

## **THE IMPACT AND REALITY OF FRAUD AUDITING BENFORD'S LAW: WHY AND HOW TO USE IT**

Benford's Law is used to find abnormalities in large data sets. Examples of the diversity of data sets include currency amounts, time durations, scientific and survey results, PIN numbers, surface areas, and numerical quantities. Learn why Benford's Law is useful and how to interpret the results it provides. Benford's Law is easily used and, best of all, not well known by fraudsters.

**GOGI OVERHOFF, CFE, CPA**  
**Investigative CPA**  
**California Board of Accountancy**  
**Sacramento, CA**

Gogi Overhoff is currently with the California Board of Accountancy, where she investigates accountants whose activities have been called into question through public complaints, referrals, or government review. Prior to her current position, Gogi was a forensic accountant and auditor with a Los Angeles CPA firm. Before that she worked in industry providing accounting and fraud prevention services to businesses in the manufacturing, banking, retail, and health care sectors. Gogi has given presentations to classes at the university level and to professional organizations including the ACFE.

“Association of Certified Fraud Examiners,” “Certified Fraud Examiner,” “CFE,” “ACFE,” and the ACFE Logo are trademarks owned by the Association of Certified Fraud Examiners, Inc.

## BENFORD'S LAW: WHY AND HOW TO USE IT

### Short Description and History

Benford's Law is a wonderful, albeit underutilized, tool in the auditor's arsenal. It looks at data in a way that is different from the normal auditor's point of view. Instead of looking for the largest amounts or a sufficient amount of coverage, Benford's Law allows the auditor to evaluate the digits that make up the numbers themselves. Benford's Law is not used to detect defective deliveries, contract rigging or off-book transactions, bribes, kickbacks, or asset thefts—but it is used to find payroll, expense, sales, accounts receivable, fixed asset, and journal entry anomalies as well as industry-specific account anomalies.

Benford's Law is based on the fact that many numbers normally used in business (and elsewhere) are not random, but rather follow some ordered progression. For example, a chart showing wealth will show that it is not uniformly distributed; a few people have much wealth and many people have less wealth. Sales, inventory, and disbursements are also not uniformly distributed. Benford's Law uses this fact to help point to fraud, inefficiencies, and other forms of data manipulation.

Back in the 1800s an astronomer, Simon Newcomb, noticed that the earlier pages in books of logarithm tables were more worn than the later pages. Logarithm books were used to multiply (and divide) large numbers. Newcomb posited that numbers beginning with lower digits were used more often than numbers beginning with higher digits. He published "Note on the Frequency of Use of the Different Digits in Natural Numbers" in 1881. He offered his observation but gave neither a use nor a proof. The article was promptly forgotten.

In the years following the Great Depression, without apparent knowledge of Newcomb or his article, Frank

### NOTES

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

Benford noticed the same thing. He was working for General Electric and had a lot of time on his hands. He decided to test his hypothesis. Benford analyzed 20 lists of large data sets (total of 20,229 data points) and 10 lists of smaller data sets (total of 2,968 data points). These lists came from random sources, such as the numbers in an issue of a magazine and death rates, as well as from sources that were not random—populations, horsepower lost, and molecular weight, and so on.

Benford published his observation and proof as “The Law of Anomalous Numbers” in 1938. Though he did not identify any uses, Benford’s article had a better reception than Newcomb’s and we now have Benford’s Law.

A few more things had to be established before Benford’s Law could be of any use to those on the finance side of things. In 1961 Roger Pinkham proved that Benford’s Law held true no matter what the unit of measurement. That means that it does not matter whether you measure items in yen, dollars, feet, miles, or meters. In 1972, Hal Varian found that you could use it to detect fraud in socioeconomic data. In the 1980s it was used to detect the reasonableness of round numbers and it was also found that invented numbers do not conform. Carslaw, in 1988, found that companies in New Zealand were not completely honest in their annual financial reports. (In the succeeding years, this finding was verified by others and for other countries). Ted Hill, also in 1988, found that people cannot create numbers and still conform to Benford’s Law. The real breakthrough for auditors came in 1994 when Mark Nigrini, a South African chartered accountant, codified a practical use. His 1992 thesis showed that accounting data conforms to Benford’s Law. In 1994 he assisted tax agencies find suspect returns. From there, he worked with companies to

## BENFORD'S LAW: WHY AND HOW TO USE IT

find fraud and continued his research to expand the applications of Benford's Law.

### **Basic Information and Why the Use of Benford's Law Is Advantageous**

Benford's Law states that many sets of numbers follow a predictable pattern, no matter what their origin or subject. The digit 1 will be the leading digit approximately 30% of the time. A leading digit is simply the left-most digit in a number. For example, the leading digit of "123" is 1 and the leading digit of "0.0552" is 5 (0 cannot be a leading digit). In fact, digits 1, 2, or 3 lead approximately 60% of the time. 9 as a leading digit appears only 4.5% of the time. Think of the following question: How much gasoline does each person put into his or her gas tank? Most of the time it will be less than \$50. Occasionally a broke college student will put \$9 in (because that is all they have in their wallet) and occasionally someone will put \$80 because they have a large tank. Think, also, of the following example. You won \$10 million in the lottery (congratulations!) and decide to invest in something that gives a guaranteed 10% return. It will take approximately 7.3 years to turn your 10 million into 20 million. When you reach 50 million it takes less than 2 years to reach 60. When you reach 90 million it will take just over 1 year to reach 100 million and then the cycle starts over again.

Current computing capabilities make Benford's Law fairly easy to implement. Research has addressed its use in many different fields and has expanded on its applicability (more on these items further). Benford's Law is admissible in U.S. courts at all levels and has been used internationally as well.

Finally, Benford's Law also helps to address the "expectation gap." As auditors, we all know that we do

## NOTES

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

NOT look for fraud, though if we find it, we will report it. We also know that the public expects us to look for fraud. Audit programs and most audit steps are not geared to find fraud. It would be like finding a needle in a haystack or having the winning lottery number. Benford's Law tests 100% of transactions, so it applies to:

- ❑ Coverage (3<sup>rd</sup> Standard of Fieldwork, AU Sec 210 and 230)
- ❑ Risk analysis (SAS 107, AU Sec 312),
- ❑ Sampling approach (SAS 111, AU Sec 350)
- ❑ Fulfills fraud requirements (SAS 99, AU Sec 316)

Best of all, Benford's Law does not require specialized (expensive) software.

Free Benford tools are also available from many different sources on the Web. For the newest version of Excel, one can purchase a TopCAATS add-in for approximately \$200. Of course, Benford's Law is included in many specialized software packages, among them ACL and CaseWare IDEA.

### **The Five Benford Tests**

There are five different options included in the original Benford's Law tests, as first promulgated by Mark Nigrini. They are: First Digit, Second Digit, First Two Digits, First Three Digits, and Last Two Digits. Each test's particular use and a short description are listed below.

Both the First and Second Digit tests are high-level tests used to check the general reasonableness of data and will only identify obvious anomalies. Their use will result in a large sample size.

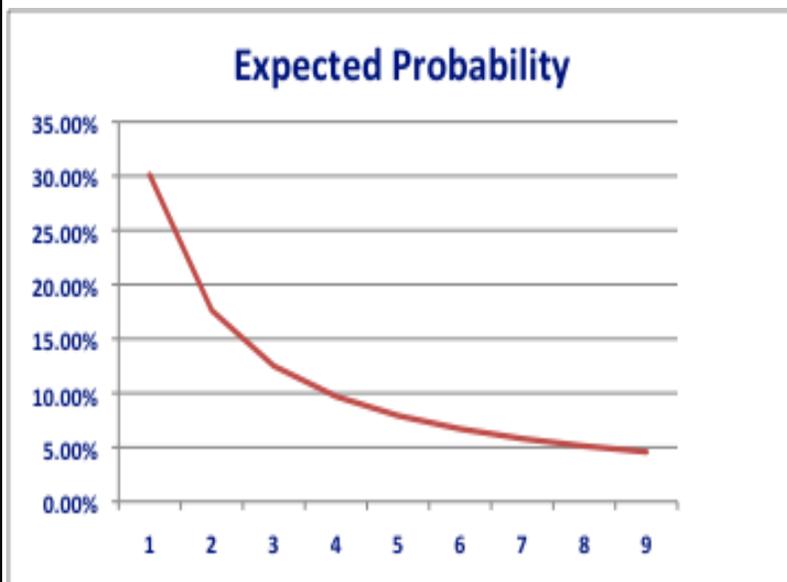
## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

The First Two Digits test is highly focused and should be used for data with less than 10,000 data points.

The First Three Digits test is also highly focused and should be used for data with more than 10,000 data points.

The Last Two Digits is used to identify overused and rounded numbers but is not generally used to find fraud (except in certain cases discussed later). This is the chart for the expected probabilities for the First Digit:

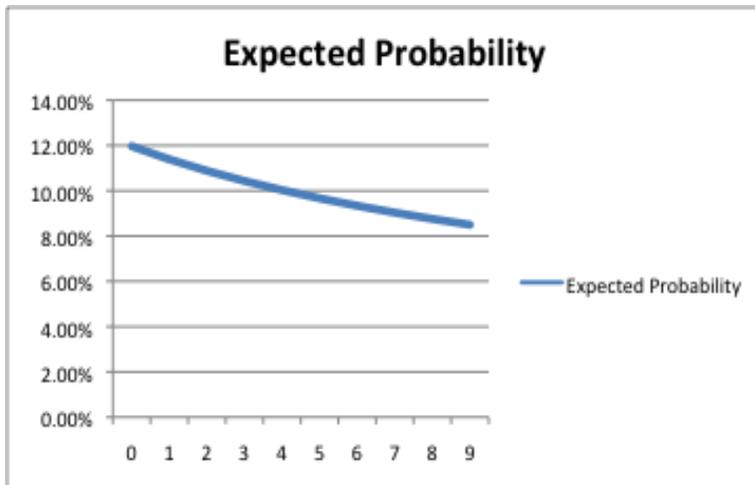


As can be seen from the chart, 1 should be the leading digit in most of the analyzed data and there should be far fewer items beginning with 7, 8, or 9. Said another way, further investigation is warranted if the data being analyzed shows a spike at, for example, 4. Remember, though, Benford's Law only looks at the digits and their placement within numbers and not at whether numbers are above or below materiality. So, the spike at 4 might turn out to be many entries for \$49.99.

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

This is the chart for the Second Digit:



Two things should be noted at this point:

1. This curve is far flatter than the curve for the First Digit
2. The horizontal axis goes from 0–9 (because 0 can be the second digit of a number)

In fact, all of the other charts have curves that are flatter than that of the First Digit. See the “Thinking Like a Fraudster” section for why this is important.

Below is a numerical chart for the expected frequencies of the first four digits:

Digit	First	Second	Third	Fourth
0	0.00%	11.97%	10.18%	10.02%
1	30.10%	11.39%	10.14%	10.01%
2	17.61%	10.88%	10.10%	10.01%
3	12.49%	10.43%	10.06%	10.01%
4	9.69%	10.03%	10.02%	10.00%
5	7.92%	9.67%	9.98%	10.00%
6	6.69%	9.34%	9.94%	9.99%
7	5.80%	9.04%	9.90%	9.99%
8	5.12%	8.76%	9.86%	9.99%
9	4.58%	8.50%	9.83%	9.98%

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

The further right a digit's placement in a number, the closer it is to having a one in ten chance of occurring, regardless of whether it is a 1 or a 9.

#### Data Requirements

Not all data works with a Benford analysis. Below is a list of rules and guidelines that should be adhered to in order to get a meaningful Benford analysis.

As required in most businesses, the data should have been reconciled and the auditor should understand the entity and its environment before embarking on any type of analysis. When doing a Benford analysis, and especially when working with a large data set, one should choose the smallest subset possible. For example, analyzing sales on a per-salesperson (or monthly) basis and then comparing the salespeople (months) to each other may be more meaningful than analyzing all salespeople together. Be aware that if amounts are controlled or constrained in some fashion (e.g., prices set by government), or if the majority of sales come from one product, then the data may have to be cleaned further to get a meaningful analysis.

Data sets *should*:

- Be composed of similar data (A/P, A/R, journal entries, sales, inventory counts, etc.)
- Consist of single transactions (not be aggregated, such as "weekly" deposits)
- Have more smaller rather than larger value items
- Be for a specific accounting period (not three-quarters or eleven months)
- Have enough items to be able to get a meaningful result
- Have an average (*mean* as defined below) that is greater than the midpoint (*median*)
- NOT have artificial minimums or maximums (no sales of the top ten salespeople, expenses over \$10 ...)

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

- NOT be a subset (except as noted above)
- NOT be broken up (test debits and credits of journal entries together, not separately or only the first ten pages of entries)
- NOT be assigned (such as Social Security numbers, phone numbers, purchase order numbers, or addresses)

### **Standard Deviation and Other Statistical Items**

Included in most Benford software is the option to see analyzed data based on statistical information. This text will not go into the details, as their understanding enhances the use of Benford's Law, but is not strictly required.

That said, here are some basic definitions.

*Confidence interval* is how sure one is that, for example, there is fraud. It is measured in terms of percent and is usually 90, 95, or 100. The number of false positives will increase when the confidence levels are lower.

*Mean* is the average of all data points.

*Median* is the data point where there are an equal number of points above and below this value.

*Outliers* are data points that are outside of the norm, usually three standard deviations away.

*Standard deviation* describes how far away the lowest data point is from the highest. The larger the range of the data being analyzed, the closer the fit to Benford's Law and the more apparent the non-conforming data will be.

The *Z-statistic* measures how large deviations are between the expected and the actual values (the margin of error).

The larger the Z-score (commonly 1% at 2.58, 5% at 1.96,

## BENFORD'S LAW: WHY AND HOW TO USE IT

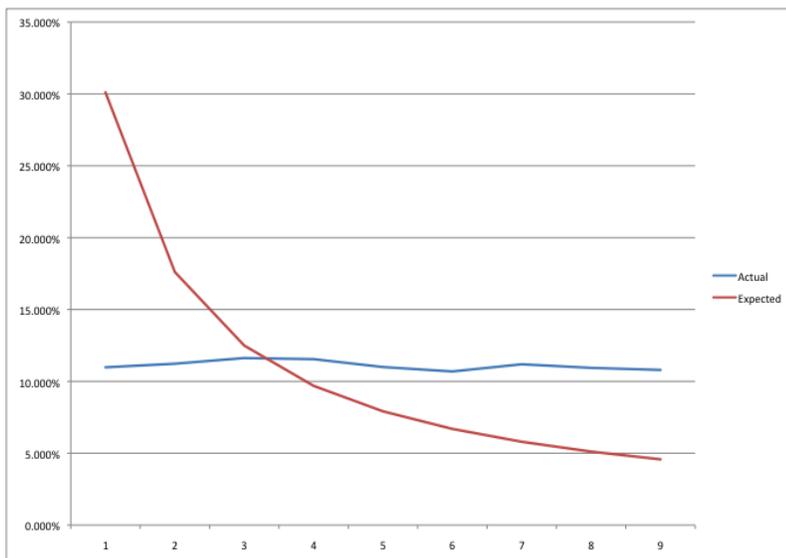
### NOTES

or 10% at 1.65) the less likely it is that the result is due to chance.

The above terms impact auditors because of their effect on the substantive testing process. In particular, the above statistical items should be taken into account to minimize dangers posed by Type I or Type II errors. Type I is when the auditor believes that fraud has occurred when there is no fraud (over-auditing). Type II is when the auditor believes that there is no fraud when, in fact, there is fraud (under-auditing).

### Selected Examples and How to Read Them

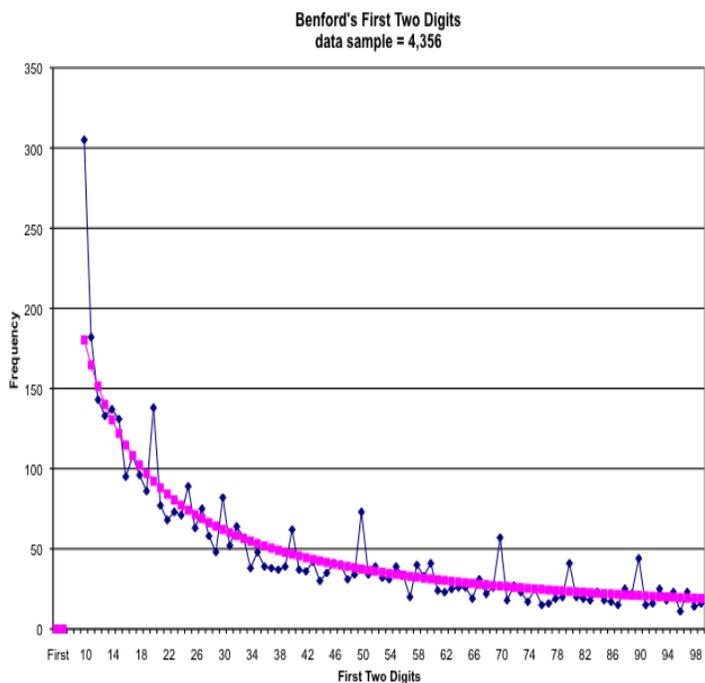
On one extreme, data is completely random if the First Digit Chart looks like this:



## BENFORD'S LAW: WHY AND HOW TO USE IT

This can be seen because the frequency of each digit is approximately 10%. A Benford analysis is not appropriate for this case.

Below is a more likely chart for a First Two Digits test:



## NOTES

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

At first glance, someone who is unfamiliar with Benford charts may be a little intimidated. After all, as auditors we deal with numbers and not with pictures (or charts). It is just a matter of practice. We have learned to recognize patterns in numbers and this is exactly what we need to look for in these charts. In this case your attention would be drawn to the spikes at the multiples of 10 and also to the great outlier at the beginning (12). Once you have recognized this pattern you would then drill down into the data to inspect the numbers beginning with 10, 20, 30.... You would also draw on your knowledge of the entity and its environment and would know that this particular chart is relatively normal because the inventory mostly consists of items bundled in groups of 10.

#### **Thinking Like a Fraudster**

Under the assumption that the reader is an honest and upright person, the following is helpful when reading Benford graphs. It is based on work done in the social sciences and mathematics.

First of all, and before getting into the graphs themselves: There are two concerns to keep in mind when using Benford's Law. The first, known as Intuitive Concern, notes that if there are only a few fraudulent transactions then they will probably not show up because there are too few to cause a spike on the chart. Other procedures would catch those items. The other, called a Statistical Concern, is the fact that the larger the size of the data set, the fewer anomalies (on a percentage basis) needed to create a spike. For example, less than 1% of entries need to be abnormal to create a spike when testing 10,000 entries but 3% of entries need to be abnormal when testing only 200 entries.

Fraudsters try to hide their shenanigans because they do not want to be caught. Therefore they often put some sort of

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

thought into the numbers they create. This ranges from (translated from German) “Douglas Adams 42,” “a bit more and a bit less,” “the keys my fingers hit,” and “accident” to “I tried to use 0s and 5s,” “not too many round numbers but a few,” “common sense,” and “avoidance of numbers that were next to each other (i.e. 4,234).”<sup>i</sup>

People cannot behave randomly. Or, better said, their idea of random does not conform to Benford's Law. If asked, people believe that digits occur randomly in numbers. That is, that the digit 9 occurs just as often as the digit 4. Benford's Law shows that that is not true, especially for the First and Second Digits but this does become truer the further one moves left. This is why curves for each of the tests become flatter.

On the fabrication of numbers: One school of thought is that people tend to have too many 0s and 5s<sup>ii</sup> and tend not to repeat digits (e.g., 333).<sup>iii</sup>

Theodore Hill has done extensive research on Benford's Law. See his chart below for the frequency of made-up (and therefore fairly random) First and Second digits.<sup>iv</sup> These do not conform to Benford's Law.

	Prob. of being 1st digit	Prob. of being 2nd digit
0	-	0.05800
1	0.14700	0.10600
2	0.10000	0.11700
3	0.10400	0.10900
4	0.13300	0.10500
5	0.09700	0.10000
6	0.15700	0.11200
7	0.12000	0.12800
8	0.08400	0.07300
9	0.05800	0.09200

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

In many cases, fraudsters tend to pass the First Digit test (another reason why this test is not so effective at finding fraud). This makes sense because they are trying to avoid detection and try to “stay under the radar.” Fraudsters tend to fail on the Second (and Third) Digit tests.<sup>v</sup> A study done on the U.S. dollar LIBOR rates found very interesting trends in Second Digit deviations (pointing to collusion among banks) during the latest economic decline.<sup>vi</sup>

Fraudsters also tend to use digits that are next to each other, “567” far more rarely than Benford’s Law suggests.

Fraudsters operating in collusion and needing to divide the spoils or people who need to keep track of entries made may tag amounts by applying unique final digits. For example, all amounts ending in “57” are Joe’s cut.<sup>vii</sup> This would be a case where the Last Two Digits test would be effective in finding fraud.

Benford’s Law is especially interesting when evaluating financial statements. In cases of net income the Second Digit test shows a surplus of 0s and a deficit of 9s. Net losses, however, have too many 9s and not enough 0s.<sup>viii</sup> Of course this makes sense if one thinks about it but it is interesting when relating this fact to its appearance on a Benford chart. It should also be noted that EPS were too often multiples of \$0.05 and rarely had 9 as a last digit.<sup>ix</sup> This has led many researchers to conclude that management tends to round net income and EPS up and round down net losses. They do also find that rounding becomes less prevalent the further away amounts are from the goal.<sup>x</sup>

Fraudsters are not necessarily safe (or more successful) if they are aware of Benford’s Law. In fact, they tend to conform less to Benford the more they try.<sup>xi</sup> A study of

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

German tax returns found that people trying to maximize refund amounts spend more time and get the same result if they know about Benford's Law than if they are unaware of it.<sup>xiii</sup> Interestingly enough, a study of U.S. returns found that amounts are not equally distributed throughout steps (amount ranges that determine tax due), but rather, more returns are concentrated in the top half of each step.

#### **Advanced Uses (A, A, & A)**

One of the requirements for applying Benford's Law is that there be no artificial maximums or minimums and that one should use a full accounting cycle. This is outside the scope of this paper but a mention is worthwhile. Lu and Boritz<sup>xiii</sup> have come up with an "Adaptive Benford" that uses an algorithm to simulate missing data. This allows a Benford analysis to be conducted despite missing or incomplete data, for example, on three quarters or one month.

Mark Nigrini has created what he calls the "2<sup>nd</sup> Order" test that detects download errors, rounded data, data generated by statistical procedures, data generated using regressions, and inaccurately ranked data.<sup>xiv</sup> The resulting charts will be "Almost Benford." If Madoff did use an algorithm to commit his fraud, Nigrini's 2<sup>nd</sup> Order would probably have detected it.

The third A is for **Artificial Neural Networks (ANN)**, a development on the horizon and currently in use on small data sets only.<sup>xv</sup> ANNs would work for frauds where there is a high degree of complexity and a sophisticated level of manipulation. ANNs work because the computer has "learned" to evaluate individual cases and thus, it takes many different things into account to arrive at a conclusion. This is how humans think. Computers are only able to arrive at conclusions by following a chain of computations

## BENFORD'S LAW: WHY AND HOW TO USE IT

and are not able to take many different things into account in parallel.

### Conclusion

Benford's Law is a different way of looking at numbers. In conjunction with other audit tools, Benford's Law helps auditors to minimize the "expectation gap" by increasing their chances of finding fraud and helps companies' bottom lines by finding inefficiencies and mistakes. It improves sampling so that auditors can concentrate on fraudulent or otherwise "suspicious" areas.

### NOTES

---

<sup>1</sup> Gambarara, F., 2004: "Benford Distribution in Science", *ETH Zürich*, available at

## BENFORD'S LAW: WHY AND HOW TO USE IT

### NOTES

[http://www.socio.ethz.ch/education/mtu/downloads/gambarara\\_nagy\\_2004\\_mtu.pdf](http://www.socio.ethz.ch/education/mtu/downloads/gambarara_nagy_2004_mtu.pdf)

<sup>ii</sup> Judge, G. and Schechter, L., 2007: "Detecting Problems in Survey Data using Benford's Law," *Journal of Human Resources*, 44:1–24.

<sup>iii</sup> McGregor, G.: "Long Arm of Benford's 'Law' Helps CRA Track Tax Cheats," *Ottawa Citizen*; April 29, 2009.

<sup>iv</sup> Hill, T., 1988: "Random-Number Guessing and the First Digit Phenomenon," *Psychological Reports* 62:967–71.

<sup>v</sup> Diekmann, A., 2007: "Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data," *Journal of Applied Statistics* 3/34:221–229.

<sup>vi</sup> Abrantes-Metz, R. and Villas-Boas, S., 201: "Tracking the Libor Rate," available at: <http://ssrn.com/abstract=1646600>

<sup>vii</sup> Nigrini, M. and Mittermaier, L., 1997: "The Use of Benford's Law as an Aid in Analytical Procedures," *Auditing: A Journal of Practice and Theory*.

<sup>viii</sup> Kinnunen, J. and Koskela, M., 2002: "Who Is Miss World in Cosmetic Earnings Management?" available at <http://ssrn.com/abstract=292806>.

<sup>ix</sup> Nigrini, op. cit.

<sup>x</sup> Lin, F., Guan, L., and Fang, W., 2009: "Heaping in Reported Earnings: Evidence from Monthly Financial Reports of Taiwanese Firms," available at [http://list.academic-journal.org/submissions/isfa2009\\_submission\\_121.pdf](http://list.academic-journal.org/submissions/isfa2009_submission_121.pdf)

<sup>xi</sup> Gambarara, op. cit.

<sup>xii</sup> Watrin C., Struffert R., Ullmann R., "Benford's Law: An Instrument for Selecting Tax Audit Targets," *Review of Managerial Science* 2 (3) (2008), available at: <http://www.springerlink.com/content/296p91r570034k25/>

<sup>xiii</sup> Lu F., Boritz E., and Covvey D., 2006: "Adaptive Fraud Detection Using Benford's Law," in Luc Lamontagne and Mario Marchand (eds.) *Advances in Artificial Intelligence: Proceedings of the 19th Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI June 2006*, pg. 347–358.

<sup>xiv</sup> Nigrini, M., Miller, S., 2009: "Data Diagnostics Using Second-Order Tests of Benford's Law," available at <http://www.tcnj.edu/~business/documents/DataDiagnosticsUsingSecond-Order-Nigrini.pdf>

<sup>xv</sup> Bhattacharya, S., Xu, D. and Kumar, K., (2011): "ANN-Based Auditor Decision Support System Using Benford's Law," *Journal Decision Support Systems*.