DOES THE APPLICATION OF BENFORD'S LAW RELIABLY IDENTIFY FRAUD ON ELECTION DAY?

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Public Policy
in Public Policy

By

Michelle S. Brown, B.A.

Washington, DC
April 13, 2012

DOES THE APPLICATION OF BENFORD'S LAW RELIABLY IDENTIFY FRAUD ON ELECTION DAY?

Michelle S. Brown, B.A.

Thesis Advisor: Andrew Wise, Ph.D.

ABSTRACT

In an attempt to bring mathematical certainty to uncertain situations, some have tried developing "election forensics" tools as a way of evaluating the quality of an election. Most election forensics tools involve applying statistical methods and underlying mathematical principles to official election results. One such tool is the application of Benford's Law to election results. In this paper, I use election data from the lowest level, that of polling station, to assess whether Benford's Law, as applied to the distribution of second-digits in vote count data, is an appropriate tool for detecting fraud. Unfortunately, my analysis shows that Benford's Law is an unreliable tool. And, as one applies more sophisticated methods of estimation, the results become increasingly inconsistent. Worse still, when compared with observational data, the application of Benford's Law frequently predicts fraud where none has occurred.

Acknowledgments

The author is grateful for the understanding and assistance of all the staff and faculty of the Georgetown Public Policy Institute.

Additionally, I must acknowledge the guidance of my thesis advisor, Professor Andrew Wise, PhD. I am grateful for his patience, advice, and enthusiasm throughout the process.

I also acknowledge the support of my colleagues at the National Democratic Institute for International Affairs (NDI) in both Washington, DC and the Republic of Georgia. They were not only supportive and gave me the flexibility needed to work on my thesis; they all indulged my various ideas for topics and nodded patiently when I droned on about digit tests. Pat Merloe and Mary O'Hagan, in particular, have continued to inspire me and encourage me in this and other endeavors.

I also recognize three individuals for their assistance and collaboration in the field. Maka Meshveliani, Anastasia Soeryadinata-Wibawa, and Eka Siradze-Delaunay made my time in Georgia more fun and full of action than I could have hoped. I will always be grateful for their humor, grace, and friendship.

Finally, I appreciate the efforts and contribution of my family.
I am thankful for the patience and support Emily Alfano has given me during this project and graduate studies more broadly; it has been more than I could have imagined. And I will always be grateful to my mother for her unwavering support and her willingness not just to indulge my curiosities but to encourage them.

With much appreciation,
Michelle S. Brown

TABLE OF CONTENTS

I. Introduction

Is the application of Benford's Law an appropriate tool for assessing electoral fraud? Elections present a unique opportunity to advance democratic progress in a country. When conducted well, elections are an opportunity for voters to express their will and make choices about their leaders. However, the leaders of some countries do not conduct their electoral process with integrity. As a result, elections can be enormously unfair. Election monitoring by nonpartisan domestic organizations helps to assess the electoral process, can raise public confidence in the election process when elections are conducted well, and can detect irregularities where electoral manipulation has occurred. There is an established international acceptance of the positive contribution of nonpartisan election monitoring by national citizen organizations and many have noted that there is no real substitute for well-trained nonpartisan domestic election observers. However, some governments place limits on the activities of election observers or deny them access to the process completely. Alternately, the security situation in a country may be such that it is too risky to allow observers the freedom of access and movement necessary to systematically assess the electoral process. In such cases, some have, as a policy, turned to election forensics as a way of drawing conclusions about the fairness of an election.

Most election forensics tools involve applying statistical methods and underlying mathematical principles to the official election results (where published) in order to assess whether fraud took place on Election Day. By focusing solely on Election Day, these forensic tools cannot assess whether the pre-election period (or post-election period) has gone well and thus cannot make a judgment about the election process as a whole. However, often these tests are touted as somehow more definitive than information available from election observers because they appear to convey a mathematical certainty. On the other hand, those with a more thorough understanding of the limitations of such mathematical tests, place such techniques within the proper context as a complementary tool, the results of which should point toward further investigation.

One such tool is the application of Benford's Law to election results. Benford's Law has been used informally in blogs and in opinion pieces to assess elections in Afghanistan, Iran, and Russia, countries where, frankly, one does not need to see the results of a digit test to draw the conclusion that the electoral process was unfair. Walter Mebane, Jr., a professor of political science and mathematics at the University of Michigan, has written extensively on the patterns in the second leading digits of vote counts and how the distribution of those digits may or may not conform to Benford's Law or what he calls "Benford-like distributions" (Mebane 2011, 2010a, 2010b, 2009, 2008, 2007). Mebane concedes that even if the distribution of the second-digits differs significantly from the distribution prescribed by Benford's Law, it does not necessarily mean that fraud has occurred. He asserts that "whether the tests are useful for detecting fraud remains an open question" (Mebane 2011). More recently, Mebane has theorized that such tests may be able to distinguish electoral fraud (defined as coercion) from strategic voting and gerrymandering (Mebane 2010b). However, academics Deckert, Myagkov, and Ordeshook (2011) have argued that Benford's Law is "problematical at best as a forensics tool when applied to elections." Though the arguments of Deckert and others are compelling, they use data aggregated at a much higher level than polling station data. Here I use election data at the lowest level, that of polling station, to assess the conditions (if any) under which the application of

Benford’s Law may be an appropriate tool for detecting fraud. In addition, I use information collected by nonpartisan election observers to test whether Benford’s Law is more applicable with regard to certain problems within the electoral process.

In the next section, I give a brief overview of Benford’s Law, the distribution of second-digits expected under Benford’s Law and two statistical tests, and discuss how others have tried to apply such tests to election data. Then in the third section, I describe my model in more detail and apply Benford’s Law to official data from the Republic of Georgia’s Parliamentary Election in May of 2008. In the fourth section, I present the results and assess whether the application of Benford’s Law detects abnormalities and outliers such as high turnout. In addition, I use data collected by nonpartisan domestic election observers about the quality of the Election Day process to see if the application of Benford’s Law predicts various types of procedural problems. Finally, I conclude with a discussion of whether the second-digit Benford tests are a useful tool for assessing elections.

II. Background

In 1938, Frank Benford, a physicist working in General Electric's research lab, noticed that the leading digits of numbers did not appear to occur with equal frequency as one might intuitively expect. Instead, he noticed that more numbers seem to have a low leading digit, such as 1 or 2, rather than a high leading digit, such as 8 or 9. In a paper written by Benford himself, he describes first observing that the pages of a book containing the tables of common logarithms provided evidence of his theory: the pages with the logarithms of low numbers were more "apt to be stained and frayed by use than those of the higher numbers" (1938). He then went on to test the first digit frequencies of various lists taken from different fields, testing data comprised of things such as the areas of rivers, street addresses, numbers appearing in Reader's Digest, the cost of concrete, and the atomic weights of elements. His results showed that on average, 30.6 percent of numbers had the first digit of 1, whereas just 4.7 percent of numbers had a first digit of 9. Benford then went on to construct an expected frequency of the first and second digits as well as the digit combinations. The formulas for the frequencies of the digits are below with D_1 representing the first digit and D_2 is the second digit where P is the probability of observing the event. The Benford's Law Formulas are:

First Digit: $P(D_1=d_1) = \log_{10}(1 + 1/d_1)$ for d_1i{1,2,…9} (1)

Second Digit: $P(D_2=d_2) = \Sigma \log_{10}(1 + 1/d_1d_2))$ for $d_1 =1$ and d_2i{0,1,…9} (2)

Later Theodore Hill elaborated a theoretical basis for the phenomena and formulated the General Leading-Digit Law:

$$P(D_1 = d_1, \dots, D_k = d_k) = \log[1 + (\textstyle\sum_{i=1}^{k} d_i \times 10^{k-i})^{-1}] \quad (3)$$

for all positive integers k, all d_1{1,2,…,9} and all d_j{0,1,…,9), j=2,…,k

For example, $P(D_1D_2D_3=455) = \log_{10}(1 + 1/455)) = 0.00095$

It is important to note that Benford's Law does not apply universally. It is meant to apply to data that have several orders of magnitude.

In order to use Benford's Law as a tool for identifying fraud, one must make two assumptions: First, one must assume that the "true" data follows the distribution as described by Benford (and Hill), and second, that the manipulated data has a different distribution. Sandron (2002) has shown that population numbers conform to Benford's Law and offers a theory that the underlying population growth accounts for why population numbers conform to the law. Nigrini (1999) asserted that Benford's Law can be used to detect fraud in accounting data and suggests several types of data that conforms to it. More recently, Tödter (2009) outlined the potential for using Benford's Law to indicate fraud in economics.

A few studies have noted that Benford's Law often "approximately describes" the *second digits* in vote counts (Mebane 2006). They theory for detecting election fraud is simple: if vote count data follow a Benford or Benford-like distribution, a deviation from that distribution would then indicate fraud. The theory behind conducting significance tests using Benford's Law is that those committing the fraud are making up the numbers (i.e., vote counts), something that is nearly impossible to do in such a way that conforms to Benford's Law. Thus, when the vote counts are artificially increased or decreased then there will be variations from the second-digit Benford distribution. However, as Deckert, Myagkov, and Ordeshook (2011) have noted, even with non-manipulated data, there is a high probability of getting a 'false positive' test result. As such, the Benford test will mark such deviations as 'significant' even though no fraud has been committed. From a practical perspective, the application of Benford's Law is acknowledged as being most appropriate when the data covers several orders of magnitude (i.e., requiring a large number of polling stations) and where the figures are a result a combinations of several distinct processes (e.g., one process determines the number eligible to vote, another process determines voter participation, another process is involved in voter preference). Under such circumstances, perpetrators of fraud who understand Benford's Law would have a difficult time coordinating fraud since that would require them to assign every single polling station a specific figure so as to not deviate from the law.

In my analysis, I hope to add to the literature in two ways: First, I use data at the level ballots are counted and reported—the polling station. Second, in addition to testing the performance of Benford's Law on abnormal results such as high turnout, I use information collected by well-trained domestic election observers that assess the quality of the process at individual polling stations. I test the performance of Benford's Law by using a series of increasingly sophisticated methods, ending with multivariate regressions more sophisticated than the tests suggested in the literature. My general intent is to give Benford's Law the best possible chance to succeed as a means of identifying election fraud, as compared to the current gold standard, professional domestic election observers. By using data collected by domestic observers, I may be able to produce a theory (or, at the very least, a foundation upon which a theory can be built) in order to guide which types of electoral irregularities may be identified by the Benford's Law. Currently, no one has put forth a detailed theory outlining which types of fraud or malfeasance Benford's Law might precisely identify.

III. Methodology

I use two tests to determine deviations from Benford's Law. The first uses the Pearson's chi-squared (χ2) goodness-of-fit test. This test is most often employed to compare an expected (Benford) distribution of data with the actual distribution. The Pearson chi-squared test statistic for Benford's Law is:

$$\chi^2_{Benford} = \sum_{j=0}^{9} \left(n_j - Nr_j\right)^2 / (Nr_j) \quad (4)$$

where N is the number of vote counts equal to, or greater than 10 (i.e., a two-digit number), n_j is the number having second digit j and r_j is determined by the Benford's Law formula equation two above).

If the counts of the digits are statistically independent, then the statistic should be compared to the chi-squared distribution with nine degrees of freedom. A test statistic greater than 16.9 indicates a statistically significant difference from Benford's Law expectation. The chi-squared statistic has a high power for large samples so that even small deviations from Benford's Law will be statistically significant.[a] The practical interpretation of the null hypothesis is that no fraud has taken place. The practical alternative hypothesis is that fraud has occurred.

A second test statistic, inspired by Grendar, Judge, and Schechter (2007) and adapted by Mebane, compares the arithmetic mean of the second digit to the expected mean if the digits are distributed according to Benford's law. The estimated mean of the second digit is denoted $\hat{\jmath}$. If

[a] It is worth noting that in the case of a small sample sizes, there is a much greater chance that the chi-squared test will result in a Type II error (i.e., there is fraud but the test does not detect it). The sensitivity for large samples noted above makes a Type I error very likely (i.e., there is no fraud but the test indicated fraud).

the counts of the second digits follow Benford's Law, then the expected value for the second-digit mean is (rounded):

$$\bar{j} = \sum_{j=0}^{9} j r_j = 4.187 \quad (5)$$

Building on what has been done previously, I then use the second statistic to create variables that measure the proportional deviation from the expected second digit mean of 4.187. I compute *P*, the percent of deviation from the expected mean:

$$Q = (\hat{j} - \bar{j}) / 4.187 \quad (6)$$

If there is no deviation from Benford's Law, *Q* will be zero. As the digit distribution is more "fraudulent" (according to this test) there should be larger deviations from the mean. Since we are subtracting the estimated mean, $\hat{j}$, from the expected mean, $\bar{j}$, a negative percentage implies that votes may have been taken away from a particular political party meaning that there are too many low second-digits such as zeros and ones and/or there are too few high-second digits such as sevens, eights, or nines. For example, if a party's vote counts are 200 or 210 when they should be 240 or 290, Q will be negative. Alternately, a positive sign could indicate that a party is benefitting from the fraud being committed by receiving additional votes.

Using the rates of deviation for various parties, I use the following model to measure how much variation in dependent variable, Y, the deviations explain:

$$Y_i = \alpha + \beta X_i + \rho_1 Q_{d\,p1} + \rho_2 Q_{d\,p2} + \rho_3 Q_{d\,p3} + \rho_4 Q_{d\,p4} + \ldots + \rho_k Q_{d\,pk} + e_i \quad (7)$$

Where Y is the dependent variable indicating electoral abnormality for polling station *i*, α is the intercept, X*i* is the vector of control variables for polling station *i* (e.g., number of registered voters), Q_{dp} is percentage by which the second-digit mean deviates from the expected mean for district *d* and party 1 *p1* … *pk* and $\rho_1, \rho_2, \ldots, \rho_k$ are the coefficients of interest that measure how much fraud the digit deviations explain, and e_i is the error term.

For my purposes, I generated the following variables from the official Republic of Georgia parliamentary election results data and used as dependent variables: the number of voters participating overall ("partic"), a continuous variable; a high percentage turnout ("dgt90"), an indicator variable that equals 1 if turnout is 90 percent or more and 0 otherwise; the number of votes recorded in the last three hours ("vtd1720sub"), a continuous variable; a high rate of votes cast in the last three hours ("dgt200in3"), an indicator variable noting a high number of voters participating per minute equal to 1 if 200 or more votes were cast in the last three hours and 0 otherwise; a high vote share percentage for one party ("vsgt90"), an indicator variable equal to 1 if a party received 90 percent or more of the votes cast and 0 otherwise.

The following variables were generated from the International Society for Fair Elections and Democracy's (ISFED) observer data and are used as additional dependent variables: serious irregularity observed ("dinc_irreg"), an indicator variable that equals 1 if an observer reported a serious irregularity and 0 otherwise; voters not consistently checked for ink ("dink"), an indicator variable that is equal to 1 if the officials in the polling station did not consistently check voters for ink as required by law and set to zero if they always checked; violation of the secrecy of the vote ("dnosecret"), an indicator variable equal to 1 if there was a violation of the secrecy of the vote and 0 otherwise; the presence of unauthorized persons ("dunauth"), an indicator variable set to 1 if a person not authorized by law was inside the polling station and 0 otherwise; an observer filed a complaint ("dcompl"), an indicator variable set to 1 if an observer filed an official complaint and 0 otherwise. Please note that the observer data is based on a sample of polling stations; therefore, the number of observations is smaller than the data from official results.

Table 1 is an overview of the estimation methods and variables used in each regression. The purpose is to give the reader an overview of the different regressions that will be discussed in detail in the Results Section. For example, the number of voters participating is the first dependent variable (indicated in column 1) and it is a continuous variable (indicated in column 2). I used ordinary least squares (OLS) to estimate participation as noted in column three. The fourth column in the table indicates which independent variables were included in each regression. For participation, I ran a total of four different regressions, starting with a simple regression using the five party vote digit deviations (noted as the "core variables" in the table). Then I ran another regression with the core variables but I also added a control variable indicating that a station counted another station. Then I ran a regression that included the core variables as well as five additional variables for age. Finally, I regressed participation on the core variables and a variable indicating whether an irregularity had been reported.

Table 1: Overview of Estimation Methods and Variables

	Dependent Variable (1)	**Dependent Variable type (2)**	**Estimation Method (3)**	**Independent Variables* used in each regression (4)**
Official Results data				
	Participation	Continuous	OLS	1. Core variables; 2. Added indicator variable for station that counted another; 3. Added five variables for age; 4. Added indicator variable noting serious irregularities were observed.

Votes recorded in final three hours	Continuous	OLS	1. Core variables; 2. Added indicator variable noting serious irregularities were observed. 3. Added five age variables
High rate of participation in last three hours	Indicator	Logit	1. Core variables; 2. Added indicator variable noting serious irregularities were observed and the five age variables
Domestic Observer Data			
Officials did not check consistently for ink	Indicator	Logit	1. Core variables; 2. Added five age variables
Unauthorized person was present	Indicator	Logit	1. Core variables; 2. Added five age variables

Note:
* Every model included the five party vote digit deviation variables that are described in detail in the Methodology Section.

Next, I turn to a detailed description of the data.

IV. Description of Data

The Republic of Georgia's 2008 Parliamentary Election provides us with real world data on which to test the performance of Benford's Law. The Parliamentary Election occurred on the 21st of May, 2008, a culmination of several months of political unrest in Georgia. On March 21, 2008, President Mikhail Saakashvili called for Parliamentary elections to be held in May of that same year. The elections were initially scheduled for late 2008 but were moved up as a result of a plebiscite held in January 2008 along with the so-called Extraordinary Presidential Election. President Saakashvili offered the plebiscite in response to demonstrations by the opposition political parties in November of 2007 as a way to end what was becoming a political crisis. The January Extraordinary Presidential Election was highly contentious and was followed by a brief period of negotiation between the ruling party (the United National Movement or UNM) and the opposition parties. Unfortunately, in March the political situation worsened and the Parliament—controlled by UNM—passed a series of amendments to the Constitution and modified the Unified Election Code. Several changes to the election system were made in direct defiance of the opposition's concerns. After the political crisis in November of 2007, both sides had agreed to change the electoral system for the 50 majoritarian seats in Parliament from a first-past-the-post system[b] to a system of regional proportional lists.[c] However, the UNM-controlled parliament, kept the first-past-the-post system kept for the majoritarian seats, and increased their total number of seats to 75 while the number of proportional seats was reduced by 25 seats down

[b] First-past-the-post is a system where the party who gets the most votes wins. Put another way, the 50 majoritarian seats are 50 single member constituencies with a simple majority system.

[c] In proportional list systems, parties put forth a list of candidates and seats are awarded according to their party's share of the vote. The idea behind proportional representation lists is to consciously reduce the disparity between a party's share of the vote and its share of the seats. This is usually done using an electoral formula or a quota which prevents too many small parties from winning seats.

to 75. These changes to the electoral system were widely viewed as unfairly favoring the ruling party. A report issued by the Parliamentary Assembly of the Council of Europe (PACE) noted that "the ruling party is considered likely to gain most [of the] majoritarian seats under a first-past-the-post system" which has led to "allegations that the ruling party intends to maintain a constitutional majority in the new parliament by manipulating the election system in its favour" (2008).

For the 2008 Parliamentary Election in May, the 150 seats were to be elected according to a mixed system for a four-year term. Under the list-based proportional system, 75 members of Parliament were elected in one nationwide constituency. The other 75 were elected in 75 single-mandate districts. As part of the changes to Election Code, the threshold of votes needed in order to be eligible for seats was lowered from 7 percent to 5 percent. The ruling party, the United National Movement, garnered the most votes in the proportional race winning 59.1 percent of the votes.

Several opposition parties united and ran as the Joint Opposition Coalition, which gained 17.7 percent. The Christian-Democrats and Labour parties cleared the threshold with 8.7 percent and 7.4 percent of the votes, respectively. Immediately prior to the election, the Republican Party broke off from the Joint Opposition Coalition and was unable to clear the threshold as they received only 3.8 percent of the votes. Seven other parties received less than one percent of the vote (see the table of results below). Turnout for the proportional election was 53.4 percent. The ruling party (UNM) won 71 of the 75 majoritarian single-mandate seats.

Table 2: Results of the Proportional System

Party/Bloc Name	Votes	Percent	Seats
United National Movement – for Victorious Georgia	1,050,237	59.18%	48
Electoral Bloc The Joint Opposition (National Council, New Rights)	314,668	17.73%	15
Giorgi Targamadze – Christian-Democrats	153,634	8.66%	6
Shalva Natelashvili – Labour Party	132,092	7.44%	6
Party of Republicans	67,037	3.78%	-
Electoral Bloc Alliance of the Rights, Topadze-Industrials (MGS, Unity, EDP)	16,440	0.93%	-
Political Union “Christian-Democratic Alliance (QDA)”	15,839	0.89%	-
Political Union of Citizens’ ”Georgian Policy”	8,231	0.46%	-
“Traditionalists - Our Georgia and Party of Women”	7,880	0.44%	-
Political Union “Union of Sportsmen”	3,308	0.19%	-
National Party of Radical-Democrats	3,180	0.18%	-
“Our Country” Party	2,101	0.12%	-
Registered Voters	3,465,736		
Voters Participating	1,850,407	53.39%	
Invalid Votes	56,077	3.16%	

The official polling station data for the proportional race of the Parliamentary Election were taken from the Central Election Commission of Georgia’s website.[d] There were 3,604 polling stations and for each station the following information was listed: the number of voters registered, the number of ballots received, the number of voters participating as of 12:00 pm, the number of voters participating as of 5:00 pm, the final number of voters participating, the total

[d] http://www.cec.gov.ge/ (accessed July 2008).

number of spoiled ballots, the total number of invalid votes, the number of unused ballots, and the total votes for each party. I first converted the polling station information into a format that I could analyze. In addition, for each polling station, the second leading digit was extracted from each party's vote count, the number of invalid ballots, and the number participating. Approximately 70 stations were "special" polling stations, such as military bases, where the ballot boxes were brought to another location for counting. The domestic observation group, International Society for Fair Elections and Democracy (ISFED), noted irregularities in the stations that counted the special stations. The results from the proportional election were annulled in 37 polling stations due to irregularities. In addition, the mobile ballot box results were annulled in 10 polling stations. Please note that any descriptive statistics not found in other tables in the body of the paper are listed in a Table A1 in Appendix 1.

Further data came from ISFED, a non-profit, non-governmental, and non-partisan organization, whose main purpose is monitoring elections. ISFED, as detailed further below, monitored the 2008 Parliamentary Election. Some background concerning ISFED's role in the Republic of Georgia may be helpful in providing context for a later discussion about their observations. ISFED has been observing elections in the Republic of Georgia since 1995 and played a key role in catalyzing the peaceful Rose Revolution after the fraudulent Parliamentary elections in 2003. In 2003, the Central Election Commission (CEC) initially announced that the party of then President Eduard Shevernadze had won a majority of votes and the opposition, which included Mikhail Saakashvili's United National Movement Party, had received only a small percentage of votes. However, ISFED had conducted an independent, parallel vote count[e] which revealed significant flaws in the election and showed that the results announced by the CEC were false. Soon after, thousands took to the streets and successfully demanded Shevernadze's resignation. Those demonstrations culminated in what was dubbed the Rose Revolution. In 2004, Mikhail Saakashvili was elected president by an overwhelming majority.

ISFED continued to monitor elections after the Rose Revolution including 2008 elections. It should be noted that I worked closely with ISFED during the 2008 Parliamentary Elections on their election monitoring methodology and provided technical assistance to ISFED during the 2008 elections. For the 2008 Parliamentary Election, ISFED's Observation Form assessed aspects of the Election Day process and asked qualitative questions such as "was the polling station prepared to open?" "were all necessary materials present?" "was the secrecy of the ballot preserved?" and "were unauthorized persons present?" ISFED also monitored the pre-election period leading up to the Parliamentary Elections and issued several updates as well as a report on the pre-election period.[f] On Election Day, ISFED deployed approximately 620 stationary observers to a nationally representative, random sample of polling stations in order to evaluate the quality of the process and project results for the proportional election. In addition, ISFED observers covered almost all of the 72 stations that would be including the results of the "special" stations in their totals and they deployed teams of mobile observers that went from

[e] The National Democratic Institute describes Parallel Vote Tabulations (PVTs) as "an advanced observation technique that has been used for two decades by domestic monitoring groups and political parties around the world to promote the integrity of elections and to detect when election results and/or election-day processes have been manipulated" (NDI, date unknown). More specifically, a PVT is the "use of rapidly transmitted quantitative and qualitative observer data from a statistically sound, representative random sample of polling stations to produce a comprehensive and systematic assessment of the Election Day process and project results (NDI, date unknown).
[f] ISFED's reports are available (in both English and Georgian) on their website: http://isfed.ge/elections/reports/ (accessed 2008, 2012).

station to station. ISFED kindly provided the data collected by observers deployed to the nationally representative sample of polling stations.

There are specific provisions in Georgia's Election Code[g] that clearly set out that domestic observer organizations such as ISFED have the right to monitor at all levels of the election administration. ISFED began observing the electoral process in the pre-election period and throughout the post-election period. In regard to the Election Day process, ISFED issued a statement affirming a general improvement in the conduct of the election but also noted a few specific problems. ISFED noted problems with turnout through the mobile or portable ballot boxes that move on Election Day to those who cannot come to a polling station (due to hospitalization or immobility). They observed and documented cases where the turnout in the mobile ballot box was 100 percent or greater. In addition, when compared with the earlier Presidential election, ISFED observed that "the number of special precincts and the number of voters registered in such precincts increased. Subsequently, the number of the precincts to which such special precincts were attached also increased. In the election precincts where special precinct voting results were tabulated, United National Movement got higher quantity of votes" (2008).

The Georgian Government also invited several international observation groups to monitor the electoral process. A general summary of the international observation organization's findings, further detailed below, shows that the 2008 Georgian Parliamentary elections are an excellent test case for Benford's Law, with some levels of fraud coexisting beside mostly clean results, all of it closely monitored. While it may seem strange that the government may intend to

[g] Article 70 of the Election Code outlines the specific rights of both domestic and international observers.

benefit from varying degrees of election fraud and still invite election observers, the current international environment makes it challenging *not* to invite observers. As Susan Hyde notes "inviting international election monitors has become an international norm" and further, if a government does not choose to invite international observers, then it is assumed that the government is attempting to hide electoral manipulation (2011). The Organization for Security and Co-operation in Europe (OSCE) has a special institution, the Office for Democratic Institutions and Human Rights (ODIHR), that monitors elections and it is viewed by many as the leading European organization in the field of election observation. The OSCE-ODIHR deployed an international observation mission on April 10, 2008, to observe the Parliamentary Elections in Georgia. The OSCE-ODIHR mission joined with delegations from the OSCE Parliamentary Assembly, the Parliamentary Assembly of the Council of Europe, the European Parliament, and the NATO Parliamentary Assembly to deploy more than 550 short-term observers for Election Day. The OSCE-ODIHR gave the following assessment: "overall, these elections clearly offered an opportunity for the Georgian people to choose their representatives from amongst a wide array of choices. The authorities and other political stakeholders made efforts to conduct these elections in line with OSCE and Council of Europe commitments." However, the mission "identified a number of problems which made this implementation uneven and incomplete." Their report acknowledged that the opening and voting process seemed generally to go well but there were "procedural shortcomings, especially with regard to inconsistent application of inking[h] procedures" (2008). The OSCE-ODIHR also noted problems with mobile voting,

[h] In 11 per cent of polling stations visited by OSCE-ODIHR observers, voters "were not always checked for traces of invisible ink, and in 9 percent, inking was not always applied" (2008). According to the law, a voter's finger must be checked for ink as they enter the station. If a voter has not been inked (indicating that they have not yet

remarking that a "relatively high number of voters were added to mobile voter lists." The counting process was assessed less positively, with "significant procedural shortcomings" observed, such as Precinct Election Commissions failing to perform basic reconciliation procedures before opening the ballot boxes, inconsistent determination of ballot validity, and several instances where voters' choices were not announced aloud during the count (2008). The OSCE-ODIHR observers assessed the process of tabulation at the District Election Commission (DEC) level. In 25 percent of the DECs, the tabulation process was assessed as negative or lacking in transparency. The IEOM observers assessed the process most negatively in the Mtskheta-Mtianeti, Shida Kartli, Kvemo Kartli, and Kakheti regions (2008).

Overall, the Parliamentary Election was evaluated as an improvement over previous elections but was still marred with specific irregularities with regard to the counting and tabulation process. Thus, the results should provide us with a good example of "noisy" election data where problems were mainly concentrated during the counting process. Except for a few cases, elections are rarely assessed as being entirely fraudulent or entirely "clean." As such, the nuanced assessments of Georgia's Parliamentary Election are a good example of the more mixed election in terms of "cleanness." I next turn to the results of my empirical estimations.

voted), then he or she proceed to have his or her name checked on the voters list and once a ballot has been issued, the voter's finger is inked in order to prevent multiple voting. Proper application of, and checking for, inking is one the fundamental safeguards in the Election Day process.

V. Results

Each of my empirical models is described alongside the results due to the large number of estimations. I move from the least sophisticated test to the most sophisticated regressions with the best possible data to see beyond what threshold, if any, Benford's Law can accurately detect fraud and act as a substitute for observers. As a general principle, the more sophisticated the method, the greater the data demands are; thus, it becomes less likely Benford's Law will function without observer data, effectively rendering it useless. I first performed the overall chi-squared test like previous studies (Mebane 2006, Mebane 2007, Mebane 2008, Mebane 2010a, Mebane 2010b, Mebane 2010c), then I carried out the more sophisticated chi-squared test of the deviation from the expected second-digit mean like Mebane (Mebane 2008, Mebane and Kalinin 2009, Mebane and Kalinin 2010, Mebane 2010a, Mebane 2010b). Then I examined the second digit distribution of selected parties, and finally I designed multivariate regressions with greater data needs. In summary, the results do not indicate Benford's Law is a suitable tool for detecting electoral fraud or an adequate substitution for observers, despite every effort to allow for success.

I first examine the simplest implementation of Benford's Law, a Pearson's chi-squared test of whether the actual results are significantly different from the expected result. When looking at the resulting Pearson's chi-squared test statistics for the distribution of the second-digits of each party's vote counts, the invalid votes, and the number participating, one might conclude that the Parliamentary Election was extremely fraudulent. Table three below shows the Pearson chi-squared test statistic, the p-value and the number of observations for each of the parties' second-digits as well as for the number of voters participating and the number of invalid votes. The second-digit distributions for nine of the twelve parties are significantly different than what is expected according to Benford's Law. The distribution for invalid votes is also significantly different though participation is not.

Table 3: Chi-squared Test Statistics for the distribution of second-digits

Party/Bloc Name	Chi-squared second-digit Benford test statistic		p-value	N	Percent of Vote
United National Movement – for Victorious Georgia	24.6	**	0.00	3560	59.18%
Joint Opposition Coalition (National Council, New Rights)	26.0	**	0.00	3184	17.73%
Giorgi Targamadze – Christian-Democrats	23.1	**	0.01	2822	8.66%
Shalva Natelashvili – Labour Party	11.9		0.22	2703	7.44%
Party of Republicans	26.6	**	0.00	2230	3.78%
Political Union of Citizens' "Georgian Policy"	22.2	**	0.01	168	0.46%
Electoral Bloc Alliance of the Rights, Topadze-Industrials	97.0	**	0.00	404	0.93%
Political Union "Union of Sportsmen"	18.6	**	0.03	11	0.19%
National Party of Radical-Democrats	9.3		0.41	20	0.18%
Political Union "Christian-Democratic Alliance (QDA)"	69.6	**	0.00	511	0.89%
"Traditionalists - Our Georgia and Party of Women"	35.6	**	0.00	188	0.44%
"Our Country" Party	10.1		0.35	13	0.12%
Participation	12.7		0.18	3600	53.39%
Invalid Votes	44.7	**	0.00	2158	3.16%

** Significant at the 95% level of confidence.

Next I examine the somewhat more precise test of deviations from the expected second-digit mean. Table 4 shows the actual second-digit means, $\hat{j}$, the standard errors, the upper and lower bounds for the 95 percent confidence level, and the amount (in percentage) that the second-digit mean differs from the expected mean (4.187). This test reveals fewer parties with a significant test statistic. As a result, the election looks a bit less fraudulent from this analysis.

Table 4: Second-digit Means

Party/Bloc Name	Mean of Second-digit		s.e.	95% CI l.b.	u.b.	Percent deviation	n
United National Movement – for Victorious Georgia	4.400	**	0.05	4.306	4.494	0.051	3560
Joint Opposition Coalition (National Council, New Rights)	3.999	**	0.05	3.901	4.097	-0.045	3184
Giorgi Targamadze – Christian-Democrats	4.052	**	0.06	3.944	4.160	-0.032	2822
Shalva Natelashvili – Labour Party	4.149		0.06	4.040	4.259	-0.009	2703
Party of Republicans	3.927	**	0.06	3.808	4.046	-0.062	2230
Political Union of Citizens' "Georgian Policy"	3.298		1.31	0.735	5.860	-0.212	168
Electoral Bloc Alliance of the Rights, Topadze-Industrials	7.376	**	0.84	5.727	9.026	0.762	404
Political Union "Union of Sportsmen"	6.992		5.34	####	####	0.670	11
National Party of Radical-Democrats	8.348		3.88	0.751	####	0.993	20
Political Union "Christian-Democratic Alliance (QDA)"	7.160	**	0.75	5.694	8.626	0.710	511
"Traditionalists - Our Georgia and Party of Women"	7.244	**	1.24	4.823	9.666	0.730	188
"Our Country" Party	7.148		4.88	####	####	0.707	13
		**					
Participation	4.285	**	0.05	4.190	4.380	0.023	3600
Invalid Votes	3.807	**	0.06	3.688	3.925	-0.091	2158

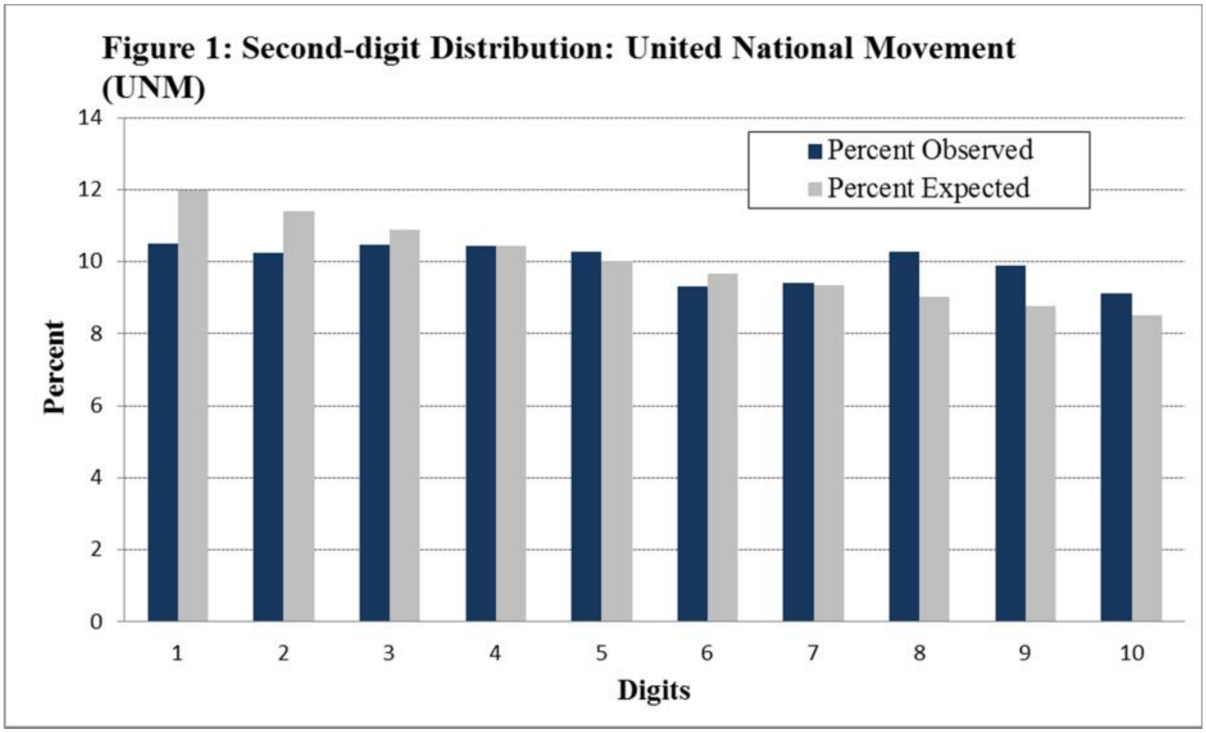
** Significant at the 95% level of confidence.

Subsequently, I examine the distribution of the second-digits for selected parties flagged by the previous tests and who gained enough votes to clear the threshold. When I look at the second-digit means for the top parties, UNM is the only party that has a mean (4.400) greater than the expected mean of 4.187. An examination of the distribution of each second-digit for UNM shows that there are more 7s and 8s than what is expected according to Benford's Law and there are fewer 0s and 1s than what is expected (see Table 5 and Figure 1 below). This distribution would imply that the vote counts for UNM are larger than they should be according to Benford's Law; therefore, they may be benefitting from the manipulation of vote counts.

Table 5: Second-digit distribution for United National Movement (UNM)

Digit	Count	Percent Observed		Percent Expected
0	374	10.5	**	12.0
1	365	10.3	**	11.4
2	373	10.5		10.9
3	372	10.4		10.4
4	366	10.3		10.0
5	332	9.3		9.7
6	335	9.4		9.3
7	366	10.3	**	9.0
8	352	9.9	**	8.8
9	325	9.1		8.5
Total	3560			

** Significant at the 95% level of confidence.



The Joint Opposition Coalition, on the other hand, would appear to be harmed as their second-digit mean is less than what is expected because their vote counts contain too many 3s and too few 8s and 9s (see Table 6 and Figure 2 below). Targamadze's Christian-Democrats also has a second-digit mean that is lower than the expected mean as their vote counts seem to contain too many 3s (see Table 7 below).

Table 6: Second-digit distribution for the Joint Opposition Coalition

Digit	Count	Percent Observed		Percent Expected
0	410	12.9		12.0
1	367	11.5		11.4
2	371	11.7		10.9
3	376	11.8	**	10.4
4	300	9.4		10.0
5	325	10.2		9.7
6	270	8.5		9.3

7	294	9.2		9.0
8	245	7.7	**	8.8
9	226	7.1	**	8.5
Total	3184			
** Significant at 95% confidence level.				

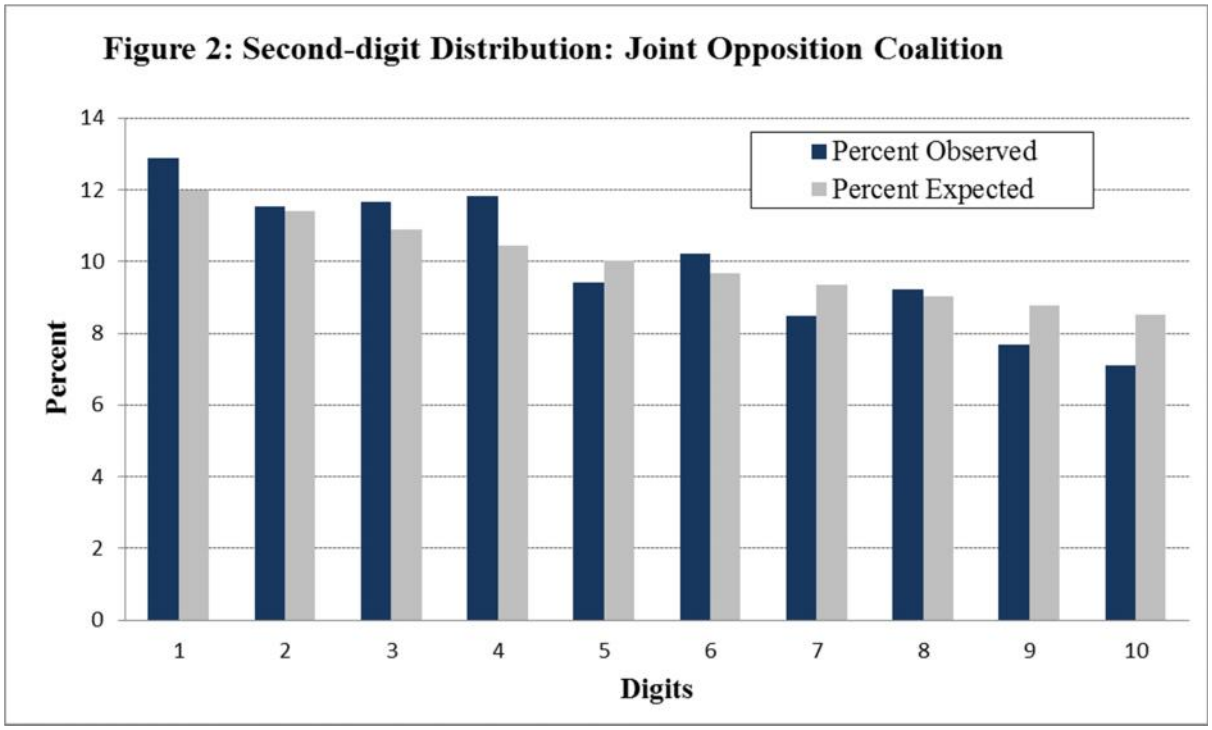
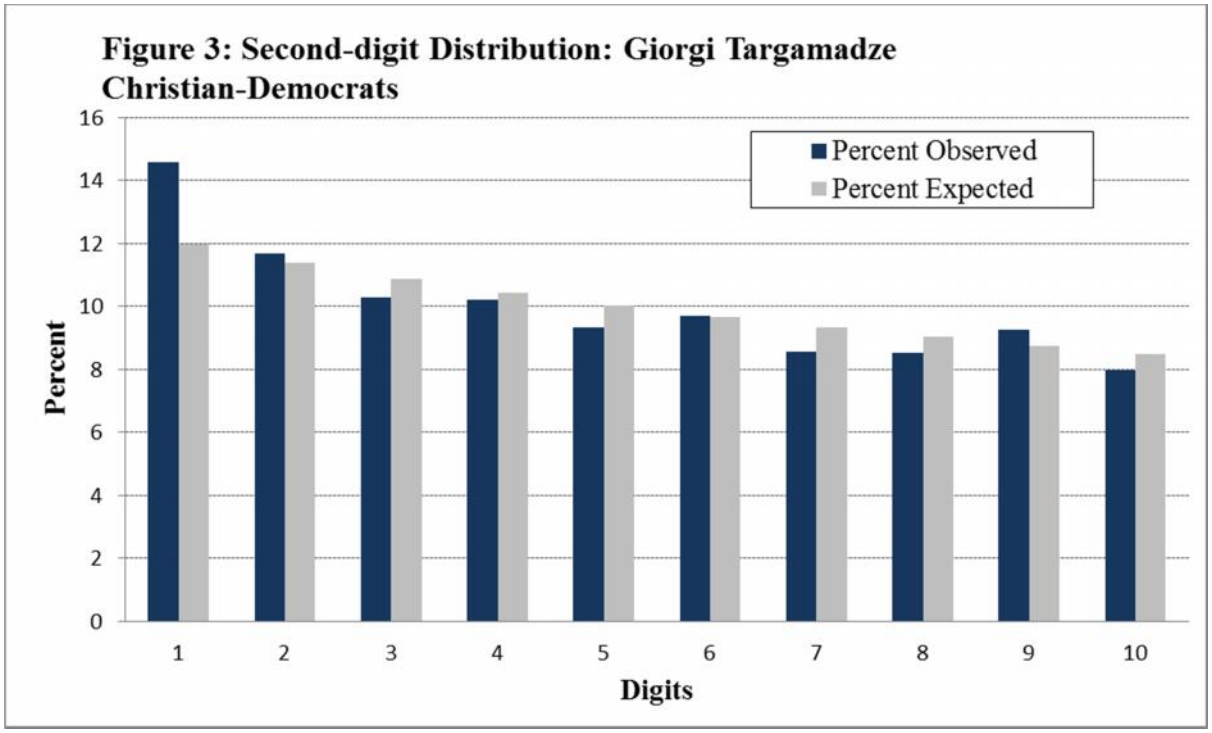


Table 7: Second-digit distribution for Giorgi Targamadze Christian-Democrats

Digit	Count	Percent Observed		Percent Expected
0	411	14.6	**	12.0
1	329	11.7		11.4
2	290	10.3		10.9
3	288	10.2		10.4
4	263	9.3		10.0
5	274	9.7		9.7
6	241	8.5		9.3

7	240	8.5	9.0
8	261	9.2	8.8
9	225	8.0	8.5
Total	2822		100

** Significant at 95% confidence level.



From the results of these test statistics alone, one might get the impression that the overall election was extremely fraudulent. In addition, one might assume that the ruling party, UNM, committed fraud by taking votes away from the Joint Opposition Coalition, the Christian-Democrats, and possibly the Republican Party. However, we know from both international and domestic observer reports that the Election Day process went well in a majority of stations. Even if we assume the Benford test statistics are correct and the election observers are wrong, we still do not have a theory for what specific type of fraud occurred, nor whether it was localized to specific areas or districts that may be skewing the digit distributions and means.

I now turn to several regressions that use the percentage at which the second-digit means deviate to see whether the regressions explain some indicators of fraud better than others. The first set of indicators concern participation levels and turnout. Several types of fraud, such as ballot stuffing or multiple voting, increase the votes of a specific party or parties and in doing so artificially inflate overall participation.

Voters Participating

Using the log of participation, I estimated the following simple regression without controls:

$$\text{Log(participation)} = \beta 0 + \beta 1(\text{UNM}) + \beta 2(\text{Joint Opposition}) + \beta 3(\text{Christian-Democrats}) + \beta 4(\text{Labour}) + \beta 6(\text{Republicans}) + \beta 6(\text{Invalid}) + \varepsilon \quad (8)$$

where β1- β6 indicate the percentage by which the second-digit means for the party vote counts and invalid votes deviate from the expected mean.

From the overall chi-squared test and the second-digit mean tests, one would expect that the coefficient for the United National Movement would be significant; but, it is not. Instead, the coefficients for the Christian Democrats and invalid votes are significant (see Table 8). The results imply that, holding other variables constant, a one percentage point increase in the digit deviation of the Christian Democrats is associated with a 72.2 percent increase in participation. Oddly, the coefficient for invalid votes is also significant and positively associated with participation. It would seem from this simple model that holding other variables constant, a one percentage point increase in the second-digit deviation of invalid votes is associated with a 96.4 percent increase in participation. During the Presidential elections in Georgia, there were allegations that the commission members purposefully invalidated the ballots of the Joint Opposition candidate, thereby increasing the number of invalid ballots and decreasing the number of votes for the Joint Opposition. This was not noted as a problem for the Parliamentary Elections but perhaps votes were invalidated. However, these results imply votes are being invalidated and instead of decreasing, the vote counts for the Christian-Democrats are inflated as participation increases. It is possible that the Christian-Democrats are benefitting from some type of fraud and another party is having its votes invalidated. However, such a conclusion is not consistent with the interpretation of the overall second-digit means. Together the six deviation variables explain 13.4 percent of the variation in the log of participation.

Table 8: Digit effects on (Log of) Participation

Explanatory Variable				
UNM	-.378	-.374	-.345	-.279
	(.350)	(.348)	(.279)	(.287)
Joint Opposition	-.085	-.058	-.060	-.149
	(. 425)	(.420)	(.277)	(.284)
Christian-Democrats	**.722 ****	**.707 ****	**.498 ****	**.428 ****
	(.203)	**(.202)**	**(.123)**	**(.130)**
Labour Party	-.207	-.211	-.001	.220
	(.439)	(.434)	(.200)	(.217)
Republicans	**.558 ***	**.549 ***	.299	**.374 ***
	(.296)	**(.295)**	(.217)	**(.213)**
Invalid Votes	**.964 ****	**.953 ****	**.458 ****	**.397 ***
	(.417)	**(.414)**	**(.207)**	**(.213)**
Intercept	6.32	6.30	.530	**3.656 ****
	(.063)	(.061)	(1.184)	**(1.615)**
Controls:				
Station counted another station	---	**.760 **** **(.075)**	**.626 **** **(.067)**	**.652 **** **(.067)**
Proportion of station	---	---	**4.70 ****	1.606

Aged 18 – 34			**(1.344)**	(1.640)
Proportion of station Aged 35 – 49	---	---	**6.22 **** **(1.132)**	**3.326 **** **(1.663)**
Proportion of station Aged 50 – 64	---	---	**8.82 **** **(1.326)**	**5.183 **** **(1.881)**
Proportion of station Aged 65 – 79	---	---	**4.39 **** **(1.660)**	**.956 **** **(2.067)**
Proportion of station Aged 95 and older	---	---	**-29.800 **** **(8.549)**	**-61.469 **** **(11.904)**
Observer noted Serious irregularities	---	---	---	**.485 **** **(.108)**
N	3,438	3,438	3390	667
R-squared	0.134	0.1602	0.3100	0.4188
F-statistic	**5.91 ****	**24.34 ****	**55.2 ****	**32.47 ****

Note:
Errors are adjusted for clustering at the district level, and are robust for unobserved heteroskedasticity for all regressions.
The variable for the proportion of stations aged 80 - 94 was omitted to avoid collinearity.
** Significant at 95% level of confidence.
* Significant at 90% level of confidence.

A more sophisticated implementation of Benford's Law is a multivariate regression with additional controls. It is possible that the inclusion of control variables will alter the coefficients and improve the accuracy and precision of fraud detection. Unfortunately, in many countries socio-economic data is not freely available, and, where it is released, it is often not available regional level let alone the level of the polling station. Regardless, proponents of digit tests argue that digit tests can be applied "to cases where data availability is poor" and in situations where there is "poor national-level information about the make-up of the voting population, and virtually no information at the subnational level" (Beber and Scacco, forthcoming 2012).

The results thus far argue against the use of Benford's Law in a data-poor environment. In this case, though, I do have some additional information about the characteristics of the polling stations. Prior to the election, the Central Election Commission released information about the polling stations designated a “special” station. Since the votes for another station were added to their votes, these stations have a much higher number of voters participating than those stations that do not include another polling station. For example, ballots cast on a military base are brought to a nearby polling station where they are comingled and counted together with the polling station’s votes. Given this information, I added an indicator variable that is equal to one for those stations that counted a “special” station and zero otherwise. I would expect this variable to be statistically significant and indeed it is (see Table 8). Holding the other variables constant, a station that counted another station increases the participation level by 76.0 percent. The coefficients for the Christian-Democrats and the invalid vote decrease slightly (from 72.1 percent and 96.4 percent to 70.7 percent and 95.3 percent, respectively) when adding this variable.

Although access to socio-economic data is not easily available in Georgia, the election commission made the list of voters available to the political parties and ISFED prior to Election Day in 2008. Thus, the next regression in Table 8 includes a set of control variables that measure the proportion of specific age groups registered at each polling station. Using the list of voters, I extracted birthdates and calculated voters’ ages at the time of the elections. For each polling station, I calculated the proportion of voters aged 18 to 34, the proportion aged 35 to 49, the proportion aged 50 to 64, the proportion aged 65 to 79, the proportion aged 80 to 94, and the proportion aged 95 and older. When adding the age controls, the explanatory power of the model goes up as the R-squared is now 0.3100 and the coefficients for the Christian-Democrats and invalid votes are still statistically significant. However, the coefficient from the Christian-Democrats decreases dramatically and now a one percentage point increase in the deviation increases participation by 49.8 percent (rather than 72.2 percent), holding all other variables constant. The effect of invalid votes decreases dramatically as well from 96 percent to 46 percent. This suggests that there was significant omitted variable bias in the simple regression and that the coefficients for the deviations were biased upwards. Again this would appear to invalidate the use of second-digit means in a data-poor environment. Regardless, this model appears to be highly sensitive to the addition of controls; thus, from a practical point of view, unless you have access to good controls, the model does not work well.

Next, I add another control variable, this one is an indicator variable denoting whether "serious irregularities" were committed in a station and reported by an ISFED observer. ISFED's observers reported "serious irregularities" in stations where incidents of violence, threats, seizure of the ballot box, or ballot stuffing occurred. Since ISFED observers were in a sample of the polling stations, this reduced the overall number of observations for the regression. By adding the "serious irregularity" observed variable, one can see in Table 8 that the explanatory power of the model increases substantially from an R-squared of 0.31 to 0.42. In this model, with controls for counting stations, age proportions and reported irregularities, the deviations for the Christian-Democrats and invalid votes are still significant and only decrease slightly. The coefficient for the Republicans is significant at the 90 percent confidence level and it would appear that for a one percentage point increase in the deviation of the Republican's second-digit mean, participation increases by 37.4 percent. From these results, we might guess that the Christian-Democrats committed fraud or at the very least that they are the beneficiary of fraud and that the Republicans also benefitted from fraud. And yet, fraud was not widespread and when irregularities were observed they favored the ruling party, UNM, not the Christian-Democrats or the Republicans.

Votes Recorded in the Last Three Hours

Perhaps part of the problem with the previous model is that high rates of overall participation do not necessarily denote fraud since voters for some parties may be more enthusiastic than those for others. It may be better to examine participation patterns throughout the day, and particularly the end of the day. Focusing on the end of the day may give a more precise indicator for irregularities. In Georgia, polling station officials are required to record participation throughout the day at noon, 5:00 pm, and at the end of the day at 8:00 pm. Historical data suggest that voters in Georgia tend to vote in the morning. If I break the day into three parts: the morning period (8:00 am to noon), the afternoon period (noon to 5:00 pm), and the evening period (5:00 pm to 8:00 pm), then I can look at the average number of voters processed per minute during each period. On average, 213 voters turned out in the morning period. During those hours, polling stations processed less than one voter (.89) per minute. Or put another way, on average, it took a little over a minute to process one voter (240 minutes/213 voters). In the afternoon period, 189 voters cast ballots meaning that two voters were processed, on average, approximately every three minutes. In the final three hours of voting, 110 voters turned out, on average. During that period it took approximately three minutes to process two voters. Often ballot stuffing, or other means of artificially inflating the vote for one party such as altering the counts, occurs at the end of the day when votes are counted and entered into the results sheet. At that point, it may be clear to the perpetrators that there is a need to inflate the votes for their favored party or candidate. Practically speaking, it is easier to inflate vote counts at the end since there is less scrutiny: voters are no longer present and observers may not be present or are fatigued since the counting process can often continue into the early morning hours of the following day. Thus, high rates of voting in the last three hours can be a very good indicator of ballot stuffing or manipulation of vote counts.

Using a variable for the number of votes recorded in the last three hours in those stations that did *not* count another station (Votes recorded in last three hours), [i] I estimated the following simple regression without controls:

$$\text{(Votes recorded in last three hours)} = \beta 0 + \beta 1\text{(UNM)} + \beta 2\text{(Joint Opposition)} + \beta 3\text{(Christian-Democrats)} + \beta 4\text{(Labour)} + \beta 6\text{(Republicans)} + \beta 6\text{(Invalid)} + \varepsilon \quad (9)$$

where β1- β6 indicate the percentage by which the second-digit means for the party vote counts and invalid votes deviate from the expected mean.

In this model, the digits explain much less of the variation in participation during the last three hours as the R-squared is only 0.06 (see Table 9). Again, the deviation for the Christian-Democrats are significant and for each one percentage point increase in their deviation, 58 more votes are recorded in the last three hours. Similarly, for each one-percentage point increase in the deviation for invalid votes, 70 more votes are recorded in the last three hours. Interestingly, the deviation for the Joint Opposition is significant at the 90 percent level of confidence. The sign of the coefficient suggests that as the second-digit mean deviates away from the mean in a negative

[i] I excluded those polling stations that counted another polling station's votes as they will have a large bulk of votes at the end of the day precisely because another stations votes are added and not necessarily because of irregularities.

way, meaning the second-digit mean is smaller than the expected mean (i.e., votes are taken away), but then 62 more votes are recorded in the last three hours. The results from this model suggest that the votes for the Christian-Democrats are being manipulated and inflated, and that the Joint Opposition's votes are being invalidated, thereby inflating the number for invalid votes and decreasing the total votes for the Joint Opposition. The digit deviation for the UNM, the governing party, appears to have no statistically significant effect.

Table 9: Digit effects on votes recorded in the last three hours

Explanatory Variable			
UNM	-41.774 (34.689)	-22.538 (43.058)	-34.494 (40.884)
Joint Opposition	**-62.024 *** **(34.907)**	-68.511 (42.846)	**-71.971 *** **(41.377)**
Christian-Democrats	**57.512 **** **(14.684)**	**37.848 *** **(19.882)**	**42.048 **** **(19.673)**
Labour Party	-2.234 (24.182)	28.733 (35.365)	23.554 (33.798)
Republicans	26.501 (19.748)	40.063 (24.475)	37.128 (23.798)
Invalid Votes	**70.447 **** **(24.544)**	**77.837 **** **(30.375)**	**73.477 **** **(31.203)**
Intercept	**118.716 **** **(5.690)**	121.307 (7.360)	359.324 (276.935)
Controls:			
Observer noted Serious irregularities	---	**97.496 **** **(28.181)**	**83.574 **** **(27.868)**
Proportion of station Aged 18 – 34	---	---	-112.54 (279.888)
Proportion of station Aged 35 – 49	---	---	-268.694 (268.038)
Proportion of station Aged 50 – 64	---	---	-313.857 (305.317)
Proportion of station Aged 65 – 79	---	---	-356.947 (370.517)
Proportion of station Aged 95 and older	---	---	**-5669.511 **** **(1763.836)**

N	3331	587	586
R-squared	0.0590	0.0981	0.1272
F-statistic	**5.33** **	**4.69** **	**4.45** **

Note:
Errors are adjusted for clustering at the district level, and are robust for unobserved heteroskedasticity for all regressions.
The variable for the proportion of stations aged 80 to 94 was omitted to avoid collinearity.
** Significant at 95% level of confidence.
 * Significant at 90% level of confidence.

Next, I add control variables and examine whether the estimates become more precise. When including the indicator variable for whether an ISFED observer reported a serious violation (i.e., serious violation equals 1, 0 otherwise), the explanatory power of the model almost doubles from an R-squared of 0.06 to 0.10. As expected, the serious violation reported variable is statistically significant and for each polling station that an ISFED observer reported a serious irregularity the number of votes recorded in the last three hours increases by 97. When controlling for this irregularity variable, the effect of the Christian-Democrats decreases from 58 votes to 38 votes, suggesting that in the simple model without controls, the coefficient for the Christian-Democrats was again biased upwards.

When adding the controls for the age proportions, the model then explains 12.7 percent of the variation in votes recorded in the last three hours. The effect of the Christian-Democrats on the number of votes recorded in the last three hours increases slightly, while the effect of invalid votes decreases slightly. The effect of the deviation in the second-digit mean for the Joint Opposition is larger than in the simple model and is statistically significant at a 90 percent confidence level.

High Rates of Participation in the Last Three Hours

In addition to looking at the number of votes recorded in the last three hours, one can choose an appropriate cutoff for what seems like a reasonable number of votes cast during the last three hour period. Consider the number of steps that a voter must proceed through upon entering the station and casting a vote: the voter must be checked for ink, present their voter identification, proceed to have their named check on the voter list, sign next to their name, have their finger inked, receive their ballot, go into the voting booth, make their choice, and deposit their ballot in the ballot box. Given that the average number of voters registered per station is high (990 voters) and there is a limited number of polling station officials, it is reasonable for the process to take several minutes. It is therefore highly unlikely that even the most efficient polling station could, on average, process more than one voter per minute. Therefore, it makes sense to create a cutoff for what is a reasonable number of voters per minute and apply that to the 3 hour time period at the end of the day. Since a turnout estimate is recorded at 5:00 pm and final turnout is recorded three hours later at the end of the voting at 8:00 pm (180 minutes), I was able to use that information to create an indicator variable with a cutoff of 200 votes in the last three hours. I chose 200 rather than 180 to allow for the fact that the 5:00pm number may have been rounded and could have been rounded down slightly. This variable is then an indicator for those stations that process an unreasonably high number of voters per minute in the last three hours (on average). The variable takes on a value of 1 for those stations where 200 or more voters turned out in the last three hours (there are 290 such stations) and takes on a value of zero for those stations with lower rates of voting. Then, I use this indicator variable as the dependent variable in the following simple regression without controls and see how much variation the second-digit deviations explain:

$$\text{Logit (200 or more votes in last three hours)} = \beta 0 + \beta 1(\text{UNM}) + \beta 2(\text{Joint Opposition}) + \beta 3(\text{Christian-Democrats}) + \beta 4(\text{Labour}) + \beta 6(\text{Republicans}) + \beta 6(\text{Invalid}) + \varepsilon \quad (10)$$

where β1- β6 indicate the percentage by which the second-digit means for the party vote counts and invalid votes deviate from the expected mean.

Only the Joint Opposition Coalition coefficient is significant at a 95 percent confidence level. A one percent increase in the deviation results in a factor of 0.019 that a station has processed an unreasonably high number of voters. It would seem from Table 10, that no single party or coalition (not the Christian-Democrats or UNM) benefits as the likelihood of a station having a high rate of participation increases. This holds true even when including controls for the age proportions and where irregularities were noted. Instead these results would lead one to believe that votes may be taken away from the Joint Opposition Coalition but the model does not show which other party benefits.

Table 10: Digit effects on high <u>rates</u> of participation (i.e. 200 or more votes for an average of more than one voter per minute) in last three hours

Explanatory Variable	Odds ratio (OR)	Odds ratio (OR)
UNM	1.704 (2.600)	2.620 (5.281)
Joint Opposition	**.019** ** **(.027)**	**.127** ** **(.023)**
Christian-Democrats	3.050 (2.624)	4.170 (4.993)
Labour Party	.580 (.666)	3.703 (6.472)

Republicans	.654 (.473)	.945 (929)
Invalid Votes	5.673 (6.265)	17.127 * (26.901)
Controls:		
Observer noted Serious irregularities		4.526 * (3.998)
Proportion of station Aged 18 – 34		.000 (.001)
Proportion of station Aged 35 – 49		.000 (.000)
Proportion of station Aged 50 – 64		.000 * (.000)
Proportion of station Aged 65 – 79		.000 (.000)
Proportion of station Aged 95 and older		.000 ** (.000)
N	3331	586
Pseudo R-Squared	0.0454	0.1296
Wald chi-squared statistic	12.85 **	42.20 **

Note:
Errors are adjusted for clustering at the district level, and are robust for unobserved heteroskedasticity for all regressions.
The variable for the proportion of stations aged 80 to 94 was omitted to avoid collinearity.
** Significant at 95% level of confidence.
* Significant at 90% level of confidence.

In addition to the variables discussed previously, I ran additional models using an indicator variable for those stations with a turnout percentage of 90 percent or more; however, the model was not significant with or without controls. I also tried using an indicator variable for high vote share by one party (90 percent or more) and the deviations did not explain any of the variations. Therefore, the results are not listed here.

Next, I change approaches slightly. In the case of the Parliamentary Elections, I also have data collected by the domestic observation organization ISFED. Since ISFED observers collected information on specific processes and not just outcomes, like turnout, then it may be that the Benford digit deviations will do a better job of predicting what the ISFED data reveal. The ISFED observers were deployed to a sample of 626 polling stations. They collected information on the process in the polling station, evaluating whether irregularities were observed, whether inking was consistently applied, if a complaint was filed, if materials were missing, whether secrecy of the vote was violated, and if unauthorized individuals were present in the station.[j] The summary statistics for these data are listed in Table 11 below.

Table 11: Summary Statistics for domestic observer data

	N	Mean	Number "Yes"	Number "No"
Were officials not consistently checking for ink?	617	0.1	40	577
Were unauthorized people present?	621	0.0	10	611
Were serious irregularities observed?	625	0.0	6	619
Was a complaint filed?	621	0.1	59	562
Was the station missing materials?	621	1.0	23	598
Was secrecy of the vote violated	617	0.1	56	561

Polling Station Officials did not consistently check for ink

From the domestic observer data, I created an indicator variable that was equal to 1 if the officials in the polling station did not consistently check voters for ink as required by law and set to zero if they always checked. The motivation behind applying ink to a voter's finger is to make it difficult for a voter to vote more than once. As such, inking is a key safeguard against multiple voting. If a voter is not checked for ink, he or she could, theoretically, get in line again and try to

[j] ISFED observers collected information on other questions as well (e.g., did the station open on time) but in most of the stations observed things went well and so there is not much variation in the answers to those questions.

cast another ballot thereby inflating a particular party's vote count. I used that indicator variable as the depending variable in the following model:

$$\text{Logit (voters \underline{not} consistently checked for ink)} = \beta0 + \beta1(\text{UNM}) + \beta2(\text{Joint Opposition}) + \beta3(\text{Christian-Democrats}) + \beta4(\text{Labour}) + \beta6(\text{Republicans}) + \beta6(\text{Invalid}) + \varepsilon \quad (11)$$

where β1- β6 indicate the percentage by which the second-digit means for the party vote counts and invalid votes deviate from the expected mean.

The results are displayed in Table 12. One would expect a party's deviation to go up as officials do not check for ink and yet this is not the case. Counter to expectation, none of the parties' deviations increase as inking was not checked consistently and only one party's coefficient is significant. A one percent increase in the deviation of the Christian-Democrats results in a factor of 0.22 as much ink checking. So, the amount of ink checking goes down as the deviations for the Christian-Democrats goes up. This seems strange since officials that purposefully disregard inking are typically trying to allow one voter to cast multiple ballots in an attempt to increase (not decrease) the votes for one party. Interestingly, when I add the controls for age proportions, none of the party's deviations are significant; instead the number invalid votes is statistically significant and would appear to have a strong effect. A one percent increase in the deviation in the second-digit mean of invalid votes increases the odds that a station will not consistently check for ink by a factor of 19.7. This would seem to imply that the more that invalid votes are purposefully manipulated then the likelihood that a station is not consistently checking voters for ink increases significantly. However, when adding the age controls, it does not appear that the vote counts for any of the parties are being manipulated.

Table 12: Digit effects on the likelihood that officials did not consistently check

voters for ink (i.e., potential for multiple voting).

Explanatory Variable	Odds ratio (OR)		Odds ratio (OR)	
UNM	.182		.103	
	(.258)		(.178)	
Joint Opposition	1.983		3.421	
	(2.820)		(5.056)	
Christian-Democrats	.220	**	.353	
	.155		(.302)	
Labour Party	.902		.828	
	(1.198)		(1.267)	
Republicans	.896		.800	
	(.643)		(.610)	
Invalid Votes	**8.696**	*****	**19.729**	******
	(10.458)		**(25.545)**	
Controls:				
Proportion of station Aged 18 – 34			431000000000 (6590000000000)	*
Proportion of station Aged 35 – 49			30329.45 (438039.2)	
Proportion of station Aged 50 – 64			2431.581 (42457.55)	
Proportion of station Aged 65 – 79			2.19e+10 (4.58e+11)	
Proportion of station Aged 95 and older			7.5e-102 (1.03e-99)	*
N	593		593	
Pseudo R-squared	0.0292		0.1014	
Wald Chi-squared statistic	8.78		33.42	**

Note:
Errors are adjusted for clustering at the district level, and are robust for unobserved heteroskedasticity for all regressions.
The variable for the proportion of stations aged 80 to 94 was omitted to avoid collinearity.
** Significant at 95% level of confidence.
* Significant at 90% level of confidence.

<u>*Unauthorized people were present in the station*</u>

According to the election code, only election officials, accredited domestic and international observers, accredited representatives of the party (i.e., party observers), and voters are allowed inside a polling station. Government officials are not supposed to be present in a station unless they are casting their vote. In addition, security officials and police are not authorized to be inside a station unless the Chairman has specifically invited them inside and asked for their help. Unfortunately, there have been cases where local government officials have been inside the polling station as a means of pressuring officials and others, especially during the counting and tabulation process. ISFED observers noted those stations where unauthorized persons were present so as to track whether intimidation or threats may have occurred. Stations where an ISFED observer noted the presence of an unauthorized person were coded as one and stations where observers did not report such a presence were coded as zero. I used that indicator variable as the depending variable in the following model:

$$\text{Logit (Unauthorized person was present)} = \beta 0 + \beta 1(\text{UNM}) + \beta 2(\text{Joint Opposition}) + \beta 3(\text{Christian-Democrats}) + \beta 4(\text{Labour}) + \beta 6(\text{Republicans}) + \beta 6(\text{Invalid}) + \varepsilon \qquad (12)$$

where $\beta 1$- $\beta 6$ indicate the percentage by which the second-digit means for the party vote counts and invalid votes deviate from the expected mean.

The results in Table 13 show that for a one percent increase in the deviations for the Republicans, the odds that a station has an unauthorized person present increase by a factor of 271, making it 271 times more likely than a station where unauthorized persons are absent. When adding age controls, the odds ratio increases substantially. Now, a one percent increase in the deviation for the Republicans increases the odds that a station has an unauthorized person present by a factor of 643. From these results, I might conclude that unauthorized people were pressuring polling station officials to inflate the vote counts for Republicans. And yet, neither domestic nor international observers reported such incidences and the Republicans did not do all that well. Indeed, the overall chi-squared and second-digit mean results led us to believe that votes were likely taken away from the Republicans and that they did not benefit from any manipulations in vote counts. So, I find a tremendous lack of precision in various Benford's Law second-digit tests, and the results contradict one another.

Table 13: Digit effects on the likelihood that an unauthorized person in present in the station (i.e., potential for intimidation).

Explanatory Variable	Odds ratio (OR)	Odds ratio (OR)
UNM	.013 (.041)	.004 (.017)
Joint Opposition	5.257 (8.504)	24.289 (57.213)
Christian-Democrats	2.066 (3.264)	1.151 (2.284)
Labour Party	4.451 (6.748)	2.493 (4.181)
Republicans	**271.330** ** **(481.674)**	**643.322** ** **(1225.248)**
Invalid Votes	18.522 (33.653)	58.554 * (136.207)
Controls:		
Proportion of station Aged 18 – 34		78.429 (2368.877)
Proportion of station Aged 35 – 49		59033.15 (1746285)
Proportion of station Aged 50 – 64		23636.48 (697525.8)
Proportion of station Aged 65 – 79		.000 (.005)
Proportion of station Aged 95 and older		2.18e+37 2.94e+39
N	597	597
Pseudo R-squared	0.0908	0.1519

Wald Chi-squared statistic	25.16**	34.16**

Note:
Errors are adjusted for clustering at the district level, and are robust for unobserved heteroskedasticity for all regressions.
The variable for the proportion of stations aged 80 to 94 was omitted to avoid collinearity.
** Significant at 95% level of confidence.
* Significant at 90% level of confidence.

The other domestic observer variables: serious irregularities observed, complaint filed, and violation of secrecy of the vote, were also used as dependent variables. However, when using the digit deviations as independent variables, the models were not statistically significant. Thus, it would seem that the digit deviations do not pick up problems related to complaints and violations in the secrecy of the vote. Perhaps the reason the digit deviations do not explain the variation in stations where complaints were filed is due to the fact that a complaint can be filed for many different irregularities and thus it may be too general to use. As for secrecy of the vote, violating that core principle of the Election Day process would appear to be a good indicator of intimidation and coercion. Unfortunately, the digit deviations do not appear to explain this type of coercion.

When looked at as a whole, the results appear to be rather mixed. The more simplistic the test such as the chi-squared test and the overall second-digit mean test, the more they over-predict fraud. In the more sophisticated regressions, the digit deviations do a better job of predicting yet they still appear to over-predict fraud and are inconsistent in their ability to identify which party is benefitting or suffering from the fraud. In the next section, I comment more expansively on the implications of my results.

VI. Conclusion and Policy Implications

I set out to examine whether Benford's Law, as applied to the distribution of the second-digits in party's votes, can accurately predict fraud. To accomplish this, I gathered data from the 2008 Georgian Parliamentary Elections and used these data to run a serious of increasingly sophisticated statistical tests. I then compared the results to independent observational data to see how the tests performed. In summary, the second-digit tests do not perform well, and are neither accurate nor consistent.

If the application of Benford's Law to second digit distributions were a good test, I would expect to see a small amount of fraud predicted. I would also clearly see which party benefits from the fraud and which party or parties are hurt by it. Instead, both the overall tests and the regressions suggest a large amount of fraud in what observers deemed a reasonably clean election. Although the results from the regression appear to yield more precision in predicting irregularities, they are completely inconsistent with regard to which parties benefit and which parties suffer. The results from the regressions may appear to be more precise than the overall tests but they are, unfortunately, more confusing and inconsistent.

It is possible that the precision might increase with the addition of further control variables. Where possible, one could try to assemble more data in advance so that such controls could be used. However, finding data at the necessary level can prove very difficult in many developing countries and directly refutes the claims of some proponents of Benford digit tests who argue that all that is needed are the vote counts themselves. Here, the addition of control variables helped enormously in accounting for variation in the dependent variables and often much more so than the deviations. This leads me to believe that the second-digit tests and regressions cannot *by themselves* do a very accurate job of predicting irregularities. As a practical matter, unless the precision can be greatly increased, an authoritarian but statistically savvy ruler can put forth as a defense that the test over-predicts fraud and thus the results of such tests are showing a false positives and do not prove fraud occurred.

Even if the results had been consistent and more precise, they would still be at best a blunt tool. Under the best circumstances, the tests might have alerted us to problems but would not able to provide a detailed diagnosis of what went wrong. As Bolton and Hand (2002) have stated in reference to statistical tools such as these, "the analysis should be regarded as alerting us to the fact that an observation is anomalous, or *more likely* to be fraudulent than others, so that it can be investigated in more detail." Second-digit Benford tests do not identify the exact stations where malfeasance occurred. Rather it identifies a *group* of polling stations where some subset of those stations might be problematic and so it does not provide enough practical information for international or domestic organizations to lodge complaints nor easily pursue the matter through the legal process. Even in the best circumstances, where it may help identify a group of stations that need further investigation, the results offer no details as to exactly what type of fraud may have taken place. Unfortunately, such tests produce little guidance on how to improve the process in the future. Indeed, even at their best, second-digit tests cannot reveal what processes went well. One might suggest that second-digit tests should not be used as a primary tool for detecting irregularities but rather a complementary tool. However, given the inconsistency of the results and the tendency to predict fraud where none exists, I remain skeptical that they are useful, even as a complementary tool.

VII. Appendix 1

Table A1: Descriptive Statistics for dependent variables and controls variables not noted in text.

Variable	mean*	standard deviation	min	max	n
Voters Participating	513.739	266.912	6.000	2195.000	3602
Votes recorded in the last three hours	102.068	81.831	0.000	750.000	3495
High rates of participation in the last three hours (i.e. 200 or more votes for an average of more than one voter per minute): Indicator Variable	290	n/a	0	1	3495
Observer noted Serious irregularities: Indicator Variable	16	n/a	0	1	694
Proportion of station Aged 18 – 34	0.312	0.061	0.020	0.698	3556
Proportion of station Aged 35 – 49	0.275	0.037	0.000	0.480	3556
Proportion of station Aged 50 – 64	0.209	0.037	0.039	0.356	3556
Proportion of station Aged 65 – 79	0.160	0.053	0.000	0.543	3556
Proportion of station Aged 95 and older	0.001	0.002	0.000	0.038	3556
Station counted another station: Indicator Variable	74	n/a	0	1	3604

Note:
* For indicator variables the count of the number of observations that equal 1 are entered into the mean column.

VIII. Bibliography

Beber, Bernd, and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* forthcoming summer 2012 issue.

Benford, Frank. 1938. "The Law of Anomalous Numbers." *American Philosophical Society* 78(4):551–572.

Bolton, Richard J. and Hand, David J. 2002. "Statistical Fraud Detection: A Review." *Statistical Science* 17(3): 235-249.

Deckert, Joseph, Mikhail Myagkov, and Peter C. Ordeshook. 2011. "Benford's Law and the detection of election fraud." *Political Analysis* 19(3):245–68.

Grendar, M., Judge, G., & Schechter, L. 2007. "An empirical non-parametric likelihood family of data-based Benford-like distributions." *Physica A* 380:429-438.

Hill, Theodore P. 2011. "A Statistical Derivation of the Law." *Statistics* 10(4): 354-363.

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

International Society for Fair Elections and Democracy. 2008. *ISFED Report on Monitoring the May 21, 2008 Parliamentary Elections.*

Mebane, Walter. 2006. "Election forensics: Vote counts and Benford's Law." Paper presented at the 2006 Summer Meeting of the Political Methodology Society, University of California, Davis, July 20–22.

———. 2007. "Election forensics: Statistics, recounts and fraud." Paper presented at the 2007 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 12–16.

———. 2008. “Election forensics: Outlier and Digit Tests in America and Russia.” Paper prepared for presentation at the American Electoral Process Conference, Princeton, NJ, May 1-3, 2008.

———.2010a. “Election Fraud or Strategic Voting?” Paper prepared for presentation at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 22–25, 2010.

———.2010b. “Election Fraud or Strategic Voting? Can second-digit tests tell the difference?” Paper prepared for the 2010 Summer Meeting of the Political Methodology Society, University of Iowa, Iowa City, IA, July 22–24.

———.2010c. “Fraud in the 2009 presidential election in Iran?” *Chance* 23:6–15.

———.2011. “Comment on ‘Benford’s Law and the Detection of Election Fraud.’” *Political Analysis* 19(3): 269-272.

Mebane, Walter R., Jr. and Kirill Kalinin. 2009a. “Comparative Election Fraud Detection.” Paper presented at the 2009 Annual Meeting of the American Political Science Association, Toronto, Canada, Sept 3–6, 2009.

———. 2010. “Electoral Fraud in Russia: Vote Counts Analysis using Second-digit Mean Tests.” Paper prepared for presentation at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 22–25, 2010.

National Democratic Institute for International Affairs. Date Unknown. *NDI and the Parallel Vote Tabulation: A Factsheet.*

Nigrini Mark J. 1999. “I’ve got your number: How a mathematical phenomenon can help CPAs uncover fraud and other irregularities.” *Journal of Accountancy* 187(5).

OSCE Office for Democratic Institutions and Human Rights (OSCE/ODIHR). 2008. *Georgia Parliamentary Elections: 21 May OSCE/ODIHR Election Observation Mission Final Report.*

Parliamentary Assembly Council of Europe. 2008. *Honouring of obligations and commitments by Georgia: Information note by the co-rapporteurs on their fact-finding visit to Tbilisi, April 2008.*

Sandron, Frédéric and Sarah R . Hayford. 2002. "Do Populations Conform to the Law of Anomalous Numbers?" *Population* 57 (4): 755-761.

Tödter, Karl-Heinz. 2009. "Benford's Law as an Indicator of Fraud in Economics." *German Economic Review* 10 (3): 339-351.