IOSR Journal of Economics and Finance (IOSR-JEF)
e-ISSN: 2321-5933, p-ISSN: 2321-5925.
PP 01-09
www.iosrjournals.org

Devising a Model for Accounting Fraud Detection Based on Benford's Law

Dr. Shalini Talwar[1], Ms. Kinjal Mehta[2],

Abstract: *The objective of this research paper is to recommend an additional due diligence tool in order to identify the possibility of fraud on the basis of the financial statements of the concerned companies. The authors have used the Benford's law to devise a model that highlights the possibility of fraud in the financial statements. The model makes use of statistical techniques to identify the companies that have a high probability of the P&L statement numbers being manipulated. The authors have analyzed the income side and the expense side of statement of profit and loss separately for the companies identified for evaluation in the context of the current study. The model was tested for a sample of 35 companies comprising of both "good" and "fraudulent companies (from accounting perspective)" to understand its efficiency. The results of two companies "First Leasing Company of India" and "TCS" are analyzed in detail in the current paper.*

Keywords: *Accounting fraud, Benford's law, Benford's (Expected) frequency,* digital analysis.

I. Introduction

Private corporations have contributed to global prosperity since past many centuries. Employment and investment opportunities created by mega corporations the world over have contributed tremendously to growth of global economy. However, time and again, instances of accounting frauds have emerged in these companies and shaken the confidence of investors, employees, lenders and regulators. Many steps are being taken at different levels to reduce the possibility and scope of such frauds. Latest research has focused not only on prevention of fraud but also on early warning mechanism for detecting such frauds. With the increase in accounting frauds done by the companies, it has become necessary to devise a model that raises a caution in case of the entries on the financial statements being manipulated and fudged.

The objective of this research paper is to recommend an additional due diligence tool in order to identify the possibility of fraud on the basis of the financial statements of the concerned companies. The authors have used the Benford's law to devise a model that highlights the possibility of fraud in the financial statements. Benford law is a phenomenological law related to the frequency distribution of the leading digits in the numerical data of many real life sets. The law states that in many naturally occurring collections of numbers, the small digits occur disproportionately, often as leading significant digits.

The model makes use of statistical techniques to identify the companies that have a high probability of the P&L statement numbers being manipulated. The authors have analyzed the income side and the expense side of statement of profit and loss separately for the companies identified for evaluation in the context of the current study. Each section has been checked for the conformity to Benford's law. In case the data fails to conform to the Benford frequency, an alert can be raised indicating the probability that the P&L numbers are fudged is high.

The model was tested for a sample of 35 companies comprising of both "good" and "fraudulent companies (from accounting perspective)" to understand its efficiency. The results of two companies **"First Leasing Company of India"** and **"TCS"**are analyzed in detail in the current paper.

The model is expected to be useful as it helps the auditors to perform digital analysis while undertaking analytical procedures. Further, it raises an alarm if the probability that the numbers of the financial statements being manipulated is high. Additionally, in order to reduce the Type-I or Type-II errors, the model gives a chart indicating the frequency of every number of P& L statement being repeated. This can be used to understand if there is an "abnormal" duplication of numbers or not.

[1]Dr. Shalini Talwar Associate Professor-Finance, K J Somaiya Institute of Management Studies and Research, Mumbai
[2]Ms. Kinjal Mehta MMS-Finance (2014-16), K J Somaiya Institute of Management Studies and Research, Mumbai

II. Literature Review

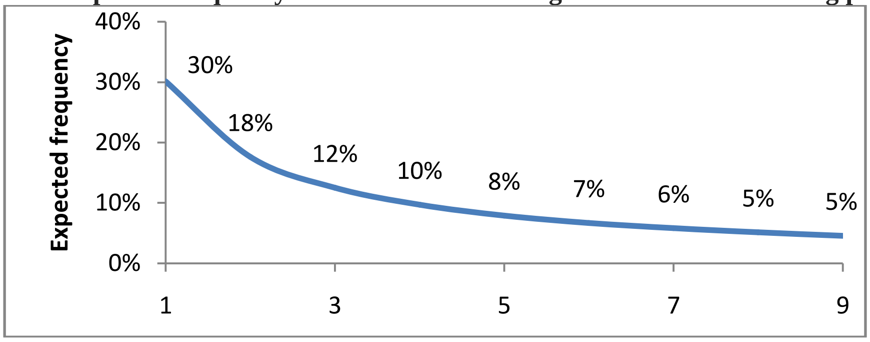
A review of existing literature was undertaken with a view to understand the meaning and implication of Benford law in addition to exploring its origin and its applicability to diverse data sets.

2.1 What is Benford's law?

Benford's law pertains to the frequency distribution of leading digits in numerical datasets. Simply put, the law states that leading-digit distribution for digits from 1 to 9 is neither random nor uniform. The chances of occurrence of lower digits in leading place are disproportionately higher than the chances of occurrence of higher digits in the leading place.

The law has been called by different names such asfirst-digit law or the law of leading digits frequencies or the significant digit law and even as digital frequencies analysis (Nigrini, 1999).An estimation of the probability of any digit to be a leading digit in a given number can be done using mathematical equation derived by Benford[3]. Beneford's equation was further analyzed by him in his paper published in 1938 and also by Pinkham (1961) and Hill (1996). Figure 1 illustrates the expected frequency of the occurrence of digits from 1 to 9 in leading place.

Figure 1: The expected frequency of the occurrence of digits from 1 to 9 in leading place



2.2 The origin of Benford's law

Benford's law, given by physicist named Frank Benford, is similar to the observations made by an astronomer and mathematician, Simon Newcomb way back in the year 1881. On the basis of certain observations, it was inferred by Necomb (1881) that low digits were more frequently used and looked up as compared to the higher digits.Newcomb had based his inferences on his observation that beginning pages of books of logarithm dealing with low digitsappeared to be more worn as compared to the pages dealing with higher digits. However, inferences drawn by Newcomb remained unnoticed, as he did not endeavor to explain his observations through any theory or formula. Almost half a century later, Benford made the same observation about the beginning pages of his own logarithmic books. He also concluded that people tend to look up numbers that began with low digits more frequently as compared to those that began with higher digits. Unlike Newcomb, Benford set about testing his hypothesis by collectingnearly 20000 observations from a variety of data sets. Through his extensive analysis, he showed that numbers invariably fell into a pattern where low digits occurred more frequently in the first place.

2.3 Application of Benford's law to real life data

Researchers have attempted to apply Beneford's law on diverse data sets. It has been found that the law can be applied to set of numbers that have certain characteristics. Wallace (2002) observed that if the mean of a given set of numbers is higher than its median and if skewness has a positive value,then that data set can be subjected to digital analysis as it would more than likely follow Benford's distribution. It was observed by Durtschi et al. (2004) that there are certain tests that can be used to ascertain whether or not Benford's law would apply to the data set under consideration.

---

[3] $p(n) = \log10(1 + 1/n)$, n= 1,2,3,...,9

To test if a random set of data conforms to Benford's law, we tested the population figures and GDP figures of countries. The results are graphically represented in figures 2 and 3.

Figure 2: Population figures from the countries

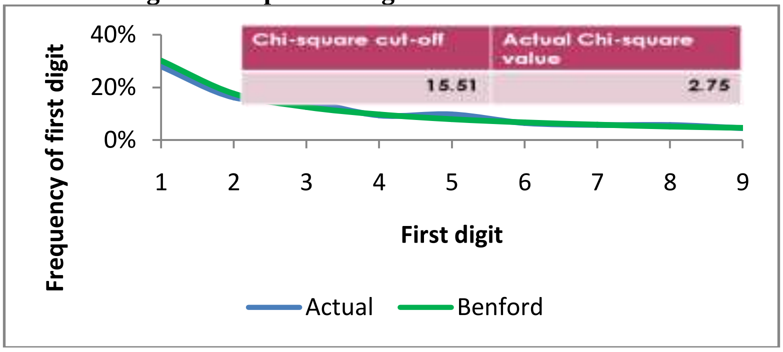
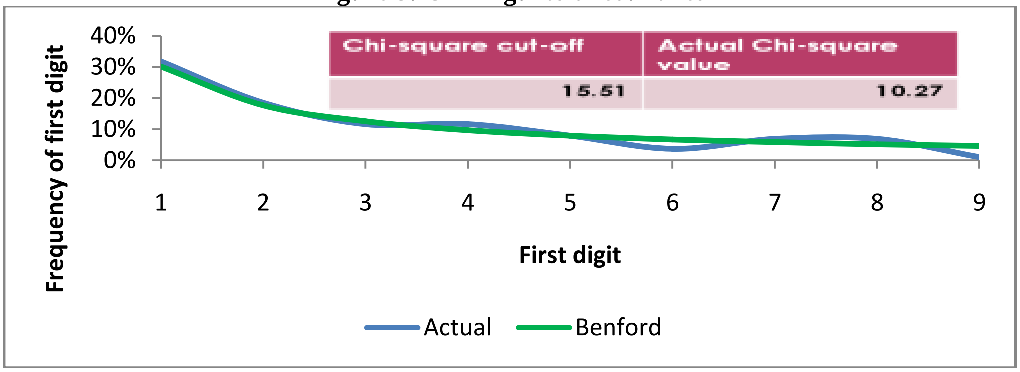


Figure 3: GDP figures of countries



2.4 Application of Benford's law in accounting fraud detection

Hill (1995) showed that this law can be applied to financial data set as well. He provided proof that this law held true in case of financial data related to stock market and certain accounting information. Hiil observed that Benford's distribution was like the normal or Gaussian distribution and most accounting-related data is more than like to conform to Benford's distribution. This being the case, accounting related data can be subjected to digital analysis. However, not all accounting data may be amenable to such analysis. Etteridge and Srivastava (1999) showed that accounting data set that does not conform to the Benford distribution might not indicate fraud. Rather, it would show certain operating inefficiencies or systemic flaws. It was suggested by Varian (1972) that Benford's law could be used to detect fraud in case of socio-economic data related to public planning decisions.Mark Nigrini (1999)suggested that Benford's law can also be used for forensic accounting and auditing to detect accounting and expenses fraud.

2.5Points to consider while applying the Benford law

While applying Benford's law, totals and sub totals should be ignored.Further, number brought from other schedules and pages should not be counted twice.Income and expense items should also be analyzed separately. Additionally, numbers that come from tables should be omitted in the analysis.

III. Methodology

To achieve the objectives of the study, a model based on the following tests was devised:

- The First Digit Test
- Second digit Test
- First Order Test
- Second Order Test
- Last Two Digits Test.

The input to the model is past 5 to 6 years data extracted from "CapitalLine".The entire data set was tested for the following statistical tests:

- Chi-Square Test
- Z-Significance Test
- K S Test

The output of the model is a score that indicates the number of tests in which the cut-off value of the statistical tests is being breached.

1.1 Calculating the Benford's (Expected) Frequency

Table 1 exhibits calculations related to **Benford's (expected) frequency.**

Table 1:Calculating the Benford's (expected) frequency

DIGIT	FORMULA
1st Digit	P(D1=d1)=log(1+1/d1) D1{1,2,3…,9}
2nd Digit	$(D2=d2)=\sum_{d1=1}^{d1=9} \log (1+1/d1d2)$ d2 {0,1,2,….,9)
1st two digits	P(D1D2 =xy) =log(1+1/xy)

3.2 Performing the basic tests

The model includes testing the basic digits to identify the possibility of "accounting fraud".
The basic digits include the following:

- First Digit
- First Two Digits
- Second Digit
- Second Order Test
- Last two digits.

3.2.1 First Digit Test

The First Digit test is a high-level test of reasonableness. There are 9 possible first digits from 1 to 9. We identify the first digit of every number on the income and the expense side using the formula:

Left (Column A, 1)

After obtaining the first digit of all the numbers, we calculate the actual proportion, i.e the frequency distribution of each digit to compare it with the Benford (expected) proportion.This is done by taking the absolute difference between the actual proportion and the Benford proportion. The difference is used to perform the statistical tests to identify if the actual proportion shows a significant deviation from the expected proportion or not.

3.2.2 Second digit test

The second digit test is a second overall test of reasonableness. There are 10 possible second digits from 0 to 9. We identify the second digit of every number on the income and the expense side using the formula:

Right (left (Column A, 2),1)

We calculate the actual proportion, i.e the frequency distribution of each digit to compare it with the Benford (expected) proportion.This is done by taking the absolute difference between the actual proportion and the Benford proportion. The difference is used to perform the statistical tests, to identify if the actual proportion shows a significant deviation from the expected proportion or not.High deviation from the Benford proportion is a flag that the data table contains abnormal duplication and anomalies.

3.2.3 First order test

The first two digit test is a more focused test and is used to detect abnormal duplications of data and possible biases in the data. In a forensic environment, the first-order test (the first-two digits test) would be run after the high level tests i.e. 1st digit test and the 2nd Digit test.There are 90 possible first two digits from 10 to 99. We identify the first two digits of every number on the income and the expense side using the formula:

Left (column A,2)

We calculate the actual proportion, i.e. the frequency distribution of each digit to compare it with the Benford (expected) proportion.Actual Proportion for 1st two digits being 10 is calculated as:

Count of the 1st two digits being "10"(can vary till 90)/ Total number of rows

The Benford proportion is calculated as follows:

Benford frequency for 1st 2 digits being 10= log (1+1/10)

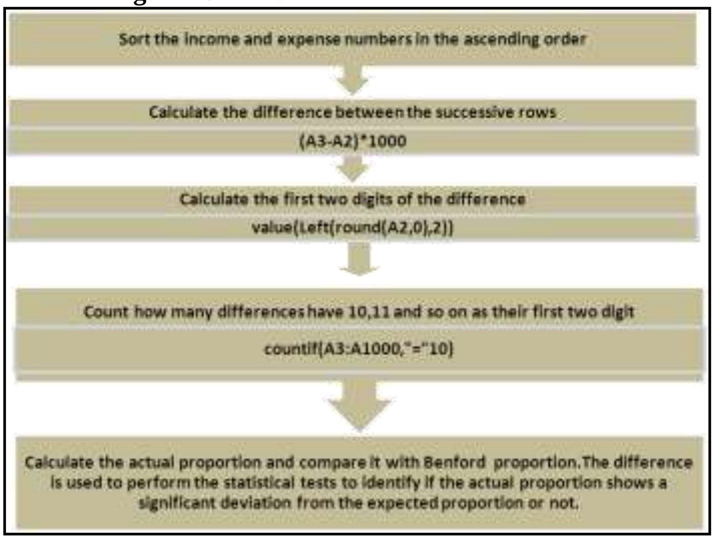
We then take the difference between the actual and the Benford proportion to run the statistical tests. This is done by taking the absolute difference between the actual proportion and the Benford proportion. The difference is used to perform the statistical tests, to identify if the actual proportion shows a significant deviation from the expected proportion or not. High deviation from the Benford proportion is a flag that the data table contains abnormal duplication and anomalies.

3.2.4The second order test

The second order test can be applied to any set of data. The second order test looks at relationships and patterns in data and is based on the digits of the difference between amounts that have been sorted from smallest to largest (ordered). The digit patterns of the differences are expected to closely approximate the digit frequencies of Benford's law.

The second order test gives few, if any, false positives in that if the results are not as expected (close to Benford), then the data do indeed have some characteristic that is rare and unusual, abnormal, or irregular.Flowchart showing the detailed explanation of the execution of the second order test is given in figure 4.

Figure 4: Execution of the second order test



The multiplication by 1000 in step 2 is so that the amounts such as 0.01 become 10.00 and so we can then use the **Left** function to calculate the first two digits.We will always have one blank (null) cell because a data set with N records only with N-1 differences. The **Round** function is to be sure that we get the correct first-two we get the correct first-two digits. There are 90 first two digits combinations from 10 to 99.

3.2.5 Last two digits test

The last- two digits test is a powerful test for number invention. This test is more appropriate when we do not want number invention or number creativity.There are 100 possible last two digit combinations from 00 to 99 and the expected proportion are equal to 0.01 for each possible last two digit combination.This is because as we move towards the right in a number, the digits are expected to be uniformly distributed. Since there are 100 possible last two digits, our expected proportion is uniform at 0.01 each.We extract the right two digits of all the numbers on the income and expense side using the formula:

Right (A2,2)

3.2.6The number duplication test

The number duplication test is a drill down test and is essentially a number hit parade. This test identifies which particular numbers were causing the spikes on the above mentioned tests. The spikes on the above mentioned tests indicate that the numbers are abnormally used.

The result of the number duplication test is a table in the form of a report showing:

- A rank
- The amount that was duplicated
- The count for each number.

The table is sorted by the count in the descending order so that the amount that occurred most often in the data is listed first.

3.3 Performing the Statistical Tests

3.3.1One digit at a time: Z-Statistic

The Z-statistic is used to test whether the actual proportion for a specific first- two digit combinations differs significantly from the expectation of Benford's law.The formula used to calculate the Z-Statistic value is:

Z= |AP-EP|-(1/2N)/sqrt((EP(1-EP))/N)

Where AP= Actual Proportion
EP = Expected Proportion
N= Number Of records.

The (1/2N) term is a continuity correction term and is only used when it is smaller than the first term in the numerator.We can see that the Z-statistic becomes larger as the difference between the observed proportion and the expected proportion becomes larger.The calculated value of Z- statistic is compared at a significance level of 5% i.e. 1.96.However the Z-statistic cannot be added or combined in some other way to get an idea of the overall extent of nonconformity.

3.3.2 Chi - Square Test

The chi – square test is used to compare an actual set of results with an expected set of return.
The chi-square value is calculated as follows:

Chi-square= (Σ (AP-EP) ^ 2)/ EP

Where AP= Actual Proportion
EP =Expected Proportion
summation is from i =1 to k
and k-1= degrees of freedom.
The cut-off value to compare the calculated value is obtained by using the formula

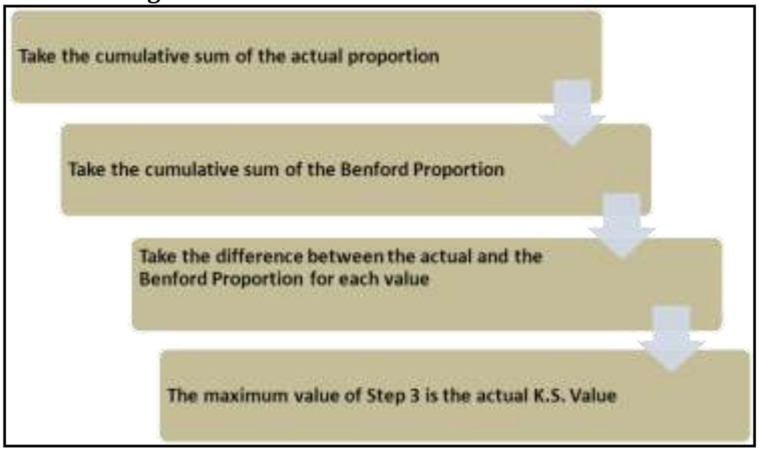
CHIINV (0.05, k-1)
Where 0.05= 5% significance level
and k-1 = degrees of freedom

One limitation of chi- square is that it suffers from the excess power problem in that when the data table becomes large, the calculated chi-square will almost always be higher than the cutoff value making us conclude that the data does not follow Benford's law.

3.3.3 Kolmogorov- Smirnoff Test (K.S.Test)

It is "all digits at once test". This test is based on the cumulative density function.We calculated the actual K.S value as illustrated in figure 5.

Figure 5: Calculation of the actual K.S value

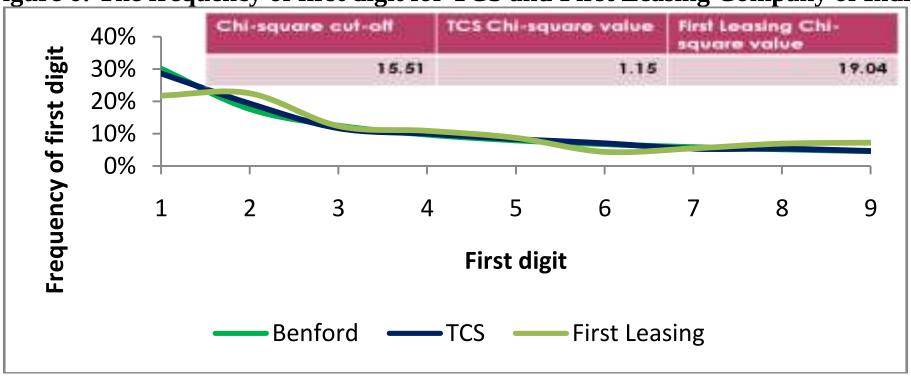


The cut-off value of K.S. Test is calculated as follows:
Kolmogorov – Smirnoff =1.36/ sqrt (N)

IV. Discussion of Results

In this section, analysis of the output of the above described tests for TCS and First Leasing Company of India is given. Figure 6 illustrates the frequency of first digit for the two companies used for the study.

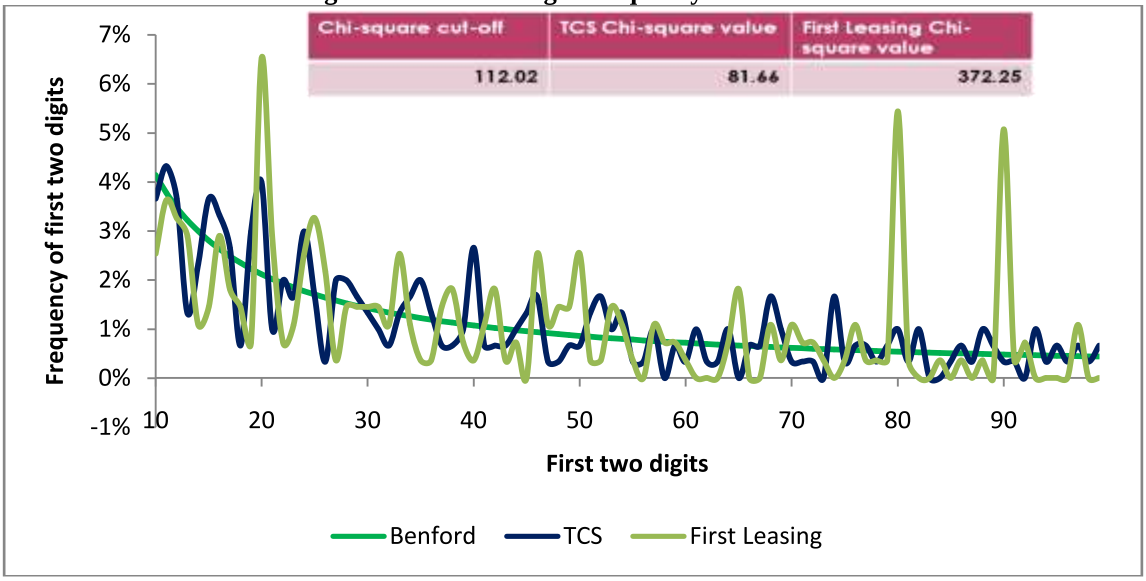
Figure 6: The frequency of first digit for TCS and First Leasing Company of India



Observation: As it can be seen the good company, TCS follows almost Benford distribution whereas the accounting fraudulent company, First Leasing Company of India, shows significant deviation from the Benford distribution.Also from the chi square statistical value, the cut off for TCS is significantly low as compared to the chi square cut-off of the first digit which is equal to 15.51.The chi square value of the First Leasing Company of India breaches the cut-off value which is in accordance with our assumption that Company Y is an accounting fraudulent company.

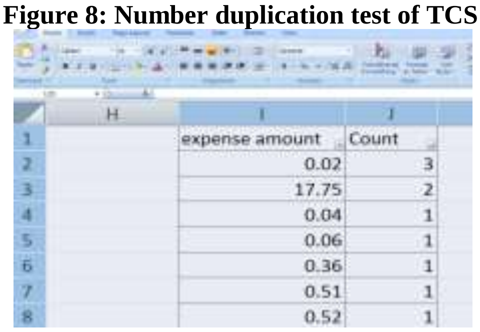
Figure 7 exhibits the first two digits frequency distribution.

Figure 7: First two digits frequency distribution



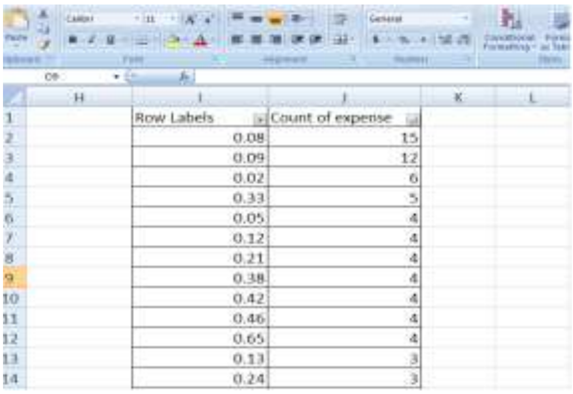
Observation: In the above graph, we have taken a sample having first two digits and compared the good company,TCS, and an accounting fraudulent company, First Leasing Company of India. The accounting fraudulent company showed several peaks and troughs, indicating significant deviation from the Benford frequency distribution. This is in accordance with chi square value which is breaching the cut-off value with a huge margin. Since, the above graph shows only a part, there is a deviation seen even in the good company, TCS. However, the chi square value of the good company, TCS,is less as compared to the cut-off.
Figures 8 and 9 illustrate number duplication test of TCS and First Leasing Company of India respectively.

Figure 8: Number duplication test of TCS



	expense amount	Count
1	expense amount	Count
2	0.02	3
3	17.75	2
4	0.04	1
5	0.06	1
6	0.36	1
7	0.51	1
8	0.52	1

Figure 9: Number duplication test of First Leasing Company of India



	Row Labels	Count of expense
1	Row Labels	Count of expense
2	0.08	15
3	0.09	12
4	0.02	6
5	0.33	5
6	0.05	4
7	0.12	4
8	0.21	4
9	0.38	4
10	0.42	4
11	0.46	4
12	0.65	4
13	0.13	3
14	0.24	3

Observation: In case of TCS, we see that the numbers are not being duplicated. However in First Leasing Company of India, we see duplication of numbers like 0.08 being repeated 15 times and 0.09 being repeated 12 numbers of times. Also we can see that odd number have occurred unusually often forFirst Leasing Company of India This can be used to understand whether there is abnormal duplication of numbers or not, or in other words whether the spikes seen in the above mentioned tests is due to “TYPE I” error or “TYPE II” error or not.Thus, number duplication test can be used as an additional test to identify if the numbers have been made up or are randomly occurring.

V. Conclusion

Every time the actual frequency on the income and expense side was breaching the expected Benford frequency, the score was increased by one.Based on the observations, a cut-off for the score was decided:

5.1 For the listed companies with 6 years data:

- Less than 5:
  A cut-off score of less than 5 means low probability of the numbers being manipulated.
- Between 5-7:
  A score between 5 and 7 is considered as grey.
- Above 7:
  High possibility of the numbers being cooked up, i.e. high possibility of **“*accounting fraud”.***

5.2 For the listed companies with 3 years data:

- Less than 3:
  A cut-off score of less than 3 means low probability of the numbers being manipulated.
- Between 3-5:
  A score between 3 and 5 is considered as grey.
- Above 5:
  High possibility of the numbers being cooked up, i.e. high possibility of **“*accounting fraud*”.**

References

[1]. Benford, F., "The law of anomalous numbers," Proceedings of American Philosophical Society, pp. 78, 551-572, 1938
[2]. Durtschi, C., Hillison, W., and Pacini, C, “The Effective use of Benford’s Law in Detecting Fraud in Accounting Data, “ Journal of Forensic Accounting, (5): 17-33,2004
[3]. Etteridge, M.L., and Srivastava, R.P., “Using digital analysis to enhance data integrity,” Issues in Accounting Education, 14(4): 675-690, 1999
[4]. Hill, T.P, “A statistical derivation of the significant-digit law," Statistical Science, pp. 10, 354-363, 1995
[5]. Hill, T. P, "The difficulty of faking data," Chance 26, 8-13. 1999
[6]. Necomb, S., "Note on the frequency of use of different digits in natural number," American Journal of Mathematics, pp. 4, 39-40, 1881
[7]. Nigrini, M.J., "The peculiar patterns of first digits," IEE Potentials, pp. 18, 24-27, 1999
[8]. Nigrini, M. J., “I’ve Got Your Number,” Journal of Accountancy, May 1999
[9]. Nigrini, M. J.,Forensic Analytics, Methods and Techniques for Forensic Accounting Investigations, Wiley, Hoboken, NJ, pp 480, 2011
[10]. Pinkham, R., "On the distribution of the first significant digits," Annals of Mathematical Statistic, pp. 32, 1223-1230, 1961
[11]. Varian, H., "Benford's Law (Letters to the Editor)," The American Statistician,26 (3): 65, 1972
[12]. Wallace, W,“Assessing the quality of data used for benchmarking and decision making,” Journal of Government Financial Management,(Fall) 51 (3): 16-22, 2002