USING BENFORD'S LAW TO THE DETECTION OF MISREPRESENTATION OF FINANCIAL STATEMENTS DATA

Radoslav Tušan[1]
[1]*Department of Finance, Faculty of Economics, Technical University of Košice, Košice, Slovak Republic*
Radoslav.Tusan@tuke.sk

Abstract: *Benford's Law is in certain situations, an important instrument to detect deliberate misrepresentation in the financial statements. Benford's Law is empirically derived and verified rule that the numerical data beginning with the digit by any of the groups 1, 2, ...9 appear in the files of large size with a certain probability. The likelihood of occurrence of numerical data in large data files according to the first numbers varies considerably. Numbers beginning with a digit 1 occur with a probability of 30.1%, numbers beginning with a digit 9 appear in this set with a probability of 4.58%. These values can be incorporated into the control procedures for assessing the quality of the reference data set. Various foreign authors have pointed out the possibility of using Benford's Law in examining the quality of primary data in the field of accounting and reporting. The authors investigated the differences between the empirical and theoretical frequency of occurrence of each number beginning with individual digits in financial reporting. Benford's analysis according them is more reliable when applied to the entire set of data. The intention of the article is to point out the possible misrepresentation in the financial statements of the registered entity whose shares are traded on the domestic stock exchange. One of the reasons for the detection of misrepresentation is that investors trust the financial statements of these entities. The research question was whether Benford's Law reveals the possible manipulation of financial statements data. The source data were financial statements of the entity for three years: balance sheet, income statement and cash flow statement. The number of each leading digit was counted and the actual frequency of leading digits to the expected distribution according to Benford's Law was compared. Potential manipulation was calculated by means of Kolmogorov-Smirnov statistics. Results of the analysis are mainly addressed to users of information of financial statements and confirmed the importance of Benford's Law in this meaning.*

Keywords: Benford's Law; reporting; financial statements.

JEL classification: C55; M41.

1. Introduction

In 1881, Simon Newcomb, an astronomer and mathematician, figured out that more numbers exist which begin with the digit one than with larger numbers. Newcomb determined that the probability that a number has a first digit, “d”, is given by:

$$Probabiliy\ (d) = Log_{10}\ \left(1 + \left(\frac{1}{d}\right)\right), for\ d = 1,2, \ldots, 9 \qquad (1)$$

In 1938, Frank Benford, a physicist, found that a large number of naturally-occurring datasets follow this pattern. He found that a pattern with low digits occur more frequently in the first position than larger digits. The first digit law discovered by Newcomb and later by Benford was called as "Benford's Law". This law gives information about the frequency of the appearance of each number as the first significant digit in a surprisingly large number of different data sets (Torres et al., 2007). Benford's Law is seeing increasing use as a diagnostic tool for isolating pockets of large datasets with irregularities that deserve closer inspection (Cho and Gaines, 2012). A comparative analysis of the bootstrap versus traditional statistical procedures applied to digital analysis based on Benford's Law realised Suh and Headrick (2011). They concluded that digital analysis based on Benford's Law is applicable in the area of reported annual earnings of S&P 1500 companies and Federal Election Commission data for example. These authors confirmed using this method over different periods of time and across small and large financial data sets. Certain sets of accounting numbers often appear to closely follow a Benford distribution (Boyle, 1994). The first researcher to apply Benford's Law extensively to accounting data with the aim to detect fraud was Mark J. Nigrini. His work used digital analysis to help identify tax evaders and to help auditors perform tests on sets of accounting numbers (Nigrini 1996; Nigrini and Mittermaier, 1997). Accounting data are mainly the result of mathematical procedures. Each accounting transaction stems out from mathematical combination of some numbers (e.g. number of items, price per item, asset value, liability value, equity). Most accounting data are expected to conform to a Benford distribution, and will be appropriate for digital analysis (Hill, 1995; Nigrini and Mittermaier, 1997). Assets accounts, liability accounts, residual equity item, revenue and expenses accounts are expected to conform. Results from Benford analysis are more reliable if the entire account is analysed rather than sampling the account. The larger number of accounting transactions or accounting data in the data set makes the analysis more precise. On the other side, some populations of accounting-related data do not conform to a Benford distribution (Durtschi et al., 2004). These are assigned numbers as check numbers, invoice numbers, purchase and payment orders, and human thought influenced numbers. Assigned numbers follow a uniform distribution than a Benford distribution. Since the late years of nineties occur to improving control tools in the field of accounting and auditing. Forensic accounting is intensively developing as a separate component of the methodology of accounting and auditing. Forensic accounting uses advanced mathematical-statistical methods and their application in solving specific fraud cases. From the audit practice in the United States are known processes that use the data quality evaluation methodology based on Benford's Law. Despite this fact, however, exact instruments are not considered to be generally applicable and the results obtained by them must always underlie to an analysis of content and formal characteristics of the surveyed sample (Durtschi et al., 2004).

Table 1 provides detailed information about items from the accounting is suitable for Benford's analysis.

Table 1: Differentiation of applicability of Benford's analysis by type of data

Suitable type of data
Set of numbers resulting from mathematical combinations: accounts receivables, liability accounts
Data from complex transactions: expenses, revenues
Large datasets: transactions throughout the year
Data files for which the mean value significantly different from the median: most sets of accounting data
Unsuitable type of data
Subset of data from a known set of data: numbers of invoices, orders
Numbers that are influenced human thinking: psychological prices
Accounts with a large number of specific numbers: account recorded a compensation from patients
Accounts for which are set maximum and minimum values: classification in assets

Source: Durtschi et al., 2004.

Probability of occurrence of the digits 1-9 in large data sets in the first place is following:

Table 2: Differentiation of applicability of Benford's analysis by type of data

Leading digit	1	2	3	4	5	6	7	8	9
Frequency	30,1%	17,6%	12,5%	9,7%	7,9%	6,7%	5,8%	5,1%	4,6%

Source: Durtschi et al., 2004.

Actual frequency of leading digits was compared with Benford's distribution of theoretical frequency and analysed by mean of Kolmogorov-Smirnov (KS) statistic. The Kolmogorov–Smirnov test (KS test) is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution, or to compare two samples (Hazewinkel, 2001). The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples:

$$KS = max\{|AD_1 - ED_1|, |(AD_1 + AD_2) - (ED_1 + ED_2)|, \dots, |(AD_1 + AD_2 + \dots + AD_9) - (ED_1 + ED_2 + \dots + ED_9)|\} \qquad (2)$$

where AD_i is actual frequency of the leading digits, and ED_i is the theoretical frequency of the leading digits.

Leading number is the first number of the absolute value of the accounting data. By comparing of cumulative values of actual and expected distribution was detected maximum difference in values of distribution (KS). Potential manipulation is recognised if KS > 1,36/√P, where P is total number of leading digits used.

2. Research and results

The research was started with assumption, that Benford's Law may be used to detect misreporting in financial statements of a company. Registered company was selected for the analysis due to the fact that investors have trust in its financial statements. Used source data were financial statements of the company for three years: balance sheet, income statement and cash flow statement. The number of each leading digit in the financial statements was counted and the actual frequency of leading digits was compared to the expected distribution according to Benford's Law. Potential manipulation in financial statements was calculated by means of Kolmogorov-Smirnov statistics. Financial statements of the registered company are in tables below. Leading digit calculation and Benford's distribution with Kolmogorov–Smirnov test are below financial statements.

Table 3: Balance Sheet, data in thousands of EUR

Data Year	2012	2013	2014
Cash	66	531	100
Receivables	3 973	2 452	2 945
Inventories	2 992	2 303	2 493
Other current assets	2	1	0
Total current assets	7 033	5 287	5 538
PPE Gross	26 834	29 304	30 907
Accumulated depreciation	18 532	20 503	21 332
PPE Net	8 302	8 801	9 575
Other assets	7	12	20
Total assets	15 342	14 100	15 133
Accounts payable	1 529	1 091	1 809
Current debt	1 591	1 445	1 890
Tax payable	158	175	128
Total current liabilities	3 278	2 711	3 827
Long term debt	4 114	3 543	3 996
Other liabilities	214	195	188
Total noncurrent liabilities	4 328	3 738	4 184
Retained earnings	4 444	4 353	3 824
Treasury stock	3 292	3 298	3 298
Equity	7 736	7 651	7 122

Source: financial statements of the company

Remark: "PPE" means Property, Plant and Equipment.

Table 4: Income Statement and Cash Flow Statement, data in thousands of EUR

Data Year	**2012**	**2013**	**2014**
Income statement			
Revenue	22 297	20 916	18 377
COGS	884	637	537
Gross profit	21 413	20 279	17 840
SGA	20 453	19 344	17 152
EBITDA	960	935	688
Depreciation	1 013	961	1 032
EBIT	-53	-26	-344
Interest	-63	-31	-52
Pre-tax income	-116	-57	-396
Income taxes	-33	-29	-85
Net income	-149	-86	-481
Cash flow statement			
Depreciation	1 013	961	1 032
Net cash flow from operating activities	671	1 955	311
Capital expenditures	-100	-1 456	-755
Net cash flow from investing activities	-89	-1 484	-721
Reduction of debt	-5	-5	-1
Dividends outflow	-140	-1	0
Net cash flow from financing activities	-145	-6	-21
Cash and cash equivalents +/-	437	465	-431

Source: financial statements of the company

Remark: “SGA” means Selling, General and administrative Expense.
“EBITDA” mean Earnings before Interest, Tax, Depreciation and Amortization.
“EBIT” mean Earnings before Interest.
“COGS” means Cost of goods sold.

Table 5: Leading digit calculation from Balance Sheet

Data Year	2012	2013	2014
Cash	6	5	1
Receivables	3	2	2
Inventories	2	2	2
Other current assets	2	1	0
Total current assets	7	5	5
PPE Gross	2	2	3
Accumulated depreciation	1	2	2
PPE Net	8	8	9
Other assets	7	1	2
Total assets	1	1	1
Accounts payable	1	1	1
Current debt	1	1	1
Tax payable	1	1	1
Total current liabilities	3	2	3
Long term debt	4	3	3
Other liabilities	2	1	1
Total noncurrent liabilities	4	3	4
Retained earnings	4	4	3
Treasury stock	3	3	3
Equity	7	7	7

Source: own elaboration

Table 6: Leading digit calculation from Income Statement

Data Year	2012	2013	2014
Revenue	2	2	1
COGS	8	6	5
Gross profit	2	2	1
SGA	2	1	1
EBITDA	9	9	6
Depreciation	1	9	1
EBIT	5	2	3
Interest	6	3	5
Pre tax income	1	5	3
Income taxes	3	2	8
Net income	1	8	4

Source: own elaboration

Table 7: Leading digit calculation from Cash flow statement

Data year	2012	2013	2014
Depreciation	1	9	1
Net cash flow from operating activities	6	1	3
Capital expenditures	1	1	7
Net cash flow from investing activities	8	1	7
Reduction of debt	5	5	1
Dividends outflow	1	1	0
Net cash flow from financing activities	1	6	2
Cash and cash equivalents +/-	4	4	4

Source: own elaboration

Table 8: Benford's distribution and Kolmogorov–Smirnov test

Leading Digit	Count	Actual Distribution	Expected Distribution	Cumulative Difference
1	37	31,4%	30,1%	1,3%
2	21	17,8%	17,6%	1,4%
3	16	13,6%	12,5%	2,5%
4	10	8,5%	9,7%	1,3%
5	9	7,6%	7,9%	1,0%
6	6	5,1%	6,7%	0,6%
7	7	5,9%	5,8%	0,5%
8	7	5,9%	5,1%	0,3%
9	5	4,2%	4,6%	0,0%
Total	118		KS	2,5%
			Cut-off	12,5%

Source: own elaboration

Maximum cumulative difference between actual and expected distribution is 2,5% (Kolmogorov–Smirnov statistic). This value is less than “cut-off” value which is 12,5% (1,36/√P, where P is total number of leading digits). Test result leads to the conclusion that the company did not manipulate the data in its financial statements.

Figure 1 point to tightness of the actual distribution and the expected distribution of leading digits according to Benford's Law.

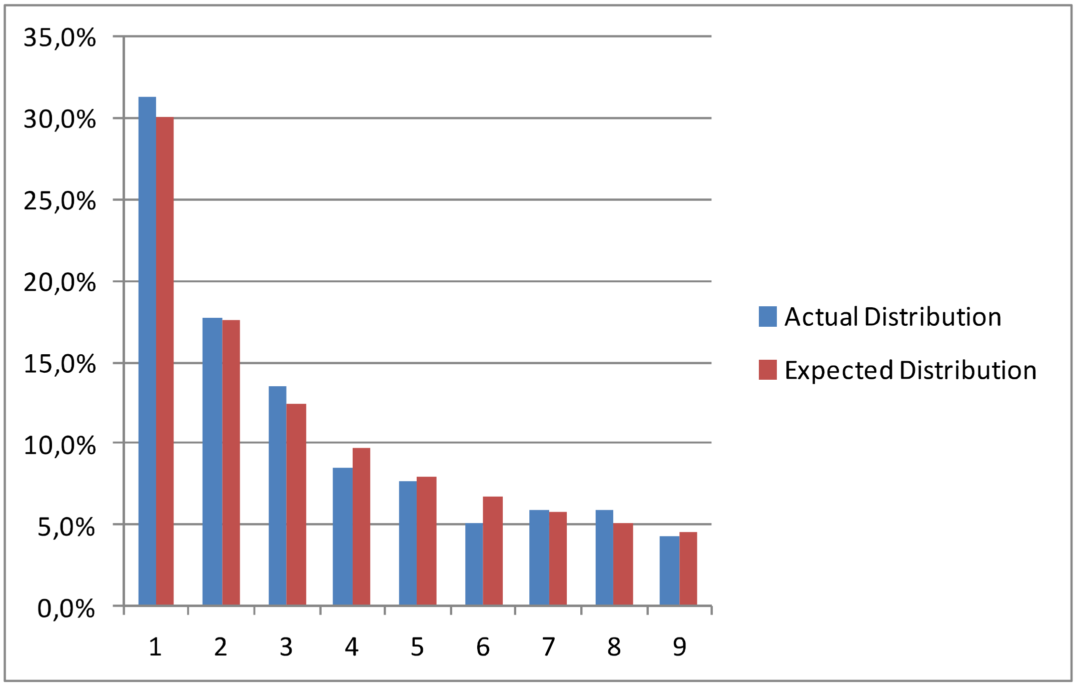

Figure 1: Actual and expected distribution of leading digits.
Source: own elaboration

3. Conclusion

Investors and lenders rely on the accuracy of data in the financial statements of companies listed on the Stock Exchange. Use of Benford's analysis helps the auditors to identify of manipulation in the financial statements. The aim of the article was to point out the possible misrepresentation in the financial statements of the listed entity whose shares are traded on the Stock Exchange. The research question was whether Benford's Law reveals the possible manipulation of financial statements data. The source data used for analysis were financial statements of the listed entity for three years: balance sheet, income statement and cash flow statement. The number of each leading digit was counted and the actual frequency of leading digits was compared to the expected distribution according to Benford's Law. Potential manipulation was calculated by means of Kolmogorov-Smirnov statistics. Maximum cumulative difference between actual and expected distribution was 2,5% (Kolmogorov–Smirnov statistic). This value was less than “cut-off” value 12,5%. Test result leads to the conclusion that the company did not manipulate the data in its financial statements.

4. Acknowledgement

This contribution has been elaborated within the project VEGA 1/0967/15: Approaches for fiscal imbalance solution in terms of the EU and in the context of the systemic crisis.

References

Benford, F. (1938) “The Law of Anomalous Numbers”. *Proceedings of the American Philosophical Society*, Vol. 78, No. 4 (Mar. 31, 1938), pp. 551-572.

Boyle, J. (1994) “An Application of Fourier Series to the Most Significant Digit Problem". *The American Mathematical Monthly,* Vol. 101, No. 9, pp. 879-886. doi: 10.2307/2975136

Cho, W.K.T. and Gaines, B.J. (2007) "Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance", *The American Statistician,* Vol. 61, Issue (3): pp. 218–223. doi:10.1198/000313007X223496.

Durtschi, C., Hillison, W. and Pacini C. (2004) "The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data". *Journal of Forensic Accounting,* Vol. 5, pp. 17–34.

Hazewinkel, M. (ed.) (2001) *Kolmogorov–Smirnov test*, Encyclopedia of Mathematics, Springer.

Hill, T.P. (1995). “A Statistical Derivation of the Significant-Digit Law”. *Statistical Science*, Vol. 10, No. 4, pp. 354-363

MPC Cessi (2016) Financial statements online. Available: http://www.registeruz.sk/cruz-public/domain/accountingentity/show/148662

Nigrini, M. (1996) "A Taxpayer Compliance Application of Benford's Law". *Journal of the American Taxation Association*, Vol. 18, No 1, pp. 72–91.

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 & Theory,* Vol. 16, No. 2, pp. 52-67.

Suh, I.S. and Headrick, T.C. (2010) "A Comparative Analysis of the Bootstrap versus Traditional Statistical Procedures Applied to Digital Analysis Based on Benford's Law". *Journal of Forensic and Investigative Accounting,* Vol. 2, Issue 2, pp. 144–175.

Suh, I. S., Headrick, T.C. and Minaburo, S. (2011) "An Effective and Efficient Analytic Technique: A Bootstrap Regression Procedure and Benford's Law". *Journal of Forensic & Investigative Accounting*, Vol. 3, Issue 3, pp. 1-22.

Torres, J., Ferndandez, S., Gamero, A. and Sola, A. (2007) “How Do Numbers Begin? (The First Digit Law)”. *European Journal of Physics*, Vol. 28, No. 3. L17. http://dx.doi.org/10.1088/0143-0807/28/3/N04