The Irrelevance of Benford's Law for Detecting Fraud in Elections

Joseph Deckert, Mikhail Myagkov and Peter C. Ordeshook *University of Oregon and California Institute of Technology*

Abstract

With increasing frequency websites appear to argue that the application of Benford's Law – a prediction as to the observed frequency of numbers in the first and second digits of official election returns -- establishes fraud in this or that election. However, looking at data from Ohio, Massachusetts and Ukraine, as well as data artificially generated by a series of simulations, we argue here that Benford's Law is essentially useless as a forensic indicator of fraud. Deviations from either the first or second digit version of that law can arise regardless of whether an election is free and fair. In fact, fraud can move data in the direction of satisfying that law and thereby occasion wholly erroneous conclusions.

The Irrelevance of Benford's Law for Detecting Fraud in Elections

Joseph Deckert, Mikhail Myagkov and Peter C. Ordeshook *University of Oregon and California Institute of Technology*

1. Introduction

Judging by the blogs that clog the internet, accusations of fraud and electoral skullduggery are now an ever-present component of democratic process. Regardless of whether voting occurs in Florida or Tehran, Ohio or Kyiv and Illinois or Moscow, the winners rejoice while the losers claim foul. And with even the most corrupt and autocratic regimes seeking the mantra of democratic legitimacy, it seems as if election observer has become a full time job. The difficulty, though, with achieving any conclusive assessment of an election on the basis of direct observation is that regimes can, as Russia's did in 2008, erect formidable administrative barriers that render any objective and viable oversight an impossibility; or, as occurred in Ukraine in 2004, both sides of a conflict can field their own cadre of observers asserting or denying fraud while the rest of us are left to debate who to believe. Indeed, with observers subject to the accusation that they operate with political agendas beyond encouraging free and fair elections, we should not be surprised to see the emergence of a market for statistical tools, applied to official election returns, that can add to the opinions and conclusions offered by first-hand observation and the journalistic account of things. Unfortunately, rather than confront the necessity for approaching the assessment of an election in the same way we approach a criminal investigation - with the careful collection and evaluation of evidence and the formulation of a theory of the crime that is sustained or disconfirmed by a careful sorting of the evidence -- that market also gives rise to the search for the "magic black box" into which we plug official election returns and out of which emerges a definitive assessment as to whether the election was or was not rigged.

The search for objective methods for detecting election fraud using the data officially provided by a state's election commission began, perhaps, with the late Dr. Alexandar Sobyanin's (1993) analysis of Russia's 1993 parliamentary elections and constitutional referendum. But while several of the indicators of fraud that he proposed

have subsequently been refined and extended as valuable tools, one indicator in particular revealed the dangers of this enterprise. Here Sobyanin took note of the fact that an especially simple empirical relationship pertains to an interestingly wide range of natural phenomena. Specifically, suppose we take the variable X, where X might measure the population of cities, the sales of corporations or even the population of different species of insects, suppose $X_1 > X_2 > ... > X_n$, and suppose we graph $log(X_i)$ against i. Then for such things as population and so on, the necessarily negatively sloped relationship is closely approximated by a straight line. In what can at best be deemed a naïve isomorphic leap, Sobyanin applied this idea to the votes won by parties in Russia's 1993 proportional representation parliamentary elections and interpreted the deviations from a linear relationship as an indicator of fraud's magnitude.

The problems here should be obvious. If voters are sophisticated and strategic and if parties must meet some threshold of representation, then we can anticipate a non-linear drop in the vote shares of parties that are not expected to win seats. Similarly, if Duverger's law applies in single mandate contests, then an even more pronounced discontinuity in shares will appear among all parties that rank third or lower. Indeed, it is precisely this sort of relationship that Cox (1997) formalizes for a range of electoral systems, which is to say that there is no theoretical reason for supposing that a log-linear relationship should hold in an election, and good theoretical reasons for believing otherwise. Put simply, the log-rank model is irrelevant as an indicator of electoral fraud. Nevertheless, the search for a readily applied indicator of fraud with a genesis in numerology did not end with Sobyanin. And no doubt the quest for an easily applied methodological tool with the attendant statistical and mathematical paraphernalia to suggest scientific rigor is the attraction of recent applications of Benford's Law to elections. However, in this essay we argue that this "law" is no more relevant to detecting electoral fraud than is Sobyanin's adaptation of the log-rank model.

2. Benford's Law

Briefly, Benford's Law states (or, more correctly, observes) that a considerable number of processes or measurements give rise to numbers (e.g., returns on investment, population of cities, street addresses, sales of corporations, heights of buildings) that

establish patterns in the digits that might otherwise seem counter-intuitive wherein lower digits are more common than larger ones. While we might expect digits to be uniformly (randomly) distributed when there is no hidden nefarious hand generating the numbers that contain them, suppose, for the simplest example, that we invest \$100 and that our investment doubles every year. If we now record the value of that investment every month, our first twelve observations will begin with the digit "1", our next seven with the digit "2", our next four with the digit "3" and so on. Thus, a graph of the distribution of first digits will look like a log-normal density. Alternatively, if we collect home street numbers at random from a telephone book, because nearly all addresses will begin with the number "1", but often renumbering will occur because a street crosses a municipal boundary or simply end before numbers beginning with higher digits appear, addresses will more often begin with a "1" than a "2", more often with a "2" than a "3" and so on.

The processes in these examples that give rise to sequences of first digits that approximate Benford's law are self-evident and statistically significant deviations from it can be taken as evidence that someone has 'cooked the books' or employed an unusual algorithm for numbering residences. Not a little effort has been devoted, then, to formalizing that law mathematically and uncovering less obvious and more general mechanism of number generation that would yield conformity to it in other contexts (Hill 1998 and Janvresse and la Rue 2004). Formally, Benford's Law can be expressed thus: The probability that the digit d (d = 0, 1, ..., 9) arises in the n-th ($n \ge 1$) position is

$$\sum_{k=10^{n-2}}^{10^{n-1}-1} \log_{10} \left(1 + \frac{1}{10k+d} \right).$$

Thus, for those processes thought to match Benford's Law, Table 1 gives the predicted frequencies for both the first and second digits (generally referred to as the 1BF and 2BF models):

Table 1: Benford Law frequencies

	0	1	2	3	4	5	6	7	8	9	mean
1 st digit	ı	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046	3.441
2 nd digit	0.120	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.088	0.085	4.187

In the realm of election forensics, it is these predictions that are then tested against official election data in the attempt to ascertain whether fraud has or has not

occurred, with emphasis increasingly placed on the second digit of official tallies (Mebane 2006, 2007, 2008, 2009; see also Pericchi and Torres 2004, Buttorff 2008, Battersby 2009, Roukema 2009 and an array of websites http://www.jgc.org/blog/2009/06/benfords-law-and-iranian-election.html). Lent the aura of scientific legitimacy by a plethora of statistical tests of significance and nonparametric curve fitting, the red light signaling fraud ostensibly turns on in the asserted 'magic black box' if election tallies – candidate precinct vote totals or turnout numbers -fail to match these distributions or the average value of first and second digits depart significantly from the predicted means.

The difficulty here is the absence of any model – any theory – that would lead us to believe that manipulated vote tallies would lead us away from the predictions of Benford's Law. Indeed, we do not even have a theory that tells us that a fraud free election should conform to 1BL or 2BL. To see what we mean by this, consider Berber and Scacco's (2008) argument for the relevance of looking at the last and next to last digits of vote tallies. Briefly, they note first that if there is little chance that the perpetrators of fraud will be prosecuted, it is not unreasonable to be suspicious of precinct or regional tallies that report a proportion of zeros and fives as the last digit in excess of what we expect by chance. The argument is that absent any legal disincentives for committing fraud, precinct and local election officials can meet their 'quotas' while saving effort by simply rounding off the numbers they report without regard to actual ballots cast. Indeed, it is precisely rounding of this sort that one finds in abundance in the turnout numbers of Russia's 2004 and 2008 presidential elections (Buzin and Lubarev 2008). And here we are reminded of the unintentionally humorous remark of Vladimir Shevchuk of Tatarstan's central election commission when commenting on Russia's 2000 presidential election that first elevated Putin to prominence: "there has been fraud of course, but some of it may be due to the inefficient mechanism used to count ballots ... To do it the right way they would have needed more than one night. They were already dead tired so they did it in an expedient way" (Moscow Times, September 9, 2000). However, as a secondary test – which is especially useful if fraud's perpetrators seek to disguise their actions -- Berber and Scacco note that if the those perpetrators enter what they regard as random (but nevertheless, for the candidate they support, inflated) numbers

in official tallies, experimental evidence suggests they will write repeated digits less frequently than what we expect from chance (Camerer 2003). That is, protocol entries ending with the digits "00", "11", "22" and so on should, in a fraud free election, occur a tenth of the time, and it is reasonable to suspect falsified protocols if observed frequencies are significantly less than this fraction (for the application of this test to a suspect election see Levin et al 2009).

There are, then, well-understood behavioral reasons for arguing that an examination of last and next-to-last digits should be one of our forensic tools. The relevance of Benford's Law to electoral processes, in contrast, and its connection to specific forms of fraud remain a mystery. The sole justification offered for its relevance rests on the finding that aggregates of numbers generated from different and uncorrelated random processes will, in the limit, fit Benford's Law (again, see Hill and Janvresse and la Rue 2004), whereupon Mebane (2006) asserts the relevance of Benford's Law with the argument that voting derives from a number of stochastic choices – for whom to vote, whether to vote, errors in voting and so on. Such verbal arguments, though, are indeed a weak reed upon which to rest a test for democratic legitimacy. Surely we can imagine stochastic processes that yield a competitive two candidate race in districts whose magnitude varies between 100 and 1000, in which case the modal 1st digit for each candidate's vote will not be 1 or 2, but rather 4, 5 or 6 (it is precisely this example that leads Brady, 2005, to reject 1BL as a test for fraud and a reason to focus, if necessary, on 2BL). Moreover, an argument such as Mebane's ignores the fact that fraud itself is often implemented in a highly decentralized way by local and regional officials, each using their own schemes and procedures – thereby adding yet another stochastic element to the mix and, arguably, encouraging an even closer fit to 1BL and 2BL.

The fact is, unlike Berber and Scacco's analysis of last and next-to-last digits, there is no corresponding behavioral or theoretical reason for supposing that the first and second digits of fraud free data will look any different from data drawn from an election permeated with instances of falsified votes and protocols. It may be that simulations of data that fit 2BL can be perturbed by falsifications of a specific sort (Mebane 2006), but that is no reason for believing that fraud free data itself will correspond to 2BL or even

that fraud cannot move official data closer in line with Benford's Law. Indeed, as we show shortly with both real and simulated data, that is precisely what can occur.

One might argue in defense of Benford's Law that it should be but one of several forensic tools employed by the analyst. But in addition to noting those innumerable essays and internet blogs taking ONLY deviations from 1BL or 2BL as evidence of fraud, we also see the Law's strongest proponents offering arguments that it is in fact virtually a magic black box: Witness the assertion that "...it does not require that we have covariates to which we may reasonably assume the votes are related across political jurisdictions. The method is based on tests of the distribution of the digits in reported vote counts, so all that is needed are the vote counts themselves" (Mebane 2006:1). Although Mebane qualifies things by stating that analyses employing 2BL might merely be used to flag suspicious data or to augment on the ground observation and vote recounts, this view remains wildly incorrect: Detecting and measuring fraud is much like any criminal investigation and requires a careful gathering of all available data and evidence in conjunction with a "theory of the crime" that takes into account substantive knowledge of the election being considered, including the socio-economic and geographic correlates of voting. To even hint that one can dispense with that data in the development of statistically based forensic indicators is a dangerous deception. Indeed, our response to the view that 1BL or 2BL is but one element of the investigator's tool kit, what we show here is that, without further elaboration of its theoretical relationship to elections and fraud – an elaboration that will, of necessity, require additional data and substantive expertise in any application -- Benford's Law as currently constituted is essentially irrelevant as a forensic indicator. Deviations from it may signal fraud when there is none and it may fail to do so when it is in fact known to exist. Our argument employs both simulations with artificial data as well as the application of the law to a sample of real world data in which fraud is known to have been pervasive as well as to instances of American elections where it is unlikely to have taken place.

3. Ohio 2008

Barak Obama's victory in 2008 did not depend on any one state, but prior to the vote, Ohio was deemed one of the states essential to his victory. Particular attention was paid here, though, not only because of its Electoral College weight, but also because of allegations of electoral impropriety in 2004 that ostensibly aided President Bush's reelection (allegations that were later disavowed by the Democratic party itself; see Democratic National Committee 2005, *Democracy at Risk: The 2004 Election in Ohio*). We now know, of course, that of the 5,698,260 people who voted for president there, Obama's electors won 2,933,388 votes to McCain's 2,674,491, or a plurality of 258,897 votes. Thus, even if we were to join forces with the conspiracy theorists of the world whose blogs report 'irrefutable evidence of fraud' for every election in which they dislike the final outcome, we'd be hard pressed to assert that Obama's minions falsified a quarter of a million votes and stole the state. Nevertheless, this is precisely what the application of Benford's Law to the second digit of precinct level tallies would lead us to suspect.

To see this, Figure 1a reports the distribution of second digits for all precincts in Ohio for McCain after eliminating those precincts in which he wins fewer than 100 votes while Figure 1b does the same for Obama. Figures 2a and 2b repeat this calculation after all precincts in which one candidate or the other wins fewer than 100 votes. Table 2 summarizes the mean value of the second digits portrayed in these figures. There is no reason to obfuscate things now with fancy statistics: While both sets of figures show that, as judged by Benford's Law, McCain's numbers give rise to few if any suspicions (the average 2nd digit equals 4.29 in both cases), Obama's numbers (means of 4.40 or 4.43 respectively) tell a different story. That is, while McCain's numbers closely approximate the 2BL value of 4.187, the second digits of Obama's precinct totals departs significantly from that law and actually form a closer match to a uniform distribution (a perfectly uniform distribution yields the average 4.5).

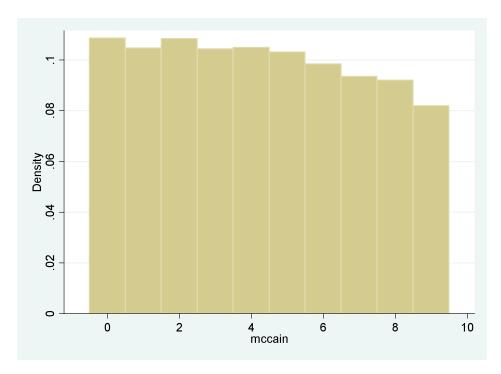


Figure 1a: 2^{nd} digit distribution, McCain, n = 10,221

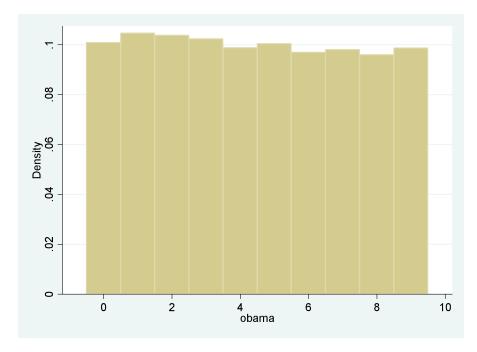


Figure 1b: Figure 1a: 2^{nd} digit distribution, Obama, n = 10,549

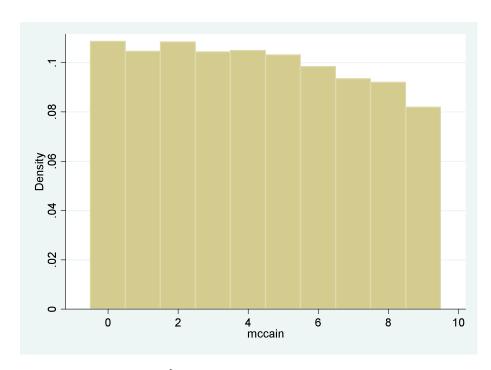


Figure 2a: Figure 1a: 2^{nd} digit distribution, McCain, n = 10,220

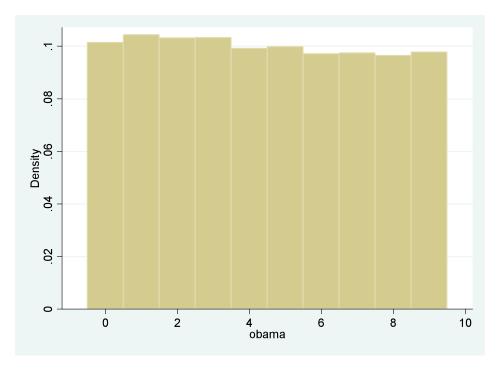
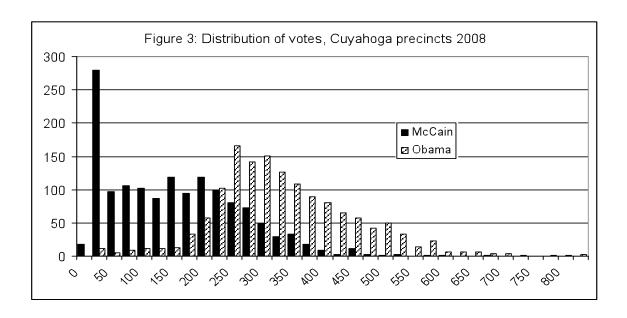


Figure 2a: Figure 1a: 2^{nd} digit distribution, Obama, n = 10,220

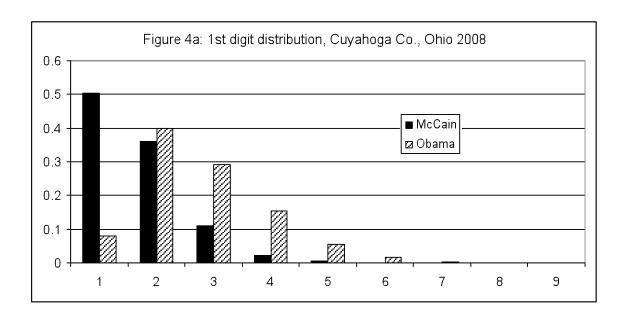
Table 2: Mean values of second digits, Ohio 2008

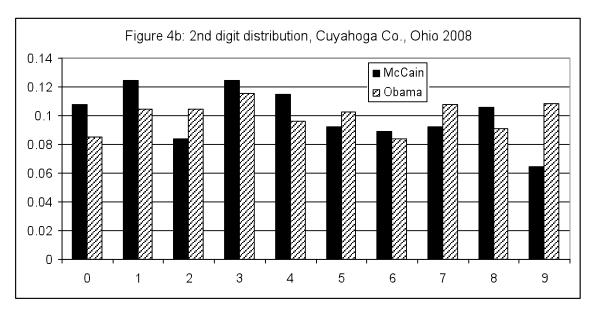
	n	mean	n	mean	
McCain	10,221	4.289	10,220	4.289	
Obama	10,549	4.439	10,220	4.432	

Rather than point a finger of suspicion at Obama because the warning red light of a magic black box is glowing brightly, there is a logical explanation for the preceding results that has nothing to do with fraud. The problems with Benford's Law here are perhaps more readily seen if we focus on data from heavily Democratic Cuyahoga County (Cleveland), which Obama carried 458,422 votes to 199,880. Thus, if one were inclined to conspiracies, we might begin here after noting that Cuyahoga provided nearly all of Obama's statewide plurality. However, note that a graph of the distributions of McCain and Obama's vote by precinct (see Figure 3, with increments of 25 votes/unit) reveals that Obama's distribution is nearly normal whereas McCain's is skewed heavily to the left. Even if we eliminate all precincts in which one candidate or the other won fewer than 100 votes, McCain's distribution remains skewed whereas Obama's continues to approximate a normal density. It is well known now that Benford's Law does not apply to the digits of numbers drawn from a normal density but instead models skewed densities that, like McCain's, approximate the log-normal.

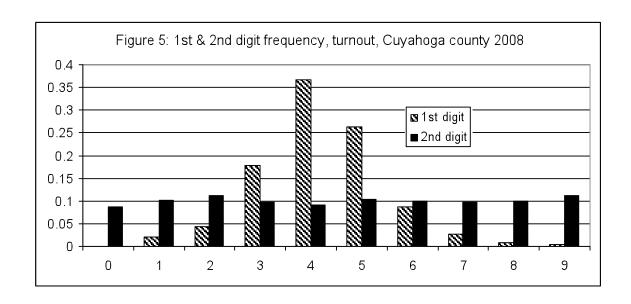


Figures 4a and 4b, now show what happens if we attempt to apply Benford's 1st and 2nd digit Law to the numbers summarized by Figure 3 (again, after we eliminate consideration of all precincts that award a candidate fewer than 100 votes). First, as Figure 4a shows, although the first digit of McCain's numbers approximate a long normal density, the mean value of digits is but 1.66 – far less that a value of 3.441 as dictated by 1BL. Obama's numbers in contrast are not long normal, but the mean value of digits here, 2.76, is actually a closer fit to 1BL. Of course, the fact that the general form of the distribution of 1st digits fits 1BL for McCain but not for Obama should come as no surprise for anyone looking at Figure 3. More interesting, then, is a test that focuses on 2BL. If we merely look at the means of second digits, it would seem that again McCain but not Obama lies above suspicion: McCain's mean is 4.19 but Obama's is 4.51. On the other hand, if we look at the actual distributions (Figure 4b) we see no relationship whatsoever to 2BL for either candidate. Both distributions are nearly uniform, or, to state matters differently, despite the fact that we begin with two radically different distributions in terms of absolute votes as well as in terms of the relative frequency of first digits, the distributions of second digits are, for all purposes, identical in form and wholly at odds with 2BL.



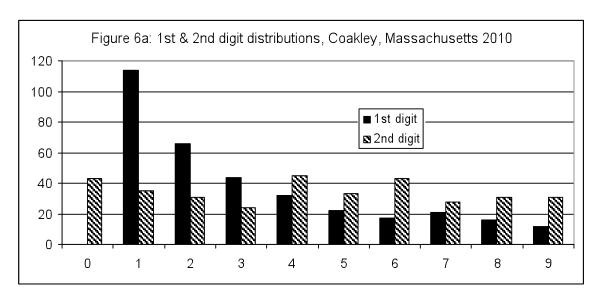


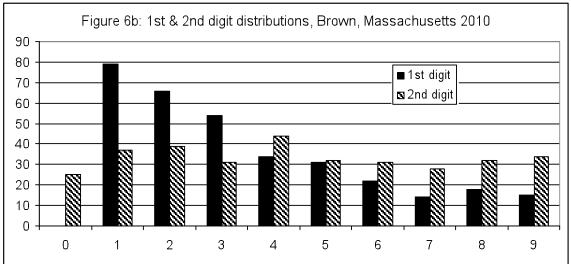
One might argue, of course, that we are looking at the wrong numbers. Since the ostensible "indicator" under consideration offers no "theory of the crime," its application leaves unspecified what sequence of numbers ought to be analyzed. The alternative to candidate vote counts, then, is turnout, since it too is subject to manipulation in a fraudulent election. Continuing, then, with our focus on Cuyahoga county's 1400+ precincts, Figure 5 gives the distribution of first and second digits among all precincts with turnout in excess of 100 votes. And as this figure reveals, evidence consisted with either 1BL or 2BL is nowhere to be found: The distribution of first digits corresponds as closely to a normal density as we are likely to find in real world data, whereas the distribution of second digits is virtually uniform (indeed, if we round our calculations to two decimal points, six of the ten digits occur with precisely the frequency of one tenth!). Thus, if we take the view that "normal" (i.e., uncontaminated) data matches Benford's 1BL or 2BL Law, then regardless of what data sequence we consider, we must conclude either that the Law, at least in northeast Ohio, is irrelevant to any assessment of the election's legitimacy or that massive fraud (presumably benefitting Obama) was pervasive there



4. Massachusetts 2010

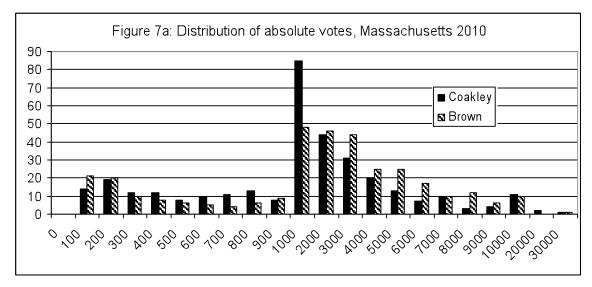
There was perhaps no more closely watched and hard fought election in the United States in the first half of 2010 than the Massachusetts special election to fill the Senate seat held by Edward Kennedy. Indeed, given the media attention paid to this contest, along with the nearly constant flow of pre-election polls, one can be certain that if significant fraud had occurred, it would not have gone unnoticed and unreported. Nevertheless, the application of Benford's Law suggests otherwise. Figures 6a and 6b give the distribution of first and second digits by township (after eliminating those towns with fewer than 100 votes for one candidate or the other) of both Coakley and Senator Brown's vote totals. What we see here, then, is that with a somewhat higher level of aggregation than we consider in Ohio, the distribution of first digits (1BL) would seem to conform closely to Benford's Law (the mean digit value for Coakley is 3.20 and for Brown 3.52). An analysis of second digits (2BL), on the other hand, would lead us to conclude that fraud permeated the votes of both candidates (the mean values for Coakley and Brown are, respectively, 4.35 and 4.47). Not only do the second digit distributions not correspond to a weak log-normal, they depart even from a uniform density and exhibit virtually no apparent pattern whatsoever.

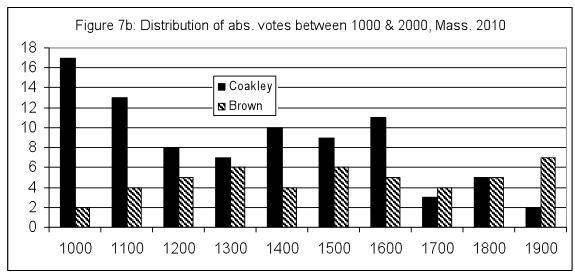




The mystery, if we can call it that, as to why the first but not the second digits of the data in Massachusetts conform to Benford's Law is revealed in part in Figures 7a and 7b. Figure 7a gives the distribution of absolute votes won by both candidates and reveals a heavy concentration of townships in which both Coakley and Brown won between 1000 and 2000 votes (48 in Brown's case, 85 in Coakley's). The first digits of the numbers for these townships, then, contribute heavily toward the "1's" that appear in Figures 6a and 6b. But what we see here is that the first digit's distribution approximates Benford's Law for the simple reason that there are few townships capable of delivering more than 4000 votes to any one candidate. If we look now at the second digits in the distribution of vote counts within this 1000 to 2000 interval (Figure 7b), we find no pattern whatsoever. While there is a heavy concentration of second digits around "0" and "1" for Coakley,

there is a concentration at "4", "5" and "6" as well. And for Brown, the least frequently observed 2nd digit is, in fact, "0". The application of 2BL, then, fails largely because the distribution among modal townships is, as we might expect, essentially random.



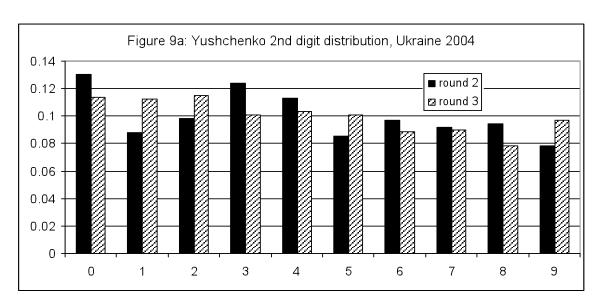


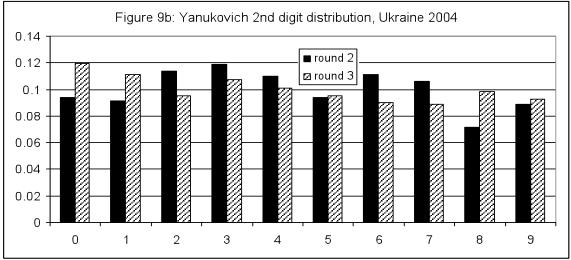
5. Ukraine 2004 and 2007

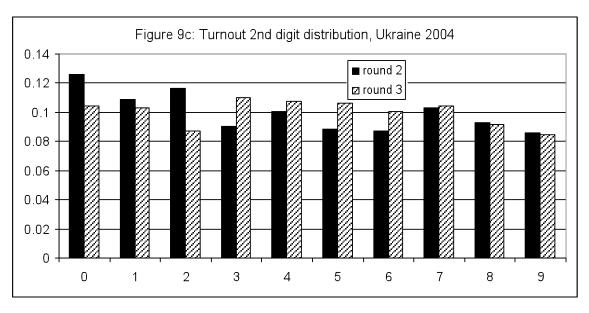
For reasons we detail elsewhere, there is no better place to evaluate an indicator of fraud than Ukraine's 2004 presidential election (Myagkov et al 2009). There is virtually universal agreement that the second, November, runoff round of that contest was replete with instances of fraud in which one candidate, Viktor Yanukovich, benefitted in the amount of between 1.5 and 3 million falsified or fraudulently cast votes (or between 5 to 10 percent of the official vote count). On the other hand, there is also universal agreement

that the court ordered rerun of the runoff -- a 3rd round of voting – was relatively free of the manipulations and artificially inflated turnout that characterized the earlier vote. Thus, if Benford's Law has any validity as an indicator of fraud, it should perform well when applied to data from these two rounds of voting. In fact, we once again find otherwise. Focusing, per Mebane's (2008) suggestion, on the second digit of returns from each of Ukraine's 755 rayons (counties), we find a mean value of 4.21 for both the first and second rounds in Viktor Yushchenko's votes. The means for Yanukovich are 4.37 in the first runoff and 4.28 in the second. Thus, one might be tempted to conclude that this difference of 0.09 signals a less fraudulent second runoff. However, we should also keep in mind that much of the fraud in the first runoff took the form of artificially inflated turnout wherein no less than 1.5 million non-existent voters were mysteriously reported by the central election commission to have cast ballots. A second digit analysis of official turnout figures, though, would seem to contest this assertion: The mean second digit in the first runoff is 4.22, and increases to 4.39 in the second runoff.

As close as all of these numbers might seem to the 4.178 value dictated by 2BL, a graph of second digit frequencies reveals very little in the way of meaningful patterns. Figure 9a graphs Yushchenko's second and third round 2nd digit distributions, Figure 9b does the same for Yanukovich, and Figure 9c does so for overall turnout across all rayons. While it is true that "0" is the most frequently observed second digit for Yushchenk in the second round (and nearly so in the third), the next two most frequently observed digits are "4" and "5". Yanukovich's 3rd round pattern seems to fit 2BL better, since there the numbers "0" and "1" are most frequently observed. But for tunrount, it is the 2nd voting round and not the 3rd that seems to fit 2BL best wherein the numbers "0", "1" and "2" are most coming in the 2nd round, but "4", "5" and "6" capturing that honor in the 3rd. In fact, detecting meaningful patterns in any of these distributions seems akin to seeing cats, dogs and cows in clouds.







Ukraine's 2007 parliamentary election presents an interesting side note to this discussion. While generally receiving high marks from international observes for meeting the standards of free and fair, one region nevertheless fell under suspicion – Donetsk, which is Yanukovich's home region and the center of support of his party, Regions. By way of background and to see the incentives for fraud, we note that following the 2006 parliamentary vote, Yanukovich became Prime Minister by forming a coalitional government with the Communists and Oskar Moroz's Socialists. Sustaining that coalition in 2007, with both Regions and the Communists guaranteed to pass the 3% threshold for representation, hinged on whether the Socialists would also exceeded that threshold. The difficulty here, though, is that the Moroz's party won their seats in 2006 as part of the Orange Coalition of then President Yushchenko. Thus, it was clear that the Socialists, viewed in Western Ukraine as turncoats, would lose their support in the West, leaving Yanukovich's coalition dependent on the Socialists' ability to make up the difference in the East. However, as the votes were counted it became increasingly evident on election day that the Socialists would fall short of the threshold, at which point a number of precincts in Donetsk held their returns until late in the day. And it was those precincts that reported an uncommonly high level of support for Moroz's party - in many instances, exceeding Regions' vote by factors of 5 and 10.

Subsequent analysis of the returns leave little doubt that, in the attempt to push the Socialists past the threshold, something in the vicinity of 100,000 votes had been fraudulently transferred in Donetsk from Regions to the Socialists (Myagkov et al 2009). The logic of this transfer was that while those votes might cost Regions a seat or two, the coalition would gain the fifteen seats that Moroz's party would secure if it inched past 3% nark. This effort failed, largely because Yanukovich's minions realized too late that the Socialists were failing in the East to offset their losses in the West. Interestingly, though, little was made of this fraud following the election since, naturally, Yanukovich was hardly going to admit to it whereas Yushchenko's coalition actually gained a net of two or three seats – enough for Yushchenko to form a new governing coalition that deposed Yanukovich as Prime Minister in favor of Yulia Tymoshenko.

For our purposes we note that Yanukovich's Regions won upwards of 1,718,600 votes in Donetsk while, officially at least, the Socialists garnered a bit more than 190,000.

Thus, the 100,000 fraudulently transferred votes barely made a dent in the numbers for Regions, but better than doubled the vote for the Socialists. The question, then, is whether Benford's Law detects this fraud if we focus our analysis on Donetsk. And more to the point, while we might expect the returns for Regions to continue to match the predictions of 2BL since fraud impacted its numbers only incrementally, we can hardly expect the same of the Socialists. Interestingly, though, we find precisely the opposite. After again eliminating those precincts in which a party wins fewer than 100 votes, the mean second digit for Regions is 3.66 whereas for the Socialists it is 4.08. One might conjecture that the low mean for Regions derives from some peculiarities in the size of precincts. But that is not the case. If we analyze separately those 1841 voting districts in which regions vote lies between 100 and 1000 votes, the mean second digit shoots up to 4.43 whereas for the remaining 530 districts, its 2nd digit score drops to 1.00 (a comparable reanalysis for the Socialists would change nothing since it secured 1000 or more votes in only eleven of the 353 districts in which it won more than 100 votes). Thus, as before, Benford's 2BL tells us nothing about voting and fraud.

6. Simulations

Clearly, Ukraine provides scant support for supposing that 2BL is a useful indicator of fraud. Nevertheless, a puzzle remains with the data we consider from the United States. Because the vote distribution for Brown and Coakley in Massachusetts can (imperfectly) be approximated by a normal density, it is perhaps not surprising that 2BL performs poorly there. McCain's vote distribution in Ohio's Cuyahoga County, on the other hand, is decidedly skewed to the left and well approximated by a long normal density. Yet even there a test employing 2BL might lead us to infer the existence of fraud. Indeed, these two empirical examples would seem to suggest a complete disconnect between Benford's Law applied to second digits and the underlying distribution generating those digits. Of course, one might seek to defend the relevance of Benford's Law against these examples by noting that they are simply that, examples – and that they merely reveal the care that must be taken when applying any ostensible forensic indicator. Our argument, however, is different than that. It is that the Law is an indicator that can consistently give the wrong signals as to the legitimacy of an election's outcome.

As further support of that argument we turn to simulations of fraud free electoral data in which voters and candidates occupy a position in a two-dimensional space, where voters, if they vote, do so for the candidate closest to their position. The positions of the candidates, in turn, are exogenously set by us so as to induce simulated elections of different degrees of competitiveness. Briefly, now, for each voter i, we let:

$$X_i = \beta_X V_i + u_{xi}$$
$$Y_i = \beta_Y V_i + u_{Yi}$$
$$T_i = \beta_T V_i + u_{Ti}$$

where $Vi \sim N(g, \sigma_V^2)$ and $g \sim N(G, \sigma_g^2)$ and where X_i is the voter's X position, Y_i is the voter's Y position, T_i is the voter's probability of voting, V_i is the value of a normally distributed random variable, β is a coefficient indicating the effect of V_i on the voter's position or turnout decision, and u is a noise term that is normally distributed. Thus, as with Mebane's (2006) justification for the relevance of Benford's Law, our simulated election data is generated by three distinct and uncorrelated stochastic processes – the voter's preferences on each of two dimensions plus the decision to vote. The parameters of interest now are precinct size and margin of victory, and their impact on adherence to the 2BL model. These two parameters were varied across eighteen sets of simulations wherein precinct size is set at 1,000, 5,000, or 10,000 voters and the winner's percentage is allowed to vary between 52% to 77%. Each set of simulations contains 1,000 precincts. Keeping in mind now that the mean value of second digits given 2BL is 4.187 (Table 1), the 2-candidate results of our simulations are given in Table 3:

Table 3: Means of 2nd digit

# voters/precinct:		1,000	5,000	10,000
	52%	4.341	4.132	4.286
	58	4.478	4.071	4.417
Candidate 1's	63	4.359	3.976	4.561
% of the vote	66	4.410	4.336	4.514
	67	4.458	4.325	4.380
	77	4.505	4.317	4.427
	48	4.434	3.879	4.183
	42	4.432	4.047	4.264
Candidate 2's	37	4.370	3.926	4.216
% of the vote	34	4.023	4.146	4.125
	33	4.273	4.067	4.114
	23	4.120	4.050	4.250

Table 3 highlights several problems with applying Benford's Law to elections. First, notice that strict adherence to 2BL appears in few of the cases. Although our simulations did not include fraud of any type, we find a wide range of values for the mean value of second digits. Indeed, of the 33 means greater than 4.0, fourteen are closer to the expected mean for a uniform distribution of second digits (4.5) than they are to what 2BL requires (4.187). Second, thirteen means fall below 4.187 and three are in fact less than 4.0. This suggests that their underlying distributions are even steeper than predicted by 2BL. Similar examples can be found empirically. In the 2008 election, for example, Barack Obama had a second digit mean of 4.04 in Minnesota. Surely fraud did not characterize voting or vote counting there, but it is surely true that we could either illegitimately add or subtract votes from Obama's counts so as to increase that mean in the direction of 4.187. In other words, if a candidate in the absence of fraud can have such a low mean to begin with, we can imagine circumstances in which fraud makes a candidate's vote counts comply with 2BL, leading to the inference that the election was free and fair. Finally, the size of precincts appears to have an effect on conformity to 2BL. With but one exception (when the 2nd ranked candidate's share equals 34%), both candidates show lower means for the second digits with 5,000-voter precincts than when precincts are smaller or larger. This indicates that the extent to which lower digits are more common is sensitive to the number of voters, but so as to belie the possibility that Banford's Law can provide any magical black box, not predictably so.

7. Conclusions

Despite our earlier critique of Sobyanin's application of the log-rank model to elections, we should understand that it is not impossible to formulate a theory that explains or predicts such a relationship in specific contexts. That is precisely what Ijiri and Simon (1977) provide with respect to firm size. But a model in one context need not apply to any other, and asserting otherwise is pseudo-science. We know, for instance, some of the conditions that, if satisfied, yield numbers in accordance with Benford's Law, the most general being a process in which numbers are generated by a variety of uncorrelated densities (Hill 1998, Janvresse and la Rue 2004). However, just as there is no basis for asserting a priori that the Ijiri-Simon model of firm growth and mergers applies to parties, candidates or anything else, there is no reason to suppose a priori that the conditions sufficient to occasion digits matching 1BL or 2BL hold any meaning for elections. Indeed, if Ijiri and Simon's model provides any guidance -- and we should take note of the fact that because they address a specific process with specific assumptions, their model does not predict a strictly linear relationship between log and rank, but rather a slightly convex one, thereby revealing the true complexity of things -- it, in combination with what we know about the stochastic processes required to occasion digits in conformity with Benford's Law, is that there cannot be any simple universally applicable magic box into which we plug election statistics and out of which comes an assessment of an election's legitimacy. This is not to say that models cannot be formulated to establish the relevance of 2BL or 1BL in specialized electoral contexts. But our analysis - especially its empirical components - suggests that the data required to validate that incidental relevance must be richer than anything any electoral commission is likely to provide.

We can illustrate what we mean here with a second indicator proposed by Sobyanin that, unlike his log-rank model, has been refined to become a valuable forensic tool. Briefly, Sobyanin noted that, absent fraud in the form of artificially inflated turnout, the relationship in otherwise homogeneous data between turnout, T, and a candidate's share of the eligible electorate, V/E, should approximately match the candidate's overall share of the vote. That is, if we estimate the relationship $V/E = \alpha + \beta T$, then absent fraud, α should approximately equal 0 and β should approximately equal the candidate's percentage of the overall vote. If, though, a candidate's vote share is padded by adding ballots in otherwise low turnout districts, estimates of α will turn negative and estimates of β increase until, in extreme cases, it exceeds 1.0 (e.g., upwards of 1.6 in places like Russia's Tatarstan). The operant clause here, though, is "in otherwise homogeneous data" since this indicator is intended to detect the heterogeneity introduced by a specific form of fraud. Any preexisting heterogeneity can only cloud the issue and distort our conclusions. It is imperative, then, that we control for those other parameters that might intervene in the relationship between turnout and candidate preference. And knowing what these things might be - things such as regional loyalties, ethnic identities or differences in voting patterns between urban and rural voters – is where substantive expertise enters and belies a blind application of this indicator. It is precisely these things, moreover, that we are unlikely to find in official election returns but that nevertheless must be taken account of before ANY forensic tool can be usefully applied.

To illustrate matters further, we note that if, after separating Republican from Democratic precincts (in recognition of the fact that these subsets are on average demographically distinct), we apply this indicator to elections drawn from Cuyahoga (Cleveland) or Franklin (Columbus) counties Ohio, nothing hinting at fraud appears. But if we do the same for Hamilton County (Cincinnati), it would seem that significant fraud has occurred in every presidential election beginning with 1992 (Myagkov et al 2009, see esp. pp.257-65). It is unreasonable to suppose, though, that fraud of the magnitude suggested by a blind application of this indicator has gone unnoticed by one party or the other thru five presidential elections. But there exists a straightforward explanation; namely, the existence of a significant number of Republican precincts in Cincinnati itself that are demographically distinct from those in the suburbs – districts that vote Republican by slim margins and with considerably lower rates of turnout than other districts. Thus, merely separating Republican from Democratic precincts, which works

well in urban areas with few Republican districts, is an imperfect control for demographic heterogeneity in Hamilton County and results in a false signal of fraud (for a nearly equivalent example of the additional data required to apply forensic indicators in Taiwan owing to the presence of military districts and districts with heavy concentrations of "aboriginal populations" see Chaing and Ordeshook 2009).

To assert that Benford's Law or any other magical indicator can somehow avoid these methodological complexities is unwarranted. Any valid analysis, regardless of the forensic indicator proposed, must include a model that either specifies the generalized impact of parameters such as district magnitude, competitiveness and regional clustering of support or it must give us sound theoretical arguments for supposing that those parameters will not impact conclusions. It must tell us how to test for the indicator's applicability and, in the process it must tell us what data we need in order to calculate what to expect in the absence of fraud along with how specific forms of fraud impact the indicator. Thus, even if we were inclined to believe that Benford's Law bore any relation whatsoever to elections and election data, we would argue that the variables likely to intervene between, say, the two parameters considered in Table 3, competitiveness and district magnitude (and keep in mind that it is common for rural election districts to be of a different size and political persuasion than urban ones), should be measured and the predictions about digit distributions adjusted in a theoretically proscribed way. This, of course, immediately negates the possibility of any variant of the Law operating as its proponents assert it does – providing the analyst with a quick and easy method in which "all that is needed are the vote counts themselves ..." But our analysis says more. As Table 3 shows and as our empirical examples illustrate, the distribution of digits in even a fraud free context is unrelated to 2BL. Thus, if our data and simulations imply anything, it is that Benford's Law is wholly irrelevant to assessing an election's conformity with good democratic practice and that effort should be directed elsewhere in the search for forensic indicators.

References:

- Alvarez, R. Michael, Thad E. Hall and Susan D. Hyde. 2008. *Election Fraud: Detecting and Deterring Electoral Manipulation*. Washington D.C.: Brookings
- Battersby, Stephen. 2009. "Statistics hint at fraud in Iranian election." *New Scientist* 24 June
- Berber, Bernd and Alexandra Scacco. 2008. "What the Numbers Say: A Digit Based Test for Election Fraud Using New Data from Nigeria," paper presented at the Annual Meeting of the American Political Science Association, Boston Ma., August 28-31.
- Brady, Henry E. 2005. "Comments on Benford's Law and the Venezuelan Election." Unpublished manuscript, Stanford University, January 19, 2005.
- Buttorf, Gail. 2008. "Detecting Fraud in America's Gilded Age," unpublished manuscript, University of Iowa.
- Buzin, Andrei and Arkadii Lubarev. 2008. "Crime Without Punishment". Moscow: Nikkolo M.
- Camerer, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.
- Chaing, Lichun and Peter C. Ordeshook. 2009. "Fraud, Elections and the American Gene in Taiwan's Democracy," unpublished manuscript, California Institute of Technology
- Cox, Gary. 1997. Making Votes Count: Strategic Coordination in the World's Electoral Systems, Cambridge: Cambridge University Press.
- Hill, T.P. 1998. "The First-Digit Phenomenon," *American Scientist*, July-August
- Ijiri, Yuri and Herbert Simon. 1977. *Skew Distributions and the Sizes of Business Firms*, N.Y.: North Holland Pub.
- Janvresse, Élise and Thierry de la Rue. 2004. "From Uniform Distributions to Benford's Law." *Journal of Applied Probability* 41"1203-1210.
- Levin, Ines, Gabe A. Cohn, R. Michael Alvarez and Peter C. Ordeshook. 2009. Detecting Voter Fraud in an Electronic Voting Context: An Analysis of the Unlimited Reelection Vote in Venezuela. *Electronic Voting Technology Workshop / Workshop on Trustworthy Elections*, Online Proceedings
- Mebane, Walter. 2006. "Election Forensics: Vote Counts and Benford's Law." Paper prepared for the 2006 Summer Meeting of the Political Methodology Society, UC-Davis, July 20-22.

- Mebane, Walter. 2007. "Election Forensics: Statistics, Recounts and Fraud," paper presented at the 2007 Annual Meeting of the Midwest Political Science Assn, Chicago, II., April 12-16
- Mebane, Walter. 2008. "Election Forensics: The Second Digit Benford's Law Test and Recent American Presidential Elections," in R.M. Alvarez, T.E. Hall and S.D. Hyde, *Election Fraud*, Washington D.C.: Brookings
- Mebane, Walter and Kirill Kalinin. 2009. "Electoral Falsifications in Russia: Complex Diagnostics Selections 2003-2004, 2007-2008." Electoral Policy REO, 57-70 (in Russian).
- Myagkov, Mikhail, Peter C. Ordeshook and Dimitri Shakin. 2009. *The Forensics of Election Fraud*. Cambridge: Cambridge University Press
- Pericchi, Luis Ra_ul and David Torres. 2004. \La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela." Reporte T_ecnico node_nitivo 2a. versi_on: Octubre 01.
- Roukema, Boudewijn F. 2009. "Benford's Law anomalies in the 2009 Iranian presidential election." http://en.scientificcommons.org/46955838
- Sobyanin, Alexandar and V. Suchovolsky. 1993. "Elections and the Referendum December 11, 1993 in Russia," unpublished report to the Administration of the President of the RF, Moscow