Paper presented in:
VII[th] International Scientific Conference on Economic Policy and EU Integration
Business Faculty, "Aleksander Moisiu" University
April, 2016

THE USE OF BENFORD'S LAW AS A TOOL FOR DETECTING FRAUD IN ACCOUNTING DATA

Ikbale TOTA, Ada ALIAJ, Jonida LAMÇJA
Business Faculty, "Aleksander Moisiu" University, Durres
Contact: e-mail: i.tota@yahoo.com

Abstract
Benford's law has been presented as an effective audit tool in identifying suspect accounts for further analysis. Many researches have focused on the validity if the Benford distribution is an appropriate tool to discriminate fraud and non-fraud accounting. According (Nigrini, 2012) 'Accounting data, in general, has a remarkable conformity to Benford's Law; It is still an open question as to whether accounting numbers from companies in emerging markets will conform to Benford's Law in their emerging stages or whether conformity comes about only in mature markets'. In this paper we attempt to describe and illustrate the implementation of fraud discovery with Benford's law analysis in an Albanian case, to check if the expected frequencies of the digits are in accordance with the accounting data. The result could be helpful for auditors to detect and prevent fraud and misrepresentation in financial statements.
Key words: *Benford's law; audit; accounting*

Introduction

It has been known for a long time that if an extensive collection of numerical data expressed in decimal form is classified according to first significant digit, without regard to position of decimal point, the nine resulting classes are not usually of equal size (Raimi, 1976). Benford's Law, also known as the first-digit law, has long been seen as a tantalizing and mysterious law of nature. (Fewster, 2009). This law has intrigued scientist and laypeople for over a century. The story began in 1881, when the American astronomer Simon Newcomb noticed that books of logarithm table's always seemed grubby on the early pages, and clean toward the back. For some reason, seemed to look up numbers beginning with the digits 1 and 2 and far more often than they looked up numbers beginning with the digits 8 and 9 (Fewster, 2009). The next development was Frank Benford's analysis (Benford, 1938). The law of probability of the occurrence of numbers is such that all mantissa of their logarithms are equally probable (Newcomb, 1881). In 1938, Frank Benford, a physicist, discovered that the digits of naturally occurring numbers such as death rates, areas drained by rivers, populations of cities, and many other phenomena are distributed in a predictable non-uniform manner. If one were to examine the leading or first digit of a large set of such data, the number '1' would appear in about 30.1 per cent of the cases, '2' would appear in about 17.6 per cent, '3' would appear in about 12.5 per cent and so on in decreasing fashion. The number '9' would occur in only about 4.6 per cent of the cases. This distribution of first digits is known as the Benford distribution, and data exhibiting this distribution are said to conform to Benford's Law (Ettredge & Srivastava, 1999). The formulas for the digit frequencies are shown next with D_1 representing the first digit, D_2 the second digit, and D_1D_2 the first-two digits of a number:

$$\Pr ob(D_1 = d_1) = \log\left(1+\frac{1}{d_1}\right); \quad d_1 \in \{1,2,...9\}$$

$$\Pr ob(D_2 = d_2) = 1\sum_{d_1=1}^{9}\log\left(1+\frac{1}{d_1d_2}\right); \quad d_2 \in \{1,2,...9\}$$

$$\Pr ob(D_1D_2 = d_1d_2) = \log\left(1+\frac{1}{d_1d_2}\right); \quad d_1d_2 \in \{10,11,...99\}$$

Where Prob indicates the probability of observing the events in parentheses (Nigrini, 2012). Using his formula, the probability that the first digit of a number is one is about 30% while the probability the first digit a nine is only 4,6 %. Table shows the expected frequencies for all digits 0 through 9 in each of the first four places in any number.

Table 1: Expected Frequencies Based on Benfrod's Law

Digit	1st place	2nd place	3rd place	4th place
0		0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.19882	0.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.09990
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982

Source: (Nigrini, 2012)

Auditors are required to use analytical procedures to identify the existence of unusual transactions, events, and trends. Benfrod's Law gives the expected patterns of the digits in numerical data, and has been advocated as a test for the authenticity and reliability of transaction level accounting data. (Nigrini & Miller, 2006). Is it possible to tell that a number is wrong just by looking at it? In some cases, you bet. Using Benfrod's law-a mathematical phenomenon that provides a unique method of data analysis-CPAs can spot irregularities indicating possible error, fraud, and manipulative bias or processing inefficiency (Nigrini, 1999). Nigrini appears to be the first researcher to apply Benford's law extensively to accounting numbers with the goal to detect fraud. He first became interested in the work on earnings manipulation and his dissertation used digital analysis to help identify tax evaders (Durtschi, Hillison, & Pacini, 2004). Accountants and auditors have begun to apply Benfrods's law to corporate data to discover number-pattern anomalies (Nigrini, 1999). The primary difficulty in applying Benford's Law to the detection of fraud is that many datasets do not naturally satisfy Benford's Law. While some datasets do largely follow the Benford behaviour, there is no "bright line" test to distinguish the two types (Pimbley, 2014). Application of Benford's Law will be fruitful in some analyses. But financial risk managers and investigators should always apply generalized common sense, curiosity, scepticism, models and diverse automated procedures to the review of data integrity (Pimbley, 2014)

It is not clear how large our numbers have to be to be reasonably large. Benford's Law assumes that each numeric amount has "many" digits. Nigrini found that conformity to Benfrod's Law requires that we have a large data table with numbers that have at least four digits. Simulations have shown that the numbers should have four or more digits for a good fit. However, if this requirement is violated, the whole ship does not sink. When the numbers have fewer than four digits, there is only a slightly larger bias in favour of the lower digits (Nigrini, 2012). A set of numbers that closely conforms to Benford's Law is called a Benford Set; this term is preferred because auditors can relate to a set of transactional data from an audit cycle (Nigrini & Miller, 2006).

(Miller, 2010) Describes some explanations for why so many different and diverse sets of data satisfy Benford's law:

(1) First explanation, the **Spread Theory**: this theory says that if a data set is distributed over several orders of magnitude, then the leading digits will approximately follow Benford's law.

(2) Second explanation, the **Geometric Explanation**: The idea is that if we have a process with a constant growth rate, then more time will be spent at lower digits than higher digits. For definiteness, imagine we have a stock that increases at 4% per year. The amount of time is takes to move from 1$ to 2$ is the same as it would take to move from 10,000$ to 20,000$ of rom 100,000,000$ to 200,000,000$. If n_d is the number of years it takes to move from d dollars to d+1 dollars then

$$d \cdot (1.04)^{n_d} = (d+1), or\ n_d = \frac{\log\left(\frac{d+1}{d}\right)}{\log 1.04}$$

It is simple algebra to show that this implies Benford behaviour. If n is the amount of time it takes to move from 1$ to 10$, then

$$1 \cdot (1.04)^n = 10 \quad or \quad n = \frac{\log 10}{\log 1.04}$$

thus we see the percentage of the time spent with a first digit of d is

$$\frac{\log\left(\frac{d+1}{d}\right)}{\log 1.04} \Big/ \frac{\log 10}{\log 1.04} = \frac{\log\left(\frac{d+1}{d}\right)}{\log 10} = \log_{10}\left(\frac{d+1}{d}\right)$$

Which is just Benford's law! There is nothing special about 4%; the same analysis works in general. This is not an isolated example. Many natural and mathematical phenomena are governed by geometric growth.

(3) For the third explanation, the **Scale Invariance Explanation**, (Miller, 2010) return to a comment from Newcomb's paper: As natural numbers occur in nature, they are to be considered as the ratios of quantities. Therefore, instead of selecting a number at random, we must select two numbers, and inquire what is the probability that the first significant digit of their ratio is the digit n. The import of this comment is that the behaviour should be independent of the units used. As the universe doesn't care what units we use for our experiments, it is natural to expect that the distribution of leading digits should be unchanged if we change our units. Benfrod's law is consistent with scale invariance.

(4) Fourth explanation, the **Central Limit Explanation**, there are many data sets in the world whose values are the product of numerous measurements. (For example the monetary value of a gold brick is a product of the brick's length, width, height and density of gold). Imagine we have some quantity X which is a product of n values, so

$$X = X_1 \cdot X_2 \cdots X_n$$

We assume the X_i are nice random variables. From our discussion above, to show that X obeys Benford's law it suffices to understand the distribution of the logarithm of X module 1. Thus we are led to study:

$$\log_{10} X = \log_{10}\left(X_1 \cdot X_2 \cdots X_n\right) = \log_{10} X_1 + \cdots + \log_{10} X_n$$

By the Central Limit theorem, if n is large then the above sum is approximately normally distributed, and the variance will grow with n; however.

The general rule is that the data set should have at least 1,000 records before we should expect a good conformity to Benford's Law. For tables with fewer than 1.000 records the Benford-related tests still can be run, but we should be willing to live with larger deviations form Benford's Law line before concluding that the data did not conform to the law (Nigrini, 2012). Another general rule is not to test the first-two digit frequencies of data sets with fewer than 300 records. The first digit test (with all its flaws) should be used on small data sets. For data sets with fewer than 300 records, the records can simply be sorted from largest to smallest and the pages visually scanned for anomalies (Nigrini, 2012).

Recent tax evasion, auditing, and forensic investigations research shows that there are practical uses of Benfrod's Law (Nigrini, 2012). Benford's analysis, when used correctly, is a useful tool for identifying suspect accounts for further analysis. Is a particularly useful analytical tool because it does not use aggregated data; rather it is conducted on specific accounts using all the data available. It can be very useful in identifying specific accounts for further analysis and investigation. (Durtschi, Hillison, & Pacini, 2004) While Benford analysis by itself might not be a "sure-fire" way to catch fraud, it can be a useful tool to help identify some accounts for further testing and therefore should assist auditors in their quest to detect fraud in financial statements.

Data and Methods

Benford's Law is likely useful when applied under several conditions. For instances, set of numbers that result from mathematical combination of numbers whereby the result come from two distributions e.g. account receivable (number sold x price); transaction-level data where sample is not needed e.g. disbursement, sales, expenses; on large database set, full year's transactions will provide more accurate result; and for account that appear to conform which the mean of a set of number is greater than the median and the skewness is positive e.g. most set of accounting numbers (Aris, Othman, Arif, Malek, & Omar, 2013). Digital analysis based on Benford's Law is an audit technique that is applied to an entire population of transactional data. Benford's Law was introduced to the auditing literature in Nigrini and Mittermaier (1997), and researchers have since used these digit patterns to detect data anomalies by testing either the first, first-two, or last-two digit patterns of reported statistics or transactional data (Nigrini & Miller, 2009). Auditors have long applied various forms of digital analysis when performing analytical procedures. For example, auditors often analyse payment amounts to test for duplicate payments. They also search for missing check or invoice numbers. Benford's law as applied to auditing is simply a more complex form of digital analysis (Durtschi, Hillison, & Pacini, 2004).

Using Benford's Law, one must start with measuring deviation. The deviation of the distribution of digits between what is observed and what is expected in many ways. One method is the "Chi Square" test, a standard statistical test for measuring the degree of similarity between elements in a table. Based upon this statistic, and the number of "degrees of freedom", it is possible to assign a probability that any variation between actual and observed is due to chance alone. The higher the Chi Square, the less likely that any difference can be explained by chance alone (Aris,

Othman, Arif, Malek, & Omar, 2013).
In this paper we attempt to describe and illustrate the implementation of fraud discovery with Benford's law analysis in an Albanian case, to check if the expected frequencies of the digits are in accordance with the accounting data. The data for this case study were taken from purchase and sale invoices of several fictitious companies. All digits of the purchases and sales were simply fabricated and not true. Since no services were offered or delivered, these companies must have invented all the numbers in his scheme, and because people are not random, invented numbers are unlikely to follow Benford's Law. The numbers seem to have been chosen so that they look randomly. Benford's Law is used in this case to verify whether it would be an appropriate tool, which if it would be used at the right time could have prevented this type of fraud.
This data set contains 2,963 records, all greater than or equal to 100 ALL. (All numbers less than 100 ALL are deleted because these numbers are usually immaterial from an audit perspective.), this data set contains 1,037 records all less or equal to 250,000 ALL.

Results and Discussion

When people fabricate data they do not choose numbers, which follow the logarithmic distribution (Hill, 1996). Even when people invent numbers without a goal such as fraud in mind, the digital frequencies do not conform well to Benford's Law

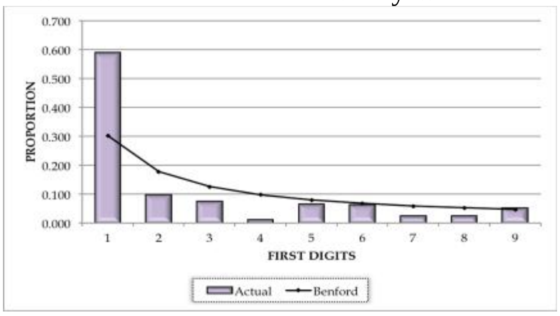

Figure 1 First Digit patterns

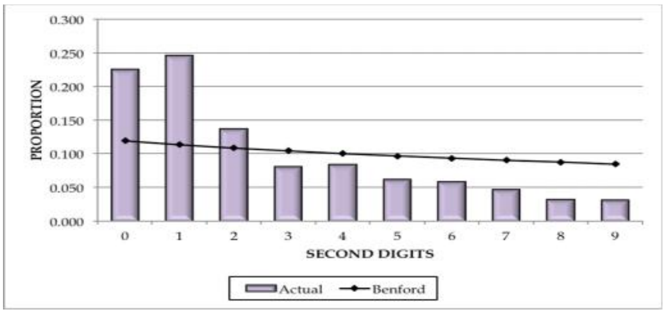

Figure 2: Second Digit Patterns

Table 2: Expected Frequencies and Actual Frequencies Based on Benfrod's Law

First and second digits							
First	Count	Actual	Benford	Difference	AbsDiff	Z-stat	Chi-square
1	1742	0,588	0,301	0,287	0,287	34,024	810,113
2	288	0,097	0,176	-0,079	0,079	11,250	104,726
3	222	0,075	0,125	-0,050	0,050	8,206	59,326
4	31	0,010	0,097	-0,086	0,086	15,875	228,491
5	193	0,065	0,079	-0,014	0,014	2,797	7,380
6	187	0,063	0,067	-0,004	0,004	0,799	0,652
7	74	0,025	0,058	-0,033	0,033	7,650	55,694
8	72	0,024	0,051	-0,027	0,027	6,593	41,762
9	154	0,052	0,046	0,006	0,006	1,575	2,501
	2963			MAD =	0,06513		
Second	Count	Actual	Benford	Difference	AbsDiff	Z-stat	Chi-square
0	666	0,225	0,120	0,105	0,105	17,596	273,433
1	727	0,245	0,114	0,131	0,131	22,498	449,672
2	404	0,136	0,109	0,028	0,028	4,782	20,634
3	239	0,081	0,104	-0,024	0,024	4,185	15,910
4	249	0,084	0,100	-0,016	0,016	2,918	7,823
5	182	0,061	0,097	-0,035	0,035	6,463	38,094
6	172	0,058	0,093	-0,035	0,035	6,577	39,590
7	139	0,047	0,090	-0,043	0,043	8,216	61,879
8	94	0,032	0,088	-0,056	0,056	10,722	105,524
9	91	0,031	0,085	-0,054	0,054	10,563	102,735
	2963			MAD =	0,05282		

(Hill T. , 1988). As expected and showed in the figure below the results of case study analysis shows that the fabricated data does not conform to expected true digital frequencies.The calculated chi-square statistic is 1308.878 which exceeds the critical value of 112.02 at $\alpha = 0.05$ with 89 degrees of freedom, and indicates that the null hypothesis of conformity to Benford's Law should be rejected. The high-calculated chi-square statistic is due in part to the number 2,963 records.

The higher the MAD, the larger the average difference between the actual and expected proportions (Nigrini M. , 2012). Compared to the Critical Values and Conclusions for Various MAD values (see: (Nigrini M. , 2012)) For the First Digits MAD Values above 0.015 signal Nonconformity and for the Second Digits MAD Values above 0.012 signal Nonconformity

Conclusions

As seen above in the graphs the results of the case study indicate the same result as (Hill T. , 1988) that the distributions of random numbers presumed by people share the following properties with the Benford distributions: (i) the frequency of numbers with the first significant digit 1 is much higher than expected; (ii) the frequency of numbers with first significant digit 8 or 9 is much lower than expected; and (iii) the distribution of the second digits is much more nearly uniform than the distribution of the first digits.

ACFE, AICPA, & IIA (2008) notes that a "Benford's Law analysis" can be used to examine transactional data for unusual transactions, amounts, or patterns of activity. Fraud detection mechanisms should be focused on areas where preventive controls are weak or not cost effective. One fraud detection scorecard factor is whether the organization uses data analysis, data mining, and digital analysis tools to "consider and analyse large volumes of transactions on a real-time basis." (Nigrini & Miller, 2009). The evidence showed that Benford's Law can be used to detect false cases or cases with fictional numbers (Nigrini M. , 2012). The case study analysed in this paper showed that Benford's Law can help to detect cases where fictional numbers are involved or at least can be used as a signal to audit more.

However, this usually requires a reasonably in-depth review and some scepticism so as not to simply dismiss a spike or a set of spikes as being normal (Nigrini M. , 2012).

Reference

1. **Aris, N. A., Othman, R., Arif, S. M., Malek, M. A., & Omar, N. (2013).** Fraud Detection: Benford's Law vs Beneish Model. *2013 IEEE Symposioum on Humanities, Science and Engineering Research (SHUSER).*
2. **Benford, F. (1938).** The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*
3. **Durtschi, C., Hillison, W., & Pacini, C. (2004).** The Effective use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting*
4. **Ettredge, M., & Srivastava, R. (1999).** Using Digital Analysis to Enhance Data Integrity. *Issues in Accounting Education*
5. **Fewster, R. (2009).** A Simple Explanation of Benfrod's Law. *The American Statician*
6. **Miller, S. (Ed.). (2010).** *Benfrod's Law: Theory and Applications.* Princeton University Press
7. **Newcomb, S. (1881).** Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics.*
8. **Nigrini, M. (1999).** I've Got Your Number. How a mathematical phenomenon can help CPAs uncover fraud and other irregularities. *Journal of Accountancy*.
9. **Nigrini, M. (2012).** *Applications for Forensic Accounting, Auditing, and Fraud Detection.* John Wiley & Sons Inc
10. **Nigrini, M. (2012).** *Benford's Law_Applications for Forensic Accounting, Auditing and Fraud Detection.* Wiley
11. **Hill, T. (1996).** A Note on Distributions of True Versus Fabircated Data. *Perceptual & Motor Skills*
12. **Hill, T. (1988).** Random-Number Guessing and The First Digit Phenomenon. Psychological Reports
13. **Nigrini, M., & Miller, S. (2006).** Data Diagnostics using Second Order Tests of Benford's Law.
14. **Nigrini, M., & Miller, S. (2009).** Data Diagnostics using second order tests of Benford's Law
15. **Pimbley, J. (2014).** Benfrod's Law and the Risk of Financial Fraud. *Quant Perspectives* **Raimi, R. (1976).** The First Digit Problem. *The American Mathematical Monthly*.