Digital Analysis of Financial Data and Detection of Fraud

Dr. Anupam Parua[♦]

Abstract

Corporate scams have become common phenomena all around the globe. In most of these scams it is the malfunctioning of accounting and auditing system which is at fault. There is no point in trying to make good for losses after the occurrence of such losses. The system should be there through which any corrupt practice is detected when it is being perpetuated and we should not leave it for the post-control mechanism to detect the fraud and corruption because by that time the real damage is done. In case of financial data the digital analysis of data can hint at the vulnerability of such reported data. Benford's law is a good technique in the field of digital analysis. This paper attempts to discuss the basic conception and acceptability of Benford's law. Simon Newcobm, an astronomer and mathematician, is thought to be pioneer for the development of this rule, although the real credit goes to Frank Benford, an American physicist, who theorized the law. this law is basically an empirical regularity commonly found in sets of data describing naturally-occurring phenomena. The regularity observed is that the frequency of the digits appearing as the first or leading digit in the numbers collected declines as the value of the digits increase. This paper also points out to the applicability issues of the use of this law. Finally the paper attempts to test the authenticity of one published data set. This law is regarded as a part of forensic accounting techniques. Adoption of forensic accounting tools does not guarantee the prevention of frauds in the matter of accounting but, for sure, it will reduce the occurrence of this menace.

Key-words: *Forensic Accounting, Corporate Fraud, Benford's Law, Accounting Practices.*

JEL Classification: *C 16, M 41, Y 10*

1. Introduction

In the recent days' corporate world financial scams have become a very much common phenomena. In world context we can cite a plethora of cases spreading all around the globe. We have the case of 'Parmalat' in 2003. Parmalat was an Italy-based producer of Ultra Hot

[♦] Assistant Professor, Dept. of Commerce (Accounting & Finance), K. D. College of Commerce & General Studies, Midnapore, e-mail ID: anupam_parua@rediffmail.com

Temperature (UHT) milk and other foods. There the accounting practices adopted by the founder of the company Calisto Tanzi created a hole of about $14 billion in company's accounting records. Then we have a very popular (?) case of 'Enron'. It was the Houston-based energy company that toppled into a spectacular bankruptcy due to a painstakingly-planned accounting fraud made by its accounting firm, Arthur Andersen. Once considered a blue-chip stock, Enron shares dropped from $90 to $0.50, which spelled disaster in the financial world, where thousands of employees and investors saw their savings vanish with the company as it filed for an earnings restatement in October 2001. Again we have the case of 'Tyco International'. That company was a diversified manufacturing conglomerate that deals with electronic components, health care, fire safety, security, and fluid control with headquarters in New Jersey, USA. In 2005, its CEO Dennis Kozlowski and CFO Mark H. Swartz were found guilty of stealing $600 million from the company. In this case there is the excess of executive compensation at shareholder's expense, where Kozlowski will be remembered for the $2 million birthday bash he gave his wife on a Mediterranean Island at the company's expense.

In India also, we have a lot of corporate scams involving accounting malfunctions. One of the most recent and very notable scams is the Satyam scam that jeopardized India's corporate image. That brings issues like business ethics, corporate governance, and accounting and business practices to the limelight in a big way. As rightly pointed out by the FICCI in the aftermath of this scam: "This fraud …… shows a systematic breakdown in audit and board oversight".

The possibility of this type of financial fraud is very much there given the system and practice of accounting and auditing in vogue. There is no point in trying to make good for losses than to prevent the occurrence of such losses. Here lies the importance of checking the authenticity of accounting practices followed in any concern dealing with large amount of money especially for those who are dealing with public money. The system should be there through which any corrupt practice is detected when it is perpetuated and we should not leave it for the post-control mechanism to detect the fraud and corruption because by that time the real damage is done.

In this context digital analysis of reported data has become all the more important. Benford's Law and Benford's Table is a very effective tool in the field of digital analysis of accounting and financial data. The rest of the paper is organized in the following way. Section 2 highlights the concept and basic issues of Benford's law; Section 3 indicates the applicability or otherwise of this law in case of certain types of data sets; Section 4 gives a practical application of this law in a given data set and finally Section 5 offers the conclusions.

2. Benford's Law

2.1. The Background and Basic Overview

Dr. T. P. Hill of Georgia Institute of Technology in United States asked his mathematics students to go home, toss a coin for 200 times, note the results and report it to him. When the students presented the results he easily pointed out which students have really tossed the coin and recorded the results and which students had presented fake results. Dr. Hill then pointed out that '*The truth is most people don't know the real odds of that exercise, so they can't fake data convincingly*".

Probability predictions are often very surprising. Dr. Hill revealed in one issue of 'American Scientists' some surprising probability. It states that the overwhelming odds are that at some point of time in a series of 200 tosses, there will a run of six or more heads or tails. The fakers don't know that and as such avoid recording long runs of any particular outcome which they mistakenly feel to be improbable.

A strange feature in such probability distribution is that it is scale invariant and also base invariant. The practical application of this in the world of finance is that whether the financial data are recorded in home currency or foreign currency the probability distribution would hold good.

In 1881, Simon Newcomb, an astronomer and mathematician, published an article in American Journal of Mathematics. In the article he noted his observation that library copies of books of logarithms were considerably more soiled in the beginning pages and progressively less worn in the subsequent pages. From that he inferred that more number exists which begin with numeral one than with larger numerals. His inference was not supported by theoretical explanation or empirical testing. As such it was better known as 'Newcomb's Conjecture'. That inference went virtually unnoticed for more than five decades.

Then in 1938, apparently independent from Newcomb's work, Frank Benford, an American physicist, also noticed that first few pages of his logarithm books were more worn out than the last pages. He came to the same conclusion as Newcomb did some 50 years ago; that people more often looked up numbers that began with lower digits than those beginning with higher digits. He also posited that there were more numbers starting with lower digits than the numbers starting with higher digits.

Though the findings of Newcomb and Benford were fundamentally same yet there were huge differences in respect of impact of their work. Newcomb's work went unnoticed as it was not supported by theoretical explanation or empirical testing. But Benford tested his hypothesis by collecting and analyzing data. For this he collected some 20000 observations from diverse fields like length of rivers, atomic weights of elements, population, cost data, X-ray volts, death rates etc.. After analysis he found that numbers consistently fall into a pattern with low digits occurring more frequently in the first position than larger digits. The mathematical tenet of this finding became known as 'Benford's Law'.

Thus, Benford's law, also called the first digit law, refers to an empirical regularity commonly found in sets of data describing naturally-occurring phenomena. The regularity observed is that the frequency of the digits appearing as the first or leading digit in the numbers collected declines as the value of the digits increase. As a result, 1 is the digit more often observed and 9 is the least frequent. That is, not all digits have the same chance of appearing as the first digit in the numbers collected.

This empirical regularity has elicited considerable interest, and even some fascination, for the following reasons. First, Benford's law contradicts what seems to be a natural intuition suggesting that any digit has the same probability of appearing as the first digit in the numbers collected. Common belief suggests that digits have a uniform distribution when appearing as leading digits in numbers. Second, observance of Benford's law has been confirmed in several datasets. Third, the statistical literature has taken a long time to provide solid formal explanations for the often observed presence of Benford's law.

2.2. Understanding the Law

Benford specifically found that the proportion of 'real world' numbers whose leading digit is d=1, 2, 3,………, 9 is approximately given by:

$P(d) = \log_{10}(1+1/d)$

In the similar way distribution of digits in other places can also be found. Thus Benford's table of distribution of probabilities of digits at different places can be had in the following way:

Table 1: Benford's Distribution

Digit	1st Place	2nd Place	3rd Place	4th Place
O	----	11.97%	10.18%	10.018%
1	30.10%	11.39%	10.14%	10.014%
2	17.61%	10.88%	10.09%	10.010%
3	12.49%	10.43%	10.06%	10.006%
4	9.69%	10.03%	10.02%	10.002%
5	7.92%	9.67%	9.98%	9.998%
6	6.69%	9.34%	9.94%	9.994%
7	5.80%	9.04%	9.90%	9.990%
8	5.12%	8.76%	9.86%	9.986%
9	4.58%	8.50%	9.82%	9.982%

Source: Nigrini (1996)

Formula:

Probability for the first digit of the number

$P(D_1=d_1) = \log_{10}[1+(1/d)]$; $d_1=1,2,3,\ldots\ldots\ldots,9$.

Probability for the second digit of the number

$P(D_2=d_2) = \log_{10}[1+(1/\, d_1d_2)]$; $d_2=1,2,3,\ldots\ldots\ldots,9,0$.

For many years the status of Benford's law was little more than a numerical curiosity. Wide discussion and practical use of it started to emerge since 1960s. In 1961 Pinkham developed what was known as scale invariance theorem. He posed the question that if there were indeed some law governing digital distributions then that law should be scale invariant. He went on to prove that Benford's law is scale invariant under multiplication. That means if any set of numbers conformed to Benford's law then the product obtained by multiplying the data set by a constant then the new data set would also follow Benford's law. Hamming (1970) suggested that the rule that almost 1/3rd of the numbers processed began with a 1 could have implications for the design of efficient computers. In economics, Varian (1972) suggested that Benford's law could help detect anomalous information in datasets used for public planning decisions.

Take one example to see why the law is good. Suppose the price of a particular corporate stock is \$1 and it grows by 1% per month. So the price will reach \$10 after 232 months. However it will take 70 months or 30.2% of total time of 232 months for the price to rise from \$1 to \$1.99. But it will take only 11 months (4.7% of total time of 232 months) to rise from \$9 to \$9.99. World data relating to share price and market capitalization are known to fit the Benford's pattern which is not surprising given the above example (Declan Laville, 2005).

Glewwe & Dang (2005) show how having computers available for data input at the district level, so that mistakes can be found more quickly and households re-interviewed sooner, can improve data quality. These computers could easily be programmed to include a Benford's law component to test for the quality of responses to different questions and from different enumerators. Judge & Schechter (2009) have demonstrated how Benford's law can be used to detect data abnormalities arising both from questions that are difficult to answer and from enumerators' errors.

Gonzalez-Garcia & Pastor (2009) found that rejection of Benford's law may be unrelated to quality of statistics, and instead may result from marked structural shifts in the data series. Hence, nonconformity with Benford's law should not be interpreted as a reliable indication of poor quality in macro-economic data.

For the effective application of Benford's law the data set should have a wide range. Otherwise the law would not work. Paldam (2010) checked Maddison's GDP Data as to their conformity with Benford's law and proposed that the deviations from the law were due to range problems. The data covered only three log-decades, the hundreds, the thousands and the ten-thousands, and two of these log-decades were not fully covered.

Benford's law has been used to study different hypotheses about datasets. For example, Nigrini (1996 and 1999) and Nigrini and Mittermaier (1997) used Benford's law to detect fraudulent data in tax payments and accounting. Hales, Sridharan, Radhakrishnan, Chakravorty, and Sinha (2008) studied the reliability of employee-reported operational data by testing the presence of Benford's law in those datasets; and Diekmann (2007) analyzed whether the first digits of regression coefficients published in scientific literature tend to be distributed according to the first digit-law.

3. Applicability of Benford's Law

Benford's law is applicable to a data set which is normally distributed. In the following circumstances Benford's distribution can help rigorously:

- When the numbers to be tested are arising out of combination of numbers from different sources Benford's distribution can be effectively utilised. In other words, combining unrelated numbers gives a distribution of distributions, a law of true randomness that is universal (Hesman, 1999);
- Data sets are expected to follow the Benford's distribution when the elements result from random variables taken from divergent sources that have been multiplied, divided or raised to integer powers (Boyle, 1994);
- Most accounting related data are expected to follow Benford's distribution (Hill, 1995). This is because typical accounts consist of transactions that result from combining numbers;
- If mean of particular set of number is greater than it median and the skewness of the distribution is positive, the data set is likely to follow Benford's distribution. Thus larger the ratio of mean divided by median, the more closely the set will follow the Benford's distribution. This is so because Benford's distribution has predominance of small numbers (Wallace, 2002);
- Data set generated by mathematical operation like repeated multiplication of current numbers by either a constant or a random number would conform to Benford's distribution ((Scott & Fasli, 2002);
- A probability in which the logarithm of the variable is distributed normally is called lognormal distribution. Lognormal distributions are unimodal, positively skewed and

have a range from 0 to ∝ and are defined by two parameter, the median of the distribution and the standard deviation of the logarithm of the variate. Lognormal distributions conform to Benford's distribution independent of scale parameter (median) and dependent on shape parameter (standard deviation) (Scott & Fasli, 2002);

His (Benford's) law is not magic, but sometimes it seems like it (Nigrini, 1998). Sometimes the Benford's predictions fail to detect frauds in data because biasness in data might arise out of some innocent reasons. So in the following cases Benford's law cannot be applied.

- For example, in United States the people who are traveling for business reasons are to submit bills for meals costing $ 25 or more. If the cost of meal is less than $ 25 they are not required to submit bill for that. So in huge number of cases we can find bills for meals costing just below $ 25 say, $ 24.90. That's why we could find so many 24s in the related data set. In this case no real fraud is committed but it is only avoidance of strict rules.
- A data set which contains data arising out of authorized but repetitive transactions would not conform to Benford's distribution. Such non-conformance does not indicate any fraud. Removal of such repetitive entries would make the data set expected to conform to the Benford's distribution (Durtschi, Hillison, Pacini, 2004);
- We can't use Benford's law to improve our chances in lottery. In lottery we are to pulls balls from a jar. The balls are labeled with certain numbers but the numbers are not really acting as numbers. The labels could be changed to use names of animals instead of numbers. The numbers in the balls are uniformly distributed and the chance of getting any number is equal. Thus, in this case Benford's law is not applicable.
- Financial indices like Dow Jones Index data or Sensex data are not expected to conform to Benford's distribution. As the change between two successive closing values will not be independent of preceding changes such data sets lose randomness and as such do not conform to Benford's distribution.
- In order to hold the Benford's pattern true the number must not the invented or assigned. For example, telephone numbers are not going to follow the pattern as these are forced to have a fixed number of digits.
- The sample wherefrom the numbers are to be taken are to be statistically large enough.

4. Benford's Law: One Practical Application on a Given Data Set

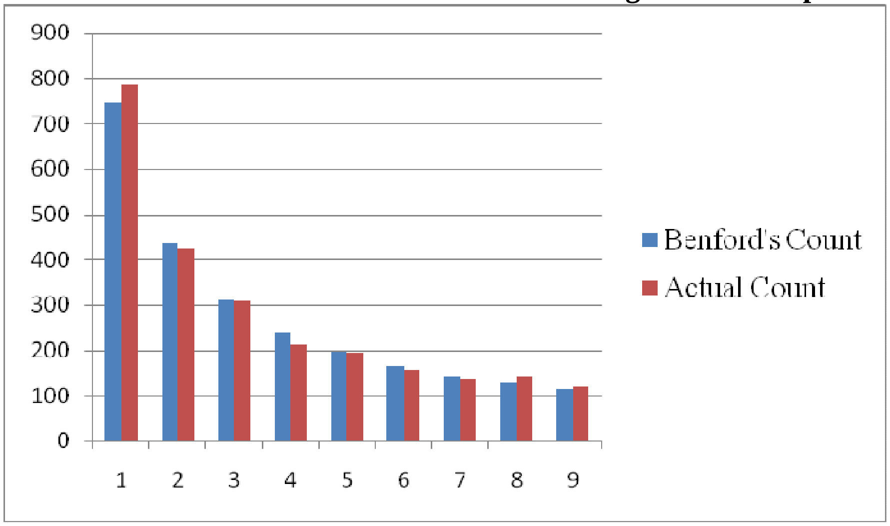
The use of Benford's law can be illustrated by means of an example. For this purpose a sample data comprising of profit after tax (PAT) data of 2478 BSE-listed companies for the accounting year 2012-13 has been taken. The extracted results are as follows:

Table 2: Proportion of Observed Proportion of Leading Digits

First Digit	Actual Count	Observed %	Benford's %	Deviation (%)
1	787	31.8	30.1	5.5
2	425	17.2	17.61	-2.6
3	307	12.4	12.49	-0.8
4	213	8.6	9.69	-11.3
5	195	7.9	7.92	-0.6
6	156	6.3	6.69	-5.9
7	135	5.4	5.8	-6.1

8	140	5.6	5.12	10.3
9	120	4.8	4.58	5.7

Figure 1: Benford's Count and Actual Count for First Digit of the Sample Data Set



From the above table it can be seen that PAT values with leading digit 1 and 8 have a proportion which is obviously higher than that of Benford's distribution. Again for PAT values with first digit of 4 the Benford's count is obviously higher than the actual count. For the first digit of 3 and 5 there is absolutely nothing to choose between. Thus PAT figures with first digit of 1 and 8 special care should be put in so as to detect whereas the reported figures represent true figures.

5. Conclusion

Now we are in all-digit world. Definitely, it gives us certain advantages but it also has brought about the undesirable feature of expediting the manipulation or even the fabrication of digital assets. In the light of recent corporate scams that were possible because of lapses and loopholes in accounting practices the necessity of detecting the perpetuation of such malfunctioning should be detected before these are committed on a larger scale and for a longer period of time. Then only the damage could be prevented or at least could be limited. In this context forensic accounting tools like application of Benflord's law and table have become very important. Adoption of forensic accounting tools does not guarantee the prevention of frauds in the matter of accounting but for sure it will reduce the occurrence of this menace.

References

[1] Benford, F. (1938), "The Law of Anomalous Numbers", *Proceedings of the American Philosophical Society*, Vol. 78, pp. 551 – 572.

[2] Boyle, J. (1994), "An Application of Fourier Series to the Most Significant Digit Problem", *American Mathematical Monthly*, Vol. 101 (9), pp. 879 – 886.

[3] Carslaw, C. A. P. N. (1988), "Anomalies in Income Numbers: Evidence of Goal Oriented Behaviour", *The Accounting Review*, Vol. LXIII (2), pp. 321 – 327.

[4] Laville, D. (2005), "Fraud Detection and Number", The Actuary, pp. 36-37

[4] Diekmann, A., (2007), "Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data," *Journal of Applied Statistics*, Vol. 34 (3), pp. 321 – 329.

[5] 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*, Vol. 5(1), 17 – 34.

[6] Glewwe, P. & Dang, H.-A. H. (2005), "The impact of decentralized data entry on the quality of household survey data in developing countries: Evidence from a randomized experiment in Vietnam,. Unpublished Manuscript.

[7] Gonzalez-Garcia, J. & Pastor, G. (2009), "Benford's Law and Macroeconomic Data Quality", IMF Working Paper, No. WP/09/10.

[8] Hamming, R. (1970), "On the Distribution of Numbers", *Bell System of Technical Journal*, Vol. 49, pp. 1609 – 1625.

[9] Hesman, T. (1999), "Cheaters Tripped up by Random Number Laws", *Dallas Morning News*, August, 22.

[10] Hill, T. P. (1995), "A Statistical derivation of the Significant Digit Law", *Statistical Science*, Vol. 10 (4), pp. 354 – 363.

[11] Hill, T. P. (1998), "The First Digital Phenomenon", *American Scientist*, Vol. 86 (4), pp. 358 – 363.

[12] Judge, G., & Schechter, L. (2009), "Detecting Problems in Survey Data Using Benford's Law, *Journal of Human Resources*, Vol. 44, pp. 1–24.

[13] Newcomb, S. (1881), "Note of the Frequency of Use of the Different Digits in Natural Numbers", A*merican Journal of Mathematics*, Vol. 4, pp. 39 – 40.

[14] Nigrini, M. J. (1996), "Taxpayer Compliance Applicationof Benford's Law", *Journal of American Taxation Association*, Vol. 18 (1), pp. 72 – 92.

[15] Nigrini, M. J. (1999), "Adding Value with Digital Analysis", *The Internal Auditor*, Vol. 56 (1), pp. 21 – 23.

[16] Nigrini, M. J. and Mittermaier, L. J. (1997), "The Use of Benford's Law as an Aid in Analytical Procedures", *Auditing: A Journal of Practice and Theory*, Vol. 16, No. 2, pp. 52 – 67.

[17] Paldam, M. (2010), "A Check on Maddison's GDP Data: Benford's Law with Some Range Problems" Economic Working Paper Series, No. 2010 – 18, School of Economics and Management, Aarhus University, Denmark.

[18] Pinkham, R. (1961), "On the Distribution of First Significant Digits", *Annals of Mathematical Statistics*, Vol. 32, pp. 1223 – 1230.

[19] Scott, P. D. & Fasli, M. (2002), "Benford's Law: An Empirical Investigation and a Novel Explanation", CSM Technical Report, Dept. of Computer Science, University of Essex, UK.

[20] Varian, H. (1972), "Benford's Law", *The American Statistician*, Vol. 23, pp. 65 – 66.

[21] Wallace, W. A. (2002), "Assessing the Quality of Data Used for Benchmarking and Decision Making", *The Journal of Government Financial Management*, Vol. 51 (3), pp. 16 – 22.