(0.020)	(0.0198)
Registered Voters	0.0075	0.118^{***}
	(0.0243)	(0.0243)
log(Driving Distance)	-0.175^{***}	-0.145^{***}
	(0.0043)	(0.00443)
PAN Observers	14.7544^{***}	15.5638^{***}
	(0.1362)	(0.1354)
PRD Observers	8.3793^{***}	8.0982^{***}
	(0.1261)	(0.1239)
PRI Observers	23.9628^{***}	27.5000^{***}
	(0.1838)	(0.1914)
65-Year-Olds	114.6258^{***}	128.4566^{***}
	(1.2901)	(1.3326)
High School	207.6447^{***}	303.4073^{***}
	(1.3329)	(1.5291)
Computer	44.1353^{***}	48.0154^{***}
	(0.4896)	(0.5022)
No Spanish	-99.8115^{***}	-175.8127^{***}
	(3.1297)	(2.8276)
Illiteracy	227.0505^{***}	319.8438^{***}
	(1.0154)	(1.3269)
PAN Observers * log(Minutes)	-2.4492^{***}	-2.6133^{***}
	(0.0223)	(0.0223)
PRD Observers * log(Minutes)	-1.3835^{***}	-1.3532^{***}
	(0.0208)	(0.0205)
PRI Observers * log(Minutes)	-3.9264^{***}	-4.5648^{***}
	(0.0294)	(0.0309)
65-Year-Olds * log(Minutes)	-19.3735^{***}	-21.6802^{***}
	(0.2218)	(0.2287)
High School * log(Minutes)	-35.0485^{***}	-51.1750^{***}
	(0.2318)	(0.2632)
Computer * log(Minutes)	-7.4468^{***}	-8.1142^{***}
	(0.0854)	(0.0873)
No Spanish * log(Minutes)	16.0395^{***}	29.1216^{***}
	(0.5039)	(0.4469)
Illiteracy * log(Minutes)	-37.7975^{***}	-53.3439^{***}
	(0.1715)	(0.2230)
θ	1.024	0.984
l-likelihood	-912221.1	-891358.8

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

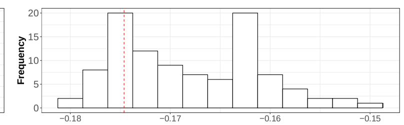
Table H: Cox model on the reporting time for vote counts in all polling stations. Mexico, 2006. Model (1) is the model used for the above analyses, whereas Model (2) excludes *all* outlying observations.



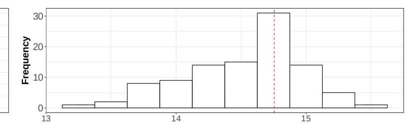
(a) Absent Poll Workers



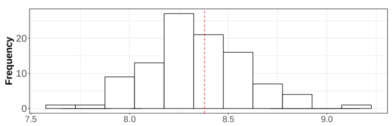
(b) Registered Voters



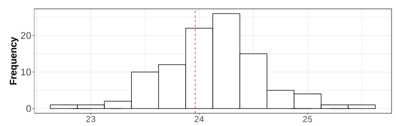
(c) Driving Distance



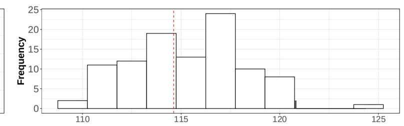
(d) PAN Observers



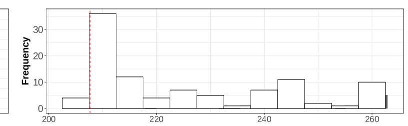
(e) PRD Observers



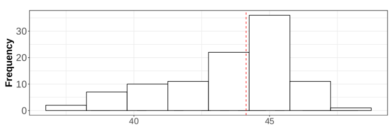
(f) PRI Observers



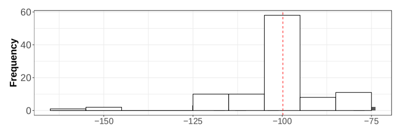
(g) 65-Year-Olds



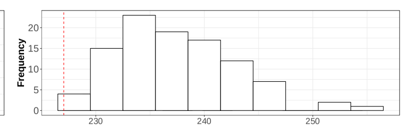
(h) High School



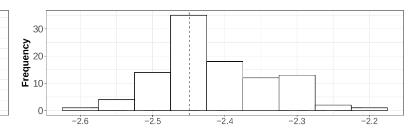
(i) Computer



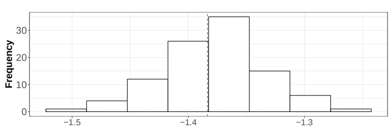
(j) No Spanish



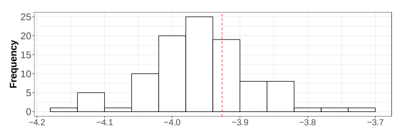
(k) Illiteracy



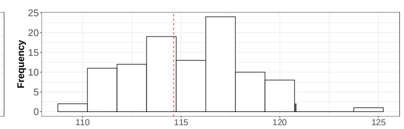
(l) PAN * log(min)



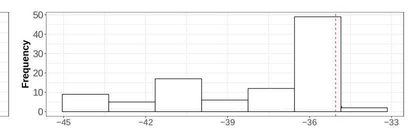
(m) PRD * log(min)



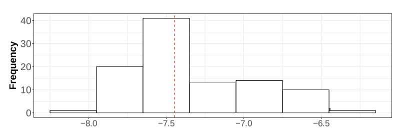
(n) PRI * log(min)



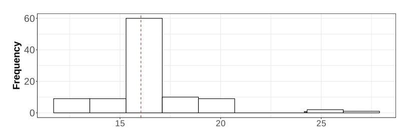
(o) 65-Year-Olds * log(min)



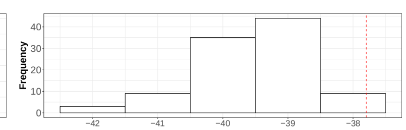
(p) High School * log(min)



(q) Computer * log(min)



(r) No Spanish * log(min)



(s) Illiteracy * log(min)

Figure W: Comparison of Coefficient Estimates from Table 2 to Coefficient Distributions of Bootstrapped Model

Note: Each figure above displays the distribution of each of the Model's Coefficients, derived from the simulations described above. The red vertical line represents the estimated coefficients from our original model.

6.5 k-means approach

This section shows the results of our identification strategy when we cluster observations using the *k-means* classifier. This approach separates observations into a given set of clusters with similar characteristics by estimating the Euclidian distance between observations. Then outliers are those observations with the largest distance from the closest location of the group's mean.

For our goal, we use this approach as follows. First, we estimate the intra-cluster distances between points and its variance to estimate the optimal number of clusters using the "elbow method." This approach finds the number of clusters that explain most of the variance. Then we group the observations into the number of clusters defined in the previous step to then estimate the Euclidean distance between each observation and the mean of its respective cluster, labeling as outliers those observations within the largest 5% distance.

We use the same battery of variables that we employ in our survival model. In this case, we standardize our time measurements to facilitate the clustering. As Figure X below shows, the elbow method suggests that the optimal number of clusters is fifteen. We then group observations in fifteen clusters to then estimate the Euclidean distance between the center of each cluster and the location of their respective observations. The summary statistics of every cluster are below. Finally, we will label as outliers those observations whose Euclidean distances with their cluster's center are in the top five percent.

We replicate the analysis described in the main body of the manuscript considering only those observations classified as outliers in both the *k-means* approach and the deviance analysis. This leaves us to 202 observations. Figure Y shows that the differences between these observations and their respective control units is statistically indistinguishable from zero.

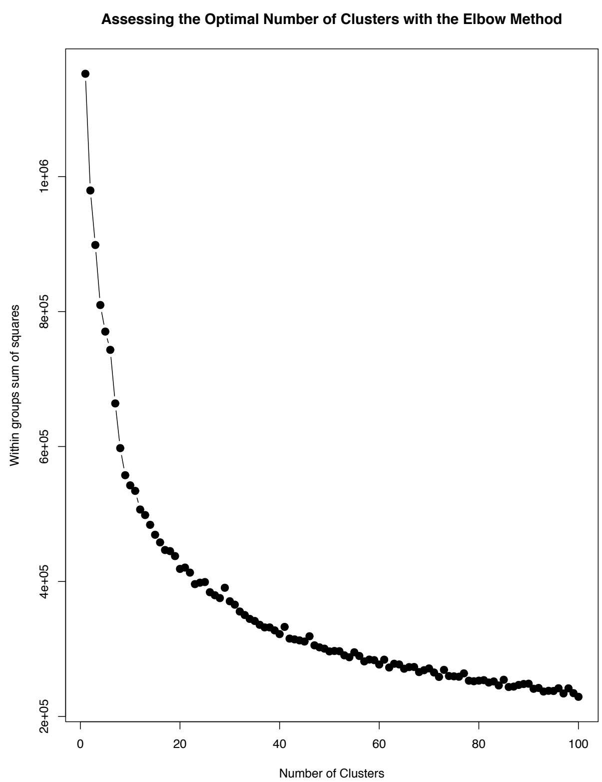


Figure X: K-means clustering Sum of Square Errors (SSE)
Note: The figure shows the Sum of Square Errors of the first 100 clusters using the *k-means* classifier.

Cluster	Duration time	Driving time	Registered voters	Absent poll workers	PAN agents	PRD agents	PRI agents	Computer	Population over 65	Illiteracy	No Spanish
1	360.90	13.99	0.79	0.56	0.97	0.96	1.00	0.18	0.05	0.07	0.00
2	351.92	14.22	0.79	0.19	1.00	0.00	1.00	0.21	0.05	0.07	0.00
3	322.93	6.37	0.80	0.12	0.98	0.92	1.00	0.51	0.05	0.03	0.00
4	348.54	13.22	0.78	0.18	1.00	0.65	0.00	0.28	0.07	0.06	0.00
5	901.62	258.33	0.69	0.18	0.90	0.75	0.93	0.12	0.07	0.10	0.00
6	293.53	5.19	0.69	0.15	0.93	0.79	0.98	0.42	0.14	0.03	0.00
7	357.91	11.71	0.87	0.12	1.00	1.00	1.00	0.16	0.04	0.07	0.00
8	593.82	80.66	0.72	0.13	0.74	0.86	0.96	0.01	0.05	0.34	0.24
9	346.77	12.70	0.79	0.20	0.00	0.00	1.00	0.23	0.06	0.06	0.00
10	508.55	76.24	0.44	0.11	0.89	0.89	0.97	0.04	0.11	0.14	0.00
11	638.74	79.44	0.77	0.14	0.97	0.97	1.00	0.09	0.07	0.11	0.00
12	336.22	9.12	0.78	0.17	0.00	0.86	0.00	0.28	0.07	0.05	0.00
13	341.27	12.10	0.79	0.18	0.00	1.00	1.00	0.21	0.05	0.07	0.00
14	270.10	10.71	0.66	0.12	1.00	1.00	1.00	0.17	0.06	0.07	0.00
15	486.34	51.33	0.73	0.12	0.89	0.93	0.98	0.02	0.06	0.22	0.01

Table I: Summary statistics of the variables by cluster.

This table shows the average values of the variables for the first fifteen largest clusters.

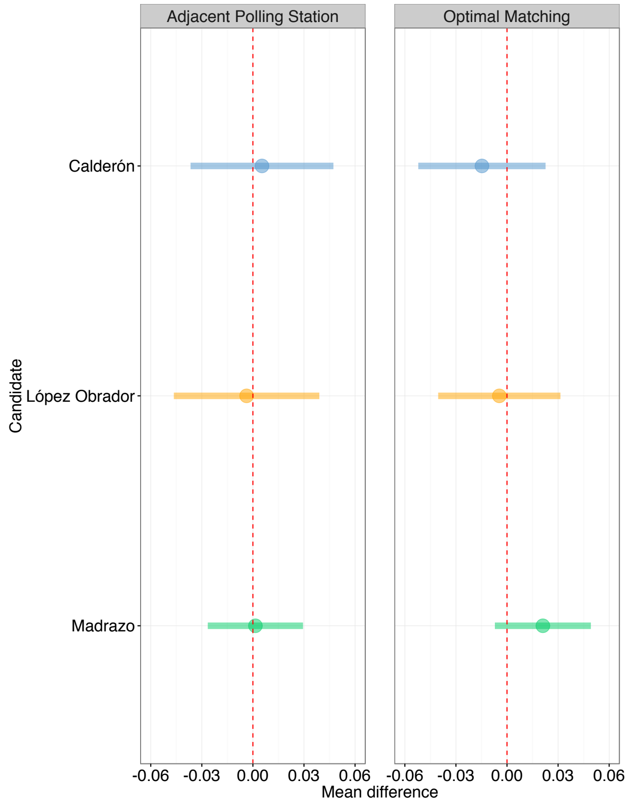


Figure Y: Results
Note: The dots and horizontal lines show the mean differences and 95% confidence intervals for comparing the vote shares of (1) those observations classified as outliers by both the deviance analysis and the *k-means* classifier and (2) their corresponding control units.