

Henselmann, Klaus; Scherr, Elisabeth; Ditter, Dominik

Working Paper

Applying Benford's Law to individual financial reports: An empirical investigation on the basis of SEC XBRL filings

Working Papers in Accounting Valuation Auditing, No. 2012-1 [rev.]

Provided in Cooperation with:

Friedrich-Alexander University Erlangen-Nuremberg, Chair of Accounting and Auditing

Suggested Citation: Henselmann, Klaus; Scherr, Elisabeth; Ditter, Dominik (2013) : Applying Benford's Law to individual financial reports: An empirical investigation on the basis of SEC XBRL filings, Working Papers in Accounting Valuation Auditing, No. 2012-1 [rev.], Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Rechnungswesen und Prüfungswesen, Nürnberg

This Version is available at:

<https://hdl.handle.net/10419/88418>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Working Papers in Accounting Valuation Auditing Nr. 2012-1

Klaus Henselmann / Elisabeth Scherr / Dominik Ditter

Applying Benford's Law to individual financial reports
An empirical investigation on the basis of SEC XBRL filings

**Friedrich-Alexander-Universität
Erlangen-Nürnberg**



Lehrstuhl für
Rechnungswesen
und Prüfungswesen

Applying Benford's Law to individual financial reports

An empirical investigation on the basis of SEC XBRL filings

Working Papers in Accounting Valuation Auditing Nr. 2012-1
www.pw.wiso.uni-erlangen.de

Klaus Henselmann^{*} / Elisabeth Scherr^{} / Dominik Ditter^{***}**

Authors: * Prof. Dr. Klaus Henselmann, ** Dipl.-Kff. Elisabeth Scherr, *** Dipl.-Kfm. Dominik Ditter, Friedrich-Alexander University of Erlangen-Nürnberg, School of Business and Economics, Department of Accounting and Auditing, Lange Gasse 20, 90403 Nürnberg, GERMANY, Phone: +49 911 5302 437, Fax: +49 911 5302 401, klaus.henselmann@wiso.uni-erlangen.de, elisabeth.scherr@wiso.uni-erlangen.de, dominik.ditter@wiso.uni-erlangen.de

Keywords: Benford's Law, Digit Analysis, XBRL, Interactive Data, SEC filings, Financial Statement Analysis, Forensic Accounting

Abstract: The business reporting standard XBRL offers the opportunity to easily extract and analyze a sufficient number of monetary items out of single annual reports for statistical analysis purposes. Using XBRL 10-K reports filed with the SEC EDGAR system, we derive first digit distributions for single companies and measure the deviation from the Benford distribution. On average, we find that for all monetary numbers that are contained in the examined XBRL reports, the first digit distribution follows Benford's Law. Furthermore, our results indicate several line items with an abnormal digit frequency potentially indicating human interaction. Taken together, the empirical results suggest that the application of Benford's Law to financial reports might be a useful analytical tool for investors. The findings also may be of interest to the SEC for planning enforcement actions, as digit analysis could be an appropriate means of identifying suspect line items carrying a higher level of risk.

Titel: Anwendung des Benford'schen Gesetzes auf Jahresabschlussdaten: Eine empirische Untersuchung auf der Basis von SEC XBRL Veröffentlichungen

Schlagwörter: Benford'sches Gesetz, Ziffernanalyse, XBRL, Interactive Data, SEC Veröffentlichungen, Bilanzanalyse, Bilanzfälschung

JEL Classification: M40, C12, C46, C81

Revised Version: December 2013

Contents

Contents..... 2

1 Introduction 3

2 Review of Related Literature 5

3 Research Design..... 7

 3.1 Sample Selection and Data Collection..... 7

 3.2 Testing Methodology 9

4 Empirical Findings 12

 4.1 Benford’s Law Applied to XBRL Accounting Data..... 12

 4.2 Suspect XBRL Line Items 14

5 Conclusions 16

Notes..... 18

Appendix 19

References 21

1 Introduction

On January 30, 2009 the Securities and Exchange Commission (SEC) released a rule mandating substantially all firms subject to SEC filing rules to file interactive data (or XBRL data) with the EDGAR (Electronic Data Gathering, Analysis, and Retrieval) system over a three-year phase-in (SEC 2009). The SEC asserts that *'interactive data can create new ways for investors, analysts, and others to retrieve and use financial information in documents filed with us'* (SEC 2009, 8). The reason is that XBRL (eXtensible Business Reporting Language) documents are structured. Every reporting element is encoded by specific 'tags' from an XBRL taxonomy, so that all types of items (e.g., monetary numbers, nonmonetary numbers, date values, text, and diagrams) contained in financial reports can be automatically derived and analyzed.

Consequently, these 'digital dictionaries' enable to extract a large number of monetary items contained in financial reports for meaningful statistical analysis in detecting manipulations or errors in accounting data. A possible analytical procedure could be digit analysis or Benford's Law (Newcomb 1881; Benford 1938). The idea behind is that for many random datasets higher deviations from the expected digit distribution according to Benford's Law might indicate a higher likelihood of human interaction. It is assumed that digits of data that are reported in accordance with the true and fair view principle will conform to Benford's Law (Nigrini 1996). Today, the application of Benford's Law in accounting data is recommended mainly for ledger data (Drake and Nigrini 2000; Durtschi et al. 2004) and is commonly used by (tax) auditors to detect fraudulent data.¹ Applying Benford's Law to financial reporting data is increasingly being discussed (Quick and Wolz 2003; Watrin et al. 2008; Nigrini 2011; Nigrini 2012). XBRL filings submitted under the SEC XBRL filing mandate provide an apt setting for building upon prior literature and enhancing knowledge on the issue.

In this paper, we analyze monetary items contained in SEC XBRL 10-K filings for conformity to Benford's Law and provide implications for future research. On the basis of about 5 million extracted digits, we derive digit distributions across line items on firm-level and compute the observed deviation from Benford's Law for each firm-year observation. Consistent with prior literature (Quick and Wolz 2003; Watrin et al. 2008), we find that, on average, the extracted monetary XBRL data follows almost exactly Benford's Law. Furthermore, we extract the most frequently reported monetary taxonomy positions and identify several items with conspicuous digit distributions. Interestingly, the item *Revenues* that prior studies found to be

the account type being most frequently affected by fraudulent transactions (Dechow et al. 1996, 10; Leng et al. 2011, 819) is among the items with the highest deviation from Benford's Law.

This study provides several insights into the usability of SEC XBRL filings. Because the XBRL mandate is still in its infancy, there is limited knowledge of new analysis opportunities that abound. We add to the XBRL literature by exemplifying digit analysis as a possible new analysis possibility arising through technological advances. Prior studies (e.g., Quick and Wolz 2003; Watrin et al. 2008; Nigrini 2011; Nigrini 2012) provide early evidence on applying Benford's Law across financial statement data. We contribute to these studies by automatically extracting all monetary line items from every single examined XBRL 10-K report and computing digit distributions across line items on firm-level. Because of the XBRL functionality, this study is able to analyze a multiple number of accounting data. Our findings provide large sample evidence of Benford's Law conformity across firm-years. Furthermore, we point out the possible usability of digit analysis to select account items as basis for the enforcement audit process.

The study proceeds as follows. The next section gives an overview of related literature on applying Benford's Law in the area of accounting, auditing and taxation. The empirical investigation in the third section is the main part of our work. We describe the research design of our study before we give a detailed explanation and discussion of our findings. The final section concludes the paper with a short summary and implications for future research.

2 Review of Related Literature

Prior studies that use abnormal digit frequencies as an indicator for human interaction in the area of accounting, auditing and taxation can be classified into two main lines of literature (see Table 1).

The first line of literature uses deviations from Benford's Law calculated for specific account items across companies as an indicator of earnings management behavior. In particular, Carslaw (1988) examines the frequency of occurrence of second digits in income numbers reported in the annual reports of 220 New Zealand listed companies and finds an abnormally high frequency of the digit zero and an abnormally low occurrence of the number nine. This supports the existence of rounding up behavior for earnings when they are just below reference points (e.g., investors may perceive earnings of \$1,900,000 being much lower than \$2,000,000). Thomas (1989) detect similar patterns for U.S. firms experiencing annual profits, whereas loss firms exhibit opposite patterns (fewer zeros and more nines). The results for positive quarterly earnings and the right-most digits of positive earnings per share data are similar to positive annual earnings but more pronounced for per share data. Niskanen and Keloharju (2000) observe rounding up behavior for Finnish companies' net income and Van Caneghem (2002) for the positive pre-tax income of British companies. Kinnunen and Koskela (2003) take a broader view and examine net income observations from almost 22,000 firms in 18 countries. Consistent with previous studies, they discover an upward rounding of net profits as well as a reversed pattern of earnings management for net losses. Das and Zhang (2003) find rounding behavior in both earnings per share before extraordinary items and net income per share. Skousen et al. (2004) begin with analyzing the earnings of Japanese firms similar to previous studies. Besides observations consistent with prior literature, they find that the probability of earnings management decreases with the distance of earnings numbers from the next reference point. In addition to other tests, Nigrini (2005) studies Enron's revenue numbers and earnings per share data between 1997 and 2000. Most numbers are scarcely above cognitive reference points suggesting cosmetic earnings management. On the basis of 24 randomly selected publicly traded companies, Johnson (2009) uses Benford's Law to identify three firm characteristics that presumably are associated with a higher risk of earnings management (low market capitalization, high insider trading and new publicly traded companies). And finally, based on a time series analysis of rounding behavior with respect to positive income numbers of publicly traded companies before and after the Sarbanes Oxley

Act of 2002, Jordan and Clark (2011) conclude that cosmetic earnings management apparently has declined after the Sarbanes Oxley Act.

Table 1: Prior Research on Abnormal Digit Frequencies in the Area of Accounting, Auditing and Taxation

Author(s)	Year	Object of research	Country
<i>Aim of the Study: Detecting Earnings Management</i>			
Carslaw	1988	net income, and ordinary income (second digit*)	New Zealand
Thomas	1989	annual and quarterly earnings before extraordinary items and discontinued operations (second digit*), and annual and quarterly earnings per share before extraordinary items and discontinued operations (second, and right-most digit*)	USA
Niskanen and Keloharju	2000	net income (second digit*)	Finland
Van Caneghem	2002	operating income, net income before extraordinary items, earnings before interest and taxes, pre-tax income, and net income (second digit*)	Great Britain
Kinnunen and Koskela	2003	net income (second digit*)	Worldwide
Das and Zhang	2003	quarterly earnings per share before extraordinary items, and quarterly net income per share (first digit after the decimal point*)	USA
Skousen et al.	2004	net income (first, second, third, fourth, and fifth digit*)	Japan
Nigrini	2005	revenues, and earnings per share (first, second, and right-most digit*)	USA
Johnson	2009	quarterly net income, and quarterly earnings per share (first digit*)	USA
Jordan and Clark	2011	income (second digit*)	USA
<i>Aim of the Study: Detecting Fraud</i>			
Nigrini	1994	fraudulent payroll check amounts (first-two digits*)	USA
Nigrini	1996	income tax return data (first, and second digit*)	USA

Notes:

This table provides an overview of literature with regard to the application of abnormal digit frequencies in the area of accounting, auditing and taxation. We generally allocate the studies according to two different aims of the study: detecting earnings management, and detecting fraud. * denote the primary digit position that is examined in the respective studies.

The second line of literature assumes that the cause of the deviation from Benford's Law is fraudulent human behavior exceeding legal boundaries, but is also based on single account items. Nigrini (1994) was the first to show that Benford's Law could be used as an indicator of fraud in the area of accounting, auditing and taxation. He analyzes the first-two digits of payroll data known to be fraudulent and finds that the fraudulent numbers deviate significantly from Benford's Law. Later, he applied the method of Benford's Law with the aim of detecting tax evasion by analyzing tax returns on the U.S (Nigrini 1996). Internal

Revenue Service Individual Tax Model Files. The results indicate that expected digit distributions could help to uncover unplanned tax evasion.

Beyond that, Quick and Wolz (2003) analyze annual reports of German listed corporations for the time period 1994-1998 (1,820 annual reports in total). The results allow them to conclude that the analyzed accounting data (line items of the Statement of Financial Position and the Statement of Income) across all examined annual reports follows Benford's Law. There is also evidence that in a set of simulation-based non-manipulated and experimental-based manipulated financial statement data, Benford's Law deviant financial statements show a higher proportion of manipulation for tax evasion purposes (Watrin et al. 2008). Nigrini (2012, 200) demonstrates Benford's Law conformity for 25 selected line items. Besides, insights into the implications of applying Benford's Law on a set of financial statement line items are still limited.

3 Research Design

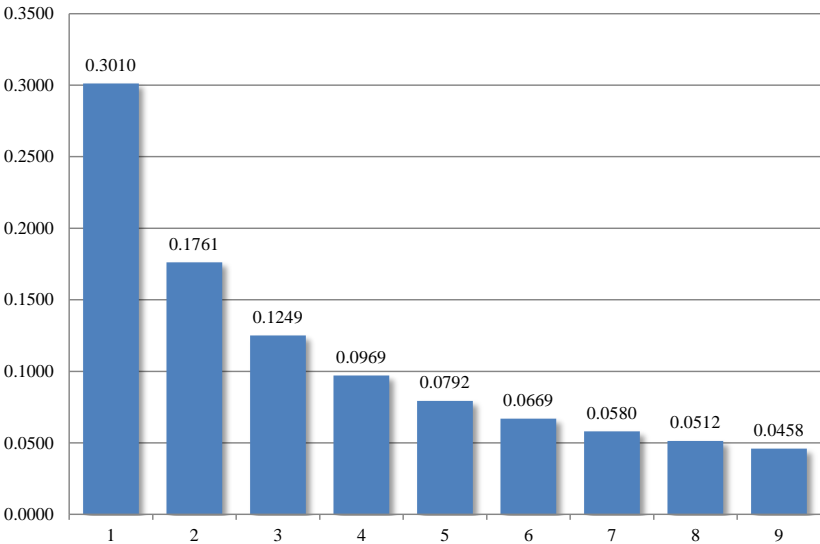
3.1 Sample Selection and Data Collection

This study analyzes digit frequencies for a sample that is based on all available SEC XBRL 10-K reports for the fiscal year 2012. We are focusing on the XBRL filings for the fiscal year 2012 because of the three-year phase-in (SEC 2009). First, substantially all firms that use U.S. GAAP had to comply with the SEC interactive data mandate by 2012, which allows us to analyze a comprehensive database. Second, for most firms the fiscal year 2012 represents at least the second year of XBRL filing, so that the respective XBRL filings are subject to detailed tagging requirements. For the first individual filing year, the SEC required only the financial statement data and not the footnote disclosure to be tagged in detail (SEC 2009, 23). From the initial 71,615 available XBRL filings submitted to the SEC between July 2009 and March 2013², we eliminate 56 firm-year observations with errors contained in the XBRL reports not enabling technical processing, as well as all duplicates and XBRL documents other than 10-K (57,616). Furthermore, we exclude all firm-years where the fiscal year (as indicated by the XBRL tag *DocumentFiscalYearFocus*) is not 2012 (8,161). The final sample consists of 5,782 firm-year observations.

The taxonomy definitions for each XBRL reported item enables the exact identification of all monetary line items. Using this taxonomy structure we extract all monetary numbers

contained in each of the examined XBRL 10-K reports and disregard all numbers which would distort the results (e.g., date specifications such as ‘2011-09-30’ and period specifications such as ‘6M’ or document identification numbers such as ‘10-K’). XBRL tags are unique in a specific context, so that monetary disclosures that are reported at different places in a financial report can be easily identified. Consequently, we regard every XBRL item only once, because otherwise the results may be biased (see Nigrini 2012, 221). As small amounts are seen to be immaterial, we also exclude all monetary numbers that have values within the spread \$-10 to \$10 (Drake and Nigrini 2000, 133). Individual XBRL tags that are not part of the standard XBRL taxonomy (i.e., XBRL taxonomy extensions) can obviate the full automatic extraction of specific target reporting contents (see Debreceeny et al. 2011). However, XBRL-based digit analysis has the advantage that these customized tags can be considered in the analysis. Therefore, numbers which are tagged using XBRL taxonomy extensions are explicitly included in the dataset. Note that we do not restrict our analysis to specific industries. Or, more technically speaking, we do not distinguish between different entrypoints (e.g., financials or commercial and industrial industry) and taxonomy versions.

Figure 1: Benford’s Law (First Digit)



Notes:
 There is a special digit distribution that can be observed for many random datasets: the so called Benford distribution or Benford’s Law (see Nigrini 2012, 6). This figure provides the distribution of the first digits as expected according to Benford’s Law (as indicated by the bars). The Y-axis represents the relative frequency of first digits $i = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. The X-axis represents the first digits i .

There exist different Benford distributions for each digit position in numbers (see Nigrini 2012, 6). For our study we focus on the first digit distribution (this logarithmic distribution is

shown in Figure 1), which is seen as an initial high-level test of reasonableness (Drake and Nigrini 2000, 133). Therefore, for all monetary numbers included in the XBRL 10-K reports, we initially determine the value of the digit in the first digit position and derive the frequency of values between one and nine. In total, our final dataset contains 4,566,310 examinable first digits.

Next, we test whether Benford's Law can be applied to our sample. Besides several qualitative factors that shall indicate a Benford dataset (Drake and Nigrini 2000, 132; Durtschi et al. 2004, 23) there are quantitative indicators suggested in prior literature. For example, Wallace (2002) recommends two quantitative indicators: (1) the mean of the number of observed digits is larger than the median, and (2) the skewness value is positive. As both indicators are positive, on average, we assume that we can apply Benford's Law to our sample.

3.2 Testing Methodology

The degree of deviation from the expected digit distribution according to Benford's Law can be assessed in several ways. We begin with tests that are commonly used in studies on Benford's Law: the Z-statistic and the Chi-square test-statistic.

(1) The *Z-statistic* measures the significance of the deviation from the expected digit distribution for each digit separately (digit-by-digit analysis; see Cleary and Thibodeau 2005) and questions, whether a single digit occurs more often or less often than it is expected according to Benford's Law. For a particular first digit ($d1$) the Z-statistic is computed as follows (Durtschi et al. 2004, 25):

$$Z(d1) = \frac{|p(o) - p(e)| \left(-\frac{1}{2n}\right)}{s(d1)} \quad (1)$$

Where: $p(o)$ is the observed proportion of digit frequency, $p(e)$ is the expected proportion of digit frequency according to Benford's Law, $s(d1)$ is the standard deviation for a particular first digit, and n is the number of observations. The continuity correction term $(1/2n)$ is only used when the term is smaller than the absolute value of the difference between the actual and the expected proportion of digit frequency.

The *standard deviation* for the expected proportion of a particular first digit is defined as:

$$s(d1) = \left(\frac{p(e)(1 - p(e))}{n} \right)^{\frac{1}{2}} \quad (2)$$

Where: $p(e)$ is the expected proportion of digit frequency according to Benford's Law, and n is the number of observations.

For a 5% level of significance (95% confidence interval), Z-statistics above 1.96 indicate that the observed frequency of occurrence of a specific digit significantly deviates from the expected digit frequency.

We also compute the 95% lower and upper tolerance limits for the expected distribution (*lower bound* and *upper bound*) in order to get a graphical representation of deviation over all digits (1 through 9). They are derived from equation (2) as follows (Nigrini and Mittermaier 1997, 59):

$$\text{Lower Bound} = p(e) - [1.96 * s(d1) - \left(\frac{1}{2n} \right)] \quad (3a)$$

$$\text{Upper Bound} = p(e) + [1.96 * s(d1) + \left(\frac{1}{2n} \right)] \quad (3b)$$

Where: $p(e)$ is the expected proportion of digit frequency according to Benford's Law, n is the number of observations, and $s(d1)$ is the standard deviation for a particular first digit.

(2) The *Chi-square test-statistic* compares the expected distribution over all digits (test-by-test analysis; see Cleary and Thibodeau 2005) with the observed distribution and indicates whether the observed distribution significantly differs from the expected distribution. It allows statements of the goodness-of-fit for all digits together and is determined through the following equation (Nigrini 2012, 153):

$$CHI^2 = \sum_{i=1}^N \frac{(n(o) - n(e))^2}{n(e)} \quad (4)$$

Where: $n(o)$ is the observed digit frequency, $n(e)$ is the expected digit frequency according to Benford's Law, and N is the number of possible digits i in the first position in numbers ($N=9$).

The numerator constitutes the residual of the compared distributions that are squared in order to get rid of negative values. The numerator is then divided by the expected number of

observations, in order to normalize bigger and smaller counts. As the conformity of the whole distribution shall be tested, the results are summed up for the digits $i = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Because we examine only the first digit, we have eight degrees of freedom ($df = 8$). We are choosing an alpha level of significance of $\alpha = 0.05$ (95% confidence interval). Accordingly, for a two sided probability of error, the critical value for rejection is 15.51. In other words, if the Chi-square value is greater than 15.51, the null hypothesis of equal distributions has to be rejected (Rafeld and Then Berg 2007, 30).

Both the Z-statistic and the Chi-square test-statistic have advantages and disadvantages. In order to avoid possible false classifications, literature suggests primarily using test-by-test analyses for signaling datasets suspicious of being manipulated, and in addition using digit-by-digit analyses for an in-depth investigation regarding possible causes for the abnormality (Cleary and Thibodeau 2005).

Note that the Chi-square test could be sensitive to the number of observations. For very large data sets already relatively small differences between the observed and expected proportions may signal a significant deviation from Benford's Law. This property is the excess power problem (Nigrini 2000, 7). Therefore, alternative test-statistics (e.g., the Mean Absolute Deviation) are suggested that presumably ignore the size of examined records.

(3) We use the *Mean Absolute Deviation (MAD)* as third conformity measure that is computed as the mean absolute difference between the observed proportion of digit frequency and the expected proportion of digit frequency based on Benford's Law (Drake and Nigrini 2000, 134):

$$MAD = \frac{1}{N} \sum_{i=1}^N |p(o) - p(e)| \quad (5)$$

Where: $p(o)$ is the observed proportion of digit frequency, $p(e)$ is the expected proportion of digit frequency according to Benford's Law, and N is the number of possible digits i in the first position in numbers.

Based on their research experience, Drake and Nigrini (2000) and Nigrini (2012) provide thresholds for testing the goodness-of-fit to the Benford distribution. However, one should be aware that there is no clear cut-off level for the MAD as compared to the traditional Chi-square test-statistic of equal distributions (Nigrini 2000, 8), so the results should be interpreted cautiously.

4 Empirical Findings

4.1 Benford's Law Applied to XBRL Accounting Data

Table 2 provides descriptive statistics of the extracted monetary digits in the first position in numbers. For each XBRL 10-K report we obtain the frequency of first digits and compute the respective degree of deviation from the Benford distribution.

Table 2: Descriptive Statistics/ Z-Statistic

	d1(1)	d1(2)	d1(3)	d1(4)	d1(5)	d1(6)	d1(7)	d1(8)	d1(9)	d1(1-9)
Total	1,381,455	811,514	569,549	441,398	366,713	301,751	262,686	227,457	203,787	4,566,310
Mean	238.96	140.38	98.52	76.35	63.43	52.20	45.44	39.35	35.25	789.88
Median	188.00	111.00	78.00	60.00	51.00	41.00	35.00	31.00	28.00	626.00
Std. Dev.	219.31	129.44	91.37	70.63	57.75	48.80	42.86	37.13	33.49	721.10
Z-statistic (mean)	1.2582	1.1791	1.1016	1.1130	1.1009	1.0353	1.0252	0.9812	0.9921	-
Z-statistic (median)	1.0472	0.9800	0.9201	0.9227	0.8901	0.8563	0.8504	0.8128	0.8202	-

Notes:

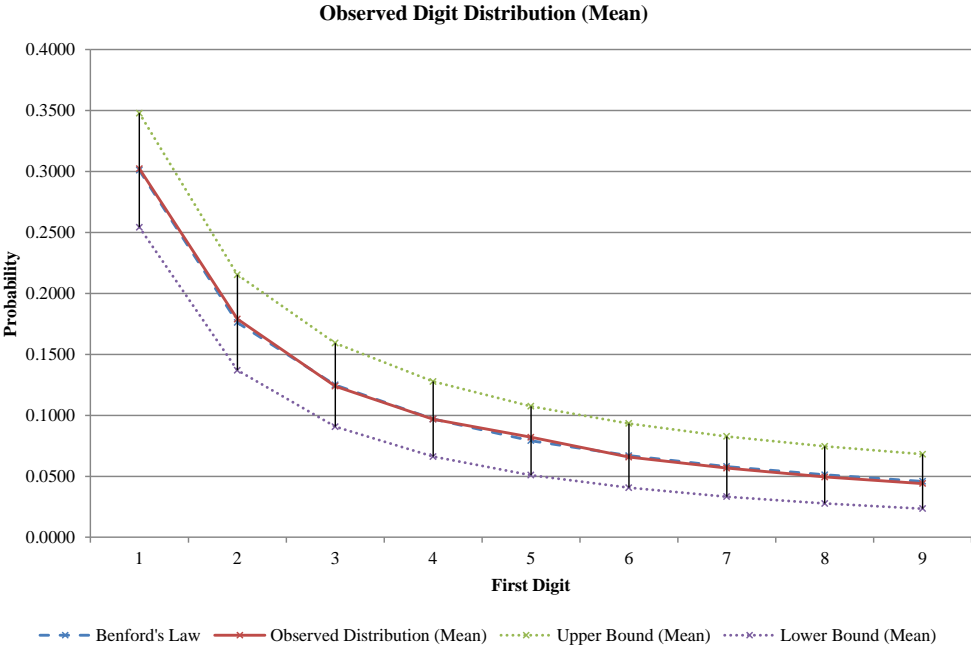
This table provides descriptive statistics about the extracted monetary digits in the first position in numbers $d1(i)$, where $i = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, that are contained in the examined 5,782 XBRL 10-K reports. It illustrates the distribution of examinable first digits as well as the mean and median number of observable first digits per firm-year. The Z-statistic is computed for each XBRL 10-K report using equation (1), respectively.

The Z-statistic does not signal a significant deviation from the expected distribution for any particular first digit. The mean observed distribution lies nearly perfect within the upper and lower bound and does not indicate any abnormalities (see Figure 2). The results indicate that the observed distribution follows Benford's Law almost exactly.

We can generally state conformity using the Chi-square value and the MAD. The median (15.08) of the computed Chi-square values is below the critical value of acceptance (based on 8 degrees of freedom and a 5 percent level of significance). The median of the MAD (0.0133) also follows Benford's Law based on the critical value of nonconformity (0.0150) as suggested by Nigrini (2012, 160). However, the mean Chi-square value (17.56) and the average MAD (0.0179) are slightly above the thresholds.

On the basis of all XBRL 10-K reports, in total, 2,783 (48.14%) financial reports have a Chi-square value above the critical value of acceptance (15.51). The results for the MAD are very similar. In total, 2,401 (41.53%) examined financial reports are above the suggested threshold of 0.0150. The Pearson correlation result between the Chi-square test-statistic and the MAD is 50.29% (significant at the 1% level).

Figure 2: Observed and Expected First Digit Distribution (Mean)



Notes:
 This figure provides the expected and the (average) observed first digit distribution for all examined 5,782 firm-year observations. The Y-axis represents the relative frequency of first digits $i = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. The X-axis represents the first digits i . The upper and lower bounds mark the limit of significant deviation at the 5% level.

Note that the size of the different datasets or in other words the overall number of records in each of the examined XBRL reports varies. Influencing factors for a higher number of reported monetary line items could be, for example, specific reporting requirements, firm size or voluntary disclosures. Interestingly, the sign (positive or negative) and extent (high or low) of the correlation between the test-statistics are different. Based on Pearson correlation results, the Chi-square test-statistic is positively correlated (3.14%, significant at the 5% level), whereas the MAD (-47.67%, significant at the 1% level) is negatively correlated with the number of examined digits. Therefore, when comparing different digit distributions on firm-level it is important to control for size effects. We also recommend using several test-statistics for measuring the goodness of fit to expected digit distributions such as Benford’s Law.

4.2 Suspect XBRL Line Items

XBRL enables to extract enough monetary numbers for computing a digit distribution on firm-level, which suggests many opportunities for risk assessment based on aggregated accounting data. However, it also offers the opportunity to select and analyze specific taxonomy positions. Consequently, in the following we apply Benford's Law on the most frequently XBRL line items contained in the 5,782 examined XBRL 10-K documents. On the basis of all reported items within the fiscal year 2012, we rank the monetary taxonomy positions based on their absolute observable occurrence. Afterwards, we focus on the 25 most frequent items, whereas we extract the fact values as described in section 3.1. Next, we obtain the first digit frequency between numeral one and nine for the first position in numbers.

All 25 XBRL line items are official US-GAAP taxonomy elements across industries. The line item with the highest observable occurrence of monetary digits is *StockholdersEquity* (76,219) which means that, on average, this taxonomy item is used about 13 times in each financial report (e.g., because of previous year values). For a detailed definition and documentation of the presented taxonomy positions please see the Appendix.

Table 3 provides the results of the computed test-statistic for measuring the degree of conformity to Benford's Law for each extracted line item, respectively. In total, 11 out of the examined 25 taxonomy items are signaling a significant deviation from the Benford deviation according to the Chi-square value. Following the MAD all positions signal conformity according to the suggested thresholds; however, the tendencies and suspect peaks are very similar to the Chi-square test-statistic. The ranking according to the Chi-square value and the MAD differs. Nevertheless, although the exact position in the ranking is different, a comparison of the distribution of the test results indicates similar tendencies. Furthermore, the same conclusions can be drawn by valuing the items according to the Chi-square test-statistic or the MAD. For example, the taxonomy position with the highest Chi-square value (see No. 17, Table 3), also signal the highest degree of deviation according to the MAD.

It is remarkable that particularly line items about the (accumulated) depreciation of real estate signal nonconformity to Benford's Law. But, also sales or revenues and available for sale securities indicate a high degree of deviation based on the expected Benford distribution. Interestingly, the most frequent type of fraud transactions according to the Accounting and Auditing Enforcement Releases (AAERs) of the SEC is concerning revenue recognition or overstatement of revenues (Dechow et al. 1996, 10; Leng et al. 2011, 819).

Table 3: Most Frequently Reported Monetary Line Items in XBRL 10-K reports

No.	XBRL Monetary Line Items (us-gaap)	n	CHI ²	MAD
1	StockholdersEquity	76.219	15,1476	0,0012
2	NetIncomeLoss	47.046	10,2200	0,0013
3	OperatingIncomeLoss	41.546	18,9464	0,0019
4	StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest	40.498	17,8684	0,0017
5	Assets	36.867	10,9068	0,0014
6	DefinedBenefitPlanFairValueOfPlanAssets	36.254	9,9962	0,0011
7	Revenues	35.265	26,5525	0,0021
8	SalesRevenueNet	32.616	24,0574	0,0024
9	PropertyPlantAndEquipmentGross	28.858	14,2534	0,0018
10	CashAndCashEquivalentsAtCarryingValue	27.006	6,5297	0,0014
11	RealEstateAndAccumulatedDepreciationInitialCostOfBuildingsAndImprovements	26.647	8,5302	0,0015
12	RealEstateAndAccumulatedDepreciationCarryingAmountOfLandAndBuildings AndImprovements	26.559	15,3051	0,0023
13	RealEstateAndAccumulatedDepreciationCarryingAmountOfBuildingsAnd Improvements	25.587	27,0156	0,0033
14	IncomeTaxExpenseBenefit	25.446	4,9499	0,0014
15	RealEstateAndAccumulatedDepreciationInitialCostOfLand	25.190	18,3130	0,0025
16	RealEstateAndAccumulatedDepreciationCarryingAmountOfLand	23.003	16,5811	0,0027
17	RealEstateAndAccumulatedDepreciationAccumulatedDepreciation	22.810	53,7354	0,0045
18	IncomeLossFromContinuingOperationsBeforeIncomeTaxesMinorityInterest AndIncomeLossFromEquityMethodInvestments	22.123	15,7552	0,0020
19	ProfitLoss	21.177	4,4082	0,0012
20	AllocatedShareBasedCompensationExpense	20.854	2,3518	0,0009
21	InterestExpense	20.620	5,1824	0,0013
22	NetCashProvidedByUsedInOperatingActivities	20.131	17,8072	0,0029
23	LoansAndLeasesReceivableGrossCarryingAmount	17.976	8,6512	0,0021
24	AvailableForSaleSecuritiesFairValueDisclosure	16.224	33,1619	0,0032
25	Goodwill	15.536	10,7052	0,0027

Notes:

This table provides the most frequently reported XBRL monetary line items contained in 5,782 examined XBRL 10-K reports. The Chi-square value (CHI²) and the MAD are computed using equation (4) and (5), respectively. Please see the Appendix for a detailed definition and documentation of the presented XBRL line items.

Obviously, much of the financial information listed in Table 3 is not fully contained ‘as reported’ or ‘unadjusted’ in databases from intermediaries. Thus, using a risk-oriented (audit) approach the combination of structured financial reports and digit analysis offers many interesting opportunities to identify suspect areas in forensic accounting. It would be interesting to see if the SEC will focus future enforcement actions in the area of the suspect XBRL line items above.

5 Conclusions

The length and complexity of annual reports is steadily increasing, which endangers the full usage of disclosed information (Paredes 2003). Whereas several studies criticize the financial reporting standards and for example demand amendments to shorten disclosure requirements (Radin 2007; Möllers and Kernchen 2010), we believe that it is important to take new technical opportunities (e.g., XBRL) into account. The eXtensible Business Reporting Language (XBRL) is intended to offer new analysis possibilities and may reduce information asymmetry that is innate in complex and extensive disclosures (see You and Zhang 2009; Miller 2010). In this study, we introduce XBRL-based digit analysis as a possible future cost-efficient analysis opportunity.

Benford's Law is a digit distribution that is expected for many random datasets, whereas deviations can indicate human interaction. We extract all monetary line items that are contained in XBRL 10-K filings submitted to the SEC for the fiscal year 2012 and find that, on average, the distribution of first digits conforms very well to the expected distribution according to Benford's Law. The results are consistent with prior research (e.g., Quick and Wolz 2003; Watrin et al. 2008) and suggest that analysis procedures based on Benford's Law are applicable to accounting data. Furthermore, we select line items with the highest total occurrence in the analyzed financial statements and analyze conformity to Benford's Law for the single line items across firm-years. Interestingly, we observe an abnormal digit frequency for the line item *Revenues*, which often is part of fraudulent transactions (Dechow et al. 1996, 10; Leng et al. 2011, 819). Besides contributing research to the XBRL literature with a comprehensive analysis of data submitted under the SEC XBRL filing mandate, we also add to the Benford's Law literature by gaining further insights on digit analysis based on aggregated accounting data.

However, our study is not without limitations. For example, the empirical findings do not enable statements about differences between firm-year observations and respective inferences. It will be interesting to see whether future studies can approve the usefulness of Benford's Law analyses based on aggregated data. In particular, further studies should find evidence for an association between human interactions (e.g., earnings management or fraud) and abnormal digit frequencies, which may give further insights into its usability across financial statement line items. Research into whether and how emerging technologies provide new dimensions for financial reporting and analysis provide an interesting field of work. With

more and more securities regulators and stock exchanges adopting XBRL worldwide, the opportunities to use XBRL data for analysis as well as the importance of research into the issue may increase.

Notes

1

For example, the audit software ACL has implemented Benford's Law functionality (Durtschi et al. 2004; Cleary and Thibodeau 2005). The German Fiscal Authority also uses Benford's Law to detect irregularities in German taxpayers' data with the help of the software IDEA (Interactive Data Extraction and Analysis). Both software packages have in common that the digit analysis can occur at a very detailed data level for single companies (Cleary and Thibodeau 2005; Watrin and Struffert 2006).

2

See <http://www.sec.gov/spotlight/xbrl/filings-and-feeds.shtml>.

Appendix

No. Documentation of XBRL Monetary Line Items (us-gaap)

- 1 | **StockholdersEquity** (*instant/credit*)
Total of all stockholders' equity (deficit) items, net of receivables from officers, directors, owners, and affiliates of the entity which are attributable to the parent. The amount of the economic entity's stockholders' equity attributable to the parent excludes the amount of stockholders' equity which is allocable to that ownership interest in subsidiary equity which is not attributable to the parent (noncontrolling interest, minority interest). This excludes temporary equity and is sometimes called permanent equity.
- 2 | **NetIncomeLoss** (*duration/credit*)
The portion of profit or loss for the period, net of income taxes, which is attributable to the parent.
- 3 | **OperatingIncomeLoss** (*duration/credit*)
The net result for the period of deducting operating expenses from operating revenues.
- 4 | **StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest** (*instant/credit*)
Amount of stockholders' equity (deficit), net of receivables from officers, directors, owners, and affiliates of the entity, attributable to both the parent and noncontrolling interests. Amount excludes temporary equity. Alternate caption for the concept is permanent equity.
- 5 | **Assets** (*instant/debit*)
Sum of the carrying amounts as of the balance sheet date of all assets that are recognized. Assets are probable future economic benefits obtained or controlled by an entity as a result of past transactions or events.
- 6 | **DefinedBenefitPlanFairValueOfPlanAssets** (*instant/debit*)
Assets, usually stocks, bonds, and other investments, that have been segregated and restricted (usually in a trust) to provide benefits, at their fair value as of the measurement date. Plan assets include amounts contributed by the employer (and by employees for a contributory plan) and amounts earned from investing the contributions, less benefits paid. If a plan has liabilities other than for benefits, those non-benefit obligations may be considered as reductions of plan assets.
- 7 | **Revenues** (*duration/credit*)
Aggregate revenue recognized during the period (derived from goods sold, services rendered, insurance premiums, or other activities that constitute an entity's earning process). For financial services companies, also includes investment and interest income, and sales and trading gains.
- 8 | **SalesRevenueNet** (*duration/credit*)
Total revenue from sale of goods and services rendered during the reporting period, in the normal course of business, reduced by sales returns and allowances, and sales discounts.
- 9 | **PropertyPlantAndEquipmentGross** (*instant/debit*)
Gross amount of long-lived physical assets used in the normal conduct of business and not intended for resale. Examples include, but are not limited to, land, buildings, machinery and equipment, office equipment, furniture and fixtures, and computer equipment.
- 10 | **CashAndCashEquivalentsAtCarryingValue** (*instant/debit*)
Amount of currency on hand as well as demand deposits with banks or financial institutions. Includes other kinds of accounts that have the general characteristics of demand deposits. Also includes short-term, highly liquid investments that are both readily convertible to known amounts of cash and so near their maturity that they present insignificant risk of changes in value because of changes in interest rates. Excludes cash and cash equivalents within disposal group and discontinued operation.
- 11 | **RealEstateAndAccumulatedDepreciationInitialCostOfBuildingsAndImprovements** (*instant/debit*)
Initial cost to the Entity for buildings and improvements.
- 12 | **RealEstateAndAccumulatedDepreciationCarryingAmountOfLandAndBuildingsAndImprovements** (*instant/debit*)
Gross amount at which land and buildings and improvements are carried at the end of the period for each property.
- 13 | **RealEstateAndAccumulatedDepreciationCarryingAmountOfBuildingsAndImprovements** (*instant/debit*)
The carrying amount at which buildings and improvements are carried at the end of the period.

- 14 | **IncomeTaxExpenseBenefit** (duration/debit)
The sum of the current income tax expense or benefit and the deferred income tax expense or benefit pertaining to continuing operations.
- 15 | **RealEstateAndAccumulatedDepreciationInitialCostOfLand** (instant/debit)
Initial cost to the Entity for land.
- 16 | **RealEstateAndAccumulatedDepreciationCarryingAmountOfLand** (instant/debit)
The carrying amount at which land is carried at the end of the period.
- 17 | **RealEstateAndAccumulatedDepreciationAccumulatedDepreciation** (instant/credit)
The amount of accumulated depreciation for the property as of the balance sheet date.
- 18 | **IncomeLossFromContinuingOperationsBeforeIncomeTaxesMinorityInterestAndIncomeLossFromEquityMethodInvestments** (duration/credit)
Sum of operating profit and nonoperating income or expense before Income or Loss from equity method investments, income taxes, extraordinary items, and noncontrolling interest.
- 19 | **ProfitLoss** (duration/credit)
The consolidated profit or loss for the period, net of income taxes, including the portion attributable to the noncontrolling interest.
- 20 | **AllocatedShareBasedCompensationExpense** (duration/debit)
Represents the expense recognized during the period arising from equity-based compensation arrangements (for example, shares of stock, unit, stock options or other equity instruments) with employees, directors and certain consultants qualifying for treatment as employees.
- 21 | **InterestExpense** (duration/debit)
The cost of borrowed funds accounted for as interest that was charged against earnings during the period.
- 22 | **NetCashProvidedByUsedInOperatingActivities** (duration/-)
The net cash from (used in) all of the entity's operating activities, including those of discontinued operations, of the reporting entity. Operating activities generally involve producing and delivering goods and providing services. Operating activity cash flows include transactions, adjustments, and changes in value that are not defined as investing or financing activities. While for technical reasons this element has no balance attribute, the default assumption is a debit balance consistent with its label.
- 23 | **LoansAndLeasesReceivableGrossCarryingAmount** (instant/debit)
Amount before allowance of loans and leases held in portfolio, including but not limited to, commercial and consumer loans. Includes deferred interest and fees, undisbursed portion of loan balance, unamortized costs and premiums and discounts from face amounts. Excludes loans and leases covered under loss sharing agreements.
- 24 | **AvailableForSaleSecuritiesFairValueDisclosure** (instant/debit)
This item represents Available-for-sale Securities which consist of all investments in certain debt and equity securities neither classified as trading or held-to-maturity securities. A debt security represents a creditor relationship with an enterprise. Debt securities include, among other items, US Treasury securities, US government securities, municipal securities, corporate bonds, convertible debt, commercial paper, and all securitized debt instruments. An equity security represents an ownership interest in an enterprise or the right to acquire or dispose of an ownership interest in an enterprise at fixed or determinable prices. Equity securities include, among other things, common stock, certain preferred stock, warrant rights, call options, and put options, but do not include convertible debt.
- 25 | **Goodwill** (instant/debit)
Carrying amount as of the balance sheet date, which is the cumulative amount paid and (if applicable) the fair value of any noncontrolling interest in the acquiree, adjusted for any amortization recognized prior to the adoption of any changes in generally accepted accounting principles (as applicable) and for any impairment charges, in excess of the fair value of net assets acquired in one or more business combination transactions.

Notes:

This Appendix contains the definition and documentation of the 25 most frequently reported XBRL line items within the 5,782 examined XBRL 10-K reports (Fiscal Year 2012). All information is provided by the official US-GAAP taxonomy and is obtained from the 'Yeti' taxonomy viewer (available at: <http://bigfoot.corefiling.com/yeti/resources/yeti-gwt/Yeti.jsp>). The information contained in parentheses are attribute values that document whether the taxonomy element is a 'debit' or 'credit' position based on double entry bookkeeping. The attribute values 'instant' or 'duration' document the underlying period type (e.g., *StockholdersEquity* is a line item of the balance sheet and presents information based on specific point-in-time (usually at the fiscal year end date), whereas the line item *Revenues* provides information about generated sales and revenues during the fiscal year).

References

- Benford, F. 1938. The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society* 78 (4): 551–572.
- Carslaw, C. A. P. N. 1988. Anomalies in Income Numbers: Evidence of Goal Oriented Behavior. *The Accounting Review* 63 (2): 321–327.
- Cleary, R., and J. C. Thibodeau. 2005. Applying Digital Analysis Using Benford's Law to Detect Fraud: The Dangers of Type I Errors. *Auditing: A Journal of Practice & Theory* 24 (1): 77–81.
- Das, S., and H. Zhang. 2003. Rounding-up in reported EPS, behavioral thresholds, and earnings management. *Journal of Accounting and Economics* 35 (1): 31–50.
- Debreceny R. S., S. M. Farewell, M. Piechocki, C. Felden, A. Gränig, and A. d'Eri. 2011. Flex or Break? Extensions in XBRL Disclosures to the SEC. *Accounting Horizons* 25 (4): 631–657.
- Dechow, P. M., R. G. Sloan, and A. P. Sweeney. 1996. Causes and Consequences of Earnings Manipulation: An Analysis of Firms Subject to Enforcement Actions by the SEC. *Contemporary Accounting Research* 13 (1): 1–36.
- Drake, P. D., and M. J. Nigrini. 2000. Computer assisted analytical procedures using Benford's Law. *Journal of Accounting Education* 18 (2): 127–146.
- Durtschi, C., W. Hillison, and C. Pacini. 2004. The Effective Use of Benford's Law in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting* 5: 17–34.
- Johnson, G. C. 2009. Using Benford's Law to Determine if Selected Company Characteristics are Red Flags for Earnings Management. *Journal of Forensic Studies in Accounting and Business* 1 (2): 39–65.
- Jordan, C. E., and S. J. Clark. 2011. Detecting Cosmetic Earnings Management Using Benford's Law. *The CPA Journal* 81 (2): 32–37.
- Kinnunen, J., and M. Koskela. 2003. Who Is Miss World in Cosmetic Earnings Management? A Cross-National Comparison of Small Upward Rounding of Net Income Numbers among Eighteen Countries. *Journal of International Accounting Research* 2: 39–68.
- Leng, F., E. H. Feroz, Z. Cao, and S. V. Davalos. 2011. The Long-Term Performance and Failure Risk of Firms Cited in the US SEC's Accounting and Auditing Enforcement Releases. *Journal of Business Finance and Accounting* 38 (7/8): 813–841.
- Miller, B. P. 2010. The Effects of Reporting Complexity on Small and Large Investor Trading. *The Accounting Review* 85 (6): 2107–2143.
- Möllers, T. M. J., and E. Kernchen. 2010. Information Overload at the Capital Market. *Europa e Diritto Private* 4: 1005–1035.
- Newcomb, S. 1881. Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics* 4 (1): 39–40.

- Nigrini, M. 1994. Using digital frequencies to detect fraud. *Fraud Magazine, The White Paper Index* 8 (2): 3–6.
- Nigrini, M. J. 1996. A taxpayer compliance application of Benford's Law. *Journal of the American Taxation Association* 18 (1): 72–91.
- Nigrini, M. J., and L. J. Mittermaier. 1997. The Use of Benford's Law as an Aid in Analytical Procedures. *Auditing: A Journal of Practice and Theory* 16 (2): 52–67.
- Nigrini, M. J. 2000. *Continuous Auditing*. Working Paper, University of Kansas.
- Nigrini, M. J. 2005. An Assessment of the Change in the Incidence of Earnings Management Around the Enron-Andersen Episode. *Review of Accounting and Finance* 4 (1): 92–110.
- Nigrini, M. J. 2011. *Forensic Analytics – Methods and Techniques for Forensic Accounting Investigations*. Hoboken, NJ: John Wiley & Sons.
- Nigrini, M. J. 2012. *Benford's Law – Applications for Forensic Accounting, Auditing, and Fraud Detection*. Hoboken, NJ: John Wiley & Sons.
- Niskanen, J., and M. Keloharju. 2000. Earnings cosmetics in a tax-driven accounting environment: evidence from Finnish public firms. *The European Accounting Review* 9 (3): 443–452.
- Paredes, T. A. 2003. Blinded by the Light: Information Overload and its Consequences for Securities Regulation. *Washington University Law Quarterly* 81 (2): 417–485.
- Quick, R., and M. Wolz. 2003. Benford's Law in deutschen Rechnungslegungsdaten. *Betriebswirtschaftliche Forschung und Praxis* 2: 208–224.
- Radin, A. J. 2007. Have We Created Financial Statement Disclosure Overload? *The CPA Journal* 77 (11): 6–9.
- Rafeld, H., and F. R. Then Bergh. 2007. Digitale Ziffernanalyse in deutschen Rechnungslegungsdaten. *Zeitschrift Interne Revision* 42 (1): 26–33.
- Securities and Exchange Commission (SEC). 2009. *Interactive Data to Improve Financial Reporting*. Release Nos. 33-9002, 34-59324, 39-2461, IC-28609. Washington, D.C.: SEC.
- Skousen, C. J., L. Guan, and T. S. Wetzel. 2004. Anomalies and Unusual Patterns in Reported Earnings: Japanese Managers Round Earnings. *Journal of International Financial Management and Accounting* 15 (3): 212–234.
- Thomas, J. K. 1989. Unusual Patterns in Reported Earnings. *The Accounting Review* 64 (4): 773–787.
- Van Caneghem, T. 2002. Earnings Management Induced by Cognitive Reference Points. *British Accounting Review* 34: 167–178.
- Wallace, W. A. 2002. Assessing the quality of data used for benchmarking and decision-making. *The Journal of Government Financial Management* 51 (3): 16–22.

- Watrin, C. and R. Struffert. 2006. Benford's Law und Chi-Quadrat Test – Chancen und Risiken des Einsatzes bei steuerlichen Prüfungen. *Der Betrieb* 33: 1748–1750.
- Watrin, C., R. Struffert, and R. Ullmann. 2008. Benford's Law: an instrument for selecting tax audit targets? *Review of Managerial Science* 2: 219–237.
- You, H., and X. Zhang. 2009. Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies* 14 (4): 559–586.



Working Papers in Accounting Valuation Auditing Nr. 2012-1

Klaus Henselmann / Elisabeth Scherr / Dominik Ditter

Applying Benford's Law to individual financial reports

An empirical investigation on the basis of SEC XBRL filings

The business reporting standard XBRL offers the opportunity to easily extract and analyze a sufficient number of monetary items out of single annual reports for statistical analysis purposes. Using XBRL 10-K reports filed with the SEC EDGAR system, we derive first digit distributions for single companies and measure the deviation from the Benford distribution. On average, we find that for all monetary numbers that are contained in the examined XBRL reports, the first digit distribution follows Benford's Law. Furthermore, our results indicate several line items with an abnormal digit frequency potentially indicating human interaction. Taken together, the empirical results suggest that the application of Benford's Law to financial reports might be a useful analytical tool for investors. The findings also may be of interest to the SEC for planning enforcement actions, as digit analysis could be an appropriate means of identifying suspect line items carrying a higher level of risk.

Impressum

Nürnberg 2012

Herausgeber, Redaktion und Druck:

Lehrstuhl für Rechnungswesen und Prüfungswesen

Rechts- und Wirtschaftswissenschaftliche Fakultät

Friedrich-Alexander-Universität Erlangen-Nürnberg

Lange Gasse 20

90403 Nürnberg

Tel +49 911 5302 - 437

Fax +49 911 5302 - 401

www.pw.wiso.uni-erlangen.de



Lehrstuhl für
Rechnungswesen
und Prüfungswesen

ISSN 1867-7932