An evaluation of last digit-based test as a tool for electoral fraud detection

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Asma Hussein
Bachelor Thesis in Statistics, 15 hp
At the Department of Economics, Fall of 2018
Supervisor: Mattias Sundén

Abstract

In the pursuit of developing reliable tools for electoral fraud detection, tools that use statistical analysis have become very popular. Specifically, methods of digits pattern analysis, of election results, based on observations such as ‘Benford’s law’ have been deemed especially promising tools in electoral fraud detection. However, some versions of this digit pattern analysis have received a fair share of scrutiny. This paper will focus on evaluating the use of ‘last place’ digit pattern analysis, a method that has been shown to be the most promising in detecting electoral fraud by previous literature. By application to the 2018 parliamentary election in Sweden, where there is no reason to suspect fraud, and to the Ugandan presidential election of 2016 where a fraud-free election is unlikely; we find that the last digit pattern analysis failed to distinguish between fraudulent and non-fraudulent elections. Giving reason to question the usefulness of last place digit analysis.

Table of content

Introduction

Free and fair election is the cornerstone of a democracy. However, not all elections are conducted without manipulation resulting in fraudulent and unfair elections. Election manipulation is a prominent problem considering how you can manipulate votes at different levels, if one controls the bureaucracies that count the votes, they can easily be manipulated with little physical trace. Additionally, in the modern day the 'traditional' route of fake ballots isn't necessarily needed to manipulate an election. Rather "All that may be needed nowadays is access to an input port and a few lines of computer code" as Walter R. Mebane puts it in "Election Forensics: Vote Counts and Benford's Law"(2006, p. 1).

Some, including Mebane, have developed and applied methods to be able to detect if electoral fraud has occurred via mathematical analysis. One such method is based on analyzing the patterns of digits in vote counts and builds on observations such as 'Benfords law'; the observations that certain place digits (in numerical data) have certain frequency distributions. Benfords law has been shown to be a promising tool for this purpose by numerous studies. The prospect of being able to, reliably, detect electoral fraud by simple statistical analysis using just the vote counts would be very useful. Not only would it cost less, since you only need to obtain vote counts in order to perform digit analysis. But unlike traditional methods of electoral fraud detection, like election commission that are often assigned by the sitting government or even independent observers such as the European Union's election monitors (known as the Election observation missions, abbreviated EUEOMs); digit-based test of electoral fraud is independent of human error and factors such as unobserved fraud.

Still, the uses of digit-pattern analysis as a form of electoral fraud detection has been shown to not always do what it aims to. Both first- and second-digit pattern analysis have been shown to detect fraud where there has been non (Shikano & Mack, 2011; Deckert, Myagkov, & Ordeshook, 2011).

This brings into question the usefulness of digit-pattern based tests of fraud. This paper will focus on evaluating the use of last digit patter analysis, specifically; a method that has been shown to be the most promising in detecting anomalies (Diekmann, 2007). When using a last digit test to detect fraudulent vote counts, it is assumed that the last digit frequency of non-fraudulent vote counts follows a certain (uniform) distribution. We should therefore be able to find such a pattern in an election where there are no real suspicions of fraud. Correspondingly, we should be able to find that digit-patterns of vote counts in an election where authentic vote counts are unlikely deviate from the 'expected' distribution of non-fraudulent vote counts. The purpose of this paper is to tests this assumption by application on real election data.

In the following sections Benfords law and digit analysis-based fraud detection will be briefly introduced, and previous research in the area will be reviewed. This will be followed by an attempt to evaluate the last digit test, as presented in Beber & Scacco (2012), by application on real-world data, namely the Swedish parliamentary election of 2018 and the Ugandan 2016 presidential election. The former does not need a digit test to determine its validity, likewise; the latter does not need a digit test to determine it is fraudulent. If the last digit test does indeed detect electoral fraud, we should expect the results to be accordingly.

Background

The distribution of certain place digits

The underlying assumption when performing a digit test is that if that digits from a 'naturally' occurring data follow certain distribution while fabricated data do not follow the same distribution.

Benfords Law is the observation that in sets of naturally occurring numerical data, the leading digits (1,2…or 9) follow a distribution that is not uniformly distributed, as one might intuitively expect. That is, the numbers 1 through 9 are not equally as likely to be a leading digit. But rather the leading digits of numerical data have a distribution where the 1 is most likely to be observed. This observation was first discovered by Simon Newcomb in the 19^{th} century and rediscover by Frank Benford (1938). Newcomb wrote that: "That the ten digits do not occur with equal frequency must be evident to anyone making use of logarithm tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener 1 than any other digit, and the frequency diminishes up to 9." (1881, p. 39)

Benford had the same realisations and continued to collect set of data of varying sort and found that many different types of data, whether it the population of countries or the square roots of natural numbers, followed the frequency distribution of what was to be known as Benfords first digit law. Benford also developed a general digit law for the frequency distribution for the kth place digit (1938), This ‘general digit law’ has been used obtain the expected frequencies of 2^{nd}, 3^{rd}, 4^{th} and later place digits. Since the development of this law researchers have found that data in more categories follow Benford’s law. A new field of forensic statistics developed where Benfords law was used for detecting fraud, by researchers and professionals alike. Mark Nigrini, a prominent researcher within this field, used Benfords law as a tool for detecting fraud in accounting and auditing namely. By using Benford’s first digit law as the expected distribution in a chi-squared goodness-of-fit test he developed a mathematical model that would detect manipulation of the data. If the data (auditing and accounting data in this case) did not follow the expected distribution one would assume that the data was not of natural occurrence and thus that it has been manipulated. Nigrini found that manipulated data not only didn’t follow Benfords law but deviates from it to quite an extent. With the 1^{st} digit frequency of number 1 being close to 0% rather than the expected 30.1% (Nigrini, 1992).

Benfords law has been used as a form of electoral fraud detection, in what some call “election forensics”(Mebane, 2006). The premise is simple: the kth place digits of non-fraudulent vote counts should follow a certain frequency distribution (depending on the place of digits being analyzed), and if the observed distribution of the digits deviates significantly from the expected distribution; the vote counts are likely fraudulent.

Various studies have explored how digit analysis can be used as a tool to detect vote manipulation. The literature within this subject is, however, split on which place digit should be tested. There are three common methods in this field: first-digit analysis, second-digit analysis and last digit analysis. Interestingly, while the expected distribution of first place digits are declining in distribution (1s are more common then 9s). The expected distribution of last place digits is uniform in distribution assuming certain criteria are met, the mathematical explanation for this is given in Dlugosz & Müller-Funk (2009) and Beber & Scacco (2012).

The expected distribution of first, second and last place digits are displayed in figure 1.

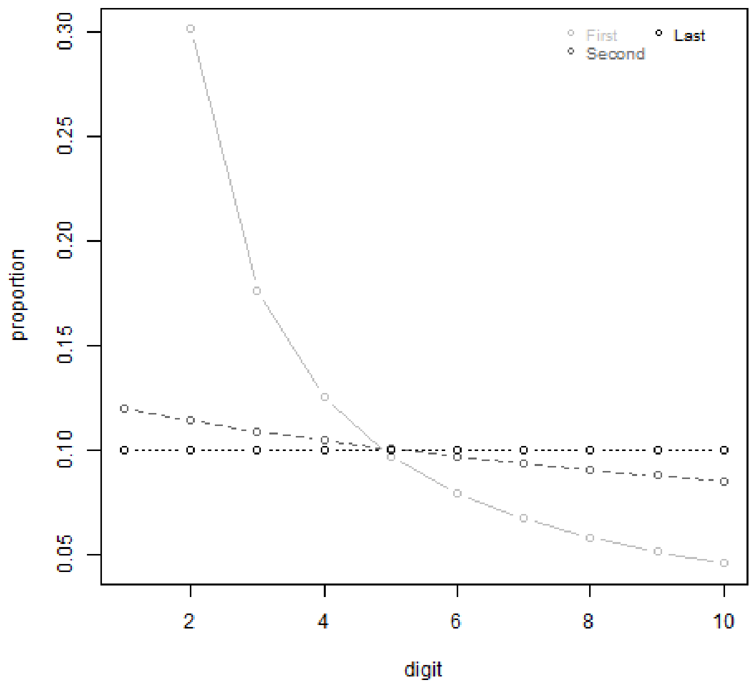

Figure 1. The distribution of 1st 2nd and last place digits

Review of previous applications of digit-based tests on elections

First digit analysis, used by, for instance, Bërdufi (2014) to detect fraud in the 2009 Albanian election, has been criticized for its proclivity to false positives; i.e. detecting 'electoral fruad' where there has been non (Shikano & Mack, 2011; Mebane, 2007). A motivation for not using the first digit law when it comes to elections was given by Mebane, saying: "Imagine a situation where all precincts contain about 1,000 voters each, and a candidate has the support of roughly fifty percent of the voters in every precinct. Then most of the precinct vote totals for the candidate will begin with the digits '4' or '5'" (2006, p. 2). Essentially, given how first-digits of vote counts will inevitable be affected by factors such as vote preferences in certain wards/precincts, they should not be expected to follow Benfords law.

The second digit Benford law(typically abbreviated as 2BL) has been used in plenty attempts at electoral fraud detection, it has been applied to the Russian election of 2007 where "extensive evidence of wide spread fraud" was found through 2BL testing (Mebane & Kalinin, 2010), and to the Iran election where fraud was also detected (Mebane, 2010). But 2BL testing has also received its share of scrutiny (Deckert, Myagkov, & Ordeshook, 2011; Shikano & Mack, 2011). Where it, like first digit testing, has been criticized for giving false positives since it has detected fraud in elections of countries where there have been no reasonable suspicions of fraud, such as the German parliamentary election (Shikano & Mack, 2011).

Furthermore, an analysis by Diekamann tested if fabricated regression coefficients can be detected by Benfords law and showed that anomalies are detected ‘better’ with later digits. That is, the rate of ‘false negatives’ (not detecting manipulation when the coefficients are indeed fabricated), also commonly known as type II error rate, was lower for ‘later’-digit analysis. The proclivity of false negatives, as well as false positives, is further elaborated on in another study by Diekmann, where the validity of using Benfords law to discriminate between manipulated and non-manipulated coefficients is discussed, it is pointed out that with an average sample size of 100 coefficients the type II error rates are high and thus whether Benford tests are powerful tools in discerning between manipulated and non-manipulated data is questioned. Even here, later digit-tests performed better (with lower rates of false negatives/positives) (Diekmann, 2010)

In like with Deikamanns studies, ‘later’ digit tests have also been seen a more appropriate choice for statistical electoral fraud analysis, last digit test being preferred by some (Beber & Scacco, 2012). For these reasons, this study will focus on last digit analysis of vote counts.

The last digits of non-fraudulent vote count (i.e. vote counts that have not been manipulated in anyway) should follow a uniform distribution. That is, 1’s should occur in equal frequency to all other possible digits. See Beber & Scacco, for a more elaborate explanation.

In previous attempts to asses a digit-analysis methods value as an indicator of electoral fraud some have applied it on simulation of fraudulent and non-fraudulent election like in Deckert, Myagkov & Ordeshook (2011). Others have applied the method on election where we know, with some certainty, if fraud was present or not, like in Brown & Wise (2012) and Shikano & Mack (2011). This paper will, similarly to the latter approach, apply digit-analysis on election where we know that there has or has not been fraud. But unlike Brown & Wise and Shikano & Mack, I will be focusing on the last digit as presented in Beber & Scacco (2012). By applying the test to election data where we know (with some certainty) that the vote counts have been tampered with or not. We can see if the test does detect vote manipulation. I will apply last digit testing to the 2018 parliamentary election in Sweden, where there is no reason to suspect fraud, and to the Ugandan presidential election of 2016 where a fraud-free election is unlikely.

If last digit testing is to be a considered a reliable fraud detection tool, we would expect the Swedish election to follow the expected distribution, while we would expect the last digits of the Ugandan vote counts to not follow the expected distribution of a ‘fair election’.

Theory

This section intends to give a theoretical overview of two aspect of this study. Firstly the conditions under which the last digit of vote counts are expected to be uniform in distribution is outlined. Secondly an overview of the use of power analysis in the context of this study is given.

Uniform distribution of last digits – in vote counts

Unlike first and second digit tests (Mebane, 2006; Shikano & Mack, 2011), the assumption that that last digits of most naturally occurring data is uniform distribution is not directly obtained by Benfords law, given that you need a digit-place specified to obtain the 'Benford distribution' according to the general digit law (Benford, 1938). However it can be observed that the Benford distribution of 'later placed digits' are approximately uniform in distribution (Diekmann, 2007). Additionally, Dlugosz & Müller-Funk (2009) and Beber & Scacco (2012) give two different mathematical proofs for the uniform distribution of last digit of number from continuous distribution.

Futher, Beber & Scacco, in addition to the mathematical proofs outlined in their study, also showed that the theoretical result(i.e. that last digits of sets of numbers are uniform in distribution)holds for numbers generated from a variety of distributions: such as normal distribution, gamma distribution and 'mixed' distributions (see section 2.2 in Beber & Scacco, 2012).

Given that last digits can generally be expected to be uniform in distribution (following the results of Beber & Scacco and Dlugosz & Müller-Funk): non-manipulated *vote counts* are expected to follow the uniform distribution, given that two conditions are met:

1. The vote counts do not cluster within a narrow range of numbers.
2. The vote counts do not contain a large portion of single- or double-digit counts.

Condition (1) is unlikely to pose a problem in application to the real-world vote counts as turnout rates and ward sizes vary enough in size (Beber & Scacco, 2012). To make sure condition (2) applies small candidates in the elections will be excluded in the analysis.

Power analysis and effect size

The test that will be used in order to determine if the last digits of the vote counts deviate significantly from a uniform distribution is the Pearson chi-squared goodness of fit test. In order to asses if the test in underpowered for the relevant sample sizes, a power analysis will be performed.

Usually the power of a chi squared test can be determined by knowing the effect size; “the degree to which the null hypothesis is false” (Cohen, 1988). In the case of this specific test the effect size would be the degree to which non-fraudulent election digits differ from the uniform distribution (of last digits). To calculate the effect size (ES), one needs to specify the alternative hypothesis, i.e. a proportion ‘p_{1j}’ needs to be chosen to specify a ‘not uniform distribution’. The ES for a chi squared goodness of fit test is given by:

$$ES = \sqrt{\sum_{j}^{n} \frac{p_{1j} - p_{0j}}{p_{0j}}}$$

Where p_{0j} is the proportion of digit j as theorized by the null hypothesis (i.e. 0.1 for all digits), p_{1j} is the proportion of digit j as theorized by the alternative hypothesis (non-uniform proportions). The ES essentially give us a measure of the difference between these proportions, i.e. it measures the difference between the paired proportions. Intuitively, one might want to say that the ES should be 0, though the power of such a test (i.e. its ability to reject the null hypothesis when the alternative is true) would presumably be low, especially considering that the sample sizes are quite large. Further, the last digits of fraudulent data have only been claimed to be approximately uniform in distribution, not exactly uniform (Beber & Scacco, 2012). However, it would be of interest to see if the power is high for small effect-sizes. Cohens definition of a ‘small’ effect size, 0.1 will be used. The power of the test will be calculated for the relevant sample sizes for different effect sizes and significance levels to get an overview of the tests power.

Methodology

As outlined in the previous sections, a last digit test the assumes that the last digits of vote counts follow a uniform distribution assuming that the vote returns are not single or double digit counts (i.e. the test can only be applied on vote counts with more than 2 digits) (Beber & Scacco, 2012).

We want to test if the last digits patterns of the real-world data deviate significantly from the uniform, expected, distribution. If we call the vector with the observed digit distribution d_i and the expected distribution D, the testing problem is:

- H_0 : d_i has distribution D
- H_a : d_i does not have distribution D

The Pearson chi-squared goodness of fit test is used to determine if the vote counts follow the expected distribution. The chi-square test statistic is:

$\chi^2 = \sum_{j=0}^{9} \frac{(d_j - D_j)^2}{D_j}$, where D_j is the expected frequency of digit j, and d_j is the observed frequency of digit j.

The test statistic is used to determine if the vote counts deviate significantly form D at $\alpha = 0.05$ which gives us the critical value 16.9 (given 9 degrees of freedom). If values greater than the critical value are obtained the null hypothesis, that the last digits of the vote counts are uniform in distribution, is rejected.

Since, according to previous applications of Benfords law, deviation from the 'expected distribution' indicates fraud: a more applicable interpretation of the null hypothesis is that the vote counts are non-fraudulent, i.e. have not been manipulated. The alternative hypothesis being that the vote counts are fraudulent.

To obtain the digit frequencies of the two elections used in this study, you first extract the last number of each row in relevant columns (where the columns consist of the vote counts for specific candidates/parties or total vote counts, and each row is vote counts obtained at each ward). An example for how this can be done in R (which was used in this study) can be found in Beber & Scacco (2012). Then the frequency at which each digit occurs is counted. These frequencies are then used to calculate the proportions used to calculate the chi-squared test statistics.

In a post analysis, the proportions retrieved for each set of vote counts are used to simulate frequencies with larger N(simply by multiplying the vector of proportions with the larger N's) in order to see how much larger N would be needed to reject the null hypothesis at $p<0.05$. This is done for sizes N up to 9000. For each 'new' vote count frequency a chi squared test is performed and the p-value is noted, these calculations are done using *for loop* in R.

The Chi squared test is followed by a power analysis, which is done in order to see if the test is underpowered, like some of the digit-based test in Deikmann (2010) were. The power analysis is done using the 'pwr' package in R (Champely, 2018). The package contains functions that perform power analysis in lines with Cohen (1998), where the same definition for ES is used. For the purpose of this analysis *pwr.chisq.test:pwr* was used. The power of the tests are calculated for ES values between 0 and 0.2 (as small effect sizes are of interest). To calculate the power for different ES: the significance level, number of observations and degrees of freedom is specified.

Description of Data

The Swedish parliamentary election of 2018 and the Ugandan presidential election of 2016 are the real-world election data to which the last digit pattern analysis will applied. The Swedish parliamentary election has had no reasonable doubt against its authenticity or any realistic accusation of fraud, this is not the case for the Ugandan election.

Ugandas 2016 presidential election was won by Yoweri Museveni, Ugandas ruler since 1986. Museveni won with 61% of the votes followed by his main opponent, Kizza Besigye, who received 35%. The results of the election were of no surprise, Museveni was expected to win, as he had the last 30 years. The election is unlikely to have had a free and fair ballot. Voters and opposition alike doubted that Museveni could ever lose, considering he controls the electoral commisioners, many calling it a 'staged election'(Abrahamsen & Bareebe, 2016). Futher, in a press statement following the election the US pointed out "delays in the delivery of voting materials, reports of pre-checked ballots and vote buying" and general "irregularities and official conduct that are deeply inconsistent with international standards and expectations for any democratic process" (US Department of State Press Statement, 2016). In summary, this election does not need a digit test to determine its inauntheticity.

The data for the Swedish election was obtianed through the The Swedish Election Authoritys(a government agency) website. Where in you can find election results dating back several years though only the most recent election was choses for this study. (2018)

The Ugandan election data was obtaineed through Development Seed (2016) who complied the ballots povided by the Ugandan Election Commission. The data from Development Seed is used for the purpose of ease (since its in cvs format rather then pdf-files) and the ‘orignal’ files its based on can be found on the Ugandan Election Commission website (2016).

The data used in this analysis will be vote counts at ward levels, where all vote counts have more than two digits. (‘valkrets’ in the Swedish data set and ‘parish’ in the Ugandan data set), and so the last digit is extracted from each of the vote counts from each ward, there are a total of 7393 wards in the Ugandan data set and 6325 in the Swedish data set. The digit frequencies for both elections can be seen in table 1 and 2.

Table 1

Digit frequencies for the Swedish election

	S	M	SD	total votes
0	621	630	590	653
1	615	641	648	607
2	623	657	599	639
3	616	623	645	602
4	645	614	604	657
5	657	625	634	611
6	647	612	651	662
7	660	633	621	621
8	639	640	650	629
9	602	650	683	644

Note: N=6325

Table 2

Digit frequencies for the Ugandan election

	Museveni	Besigye	registered voters	total votes
0	736	774	731	752
1	712	725	726	802
2	758	733	754	739
3	761	714	713	709
4	663	727	770	787
5	764	741	770	711
6	792	759	734	714
7	744	740	735	715
8	728	723	734	724
9	735	757	726	740

Note: N=7393

Results

2018's Swedish parliamentary election

The chi-squared statistics for the vote counts of the Social Democrats (S), Moderate Party (M) and Swedish Democrats (SD) and the total vote counts (see table 3) are all beneath the critical value (16.9) and thus the vote counts do not deviate significantly from the expected uniform distribution. In other words, it is reasonable to assume that the vote counts of the election are non-fraudulent, according to the last digit test. (see figure 2 and table 3)

Table 3

Test statistics for the Swedish election

	Chi-squared statistic	p-value	N
S	5.528	0.786	6325
M	3.084	0.961	6325
SD	11.811	0.224	6325
total votes	6.723	0.666	6325

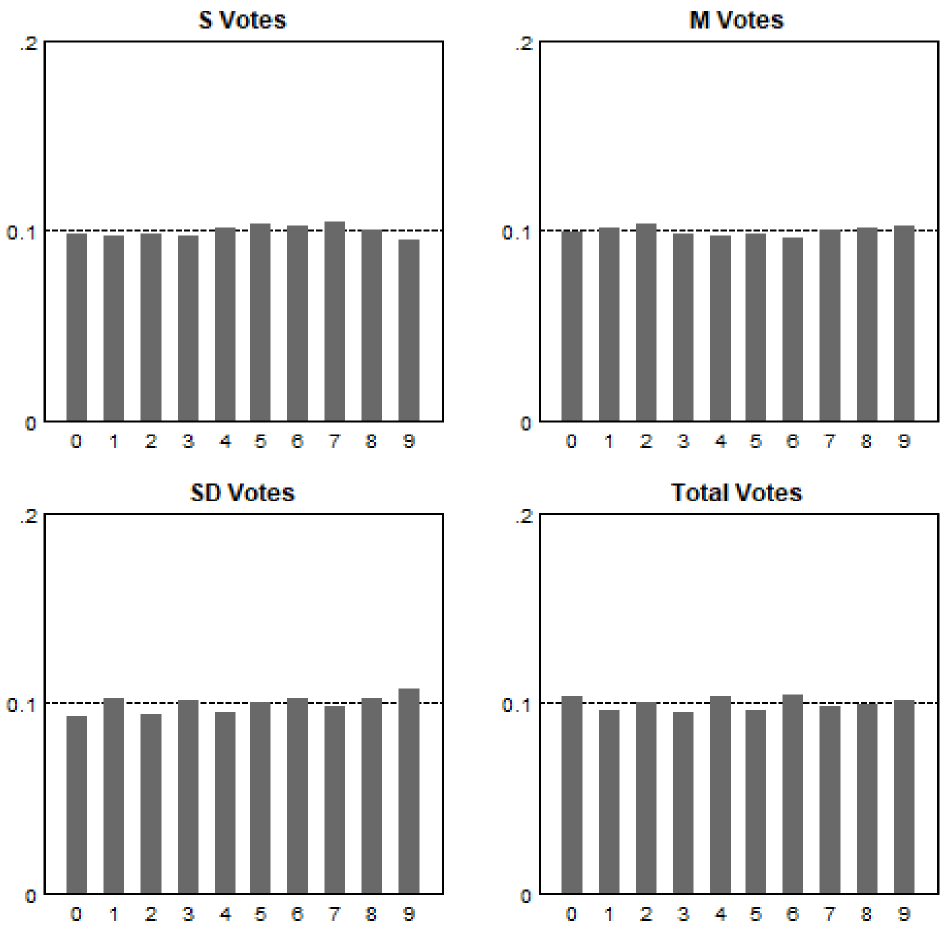

Figure 2. digit frequencies for the Swedish election

2016’s Ugandan presidential election

The chi-squared statistics for the vote counts of the Museveni, Besigye and the registered voters and the total vote counts (see table 4) are all beneath the critical value (16.9) and thus the vote counts do not deviate significantly from the expected uniform distribution. In other words, it is reasonable to assume that the vote counts of the Ugandan election are non-fraudulent, according to the last digit test. (see figure 3 and table 4)

Table 4

Test statistics for the Ugandan election

	Chi-squared statistic	p-value	N
Museveni	14.817	0.096	7393
Besigye	4.342	0.887	7393
registered voters	4.450	0.879	7393
total votes	12.920	0.166	7393

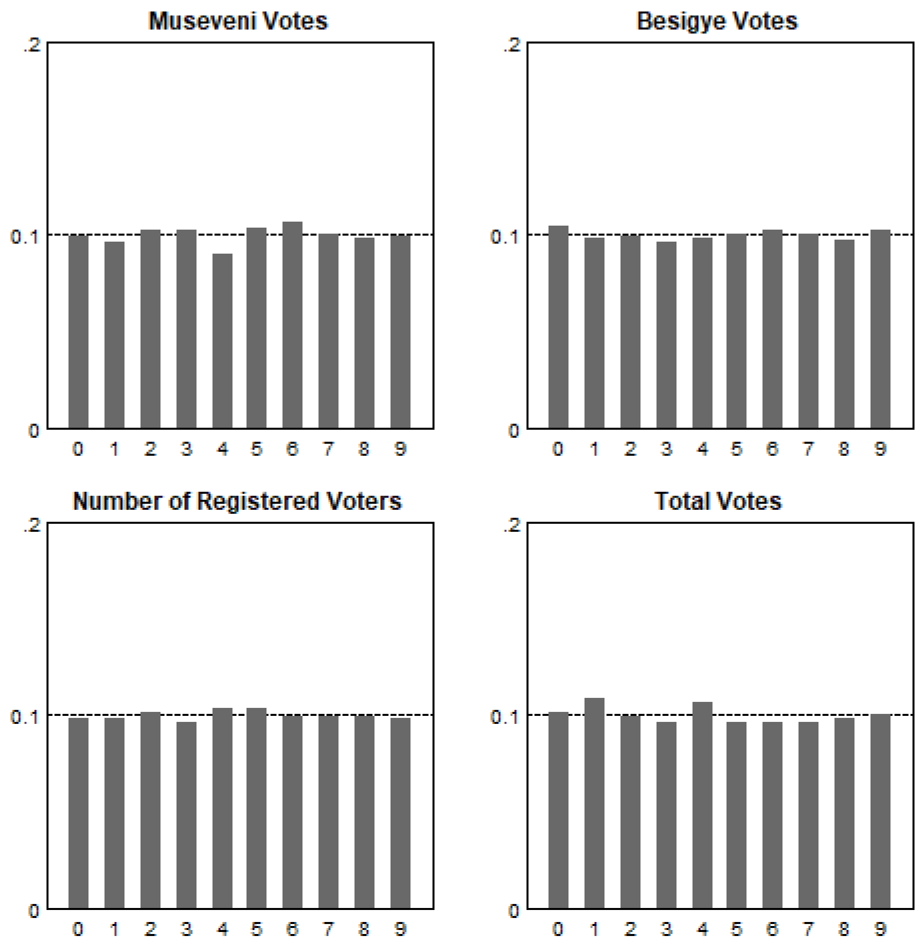

Figure 3. digit frequencies for the Ugandan election

Further, it is worth noting that the proportion of the vote counts in the Swedish election is indeed closer to an exact uniform distribution than the vote counts of Museveni (see table 3 and 4), where the Museveni vote counts would have been considered significantly non uniform in distribution at $p<0.1$. Additionally; through simulating last-digit frequencies with the same proportion as the Museveni vote counts but with larger number of observations (in order to see how much larger N would be needed to reject the null hypothesis as $p<0.05$). We see, through the p-values of the chi square tests performed on the simulated data (illustrated in figure 4) that the null hypothesis would be rejected at number of observations more than N=8441; which corresponds to an increase of 1048 observations.

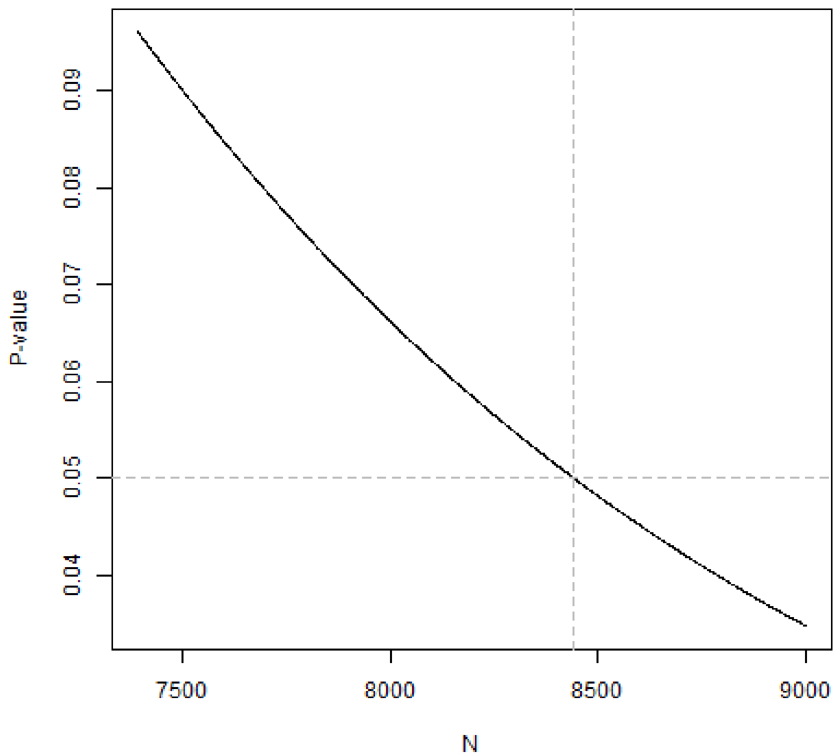

Figure 4. P-values of chi-squared goodness of fit tests with different N, with observed proportions based on Museveni vote counts.

A similar simulation was done for the rest of the Ugandan vote counts for which the null hypothesis was not rejected for N less than 9000. Similarly, for the proportions of last digits of the Swedish parties the null hypotheses the null hypothesis would not be rejected even when the number of observations is increased to 9000 (an increase by 2675). It goes without saying, however, that if N was increased more the null hypothesis would be rejected for all last digit frequencies eventually, given that even *very* small differences would be detected eventually as $N \rightarrow \infty$. This was however not the point of the simulation, rather it shows that a relatively small increase in N for the Museveni data results in a rejection of the null hypothesis at $p<0.05$; i.e. the assumption that the null hypothesis is non fraudulent would no longer hold. We could therefore say that the Museveni-data, while being the only vote counts reasonable suspected of fraud, is also the only set of vote counts with last digits frequencies relatively close to being significantly different from uniform in distribution.

Power analysis

The results from the power analysis (illustrated in figure 5 and 6), show that at least for sizes N like that used in this study, the test can discriminate between small differences in distribution with high power(1); given the arbitrary definition of a 'small' effect size, 0.1 (Cohen, 1988). Further, the power of the test is higher than 0.8 for effect sizes above 0.05 even at a 0.01 significance level.

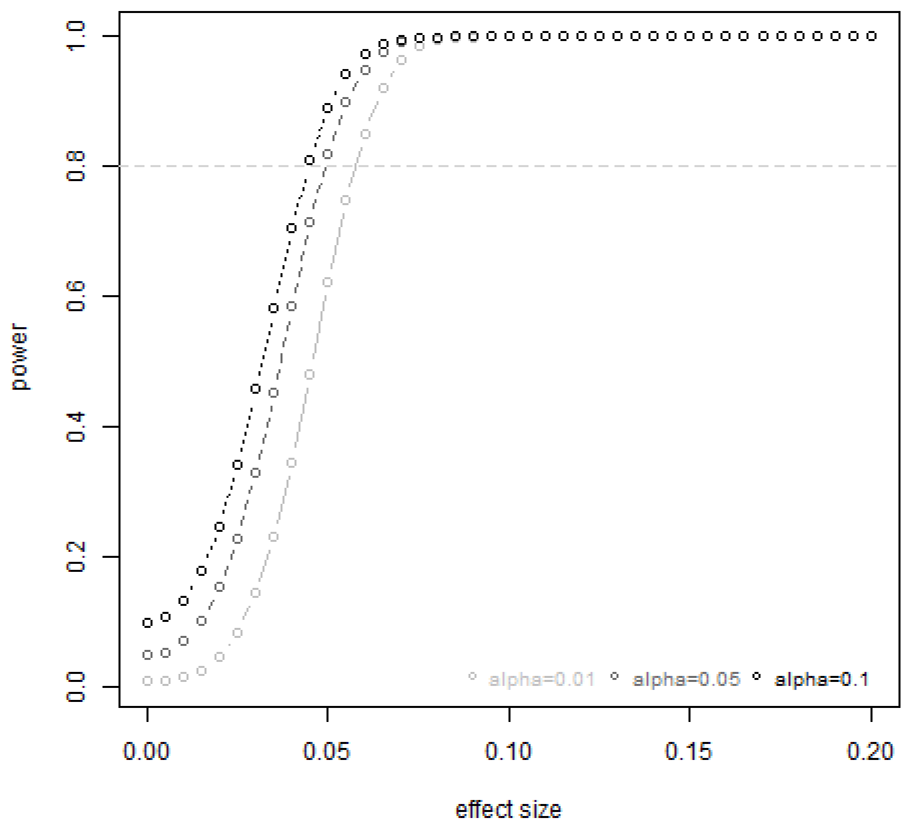

Figure 5. Power for chi-squared test at different effect-sizes and significance levels (with N=6325)

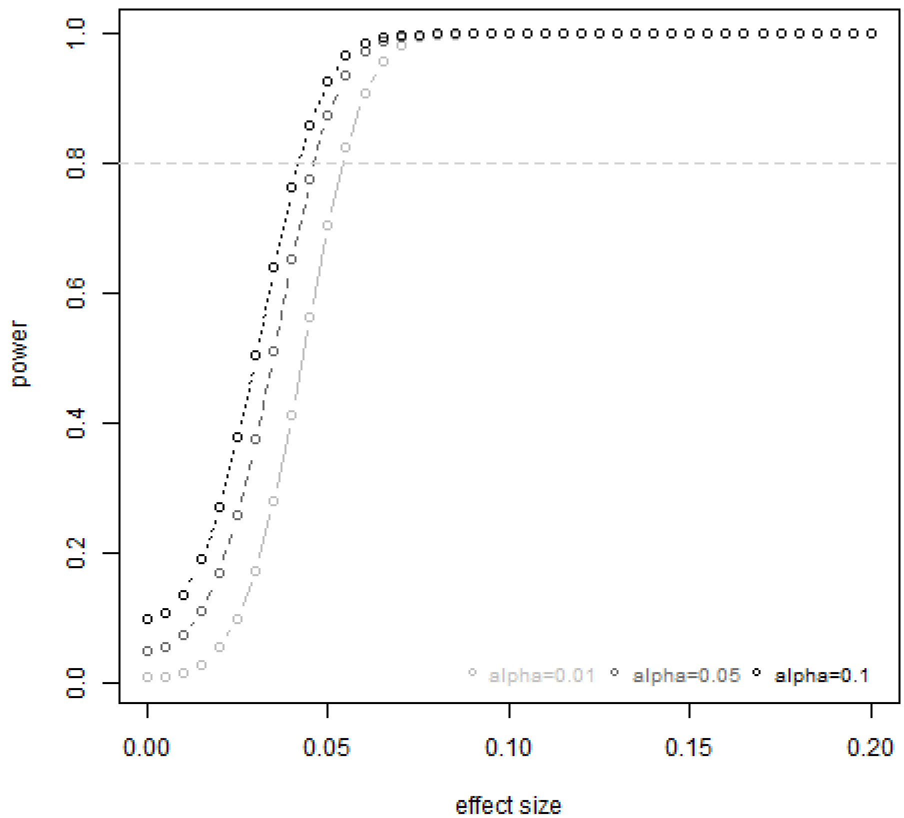

Figure 6. Power for chi-squared test at different effect-sizes and significance levels (with N=7393)

To be able to say that the test is a powerful test in discriminating between fraudulent and non-fraudulent vote count we would have to assume that non-fraudulent data does not differ with less than 0.05 effect size from the expected uniform distribution. Because if differences in distribution smaller then 0.05 (in effect size) are to be deemed fraudulent, the test would have a type II error rate higher then 0.2: i.e. it would not detect fraud, when fraud has occurred more than 20% of the time.

Still, if differences smaller than 0.1 in effect size are to be deemed negligible, the test could by no means be considered underpowered for the relevant sample sizes (the power being 1 in both cases).

Conclusion

If the last digit frequency of non-fraudulent, authentic vote counts is uniform in distribution, as claimed in Beber & Scacco (2012), then the Swedish and Ugandan election are free of fraud according to the last digit-pattern analysis. Since both last digits frequencies of both elections do not deviate significantly from uniform distribution, we can reliably assert that both elections have last digit frequencies of uniform distribution, at a significance level of 0.05, The problem here is apparent: while the test can be expected to reliably distinguish between uniform and non-uniform distributions; the last digit analysis failed to indicate fraud where electoral fraud likely occurred.

Further, given the results of power analysis, the test seems to not be underpowered when it comes to distinguishing between uniform and non-uniform distributions for the sample sizes in this study; rather it has a high power when distinguishing 'small' differences. The definition of a 'small' effect size is, however, not a given in every case; and it ought to be decided on a case-to-case bases, usually by knowing how a given theoretical phenomena's expected distribution differs in the 'real world' (Cohen, 1998). What size difference between an exact uniform distribution of digits and real-world (non-fraudulent) vote count distributions should be deemed negligible is hard to answer without basing it on measuring the real discrepancy between the last digit vote counts proportions of a large sample of known non-fraudulent vote counts and an exact uniform distribution, and thus specifying an alternative hypotheses to calculate the appropriate effect size. However, given the lack of access to a more suitable definition of 'small' effect size for the scope of this paper, Cohen's definition is relied on, and thus the tests are not considered underpowered.

The results of the digit pattern analysis for the Swedish election is as expected, since other obviously non-fraudulent elections have been shown to have uniformly distributed last digit frequencies (Beber & Scacco, 2012). But if this last digit analysis is to be trusted, the Ugandan vote counts should be deemed authentic, however unlikely that is, and thus we should ignore valid suspicions of vote manipulation as reported by numerous sources (Abrahamsen & Bareebe, 2016). This raises the question: if an election obviously riddled with vote tampering and irregular conduct is 'non-fraudulent' according to the last digit analysis, then is the test useful? While the small scope of this paper cannot give a definite answer to this, it at the very least calls for the need to further asses the value of this test. Further, it could perhaps even be the case that the vote counts are indeed authentic, despite the accusations of vote manipulation. One could argue, however, that a case like Uganda, where suppression of opposition prior to elections is common (Abrahamsen & Bareebe, 2016), the authenticity of vote counts still does not indicate a 'fair' elections. This is the most obvious limitation of the digit analysis; it only indicates direct vote manipulation.

It is worth considering however, that if the chi-squared goodness of for test was to be done at a 0.1 significance level, the null hypothesis would have been rejected for Museveni vote counts and they would have been deemed fraudulent. Additionally, the Museveni vote counts would have been rejected at $\alpha = 0.05$ if the number of observations where increased (by 1048). So, one could say that the Ugandan vote counts deviate 'more' from the expected last-digit distribution. This could however just be by chance, and the last digit test (as initially constructed with $\alpha = 0.05$) could be correct in deeming the Ugandan vote count non-fraudulent. This bring into question however, what significance level is appropriate for this type of test. Do we want to increase the likelihood of type I error? Or should we, if anything, have a higher (say, 0.01) significance level? Surely, detecting fraud where there has been none is undesirable, but so is the reverse; if the test is to be useful.

A problem with my analysis is that only one election of, most likely, fraudulent vote counts is used. A study of larger scope would be preferable, an obvious issue being the unavailability of obviously fraudulent election data. Perhaps then simulations of fraudulent vote, like in Deckert, Myagkov, & Ordeshook (2011) and Mebane (2006), would be a better approach. Although, both studies had vastly different approaches in what they deem constitutes 'fraudulent vote counts' simulations. Which begs the question: what does fraudulent vote counts look like? Or rather, what is the appropriate way of simulating them? Ideally, you could approach this question by looking at the distribution of the last digits of several fraudulent elections, this of course assumes such data could be available.

To finish, confirming that of previous studies, last digits of non-fraudulent vote counts seem uniform in distribution. Yet from the results of my analysis, it does not seem obvious that fraudulent vote counts do not *also* follow this uniform distribution. Of course, further exploration is needed in order to make any conclusive statements about the distribution of fraudulent vote counts. This would be an interesting approach for further research.

References

Abrahamsen, R., & Bareebe, G. (2016). Uganda's 2016 elections: Not even faking it anymore. *African Affairs, Volume 115, Issue 461*, 751–765.

Beber, B., & Scacco, A. (2012). What the numbers say: A digit-based test for election fraud. *Political analysis 20(2)*, 211-234.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American philosophical society*, 551-572.

Bërdufi, D. (2014). Statistical Detection of Vote Count Fraud (2009 Albanian Parliamentary Election and Benford's Law). *Mediterranean Journal of Social Sciences 5(2)*, 755.

Brown, M., & Wise, A. (2012). Does the Application of Benford's Law Reliably Identify Fraud on Election Day? *ProQuest Dissertations and Theses*.

Champely, S. (2018). pwr: Basic Functions for Power Analysis. R package version 1.2-2. https://CRAN.R-project.org/package=pwr

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: L. Lawrence Earlbaum Associates.

Deckert, J., Myagkov, M., & Ordeshook, P. C. (2011). "Benford's Law and the detection of election fraud.". *Political Analysis 19.3* , 245-268.

Development Seed. (2016). Developmentseed/open-elections: Uganda 2016 Presidential Elections. Available at: https://github.com/developmentseed/open-elections/tree/master/Uganda, based on data in: http://www.ec.or.ug/ecresults/0-Final_Presidential_Results_Polling%20Station.pdf. (Last visited: 2018-11-02)

Diekmann, A. (2007). Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *Journal of Applied Statistics 34(3)*, 321-329.

Diekmann, A. J. (2010). Benford's law and fraud detection: Facts and legends. *German Economic Review, 11(3)*, 397-401.

Mebane, J. W. (2007). Statistics for digits. *In Summer Meeting of the Political Methodology Society* (pp. 18-21). Pennsylvania State University.

Mebane, J. W. (2010). Fraud in the 2009 presidential election in Iran? *Chance 23*, 6–15.

Mebane, J. W. (July. 2006). Election forensics: Vote counts and Benford's law. *Summer Meeting of the Political Methodology Society.* UC-Davis.

Mebane, J. W., & Kalinin, K. (2010). Electoral fraud in Russia: Vote counts analysis using second-digit mean tests. *Paper prepared for the 2010 Annual Meeting of the Midwest Political Science Association*, (pp. 22-25). Chicago, IL.

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of mathematics 4*, no. 1(39-40).

Nigrini, M. J. (1992). The detection of income tax evasion through an analysis of digital distributions. *(Doctoral dissertation)*.

Shikano, S., & Mack, V. (2011, 01 1). When Does the Second-Digit Benford's Law-Test Signal an Election Fraud? . *Jahrbücher für Nationalökonomie und Statistik*, pp. Vol.231(5-6).

Statement, US Department of State Press. (2016, February). U.S Embassy in Uganda. Available at: https://ug.usembassy.gov/press-statement-on-the-results-of-ugandas-presidential-elections/. (Last visited: 2018-11-01)

Valmyndigheten. Valpresentation 2018. Avilable at: https://data.val.se/val/val2018/statistik/index.html. (Last visited: 2018-11-01)