Using Election Forensics to Detect Election Fraud in the Philippine Elections, 2016*

Kirill Kalinin† Allen Hicken‡

November 13, 2017

DRAFT: DO NOT CITE WITHOUT PERMISSION

*Prepared for presentation at the 2017 Annual Meeting of the American Political Science Association, San-Francisco, CA, August 31 – September 3, 2017. The authors are thankful to Walter Mebane for his help with the finite mixture model's computations.

†W. Glenn Campbell and Rita Ricardo-Campbell National Fellow at the Hoover Institution, Stanford University (E-mail: kkalinin@stanford.edu)

‡Professor, Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: ahicken@umich.edu).

Abstract

This paper uses the tools of election forensics to investigate electoral fraud in the most recent Philippine national elections, 2016. We first focus on digit tests, finite mixture model and its equivalents. We pay particular attention to the measurement of stolen votes and geographic allocation of election fraud across national elections. We then focus on *Marcos v. Robredo* court case, which helps us to validate some of our research findings for the vice-presidential election.

Keywords: election forensics, election fraud, election forensics toolkit, finite mixture model, Philippines.

Introduction

Election forensics adds distinctive value to current efforts to promote the integrity of elections around the world by developing forensic tools and techniques designed to detect the presence of election fraud and to estimate its magnitude based on the reported results of elections. By utilizing electoral data, election forensics seeks to provide statistical evidence, which could refute or support various sorts of accusations related to electoral quality. Over the years the availability of election forensics tools has significantly expanded starting from digit tests, such as the first digits of aggregate vote totals (Cantu and Saiegh 2011), second significant digits (Pericchi and Torres 2011; Mebane 2011), the last digits in vote counts (Beber and Scacco 2008) or last digits in percentages (Kalinin and Mebane 2013), to various distribution methods proposed by Myagkov, Ordeshook and Shaikin (2009); Shpilkin (2011) and regression analysis (Sobyanin and Sukhovolsky 1995). Most recently, Mebane has developed a positive empirical model of election frauds which enables researchers to parametrically estimate the precinct-level frauds probabilities obtained for an entire election (Mebane 2016). This method, dubbed the finite mixture model, uses sophisticated statistical algorithms in and attempt to move election forensics away from merely descriptive analysis while addressing what are often debatable assumptions underlying existing approaches.

In this paper we employ a variety of these election forensics techniques to analyze the results of the 2016 general elections in the Philippines. Specifically, we investigate the following questions: 1. How clean were the Philippine 2016 elections? 2. Where were the most suspicious patterns in electoral data located? 3. What evidence, if any, is there to support allegations of fraud in the closely fought vice-presidential race, allegations now being played out in a pending court case, *Marcos v. Robredo*?

The structure of this paper is as follows. Section 1 describes the electoral context of 2016 elections. Section 2 explores the literature on election forensics methodology and the methodological tools used in this analysis. Section 3 presents our general findings from empirical data analysis. In the final part, we draw conclusions on the basis of our findings.

1 Context

On May 9, 2016 more than 55 million voters went to the polls to select candidate for an array national and local offices. At the national level contests included the offices of president, vice-president, all members of the House of Representatives and one half of the Senate. At the provincial and municipal levels elections were held for all governor and mayor posts and deputies along with provincial boards and municipal/city councils. Turnout for the election remained high, as is the historical pattern in the Philippines, at just under 82 percent.

The Philippines uses the common single member district plurality system for all executive offices (president, vice-president, governor, vice-governor, mayor, vice-mayor). Unusually, however, voters cast a separate vote for each executive and vice-executive position. Single member district plurality is also used to elect 80 percent of the members of the House of Representatives. The remaining 20 percent are elected via an unusual national party list in which each voter casts a vote for a single party, and each party gets 1 seat for every 2 percent of the vote they obtain, with a maximum of three seats per party. The electoral system employed for the Senate, provincial boards, and municipal or city councils contests is multi-seat plurality. Under this system electoral constituencies have multiple seats, voters can cast as many votes for individual candidates as there are seats, and seats are awarded on a plurality basis. For example, for the Senate election, voters can vote for up to 12 candidates, and the 12 contenders with the most votes each win a seat.

The focal point for the election was, of course, the presidential contest. Without an incumbent or a clear frontrunner, the five-way race was highly competitive, with four serious contenders: Manuel “Mar” Roxas II, standard bearer for the ruling Liberal Party; Jejomar “Jojo” Binay, the sitting vice-president running under the UNA party banner; Grace Poe, an independent candidate and daughter of 2004 presidential hopeful Fernando Poe Jr.; and late-comer, Rodrigo Duterte, standard-bearer for the PDP-Laban party. Prior to Duterte’s entry into the race Roxas, Binay, and Poe were locked in a close battle but Duterte’s entry altered the dynamics of the race. Unusually course and controversial in his rhetoric, Duterte positioned himself as an outsider populist who was willing to do whatever it took to eliminate corruption and crime (especially the scourge of illegal drugs). Backed by an online army of aggressive acolytes, Duterte promised to do for the Philippines what he had done for Davao City as mayor. Early polling placed Duterte last among the four chief contenders, but by May Duterte had established a 5-10 point lead in most national polls. Supporters of Roxas hoped the ruling party's machinery would be sufficient to overcome Duterte's popularity advantage but in the end those hopes were misplaced. Duterte cruised to victory with just under 40 of the vote. Roxas and Poe were distant runners-up at 23.5 and 21.4 percent of the vote respectively (See Table 1).

*** Table 1 about here ***

The race for vice-president proved to be a much closer contest. Throughout the campaign polls showed a tight contest between Congresswoman Maria Leonor "Leni" Robredo, running mate to Roxas, and Ferdinand "Bongbong" Marcos Jr., son of the late dictator and running mate to Miraim Defensor Santiago. The final vote tally was incredibly close, with Robredo winning by a margin of only 263,473 votes (35.1 percent of the votes to Marcos' 34.5 percent).

In the elections for the House of Representatives the ruling liberal party captured a strong plurality of 38.7 percent of the seats (115 seats). The remaining seats were split between 25 other parties, with the second place NPC party securing only 14.1 percent of the seats. Notably, the party of winning presidential candidate Rodrigo Duterte, PDP-Laban, won only three seats. However, once the election was concluded legislators followed the now-familiar pattern of the flocking to the president's party. Once the party-switching dust had cleared PDP-Laban was the largest party in the house, boasting 93 members. Most of the turncoats came from the Liberal party, which fell from 115 seats to 35.

Most observers judged the 2016 election to be a step forward in terms of the electoral integrity. The National Citizens' Movement for Free Elections (NAMFREL) stated that the conduct of the May 9th election was "generally perceived credible and orderly"(NAMFREL 2016). This was echoed by the Carter Center, which stated, "the conduct of the polling, counting and tabulation processes was generally satisfactory" (Center 2016) and the Electoral Integrity Project, which gave counting integrity score (Count) for the election of 74 out of 100 (Norris and Grömping 2017). While the overall electoral integrity picture is good, observers reported electoral malfeasance in certain areas, particularly in parts of Mindanao(Center 2016). For example, 116 polling stations where turnout was 100 percent, polling stations where all voter signatures appeared to be in the same hand, and polling stations where candidates received more than 95 percent of the vote (Center 2016).

Nationally the most notable allegations of electoral fraud came from the Marcos camp. Supporters Marcos noted that the candidate had been leading in the election count for much of the election night, but then saw his lead slip away after a new computer command was inserted into the Commission on Elections' (COMELEC) transparency server during the transmission and counting of the results. Election officials explained that the new code was to fix an error in the printing of candidate names (specifically, changing a "?" to "ñ") but Marcos and his supporters were unconvinced (Santos 2016). They alleged that the administration had used the coding change to ensure victory for their candidate, Leni Robredo. In addition, Marcos claimed that he was the victim of fraud in Lanao del Sur, Basilan, and Maguindanao, where there were seven municipalities in which he received no votes, as well as in 22 other provinces and 5 cities where there was a large number of under votes (ballots where either voters did not cast a vote for vice president, or where there they voted for more than one candidate). Marcos' legal team argued that all of the under votes should be credited to their candidate (Masigan 2017).

Marcos followed up his allegation by filing a formal election protest, calling into question the results from 39,221 clustered precincts composed of a total of 132,446 precincts. Robredo then filed a counter-protest, alleging irregularities in 8,042 clustered precincts composed of 31,278 precincts (Reformina 2017). The protest is currently before the Presidential Electoral Tribunal, with a partial recount set to commence soon.

2 Methodology

Our aim with this paper is to generate estimates of election fraud and then specify the distribution of any fraud detected. There are a variety of measures that can help us to address the difficulty of distinguishing the effects of frauds from the effects of strategic behavior. In our election forensics analysis we use the *Election Forensics Toolkit* a website developed by Walter Mebane and Kirill Kalinin with initial funding from USAID. The website has been specifically designed to make election forensics accessible for the use of policymakers, practitioners, and scholars.[1] While most of the empirical analysis presented here utilizes the Toolkit, we also use additional scripts provided by Walter Mebane to expand our analysis.

Our electoral data comes from the 2016 presidential, vice-presidential and Congress elections in the Philippines. For the purpose of this analysis we use three components of the electoral data: the number of people registered in the elections (the number of people eligible to vote or registered voters), the number of ballots cast (total number of voters), and the number of people supporting the candidate of interest (vote counts). We also calculate *turnout percentage* as the votes cast divided by the number of eligible voters and *candidate's percentage* as the number of electoral votes cast for a candidate divided by the total number of ballots cast.

To assess the extent to which there was fraud in the 2016 election we utilize a variety of statistical methods, which we describe below.

2.1 Digit tests

Digit tests treat the distribution of the digits of the vote count values as indicators of anomalies. They are built on a comparison of the empirical digit distributions with pre-specified theoretical distributions. Among the most popular version of this test are the first digit test of aggregate vote totals (Cantu and Saiegh 2011), the second significant digits test (Pericchi and Torres 2011; Mebane 2011), the last digits in vote counts test (Beber

[1] The link to the website: http://electionforensics.ddns.net:3838/EFT_USAID/

and Scacco 2008) and last digits in percentages test (Kalinin and Mebane 2013). For the first and second digits analysis the distribution of digits is compared to the Benford's Law distribution (i.e. numerical digits appear with different frequencies prespecified by the law). For the last digits analysis the distribution is expected to be uniform.

Beber and Scacco (2012) propose the last-digit test based on the idea that clean vote counts have uniformly distributed 0-9 last digits. The authors also list several conditions, which need to be met in order for the test to work: a) vote counts do not cluster within a narrow range of numbers, and there is minor variation in election unit sizes, electoral support or turnout; b) vote returns must not contain many single- and double-digit counts, i.e. the method should not be applied to the minor candidates with small vote counts or to small polling stations. Once these conditions are met any statistically significant divergence from the uniform distribution can be attributed to fraudulent electoral outcome.

The last digit approach was further extended with a new test of 0s and 5s appearing in the last digit of percentages. This approach argues that the presence of an abnormal proportion of 0s and 5s can be an indication of election fraud. In the case of Russia, for example, 0s and 5s were a mechanism for signaling the loyalty of regional bosses' to the center and of their ability to mobilize the administrative resources to the center's benefit (Kalinin and Mebane 2013). Also consistent with the theory would be political clients (e.g a precinct official) receiving directions about desired vote totals from local bosses, and responding in kind to signal their responsiveness. Regardless, data manipulation is most likely to take place with rounded percentages of turnout and electoral support as this is the easiest to give direction to political clients ("we need 1,200 votes") and the most readily detectable way to signal responsiveness to political principals.

The second digit methods are based on the idea that nonanomalous vote counts must follow the second-digit Benford's Law ($2BL$) distribution. According to Benford's law, the expected relative frequency of the second digit is as follows:

$$\mathrm{q}_0, \ldots, q_9 = (0.120, 0.114, 0.109, 0.104, 0.100, 0.097, 0.093, 0.090, 0.088, 0.085).$$

This approach was developed in the works of Pericchi and Torres (2011); Mebane (2011, 2006) but the method has its critics. For instance, Deckert, Myagkov and Ordeshook (2011) use both simulations and empirical data and find that the 2BL test is susceptible to false positives. Due to the lack of solid underlying theory, it remains unclear how its performance is affected by differences in election rules, strategic voting patterns, the number of competing parties or candidates, etc. In his later work Mebane (2012); Mebane and Klaver (2015) also disavows the 2BL test and finds that the second digits are sensitive to strategic, gerrymandered, and coerced votes, making it hard to disentangle strategic behavior from the effects of election fraud.

We make cautious use of each of the digit methods in this paper. Specifically, to assess the possibility of signaling mechanisms we apply to 0s and 5s in turnout (counts of registered voters and turnout percentage), vote counts (C05s) and vote percentages (P05s). We also utilize 2BL to investigate patterns related to election fraud and/or strategic voting. Since digit tests for the minor candidates can produce biased results we limit our tests to the two candidates with the most votes in presidential and vice-presidential elections and Liberal Party for the Congress elections. Using the Election Forensics Toolkit for each test, we also construct 95% confidence intervals based on the nonparametric bootstrap errors. The signaling analysis utlizes a resampled kernel density method developed by Rozenas (2017).

To detect region-level signaling patterns we also perform a series of tests for the regions, using the Election Forensics Toolkit. Specifically, we use the Election Forensics Toolkit to detect region-level signaling patterns based on vote counts for the winning candidates/parties (unfortunately, the model does not allow to estimate anomalies for non-winning candidates/parties). For cross-regional and cross-district analysis we aggregated the computed precinct-level finite mixture estimates, i.e. probabilities, to the appropriate unit of analysis. Finally, using Mebane (2016) we also produce an estimate of the magnitude of election frauds for all three elections.

2.2 Finite Mixture Likelihood Model

A recent breakthrough in election forensics methodology is the development of a positive empirical model of election frauds proposed by Klimek, Yegorov, Hanel and Thurner (2012). The model suggests that there is a winning party or candidate, i.e. the one with most votes, which benefits from fraud. That is, votes are transferred to the winning party/candidate from other parties/candidates and nonvoters. Mebane (2016), based on Klimek et al. (2012), further develops the model by utilizing a finite mixture likelihood model with three distinct components: probabilities of incremental fraud, extreme fraud and no fraud, calculated for each precinct. The two kinds of election fraud refer to how many of the opposition and nonvoters' votes are transfered: with "incremental fraud" moderate proportions of the votes are transfered; with "extreme fraud" almost all of the votes are transfered.

$$F(\mathbf{W}, \mathbf{O}, \mathbf{A} \mid \mathbf{N}; \mathbf{\Psi}) = \sum_{j \in \{0,i,e\}} f_j \prod_{i=1}^{n} g_{jW}(W_i \mid N_i; \Psi) g_{jA}(A_i \mid N_i; \Psi) \tag{1}$$

In the model f_0, f_i and f_e are the probabilities of no fraud, incremental fraud and extreme fraud, where $f_0 + f_i + f_e = 1$. The model describes the joint density of the observed vote counts for the winning party/candidate W_i, the observed sum of votes cast for all other parties/candidates O_i and the number of observed nonvotes A_i, i.e. $(W_i; O_i; A_i)$ as being conditioned on the number of eligible voters in each precinct N_i and estimated parameters $\mathbf{\Psi} = (\alpha, \nu, \tau, \sigma_\nu, \sigma_\tau, \phi)'$ defined by Klimek's model. The estimated precinct-level parameters are α – the intensity of incremental fraud, ν – the winner's vote proportion and its standard deviation, σ_ν, τ – the turnout and its standard deviation, σ_τ, and θ – incremental fraud garnering a higher number of votes for the leading party. The analytic integration of the finite mixture estimates with the estimates computed from alternative data sources, such as election monitoring reports or postelection complaints, provide the most fruitful strategy for election forensics research (Mebane 2016).

Based on Mebane (2016), we use the finite mixture model's estimates to compute the number of stolen votes for each election race: $M_i = \sum_{i=1}^{n} N_i \hat{f}_{ii} \xi_i$, $M_e = \sum_{i=1}^{n} N_i \hat{f}_{ei} \xi_e$, where $\xi_i = E[x_i(1-\tau_i) + x_i^{\alpha}(1-\nu_i)\tau_i|\hat{\Psi}]$ and $\xi_e = E[(1-y_i)(1-\tau_i) + (1-y_i)^{\alpha}(1-\nu_i)\tau_i|\hat{\Psi}]$[2]

. The computation of the proportion of stolen votes is thus: $p_i = M_i(\sum_{i=1}^{n} V_i)^{-1}$ and $p_e = M_e(\sum_{i=1}^{n} V_i)^{-1}$

2.3 Revised Shpilkin's Method

The third forensics approach used in this paper was developed by Sergey Shpilkin and his associates. This approach holds that number of ballots received by each party is a function of turnout and respective vote share at each polling station. Shpilkin's method] uses a histogram with turnout broken into a series of intervals or bins (x-axis). For each interval the level of electoral support is calculated (y-axis). Even though the original script for this method is unavailable, we managed to replicate its major parts and derive estimates that are approximately similar to Shpilkin's estimates. The algorithm used here is as follows. First, the histogram is constructed using 1% bin for turnout(x-axis) within which the number of votes for a given party is calculated(y-axis). Second, the highest mode located on the left from the official turnout is identified m, which is used as an approximation of the winner's clean votes p_m and the level of real turnout t_m. Third, for each bin the true winner's electoral support is calculated by weighting the votes for all other parties (except the winner's) by the proportion of the incumbent's clean votes p_m across all the bins. In the figure the number of anomalous votes is displayed as the difference between the areas under the curve representing the official electoral support of the winner and calculated "theoretical" curve obtained from the weighting procedure. The numeric measure of election fraud can be computed using the formula: $F_w = N_w - N_{all} \cdot p_m$ where: F_w is the number of fraudulent votes received

[2]For incremental fraud: x – a proportion of what should have been nonvotes are counted for the leading party; x^{α} – a proportion of vote that should have gone to opposition instead go to the leading party. For extreme fraud: $1-y$ – a proportion of the nonvotes counted for the leading party $(1-y)^{\alpha}$ – proportion of opposition going to the leading party.

by the winner, N_w is the number of votes received by the winner, N_{all}–the total number of votes received by all the parties except the winner, p_m is the proportion of the winner's clean votes. The approach is built on too important elements: a) identification of the highest mode located on the left from the official turnout, which serves as an approximation of the winner's clean votes; b) computation of the true winner's electoral support calculated by weighting the votes for all other parties (except the winner's) by the proportion of incumbent's clean votes across all the bins. This approach contains two weaknesses: first, it is highly sensitive to the algorithm by which the highest mode is identified; second, we think that this approach can provide inflated estimates since the observed multimodality can be attributed to the electorate's heterogeneity and differences in electoral behavior by voters across the parties.

In this paper we extend Shpilkin's model by adding several features. First, the algorithm selects all modes $m_{1...m}$ located on the left from the official turnout and computes the average across all the modes. This feature enables us to address the multimodality issues, and the difficulty in selecting the "right" mode, $F_w = \sum_{n=1}^{m}(N_w - N_{all} \cdot p_m) \cdot m^{-1}$. Second, to measure the accuracy of our estimates of anomalies we apply a parametric and nonparametric bootstrap. In the parametric bootstrap both *turnout* and *vote shares* are assumed to be binomial random variables with parameters n and p. The *turnout* is simulated from the binomial distribution with n – number of votes, and p – the precinct-level probability of turnout. The *vote shares* are simulated from the binomial distribution with n equal to the simulated turnout and p equal to the precinct-level probability of vote share (For more detail see Rozenas (2017)). In the nonparametric bootstrap we compute the empirical distribution of the estimates, constructed from random sampling with replacement of the observed dataset. For each simulation we: 1) derive a histogram with turnout broken into a series of intervals, within which the level of electoral support falling into each interval is calculated (y-axis); 2) compute the average value across the modes $m_{1...m}$ located on the left from the official turnout; 3) compute bootstrapped standard errors based on 1000 simulations.

2.4 Geographic clustering tests

Geographic clustering is another forensics tool that can show where there is cooperation or collaboration consistent with electoral frauds. The clustering may suggest the presence of shared data-generating process for a specific locality, such as the home base of a political leader or an area in which the leader's political party or ethnic group is predominant. For our cluster analysis we use local Moran's I_i Anselin (1995), measuring whether the value at observation i differs from the mean of values geographically close to the observation i. To estimate p-values we use permutation test methods and correct them for multiple testing using false discovery rate procedures (Mebane 2015). All geographic clustering tests were implemented using the Election Forensics Toolkit.

2.5 Final thoughts

Each of the methods and tools described above have their disadvantages and drawbacks. Our approach is to examine the electoral results using each of these tools, looking for patterns both within and across the different analyses. A conclusion that fraud likely occurred is more convincing if a) the evidence is strong in a given test and b) if indications of fraud are observed across many different kinds of tests.

3 Analysis

We begin by looking at the presidential race using digit tests. The panel for presidential election in Table 2 indicates the presence of statistically significant anomalies in Turnout (colored in red) when looking at the proportion of last digit 0s and 5s (Hicken and Mebane 2015). For Roxas two forensics indicators are significant: the mean of last digits LastC and C05s. The 2BL indicator is also statistically significant for across for all three levels: turnout, Duterte and Roxas. However, a note of caution is warranted. Recall the 2BL test is sensitive to both election fraud and strategic voting and it is hard to distinguish between the two.

All we can conclude is that the distribution of digits departs from the normal pattern one would expect.

*** Table 2 about here ***

The digit tests for the vice-presidential election are also interesting. The 2BL test is only significant for turnout and the LastC indicator is significant for both of the two leading candidates, Bongbong Marcos and Leni Robredo. The vote counts indicator C05 is also significant for Robredo

Finally, turning to elections for the House of Representatives we some evidence of election anomalies in the votes for the Liberal Party, with significant results for both the 2BL and P05s indicators.

Thus, election anomalies in turnout and votes for candidates of interest are indicative of basic "signaling" mechanism by which subnational actors demonstrate their loyalty to political principals. In the case of the Philippines these principals would most likely be local or provincial political bosses. In order to localize the signaling sources, we compute region-level estimates of election fraud using the Election Forensics Toolkit, and marking only those regions for which statistically significant anomalies are observed.

*** Figures 1 and 2 about here ***

According to Figure 1 anomalies associated with P05 are detected for Turnout in Central Visayas (Region VII), MIMAROPA (Region IV-B) and Northern Mindanao (Region X); for Duterte in the Davao Region (Region XI) and the Autonomous Region of Muslim Mindanao (ARMM); for Roxas in the Autonomous Region of Muslim Mindanao (ARMM), the Davao Region (Region XI), the Ilocos Region (Region I), and MIMAROPA (Region IV-B). Given that Duterte served as a mayor of Davao City prior to running for president and his strong level of support in Mindanao and the Visayas these results are perhaps not surprising.

Figure 2 displays the pattern for the vice-presidential contest. For Robredo anomalies associated with P05s are located in the Cordillera Administrative Region (CAR) and Davao

Region (Region XI); for Marcos in the Autonomous Region of Muslim Mindanao (ARMM), Bicol Region (Region V), and the Ilocos Region (Region I). Finally, for Turnout P05 is significant for the following regions: Autonomous Region of Muslim Mindanao (ARMM), Cordillera Administrative Region (CAR), Davao Region (Region XI), Eastern Visayas (Region VIII), Ilocos Region (Region I) and Western Visayas (Region VI). For Marcos one of the three regions displaying anomalies includes his family's provincial bailiwick, Ilocos Norte, where he served three terms as Governor.

*** Figures 3 and 4 about here ***

Using a resampled kernel density method (Rozenas 2017) we find additional support for the signaling interpretation. As Figures 3(a)–3(d) show, the percentage of precincts with "fraudulent" election results is non-zero. In the presidential elections there are 0.12 percent of the "suspicious " vote shares received by Duterte in individual precincts which are 0.35 and 0.75; for Roxas the percent of anomalous precincts is about 0.07 percent. In the vice-presidential elections the percent of anomalous vote shares for Robredo and Marcos is 0.10% and 0.05%, respectively. Finally, Figure 4 illustrates the presence of anomalies in 0.42 percent of precincts partly due to extreme electoral support for Liberal Democrats in a few precincts. Thus, Rozenas (2017)'s method provides us with extra evidence about the presence of small "signaling" patterns which are relatively higher for the winning candidates than for the losing candidates. This finding implicitly indicates that the former were better in mobilizing their political machines than the latter, as we would expect. Note, however, that this does not mean that the central Liberal Party was the organization doing the coordinating or the receiving the signal. Rather, what we observe is likely a byproduct of the operations of local and provincial machines, the bulk of which were aligned with the Liberal party in 2016.

*** Table 3 about here ***

The results from the digit tests provide some grounds for concern, but recall that these models may be picking up strategic rather than fraudulent behavior. For these reasons we validate these initial results with other methods, starting with the finite mixture model. In contrast to the digit tests, estimates of the finite mixture frauds model Mebane (2016) reported in Table 3 demonstrate the presence of only minor anomalies in presidential and vice-presidential elections. The table reports point estimates $\hat{f}_{\mathrm{i}}$, $\hat{f}_{\mathrm{e}}$, $\hat{\alpha}$, $\hat{\theta}$, $\hat{\tau}$ and $\hat{\nu}$. f_{i} and f_{e} are respectively the probabilities that each precinct is affected by incremental or extreme frauds: for both presidential and vice-presidential only 2 percent of precincts is affected by incremental fraud, while none show any evidence of extreme fraud. The results also suggest ($\alpha > 1$) that where fraud does occur manufacturing votes from nonvoters is more common than vote stealing. A higher value of θ implies that in the few places that incremental fraud does occur it garners a higher number of votes for the leading candidate. τ and ν are respectively the mean turnout proportion and mean proportion of votes for the leading party in the absence of fraud.

*** Figure 5 about here ***

Our analysis using the finite mixture model for the Congress election is much more complicated. Since the Philippines uses a modified mixed-member system in which 238 of the seats are allocated to district representatives, and 59 to party-list representatives, the finite mixture model was estimated using only district candidates (the model's convergence issues in some of districts reduced this number to 225). Since the volatility in the parties' electoral support across districts is extremely high, the finite mixture model was estimated for each district's winner separately. In Figure 5 red color is assigned to the district won by the Liberal Party and black color to everyone else. The figure illustrates no particularly systematic pattern – observed anomalies seem to be only weakly associated with the Liberal Party.

According to Figure 5(a) even though the majority of districts display the absence of election fraud, there is a a subgroup of districts for in which observe patterns of vote stealing patterns(for which $\hat{\alpha} < 1$). These are: 11184, 11256, 11248 etc. These is also a small subset districts displaying evidence of votes manufacturing: 11181, 11216, 11233, 11285. In Figure

5 (b), (c) display districts where there were high probabilities of election fraud. Extreme fraud, $\hat{f}_{\mathrm{e}}$, is extremely rare with only three districts demonstrate a negligible proportion of anomalies (11252, 11346, 11354). For incremental fraud $\hat{f}_{\mathrm{i}}$ there is much more variation than for $\hat{f}_{\mathrm{e}}$; districts with the largest anomalies include 11281, 11263, 11381.

Given that the electoral contests in question were held simultaneously, it is worthwhile to examine whether there are correlations across all three races in terms of the winner's vote proportions, incremental fraud (f_i) and extreme fraud probabilities (f_e). As we would expected, the data on turnout is highly correlated. There is also a moderate positive correlation between f_i measures for the presidential and vice-presidential elections, and a moderate positive correlation between LP support in the House elections and f_i in the vice-presidential elections. All other correlations are statistically or substantively negligible.

*** Figure 6 about here ***

We next examine the geographic distribution of the finite mixture model estimates by calculating the local Moran's I_i for each municipality (the lowest level for which we have the requisite geographic data). Figure 7 displays regional averages of frauds probabilities in Northern Luzon and Mindanao: brighter red color indicates higher levels of election fraud. The Figure illustrates that almost all red spots are uniformly scattered across the areas suggesting that anomalies are clustered in many small localities without any visible distribution patterns. Also, for all three elections there is not much difference in the distribution of clusters between north and south. As expected, Figure 8 demonstrates the absence of extreme frauds, though there are some small clusters in the Illocos and Cordillera regions in Luzon and in Central Mindanao.

*** Figures 7 and 8 about here ***

Using the finite mixture model's estimates for each election, we also compute the number of stolen votes.

*** Table 4 about here ***

Table 4 demonstrates that assuming our finite mixture model is correct, the magnitude of election fraud in all three studied elections is quite small. The vice-presidential race displays the largest magnitude of election fraud amounting to an estimates 91,414 stolen votes, followed by Congress election with 79,799 stolen votes, and finally the presidential election only about 38,704 votes.

*** Tables 5, 6, 7 and 8 about here ***

The precise breakdown of stolen votes into the regions and districts can be found in Tables 5, 6, 7 and 8. For the all three contests the Autonomous Region of Muslim Mindanao (ARMM) stands out as the region with the most stolen votes, in most cases by a large margin. For the presidential election ARMM contains the most stolen votes ($M_i = 15863$, $M_e = 15620$) and followed by REGION XI (Davao Region), ($M_i = 7778$). Both regions were won handily by President Duterte. For the vice-presidential election the region with the most detected frauds is again ARMM ($M_i = 17936$), followed by REGION V (Bicol Region) ($M_i = 16286$) and REGION VI (Western Visayas) ($M_i = 14228$). Robredo was received the most votes in all three regions. Finally, for elections to Congress the regions with most election fraud are ARMM ($M_i = 17706$), REGION X (Nortern Mindanao) ($M_i = 11333$), the National Capital Region (NCR) ($M_i = 8751$) and REGION XII (SOCCSKSARGEN) ($M_i = 8395.00, M_e = 855.00$).

*** Table 9 and Figure 9 about here ***

Finally, we turn to Shpilkin's revised model and for another estimate of the magnitude of election fraud in the 2016 election. According to Table 9, the confidence intervals for all elections are pretty wide, enabling us to conclude that there is not much supporting evidence that the Philippine elections were seriously polluted by the election fraud in 2016.

All together, the presented evidence suggests that the amount of election fraud was quite small for all three studied elections.

4 Robustness check: *Marcos v. Robredo*

As discussed above, the vice-presidential elections has produced and electoral challenge and associated court case: *Marcos v. Robredo*, in which Marcos argues that the Robredo was the beneficiary of electoral fraud perpetrated by the government, to the disadvantage of Marcos.

The following is an extract from the original 1045-pages complaint filed with court: "These irregularities, anomalies, and electoral fraud affected the conduct of the elections and the results in thirty-nine thousand two hundred twenty one (39,221) clustered precincts for the 2016 Elections in the following areas: Lanao Del Sur, Maguindanao, Basilan, Cebu Province, Leyte, Negros Occidental, Negros Oriental, Masbate, Zamboanga Del Sur, Zamboanga Del Norte, Bukidnon, Iloilo Province, Bohol, Quezon Province, Batangas, Western Samar, Misamis Oriental, Camarines Sur, 2nd District of Northern Samar, Palawan, Albay, Zamboanga Sibugay, Misamis Occidental, Pangasinan, Isabela, Iloilo City, Bacolod City, Cebu City, Lapu-lapu City, and Zamboanga City. ...Hence, protestant Marcos files this Election Protest, and hereby seeks that the proclamation of protestee Robredo as the duly elected Vice-President be annulled and set aside for having been made on the basis of COCs[3] that were not authentic. Furthermore, said proclamation constituted an evasion by Congress of its constitutional duty to determine the authenticity and due execution of the COCs."(Protest 2017)

From the list of $39,221$ clustered precincts, we managed to extract $33,751$.[4] Using this auxiliary data from the court case we can determine the decision rule by which the precincts were selected onto the list and investigate the level of fraud in those precincts. This presents an opportunity for us to validate the finite mixture model. Obviously, the decision rule will reflect Marcos's strategic interest to undermine Robredo's victory and expose anomalies in a select group of precincts. In other words, in addition to a validation opportunity this exercise allows us to investigate whether Marcos's claims are supported by our empirical analysis.

[3] Certificates of Canvass

[4] We are still investigating the reason why data for the remaining 5,471 clustered precincts could not be successfully extracted

Suppose the court eventually agrees with Marcos and nullifies the results of in all listed precincts. The result of this action would increase Marcos's electoral support to 39.38% while lowering Robredo's to 29.87%, giving Marcos's a decisive victory. Obviously, this would indicate that anomalies are substantial and contradiction to our election forensics findings. Among the selected precincts how much fraud can Marcos actually claim? To begin with, our analysis indicates that the precincts Marcos included in his challenge were not systematically more fraud prone than those excluded. Selected precincts contain only 10,000 more fraudulent votes than unselected precincts–not enough to change the outcome of the election.

*** Figure 10 about here ***

Figure 10 confirms our expectation of Marcos's decision rule. Regions where Marcos did relatively poorly were more likely to be selected into the list than those regions where he did relatively well.

*** Figure 11 about here ***

Our municipal-level digit analysis shown in Figure 11, doesn't seem to suggest anomalies or and systematic attempt to manipulate the vote for Robredo. The few red spots in the map are more or less uniformly distributed across the municipalities without any suspicious patterns or large signaling clusters to be found.

Our observation of Marcos's precinct-selection strategy is supported by a logistic regression analysis. Our logit model includes Marcos's and Robredo's vote shares, turnout, incremental fraud(f_i)[5] and a set of region dummies (results from regression analysis can be found in Table 11). Moreover, to improve interpretation of the regression coefficients we estimated the first differences using the Zelig package (Imai, King and Lau 2008).

*** Table 10 about here ***

[5]The extreme fraud(f_e), was excluded due to the lack of variation

The findings are reported in Table 10. Controlling statistically for the effects of other variables, the probability that the precinct is selected into the list increases by about 68 percentage points when Robredo's vote shares are shifted from minimum to maximum. Conversely, the probability that the precinct is selected into the list decreases by about 13 percentage points if Marcos's vote shares are shifted from minimum to maximum. The likelihood of precinct being selected to the list decreases by 57 percentage points if turnout changes from minimum to maximum. Finally, the presence of incremental election fraud(f_i) reduces the likelihood of including precinct into the list by about 9 percentage points.

Our findings with respect to incremental fraud chance when we include region dummies. According to model 2 the presence of incremental fraud, f_i increases the likelihood of precinct being selected to the list by about 35 percentage points. Hence, according to this model Marcos's list, indeed, includes more precincts identified by the algorithm as anomalies. The effect of the other variables in the model remain more or less the same. The probability that a precinct is selected into the list increases by about 53 percentage points when Robredo's vote shares are shifted from minimum to maximum and it drops by 32 percentage points when Marcos's vote shares are shifted from minimum to maximum. The effect of turnout on precinct's inclusion reduced to 7 percentage points.

*** Figure 12 about here ***

Our key findings are also displayed in Figure 12. In this Figure subfigures (a) and (b) are based on model 1 and subfigures (c) and (d) are based on model 2. Let's focus on the bottom subfigures. According to subfigure (c) as Marcos's electoral support increases the probability of non-fraudulent precinct being selected into his list ($f_i = 0$) decreases; for the fraudulent precinct the observed effect is statistically insignificant. Now let us turn to subfigure (d). According to this figure, as Robredo's vote shares increase so does the probability of the precinct being included to the list; for the fraudulent precincts, however, the probability of inclusion is much higher than for non-fraudulent precincts. Interestingly, the rates of inclusion are different: as Robredo's support increases, clean precincts are included in the list at a higher rate than the fraudulent precincts – this, perhaps, indicates that the selection of precincts wa based on Robredo's vote shares rather than detected anomalies. This finding confirms our expectations that Marcos's strategic interest is to undermine Robredo's victory by including not only fraudulent precincts, but also clean ones. Since 95% confidence intervals for both types of precincts intersect at Robredo's 100%, their inclusion probability at such extreme values becomes statistically indistinguishable.

5 Conclusion

Our key findings suggest that the Philippine 2016 election are relatively clean. Even though there is some limited evidence suggesting the presence of election fraud their effect on the electoral outcome for the national races is insignificant. In addition to election forensics analysis we also performed validity check by employing the data from *Marcos v. Robredo* court case. Our original findings indicate that Marcos's selection of "suspicious" precincts into the list was supported election forensics analysis. However, to undermine Robredo's victory he seems to be interested in including clean precincts along with fraudulent ones.

Table 1: Election Results

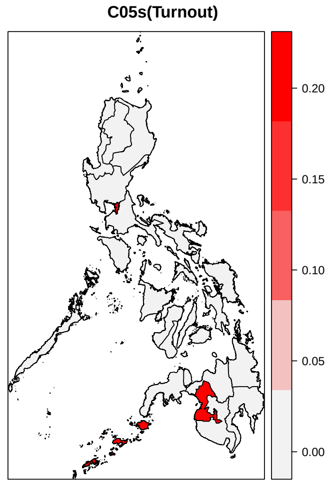
	President	Vice-President	House Seats
PDP-Laban	**Duterte**	Cayetano	
	39%	14.4%	1%
Liberal Party	Roxas	**Robredo**	
	23.5%	**35.1%**	38.7%
UNA	Binay	Honasan	
	12.7%	1.9%	3.7%
PRP/Independent	Santiago	Marcos[a]	
	3.4%	34.5%	—
Independent	Poe	Escudero	
	21.4%	12%	—
Independent		Trillanes	
	—	2.1%	—

Notes: [a]Marcos ran as an independent, but as running mate to Santiago.

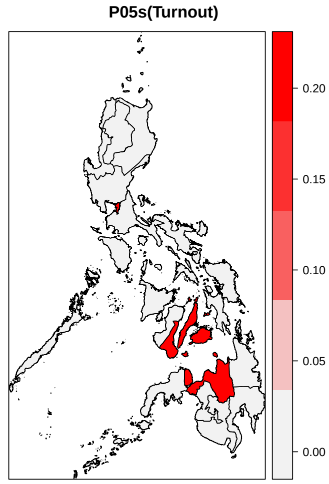
Table 2: Digit Tests

Candidate's Name	2BL	LastC	P05s	C05s
		Presidential Election 2016		
Turnout	4.476	4.495	0.198	0.201
	(4.458, 4.497)	(4.475, 4.513)	(0.195, 0.2)	(0.198, 0.204)
RODY DUTERTE	4.259	4.494	0.2	0.2
	(4.241, 4.279)	(4.474, 4.513)	(0.198, 0.203)	(0.197, 0.202)
MAR ROXAS	4.101	4.463	0.201	0.205
	(4.084, 4.12)	(4.445, 4.481)	(0.198, 0.203)	(0.202, 0.208)
		Vice-Presidential Election 2016		
Turnout	4.46	4.489	0.203	0.201
	(4.44, 4.477)	(4.469, 4.505)	(0.2, 0.205)	(0.198, 0.203)
BONGBONG MARCOS	4.172	4.481	0.199	0.201
	(4.153, 4.191)	(4.461, 4.5)	(0.196, 0.202)	(0.198, 0.204)
LENI ROBREDO	4.172	4.473	0.2	0.203
	(4.154, 4.19)	(4.454, 4.491)	(0.197, 0.203)	(0.2, 0.205)
		Congress Election 2016		
Turnout	4.473	4.554	0.201	0.262
	(4.45, 4.498)	(4.534, 4.575)	(0.197, 0.204)	(0.258, 0.265)
LP	4.354	4.495	0.288	0.203
	(4.333, 4.376)	(4.473, 4.519)	(0.284, 0.291)	(0.199, 0.206)

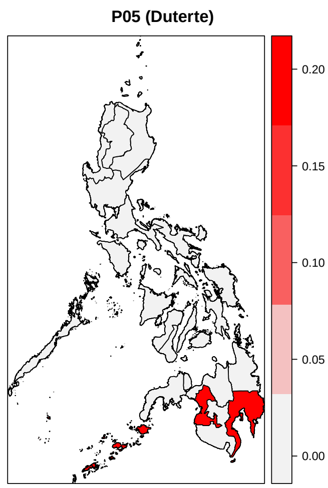
Figure 1: Digit tests, by Region



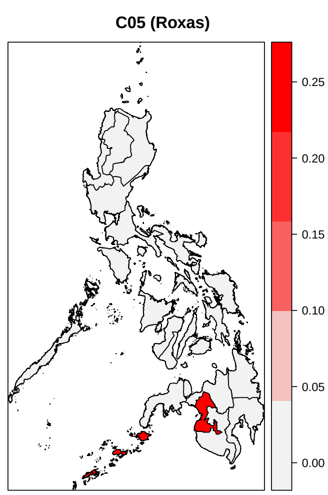
(a)



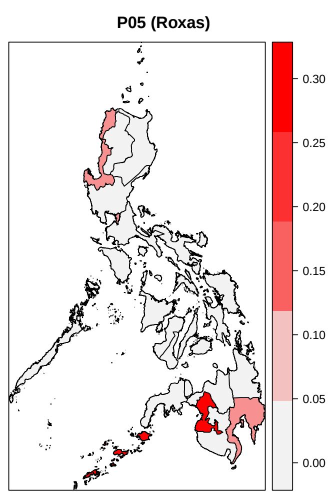
(b)



(c)

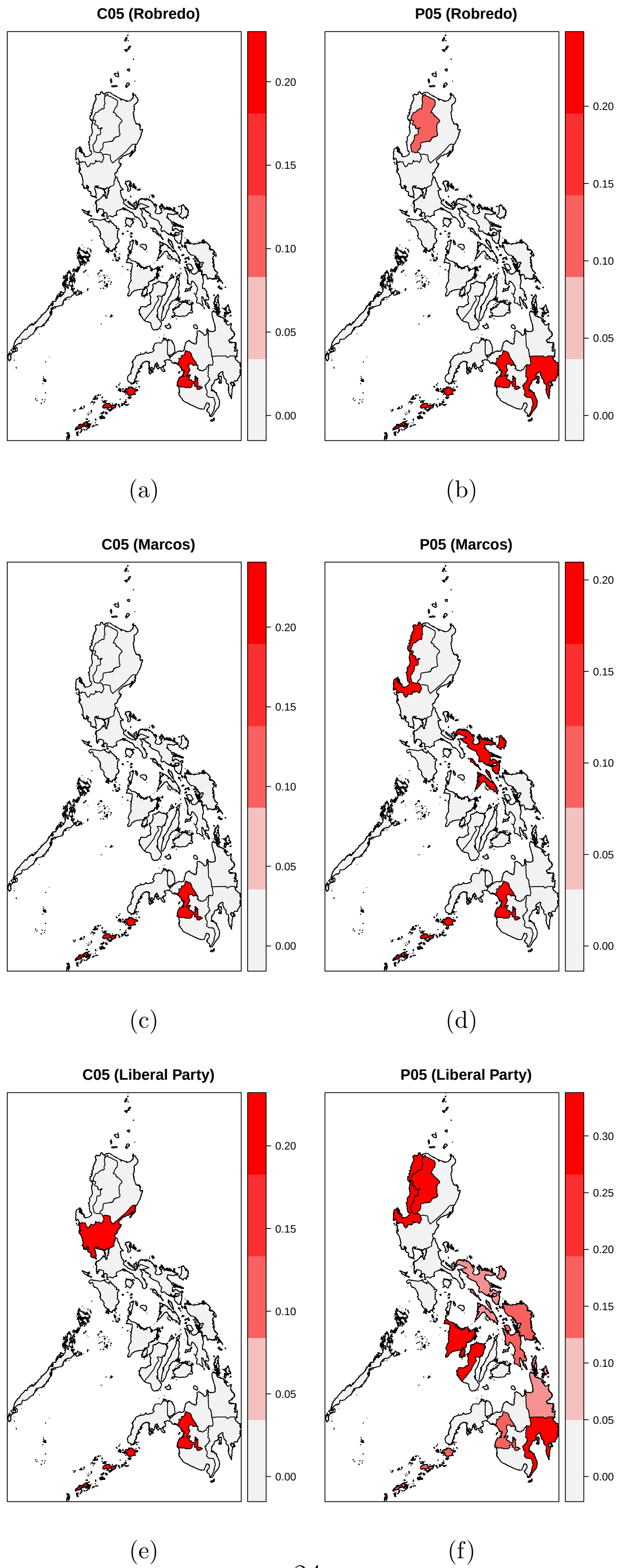


(d)



(e)

Figure 2: Digit tests, by Region

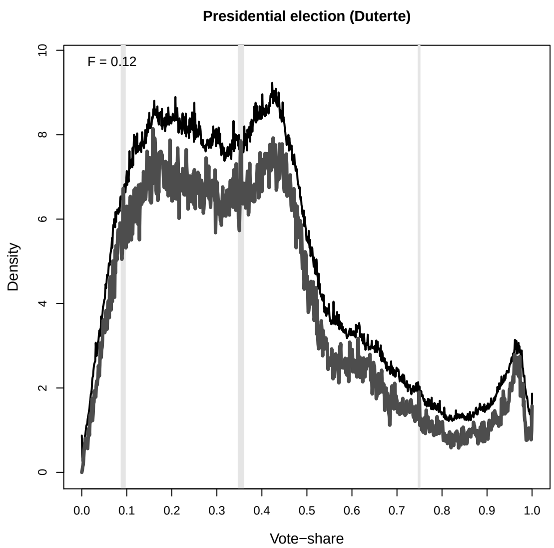


(a) (b)

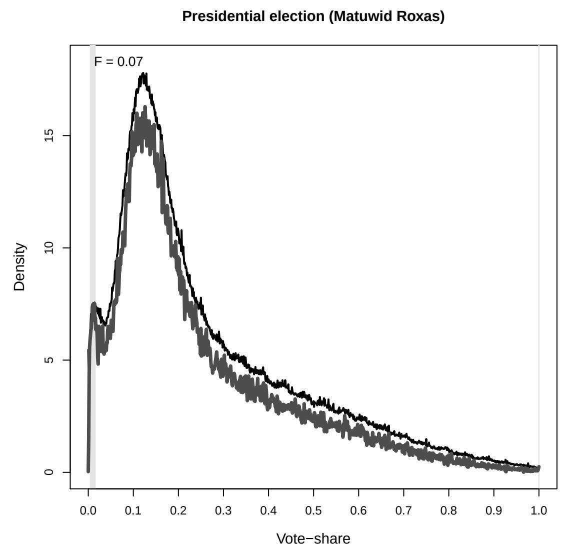
(c) (d)

(e) (f)

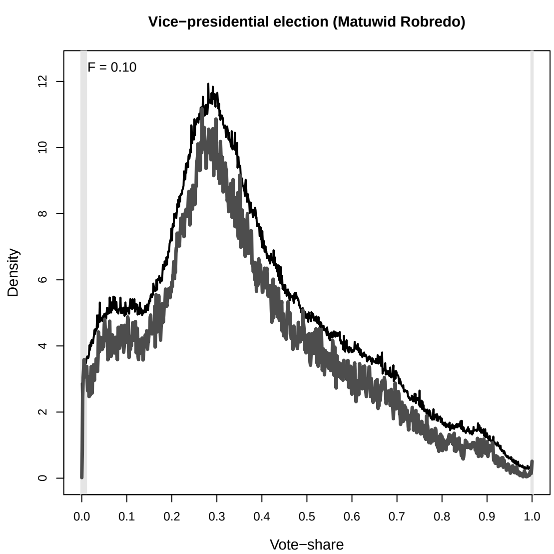
Figure 3: “Spikes” Tests for Vote Proportions



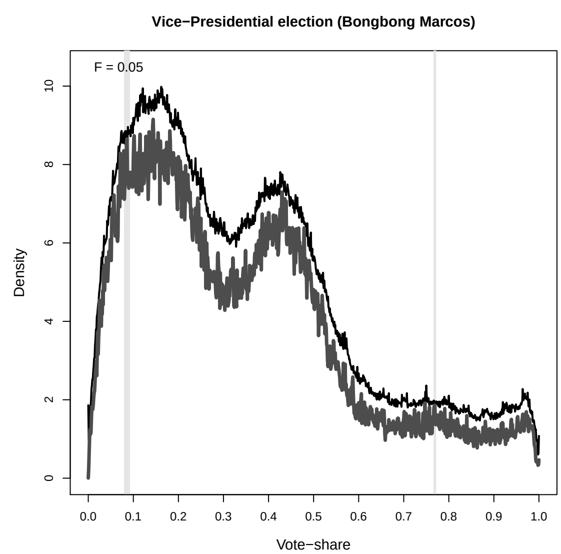
(a)



(b)



(c)



(d)

Figure 4: "Spikes" Tests for Vote Proportions

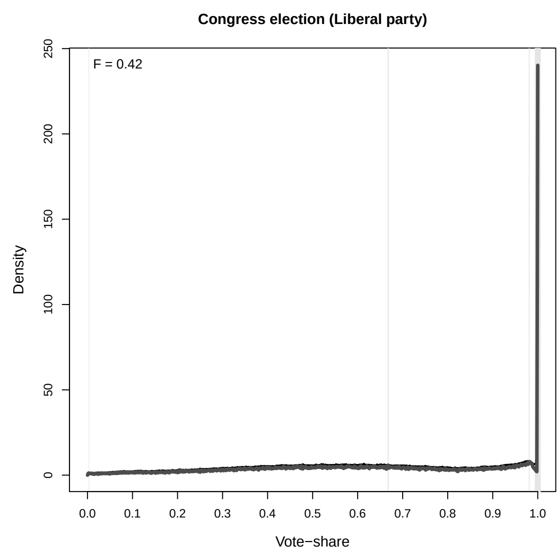
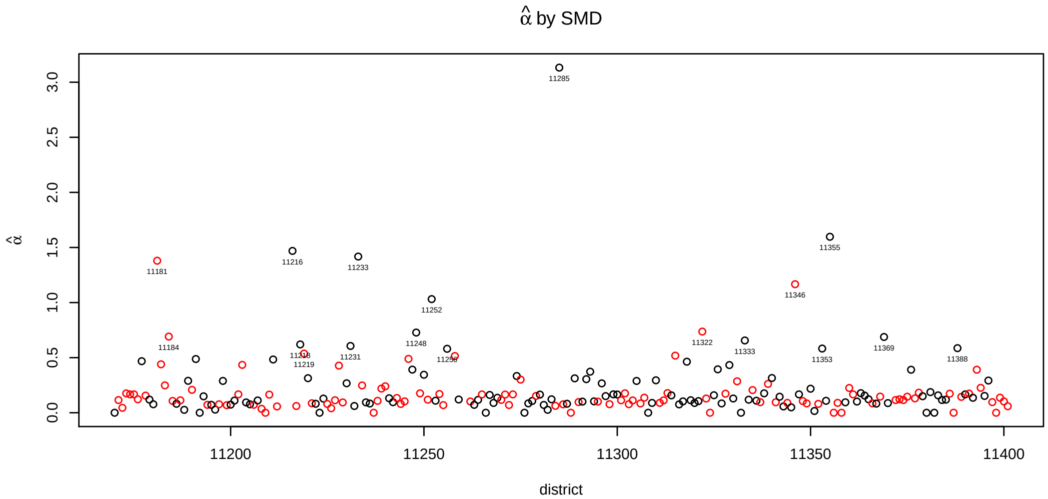


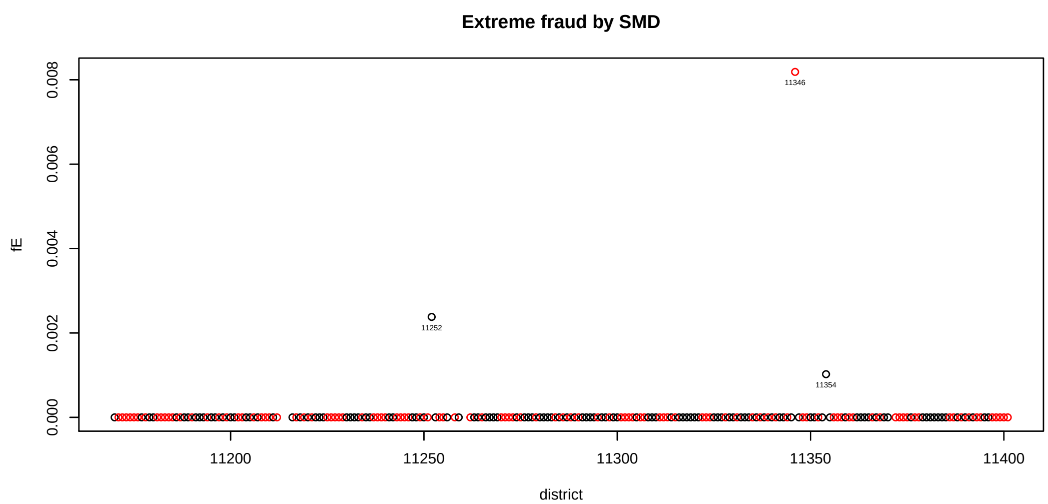
Table 3: Finite Mixture Estimates for Presidential and Vice-Presidential Election 2016

	Presidential	Vice-Presidential
$\hat{f}_i$	0.02	0.02
$\hat{f}_e$	0.00	0.00
$\hat{\alpha}$	3.3	1.7
$\hat{\tau}$	0.78	0.75
$\hat{\nu}$	0.34	0.33
σ	0.17	0.18
$\hat{\theta}$	0.44	0.44
LR	-889604.19	-918563.02
n	89816	90394

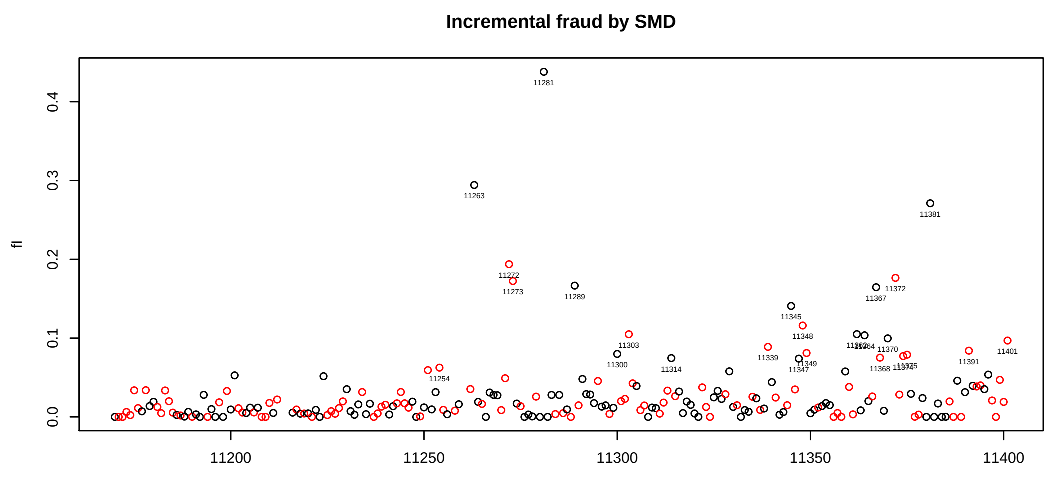
Figure 5: Finite Mixture Model Estimates, Philippines 2016 SMD



(a)



(b)



(c)

Figure 6: Correlations Between Turnout, Winner's Vote Support and Estimated Election Fraud

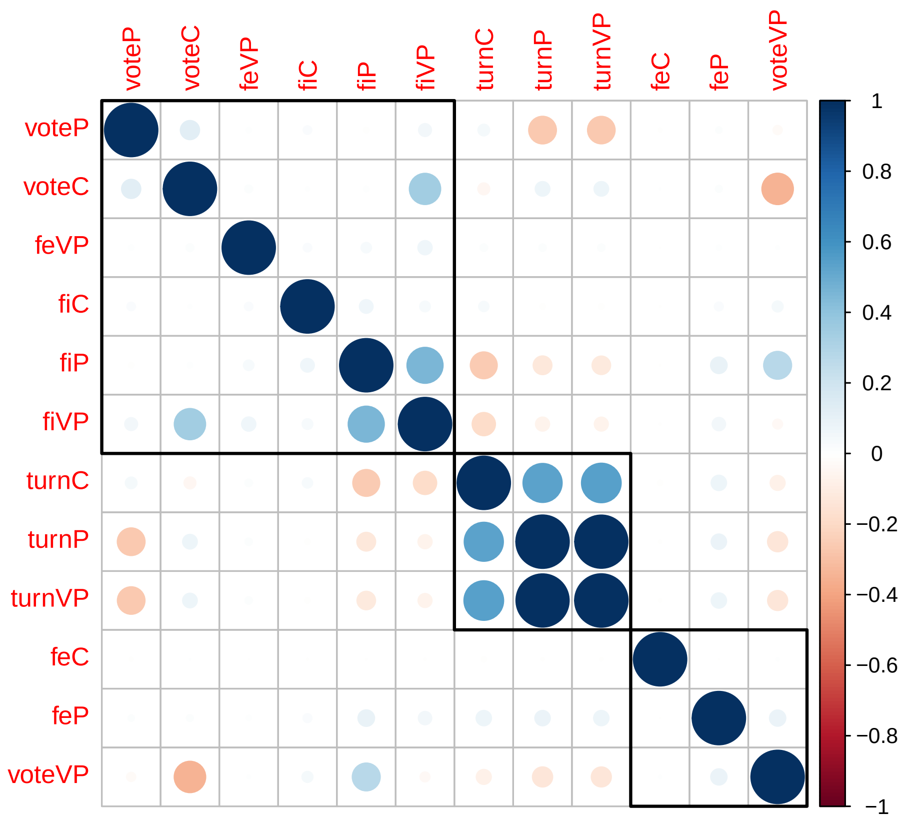
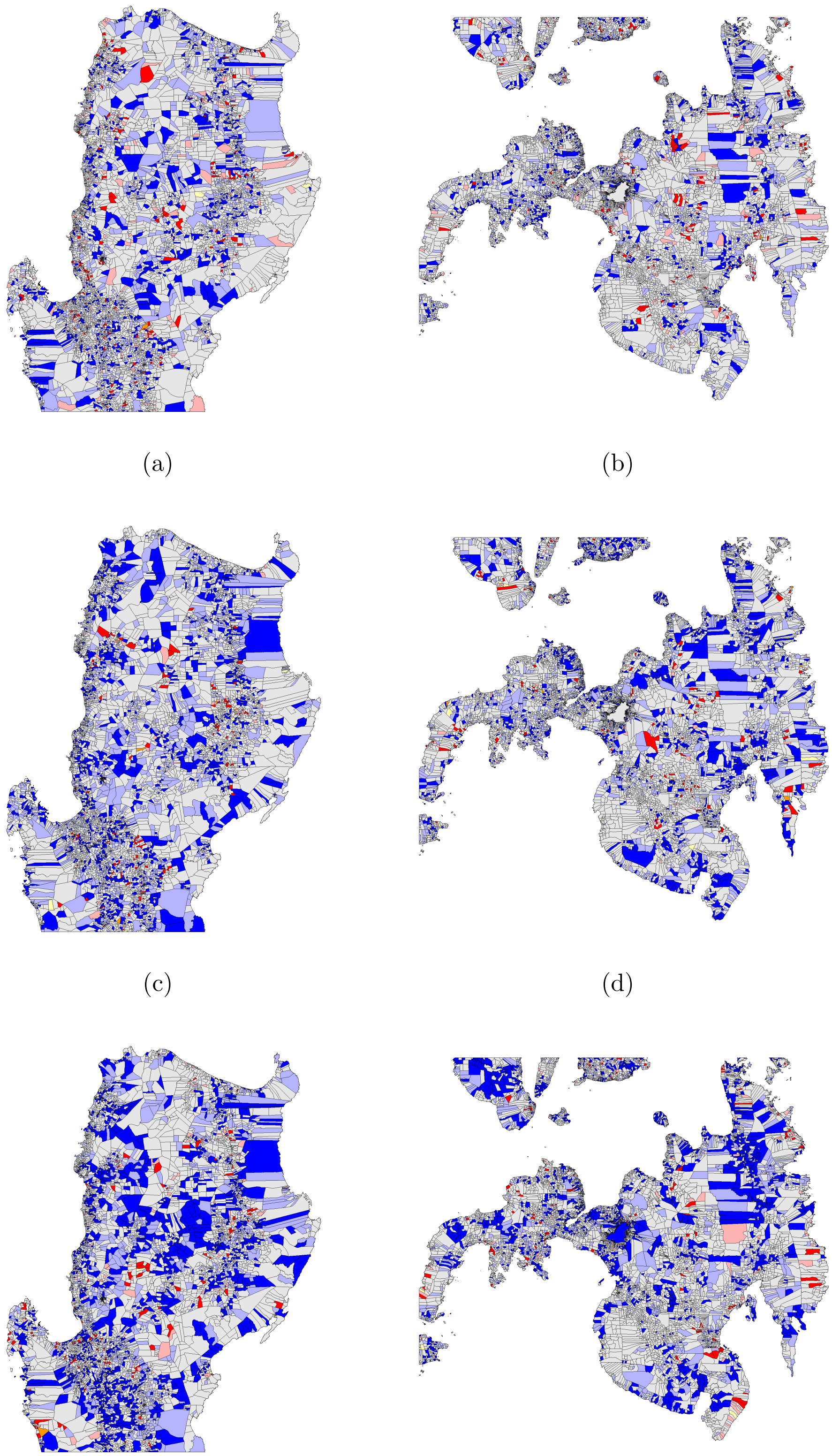


Figure 7: Finite Mixture Model Estimates, f_i, Municipalities



(a) (b)

(c) (d)

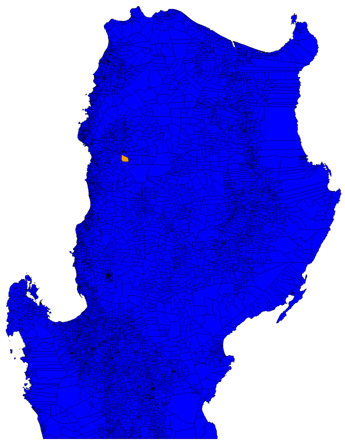
(e) (f)

Notes: (a,b) – vice-presidential election; (c,d) – presidential elections; (e,f) – elections to Congress.

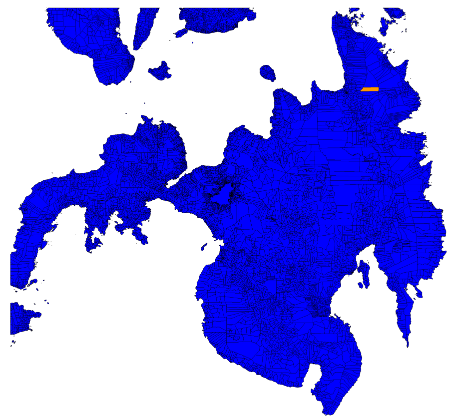
Table 4: Estimated Fraudulent Vote Counts and Proportions for Philippine Elections

Election	M_i	M_e	SE_i	SE_e	p_i	p_e	p_{ie}
Congress	79799	1105	7.16	0.29	0.002	0.00	0.002
Presidential	38704	16381	5.29	0.95	0.001	0.00	0.001
Vice-Presidential	91414	20	6.07	0.006	0.002	0.00	0.003

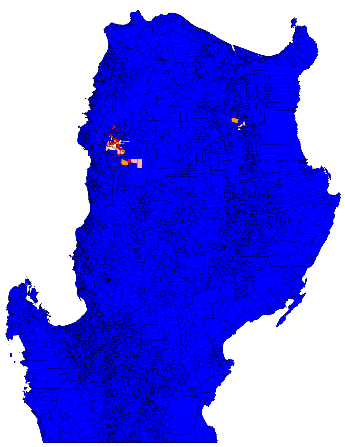
Figure 8: Finite Mixture Model Estimates, f_e, Municipalities



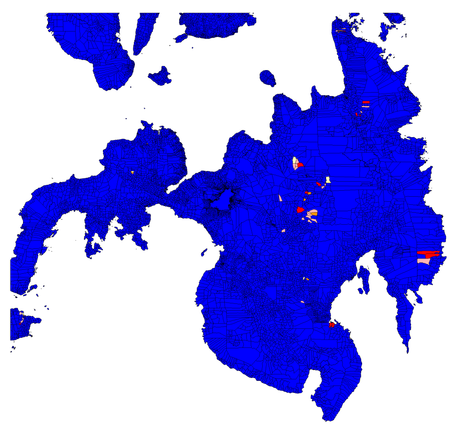
(a)



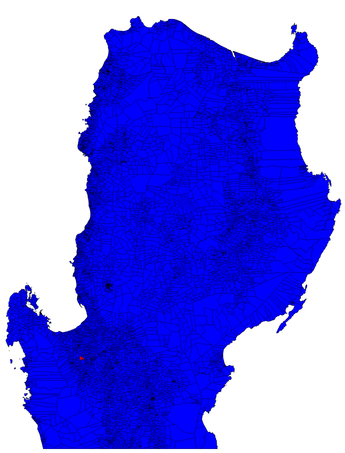
(b)



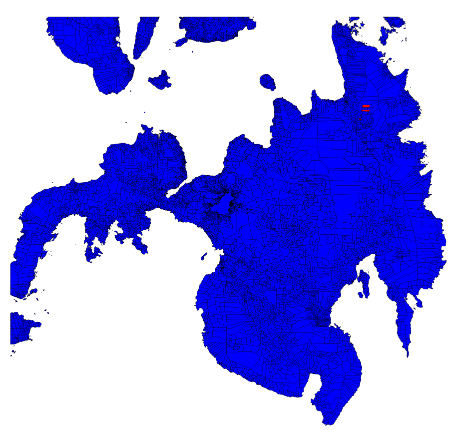
(c)



(d)



(e)



(f)

Notes: (a,b) – vice-presidential election; (c,d) – presidential election; (e,f) – election to Congress.

Table 5: Estimated Fraudulent Vote Counts and Proportions for Philippine Congress Elections, by region

Region	M_i	M_e	p_i	p_e
REGION VI (WESTERN VISAYAS)	1092.00	0.00	0.01	0.00
REGION V (BICOL REGION)	4050.00	0.00	0.02	0.00
AUTONOMOUS REGION IN MUSLIM MINDANAO (ARMM)	17706.00	0.00	0.07	0.00
REGION III (CENTRAL LUZON)	1863.00	135.00	0.01	0.00
REGION II (CAGAYAN VALLEY)	3587.00	111.00	0.03	0.00
REGION IV-A (CALABARZON)	3607.00	0.00	0.01	0.00
CORDILLERA ADMINISTRATIVE REGION (CAR)	973.00	0.00	0.02	0.00
REGION VII (CENTRAL VISAYAS)	1272.00	0.00	0.01	0.00
REGION X (NORTHERN MINDANAO)	11333.00	0.00	0.04	0.00
REGION XI (DAVAO REGION)	1667.00	0.00	0.01	0.00
REGION VIII (EASTERN VISAYAS)	2146.00	0.00	0.02	0.00
REGION I (ILOCOS REGION)	1236.00	0.00	0.01	0.00
NATIONAL CAPITAL REGION (NCR)	8751.00	0.00	0.03	0.00
REGION IV-B (MIMAROPA)	1771.00	0.00	0.03	0.00
NEGROS ISLAND REGION (NIR)	2653.00	0.00	0.03	0.00
REGION XII (SOCCSKSARGEN)	8395.00	855.00	0.04	0.00
REGION XIII (Caraga)	1008.00	0.00	0.03	0.00
REGION IX (ZAMBOANGA PENINSULA)	6048.00	0.00	0.05	0.00

Table 6: Estimated Fraudulent Vote Counts and Proportions for Philippine Presidential Elections, by region

Region	M_i	M_e	p_i	p_e
AUTONOMOUS REGION IN MUSLIM MINDANAO (ARMM)	15863.00	15620.00	0.23	0.00
CORDILLERA ADMINISTRATIVE REGION (CAR)	58.00	0.00	0.00	0.00
NATIONAL CAPITAL REGION (NCR)	1937.00	0.00	0.01	0.00
NEGROS ISLAND REGION (NIR)	149.00	0.00	0.00	0.00
REGION I (ILOCOS REGION)	396.00	0.00	0.00	0.00
REGION II (CAGAYAN VALLEY)	96.00	0.00	0.00	0.00
REGION III (CENTRAL LUZON)	1033.00	0.00	0.00	0.00
REGION IV-A (CALABARZON)	1665.00	0.00	0.00	0.00
REGION IV-B (MIMAROPA)	57.00	0.00	0.00	0.00
REGION IX (ZAMBOANGA PENINSULA)	2013.00	52.00	0.02	0.00
REGION V (BICOL REGION)	85.00	0.00	0.00	0.00
REGION VI (WESTERN VISAYAS)	73.00	0.00	0.00	0.00
REGION VII (CENTRAL VISAYAS)	1411.00	2.00	0.01	0.00
REGION VIII (EASTERN VISAYAS)	358.00	0.00	0.00	0.00
REGION X (NORTHERN MINDANAO)	2387.00	2.00	0.02	0.00
REGION XI (DAVAO REGION)	7778.00	0.00	0.06	0.00
REGION XII (SOCCSKSARGEN)	3025.00	705.00	0.03	0.00
REGION XIII (Caraga)	320.00	0.00	0.01	0.00

Table 7: Estimated Fraudulent Vote Counts and Proportions for Philippine Vice-Presidential Elections, by region

Region	M_i	M_e	p_i	p_e
AUTONOMOUS REGION IN MUSLIM MINDANAO (ARMM)	17936.00	20.00	0.14	0.00
CORDILLERA ADMINISTRATIVE REGION (CAR)	331.00	0.00	0.01	0.00
NATIONAL CAPITAL REGION (NCR)	4010.00	0.00	0.01	0.00
NEGROS ISLAND REGION (NIR)	7820.00	0.00	0.04	0.00
REGION I (ILOCOS REGION)	1058.00	0.00	0.00	0.00
REGION II (CAGAYAN VALLEY)	528.00	0.00	0.00	0.00
REGION III (CENTRAL LUZON)	4708.00	0.00	0.01	0.00
REGION IV-A (CALABARZON)	5987.00	0.00	0.01	0.00
REGION IV-B (MIMAROPA)	1588.00	0.00	0.01	0.00
REGION IX (ZAMBOANGA PENINSULA)	2186.00	0.00	0.02	0.00
REGION V (BICOL REGION)	16286.00	0.00	0.06	0.00
REGION VI (WESTERN VISAYAS)	14228.00	0.00	0.07	0.00
REGION VII (CENTRAL VISAYAS)	4180.00	0.00	0.02	0.00
REGION VIII (EASTERN VISAYAS)	3084.00	0.00	0.02	0.00
REGION X (NORTHERN MINDANAO)	3732.00	0.00	0.02	0.00
REGION XI (DAVAO REGION)	615.00	0.00	0.00	0.00
REGION XII (SOCCSKSARGEN)	2790.00	0.00	0.02	0.00
REGION XIII (Caraga)	347.00	0.00	0.01	0.00

Table 8: Estimated Fraudulent Vote Counts and Proportions for Philippine Congress Elections, by district

District	M_i	M_e	p_i	p_e
N11170	0.00	0.00	0.00	0.00
N11171	0.00	0.00	0.00	0.00
N11172	0.00	0.00	0.00	0.00
N11173	6.00	0.00	0.01	0.00
N11174	38.00	0.00	0.00	0.00
N11175	686.00	0.00	0.03	0.00
N11176	50.00	0.00	0.01	0.00
N11177	24.00	0.00	0.01	0.00
N11178	0.00	0.00	0.03	0.00
N11181	471.00	0.00	0.01	0.00
N11179	299.00	0.00	0.01	0.00
N11183	430.00	0.00	0.03	0.00
N11180	184.00	0.00	0.02	0.00
N11182	47.00	0.00	0.00	0.00
N11184	67.00	0.00	0.02	0.00
N11185	138.00	0.00	0.01	0.00
N11186	0.00	0.00	0.00	0.00
N11187	15.00	0.00	0.00	0.00
N11189	59.00	0.00	0.01	0.00
N11188	11.00	0.00	0.00	0.00
N11190	0.00	0.00	0.00	0.00
N11192	0.00	0.00	0.00	0.00
N11191	0.00	0.00	0.00	0.00

N11193	346.00	0.00	0.03	0.00
N11196	1.00	0.00	0.00	0.00
N11195	0.00	0.00	0.01	0.00
N11194	0.00	0.00	0.00	0.00
N11197	135.00	0.00	0.02	0.00
N11198	0.00	0.00	0.00	0.00
N11200	123.00	0.00	0.01	0.00
N11199	696.00	0.00	0.03	0.00
N11201	927.00	0.00	0.05	0.00
N11203	50.00	0.00	0.01	0.00
N11202	39.00	0.00	0.01	0.00
N11208	2.00	0.00	0.00	0.00
N11206	1.00	0.00	0.01	0.00
N11204	14.00	0.00	0.00	0.00
N11207	104.00	0.00	0.01	0.00
N11205	89.00	0.00	0.01	0.00
N11209	0.00	0.00	0.00	0.00
N11211	19.00	0.00	0.01	0.00
N11210	150.00	0.00	0.02	0.00
N11212	190.00	0.00	0.02	0.00
N11219	17.00	0.00	0.00	0.00
N11218	24.00	0.00	0.00	0.00
N11217	193.00	0.00	0.01	0.00
N11216	131.00	0.00	0.01	0.00
N11226	145.00	0.00	0.01	0.00
N11221	0.00	0.00	0.00	0.00
N11222	203.00	0.00	0.01	0.00

N11223	0.00	0.00	0.00	0.00
N11224	413.00	0.00	0.05	0.00
N11220	29.00	0.00	0.00	0.00
N11227	32.00	0.00	0.00	0.00
N11228	44.00	0.00	0.01	0.00
N11225	0.00	0.00	0.00	0.00
N11229	128.00	0.00	0.02	0.00
N11230	570.00	0.00	0.04	0.00
N11231	15.00	0.00	0.01	0.00
N11232	71.00	0.00	0.00	0.00
N11234	321.00	0.00	0.03	0.00
N11235	39.00	0.00	0.00	0.00
N11233	513.00	0.00	0.02	0.00
N11236	105.00	0.00	0.02	0.00
N11237	0.00	0.00	0.00	0.00
N11238	110.00	0.00	0.00	0.00
N11239	50.00	0.00	0.01	0.00
N11240	8.00	0.00	0.02	0.00
N11241	3.00	0.00	0.00	0.00
N11243	119.00	0.00	0.02	0.00
N11242	138.00	0.00	0.01	0.00
N11248	0.00	0.00	0.00	0.00
N11245	125.00	0.00	0.02	0.00
N11247	323.00	0.00	0.02	0.00
N11246	75.00	0.00	0.01	0.00
N11244	296.00	0.00	0.03	0.00
N11249	3.00	0.00	0.00	0.00

N11252	67.00	110.00	0.01	0.00
N11251	1230.00	0.00	0.06	0.00
N11250	36.00	0.00	0.01	0.00
N11253	370.00	0.00	0.03	0.00
N11254	606.00	0.00	0.06	0.00
N11256	29.00	0.00	0.00	0.00
N11255	137.00	0.00	0.01	0.00
N11259	37.00	0.00	0.02	0.00
N11258	321.00	0.00	0.01	0.00
N11262	531.00	0.00	0.04	0.00
N11264	65.00	0.00	0.02	0.00
N11263	8050.00	0.00	0.29	0.00
N11266	0.00	0.00	0.00	0.00
N11265	356.00	0.00	0.02	0.00
N11271	495.00	0.00	0.05	0.00
N11267	320.00	0.00	0.03	0.00
N11270	15.00	0.00	0.01	0.00
N11268	142.00	0.00	0.03	0.00
N11269	94.00	0.00	0.03	0.00
N11273	4298.00	0.00	0.17	0.00
N11272	10041.00	0.00	0.19	0.00
N11274	74.00	0.00	0.02	0.00
N11275	18.00	0.00	0.01	0.00
N11276	0.00	0.00	0.00	0.00
N11277	44.00	0.00	0.00	0.00
N11279	6.00	0.00	0.03	0.00
N11278	0.00	0.00	0.00	0.00

N11280	0.00	0.00	0.00	0.00
N11282	0.00	0.00	0.00	0.00
N11281	2609.00	0.00	0.44	0.00
N11283	352.00	0.00	0.03	0.00
N11284	57.00	0.00	0.00	0.00
N11285	212.00	0.00	0.03	0.00
N11287	214.00	0.00	0.01	0.00
N11286	0.00	0.00	0.00	0.00
N11289	1898.00	0.00	0.17	0.00
N11288	0.00	0.00	0.00	0.00
N11290	0.00	0.00	0.01	0.00
N11297	69.00	0.00	0.01	0.00
N11294	2.00	0.00	0.02	0.00
N11295	82.00	0.00	0.05	0.00
N11292	175.00	0.00	0.03	0.00
N11291	646.00	0.00	0.05	0.00
N11296	43.00	0.00	0.01	0.00
N11293	751.00	0.00	0.03	0.00
N11299	212.00	0.00	0.01	0.00
N11298	77.00	0.00	0.00	0.00
N11300	596.00	0.00	0.08	0.00
N11301	538.00	0.00	0.02	0.00
N11303	2880.00	0.00	0.10	0.00
N11302	252.00	0.00	0.02	0.00
N11304	37.00	0.00	0.04	0.00
N11305	46.00	0.00	0.04	0.00
N11306	196.00	0.00	0.01	0.00

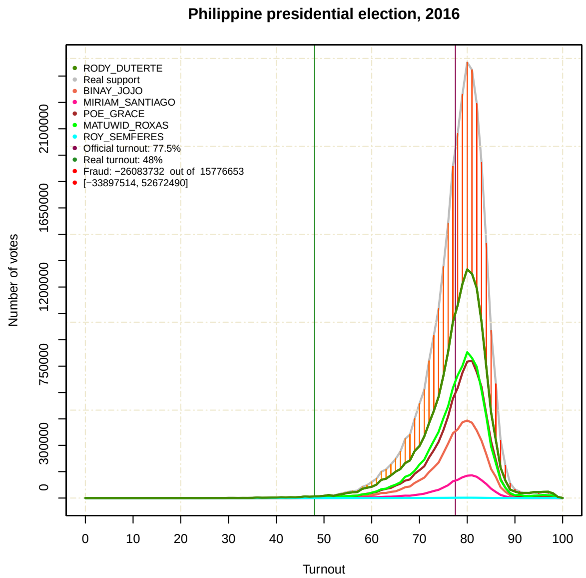
N11308	0.00	0.00	0.00	0.00
N11309	196.00	0.00	0.01	0.00
N11307	210.00	0.00	0.01	0.00
N11310	127.00	0.00	0.01	0.00
N11311	61.00	0.00	0.00	0.00
N11312	88.00	0.00	0.02	0.00
N11313	203.00	0.00	0.03	0.00
N11316	69.00	0.00	0.03	0.00
N11314	745.00	0.00	0.07	0.00
N11315	482.00	0.00	0.03	0.00
N11317	84.00	0.00	0.00	0.00
N11320	17.00	0.00	0.00	0.00
N11319	293.00	0.00	0.02	0.00
N11318	104.00	0.00	0.02	0.00
N11321	0.00	0.00	0.00	0.00
N11322	584.00	0.00	0.04	0.00
N11325	127.00	0.00	0.02	0.00
N11326	14.00	0.00	0.03	0.00
N11323	75.00	0.00	0.01	0.00
N11324	0.00	0.00	0.00	0.00
N11329	245.00	0.00	0.06	0.00
N11330	106.00	0.00	0.01	0.00
N11327	81.00	0.00	0.02	0.00
N11328	662.00	0.00	0.03	0.00
N11331	1.00	0.00	0.01	0.00
N11332	0.00	0.00	0.00	0.00
N11335	154.00	0.00	0.03	0.00

N11334	6.00	0.00	0.01	0.00
N11333	132.00	0.00	0.01	0.00
N11336	126.00	0.00	0.02	0.00
N11337	12.00	0.00	0.01	0.00
N11338	286.00	0.00	0.01	0.00
N11339	206.00	0.00	0.09	0.00
N11341	26.00	0.00	0.02	0.00
N11340	535.00	0.00	0.04	0.00
N11343	1.00	0.00	0.01	0.00
N11342	3.00	0.00	0.00	0.00
N11344	366.00	0.00	0.01	0.00
N11346	417.00	862.00	0.03	0.00
N11345	3794.00	0.00	0.14	0.00
N11347	1358.00	0.00	0.07	0.00
N11348	1583.00	0.00	0.12	0.00
N11350	92.00	0.00	0.00	0.00
N11349	405.00	0.00	0.08	0.00
N11352	96.00	0.00	0.01	0.00
N11351	335.00	0.00	0.01	0.00
N11353	58.00	0.00	0.01	0.00
N11355	26.00	0.00	0.01	0.00
N11354	412.00	133.00	0.02	0.00
N11356	0.00	0.00	0.00	0.00
N11358	0.00	0.00	0.00	0.00
N11357	22.00	0.00	0.00	0.00
N11359	379.00	0.00	0.06	0.00
N11360	343.00	0.00	0.04	0.00

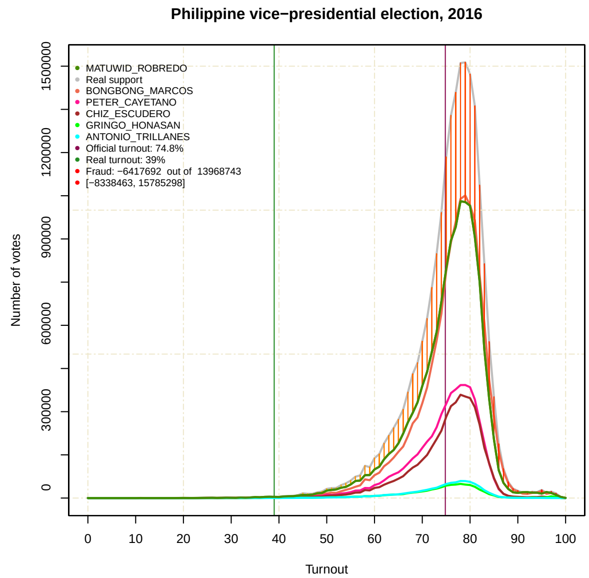
N11361	37.00	0.00	0.00	0.00
N11362	2332.00	0.00	0.11	0.00
N11363	216.00	0.00	0.01	0.00
N11365	495.00	0.00	0.02	0.00
N11364	1408.00	0.00	0.10	0.00
N11366	262.00	0.00	0.03	0.00
N11367	323.00	0.00	0.16	0.00
N11368	373.00	0.00	0.08	0.00
N11369	101.00	0.00	0.01	0.00
N11370	766.00	0.00	0.10	0.00
N11372	721.00	0.00	0.18	0.00
N11375	553.00	0.00	0.08	0.00
N11373	164.00	0.00	0.03	0.00
N11374	332.00	0.00	0.08	0.00
N11376	111.00	0.00	0.03	0.00
N11377	6.00	0.00	0.00	0.00
N11378	2.00	0.00	0.00	0.00
N11379	75.00	0.00	0.02	0.00
N11380	0.00	0.00	0.00	0.00
N11381	3196.00	0.00	0.27	0.00
N11382	0.00	0.00	0.00	0.00
N11383	34.00	0.00	0.02	0.00
N11384	0.00	0.00	0.00	0.00
N11385	0.00	0.00	0.00	0.00
N11386	196.00	0.00	0.02	0.00
N11387	0.00	0.00	0.00	0.00
N11388	906.00	0.00	0.05	0.00

N11389	0.00	0.00	0.00	0.00
N11391	476.00	0.00	0.08	0.00
N11390	79.00	0.00	0.03	0.00
N11392	25.00	0.00	0.04	0.00
N11393	234.00	0.00	0.04	0.00
N11394	65.00	0.00	0.04	0.00
N11395	493.00	0.00	0.04	0.00
N11396	180.00	0.00	0.05	0.00
N11397	37.00	0.00	0.02	0.00
N11398	0.00	0.00	0.00	0.00
N11399	627.00	0.00	0.05	0.00
N11400	180.00	0.00	0.02	0.00
N11401	67.00	0.00	0.10	0.00

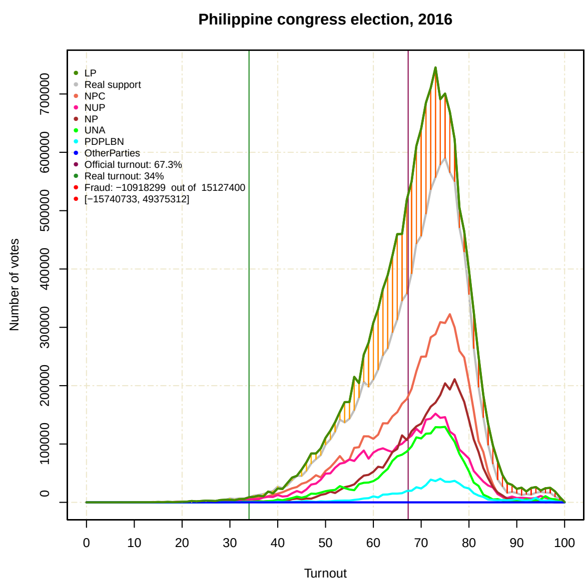
Figure 9: Shpilkin's Modified Method



(a)



(b)



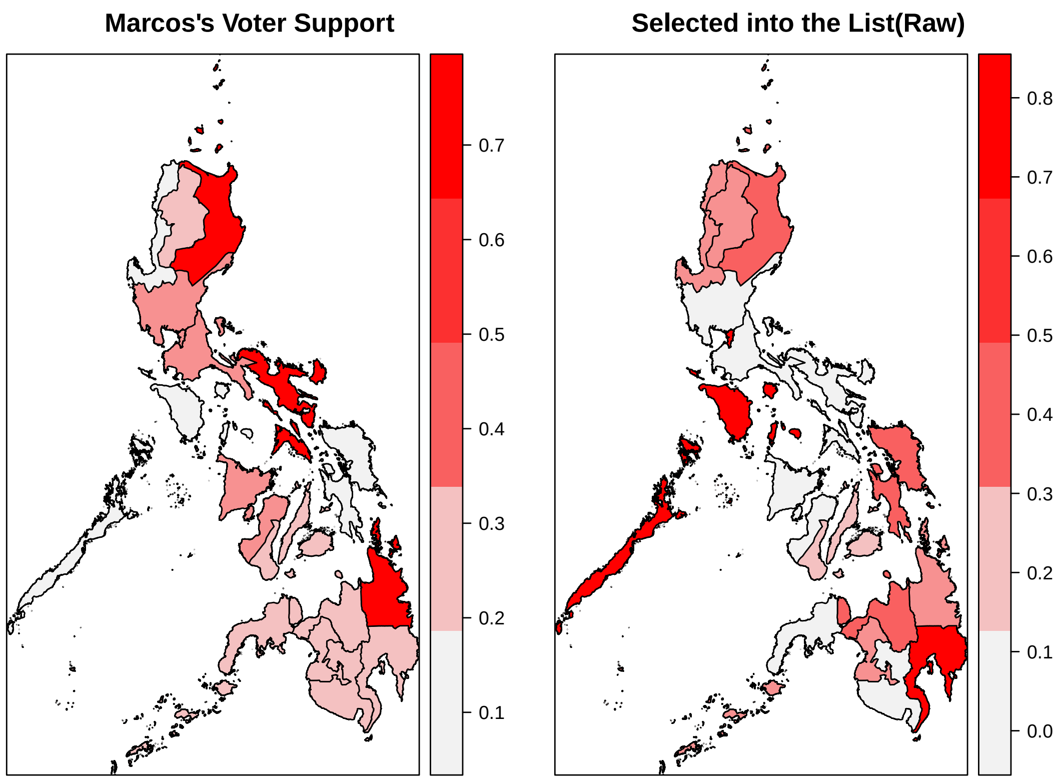
(c)

Table 9: Shpilkin's Modified Method

Name	**Method**	M_v	**95% CI**		$\mathbf{M_t}$	**95% CI**	
			Presidential Election 2016				
DUTERTE	nonp	-26083732	-33897514	52672490	0.48	0.44	0.5
DUTERTE	param	-34347080	-46715294	51675316	0.49	0.48	0.54
			Vice-Presidential Election 2016				
ROBREDO	nonp	-6417692	-8338463	15785298	0.39	0.33	0.4
ROBREDO	param	-5510106	-7166414	12929195	0.4	0.36	0.42
			Congress Election 2016				
LIBERAL PARTY	nonp	-10918299	-15740733	49375312	0.34	0.3	0.36
LIBERAL PARTY	param	-10464511	-19733809	22220895	0.3	0.24	0.3

Notes: Results from 1000 bootstrap parametric/nonparametric simulations (param/nonp) using Shpilkin's method. Computed anomalies with respect to the candidate: M_v – a point estimate for the number of votes produced by "election frauds"; M_t – "true" turnout estimated by the algorithm.

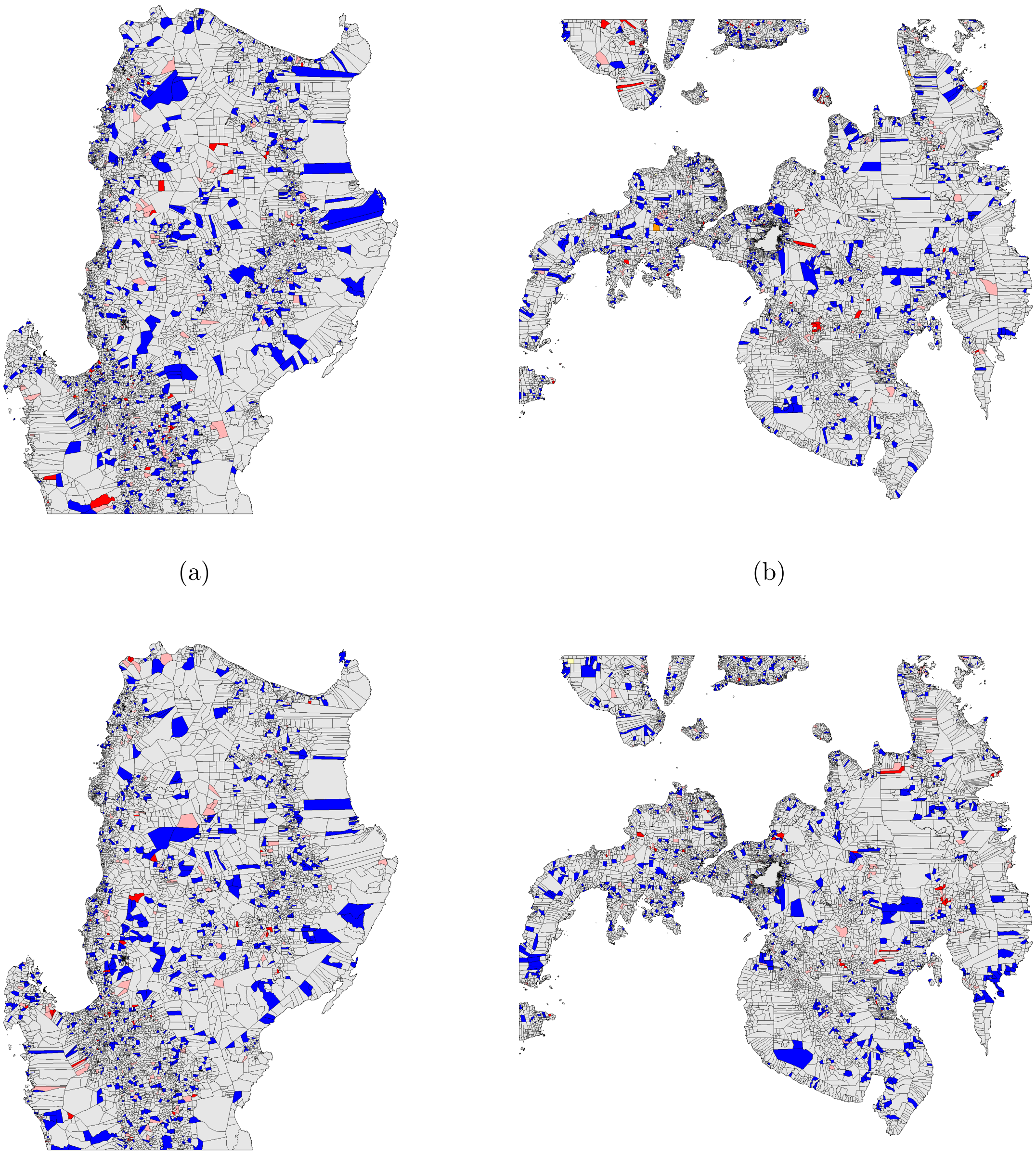
Figure 10: Proportion of Selected Precincts



(a)

(b)

Figure 11: Finite Mixture Model Estimates, Signaling Strategies(Robredo), Municipalities



(a) (b)

(c) (d)

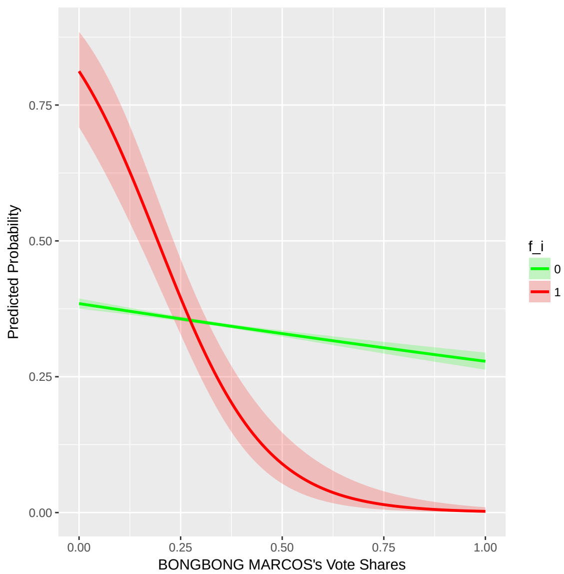
Notes: (a,b) – 0 and 5s in Robredo's vote counts; (c,d) – 0 and 5s in Robredo's vote percentages.

Table 10: Effect of Vote Shares, Turnout and Estimated Fraud on Being Selected into the MARCOS's list, First Differences

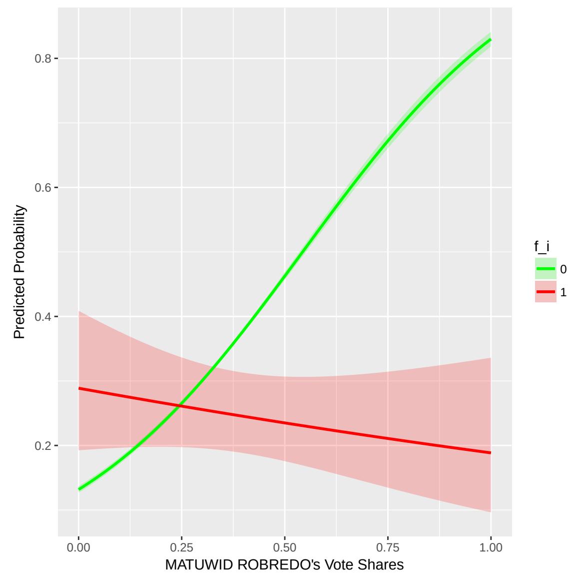
	Mean	SD	50%	2.5%	97.5%
	Regression without region dummies				
MARCOS	-0.134	0.012	-0.134	-0.159	-0.111
ROBREDO	0.683	0.008	0.684	0.667	0.699
Turnout	-0.573	0.016	-0.573	-0.604	-0.542
f_i	-0.093	0.0327	-0.094	-0.155	-0.029
	Regression with region dummies				
MARCOS	-0.321	0.013	-0.321	-0.349	-0.296
ROBREDO	0.525	0.017	0.524	0.491	0.561
Turnout	-0.066	0.026	-0.066	-0.118	-0.014
f_i	0.348	0.050	0.349	0.246	0.440

Notes: Results from 1000 simulations. Original model: $logit(Claim) = \beta_0 + \beta_1 VoteShare_{MARCOS} + \beta_3 VoteShare_{ROBREDO} + \beta_4 Turnout + \beta_5 f_i \cdot VoteShare_{MARCOS} + \beta_6 f_i \cdot VoteShare_{ROBREDO} + \beta_7 f_i \cdot Turnout$

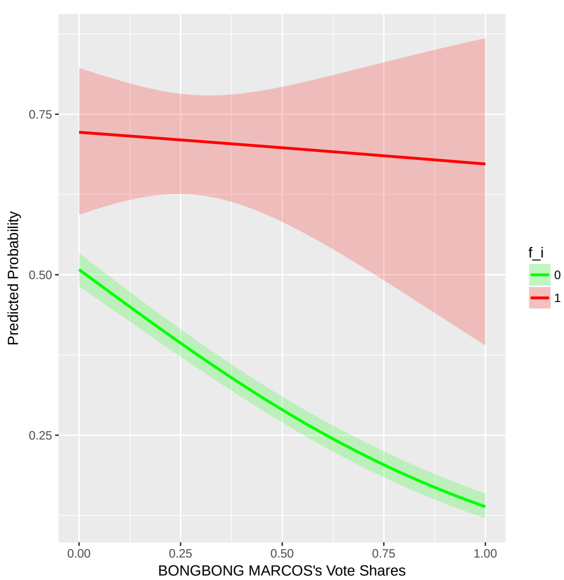
Figure 12: Effect of Candidate's Vote Shares on Being Selected Into the List Conditional on Estimated Fraud f_i



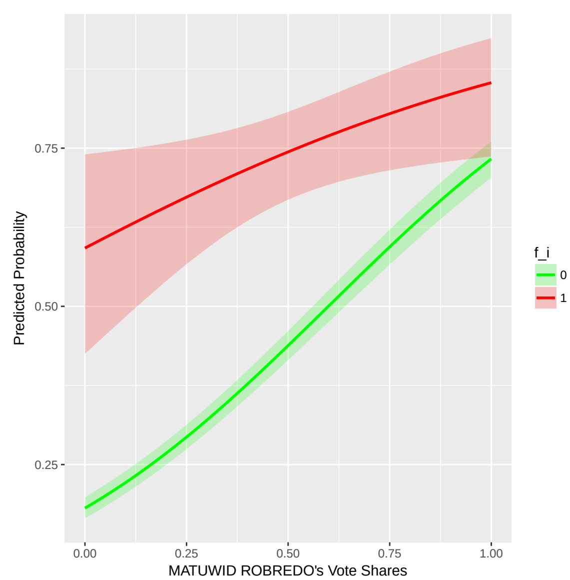
(a)



(b)



(c)



(d)

Table 11: Effect of Vote Shares, Turnout and Estimated Fraud on Being Selected into the MARCOS's list, Binary Logit

	M(01)	M(02)	M(03)
Intercept	0.31	-0.6	-0.41
	(0.07)	(0.11)	(0.11)
MARCOS	-0.48	-1.86	-1.71
	(0.06)	(0.09)	(0.1)
ROBREDO	3.47	2.52	2.54
	(0.06)	(0.09)	(0.09)
Turnout	-2.73	-0.38	-0.66
	(0.1)	(0.13)	(0.13)
f_i	0.23	-0.86	-0.41
	(0.44)	(0.46)	(0.48)
MARCOS X f_i	-7.08	1.62	1.27
	(1)	(0.8)	(0.82)
ROBREDO X f_i	-4.03	-1.13	-0.99
	(0.58)	(0.63)	(0.66)
Turnout X f_i	4.24	2.93	1.6
	(0.37)	(0.45)	(0.47)
Region-dummies	No	Yes	Yes
Interactions:			
Region-dummies X f_i	No	No	Yes
Log-likelihood	-52794	-34026	-33702
Obs.	90399	89960	89960

References

Anselin, Luc. 1995. “Local Indicators of Spatial Association – LISA.” *Geographical Analysis* 27(2):93–115.

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

Beber, Bernd and Alexandra Scacco. 2012. “What the Numbers Say: A Digit-Based Test for Election Fraud.” *Political Analysis* 20(2):211–234.

Cantu, Francisco and Sebastian M. Saiegh. 2011. “Fraudulent Democracy? An Analysis of Argentina’s Infamous Decade Using Supervised Machine Learning.” *Political Analysis* 19(4):409–433.

Center, Carter. 2016. Limited Election Observation Mission to the Philippines. Technical report The Carter Center.
URL: *http://bit.ly/2vlAUQu*

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

Hicken, Allen and Walter R. Mebane, Jr. 2015. “A Guide to Election Forensics.” Working paper for IIE/USAID subaward #DFG-10-APS-UM, “Development of an Election Forensics Toolkit: Using Subnational Data to Detect Anomalies”.

Imai, Kosuke, Gary King and Olivia Lau. 2008. “Toward A Common Framework for Statistical Analysis and Development.” *Journal of Computational and Graphical Statistics* 17:892–913.

Kalinin, Kirill and Walter R. Mebane, Jr. 2013. “Understanding Electoral Frauds through Evolution of Russian Federalism: the Emergence of Signaling Loyalty.” Paper prepared for the Annual Meeting of Midwest Political Science Association, Chicago.

Klimek, Peter, Yuri Yegorov, Rudolf Hanel and Stefan Thurner. 2012. "Statistical Detection of Systematic Election Irregularities." *Proceedings of the National Academy of Sciences of the United States of America* 109(41):16469–16473.

Masigan, Andrew J. 2017. "Marcoss election protest: Who really wins?" *Business World Online* .
URL: *http://bit.ly/2vuvfqE*

Mebane, Jr., Walter R. 2006. "Election Forensics: Vote Counts and Benfords Law." Paper prepared for the 2006 Summer Meeting of the Political Methodology Society, UC-Davis, July 20-22.

Mebane, Jr., Walter R. 2011. "Comment on "Benford's Law and the Detection of Election Fraud"." *Political Analysis* 19(3):269–272.

Mebane, Jr., Walter R. 2012. "Second-digit Tests for Voters Election Strategies and Election Fraud." Prepared for presentation at the 2012 Annual Meeting of the Midwest Political Science Association, Chicago, April 1215, 2010.

Mebane, Jr., Walter R. 2015. "Election Forensics Toolkit." Working paper for IIE/USAID subaward #DFG-10-APS-UM, "Development of an Election Forensics Toolkit: Using Subnational Data to Detect Anomalies".

Mebane, Jr., Walter R. 2016. "Election Forensics: Frauds Tests and Observation-level Frauds Probabilities." Prepared for presentation at the 2016 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 7–10.

Mebane, Jr., Walter R. and Joseph Klaver. 2015. "Election Forensics: Strategies versus Election Frauds in Germany." Prepared for presentation at the 2015 Annual Conference of the European Political Science Association, Vienna, Austria.

Myagkov, Mikhail, Peter C. Ordeshook and Dimitry Shaikin. 2009. *The Forensics of Election Fraud: With Applications to Russia and Ukraine*. New York: Cambridge University Press.

NAMFREL. 2016. NAMFREL Assessment of the May 9, 2016 Elections. Technical report National Citizen's Movement for Free Elections (NAMFREL). Press Statement.
URL: *http://bit.ly/2wD1JUL*

Norris, Pippa and Max Grömping. 2017. "Perceptions of Electoral Integrity, (PEI-5.0)", doi:10.7910/DVN/KI5WB4, , V1, UNF:6:RIRLjEG5jWANoJgXAG7lMQ==." *Harvard Dataverse, V1* . UNF:6:RIRLjEG5jWANoJgXAG7lMQ==.

Pericchi, Luis Raul and David Torres. 2011. "Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela." *Statistical Science* 26(4):502–516.

Protest, Election. 2017. "Election Protest: Marcos vs. Robredo." Court Case.

Reformina, Ina. 2017. "Marcos pays full P66-M for electoral protest vs Robredo." *ABS-CBN News* .
URL: *http://bit.ly/2ufQXjf*

Rozenas, Arturas. 2017. "Detecting Election Fraud from Irregularities in Vote-Share Distributions." *Political Analysis* 25(1):41–56.

Santos, Eimor P. 2016. "Hush-hush: The hash code explained." *CNN* .
URL: *http://bit.ly/2w9w9e2*

Shpilkin, Sergei. 2011. "Statistika issledovala vybory: Statisticheskij analiz vyborov v Gosdumu 2011 goda pokazyvaet vozmozhnye fal'sifikacii (Statistics examined elections: Statistical analysis of elections to the State Duma in 2011 indicates possible fraud)." gazeta.ru (in Russian).
URL: *http://bit.ly/2q4Jdgc*

Sobyanin, Alexander and Vladislav Sukhovolsky. 1995. *Demokratiya, ogranichennaya fal'sifikatsiyami: vybory i referendumy v Rossii v 1991-1993 gg.* Moscow: .