MA615 Project Report

Using Benford's Law to Study the Current Ratio of Corporate Bankruptcy

Xuan Zhu

12/14/2018

Contents

1 Abstract

The current ratio is one of the indicators that analysts use to predict corporate bankruptcy in a financial forecasting period. In this project, Benford's law is applied to explore whether the distribution of the first leading digits has a pattern that helps analysts be more alert to unusual reports.

2 Introduction

This chapter aims to introduce some financial concepts and mathematical techniques that are related to the project. Section 2.1 breaks down the concept of the current ratio, discusses its significance in predicting corporate bankruptcy, and states its limitations. Section 2.2 gives a general explanation of Benford's Law with formulas and examples, especially of the part First Digit Test.

2.1 Current Ratio

The current ratio is a term often used in the area of financial analysis. It measures one company's ability to pay short-term liability and long-term obligations. To calculate the ratio, analysts apply the following formula:

Current Ratio = Current Assets / Current Liabilities

The current assets include the assets that can be transformed into cash in less than a year, such as cash and inventory. The current liabilities are how much the company needs to pay to other subjects in the future, such as taxes payable and wages.

It is hard to say how good or bad a current ratio is, as the standard varies a lot across industries. To avoid this, we only select out the data of companies within the same industry (Polish companies in the manufacturing sector, which will be discussed with details later in Chapter 2.) But even in the same industry, getting a current ratio above the industry average line is not the same as the signal of well management, because it is possible that the company resources are not efficiently used. Similarly, getting a current ratio below one is typically an alarm that now the company is not able to pay all its obligations if they are all due at one, but the company may be struggling in one financial reporting period and then becomes healthy in the future. In a short summary, the current ratio itself is limited in

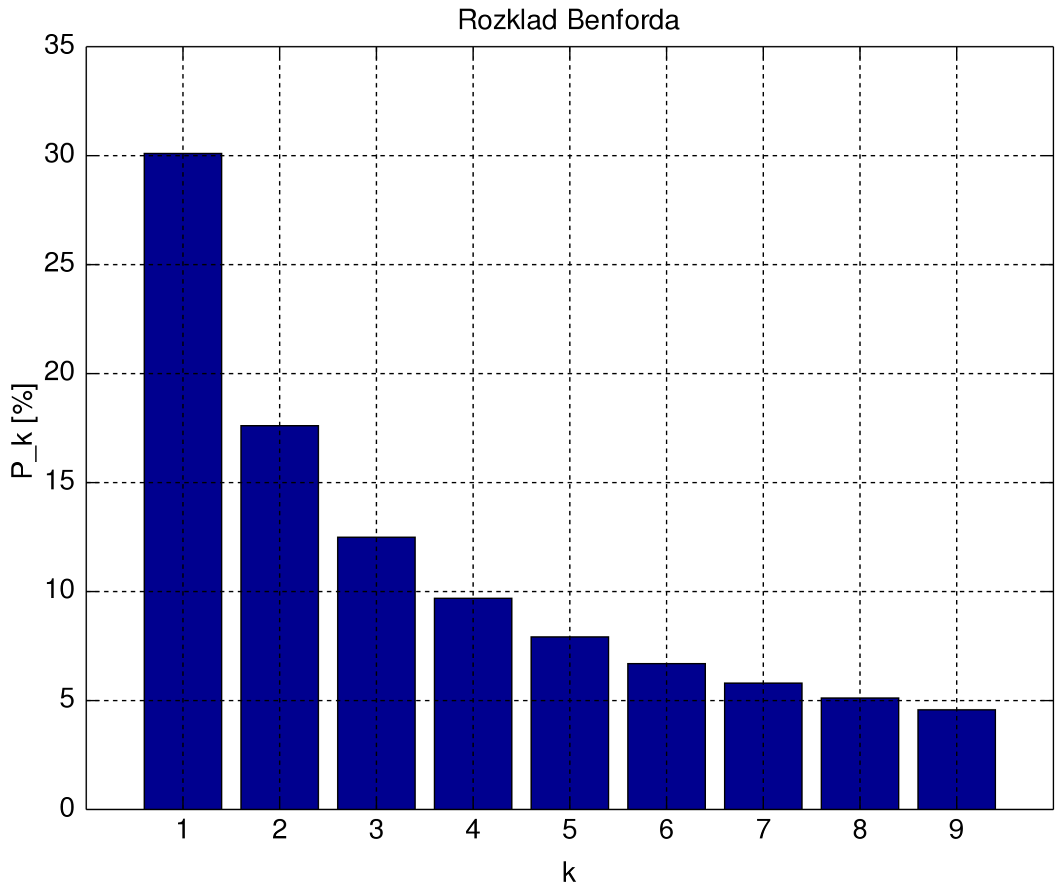


Figure 1: The distribution of first digits, according to Benford's law. Each bar represents a digit, and the height of the bar is the percentage of numbers that start with that digit.

comparison between different companies without other liquidity ratios' help, and we should always look at the trending information rather than focusing on analyzing data from a single financial period.

2.2 Benford's Law

Benford's Law is an observation about the frequency distribution of leading digits in any large, randomly produced data sets of numbers. According to the definition, a set of numbers is said to satisfy Benford's law if the leading digit d (d $\in \{1, \ldots, 9\}$) occurs with probability

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

Visually, the theoretical distribution of leading digits is as the following(*1):

Data set	Features from	Bankruptcy after	No. bankrupt	No. not bankrupt	Sum
1stYear	1st year	5 years	271	6,756	7,027
2ndYear	2nd year	4 years	400	9,773	10,173
3rdYear	3rd year	3 years	495	10,008	10,503
4thYear	4th year	2 years	515	9,277	9,792
5thYear	5th year	1 years	410	5,500	5,910

Figure 2: a summary of source data

In reality, the Benford's Law is widely applied to many areas such as detecting fraud. After testing whether one data set is consistent with the law, analysts narrow down their search on numbers with suspect leading digits that may be manipulated manually. In this project, we are focusing on the First Digit Benford Test, and we expect that the current ratio reported by those companies going bankrupt in five years is manipulated and thus the distribution of first leading digits deviates from the law.

3 Source Data

According to UCI Machine Learning Repository, the data is first extracted and used by Zieba, M., Tomczak, S. K., and Tomczak, J. M. (2016)(*). This chapter describes the basic information about the data set that is related to the project.

The dataset contains five subsets, each has 64 indicators predicting corporate bankruptcy from the first year of forecasting period to the fifth year of it. The companies classified as "0" in the dataset remained healthy financial status after the forecasting period, while those classified as "1" did not survive successfully. The explanation of 64 indicators is appended in the Appendix chapter. In the project, I extract out the column of X4/Attr 4 as the subject of the study.

Two things about the quality of the data set need to be mentioned. First, the data is heavily imbalanced. There are much more healthy companies existing than the problematic ones, as you can see in the above table. Second, the data is not comprehensive about all the companies operating in the forecasting period, because some information is not public or missing from the original database.

4 Method

4.1 Choice of the indicator

The reason why I choose to test the distribution of the leading digits by current ratio rather than picking other 63 indicators is that the current ratio aims to reflect a company's ability to pay all its liabilities in a time period as short as less than one year. Therefore, things become interesting when we can compare the reports from those companies which would claim bankruptcy in only one year with the reports they produced in the previous years.

In the fifth data set, we know that those companies would no longer exist in the market one year later and certainly, the numbers were revised upward. But do they display a similar pattern of manipulation to the pattern in the first data set where companies were in the beginning year of the forecasting period? Second, since we also have data on healthy companies, how is the distribution of the healthy different from that of the bankrupt?

4.2 First Digit Analysis

When we browse through the data, we find out some outliers below 0. Usually the current ratio will not be less than zero, so we will remove the unusual points before conducting the First Digit Test.

Below shows all what we exclude from the data sets.

```
##        CA_SL class
## 1 -0.045319     0
```

```
##       CA_SL class
## 1 -0.40311     1
```

There are only two data points existing as outliers, so we assume that the effect of removing negative values is small enough to neglect it.

The first digit test compares the actual first digit frequency distribution of the data set with the theoretical Benford's distribution.

Then we can perform the test on data.

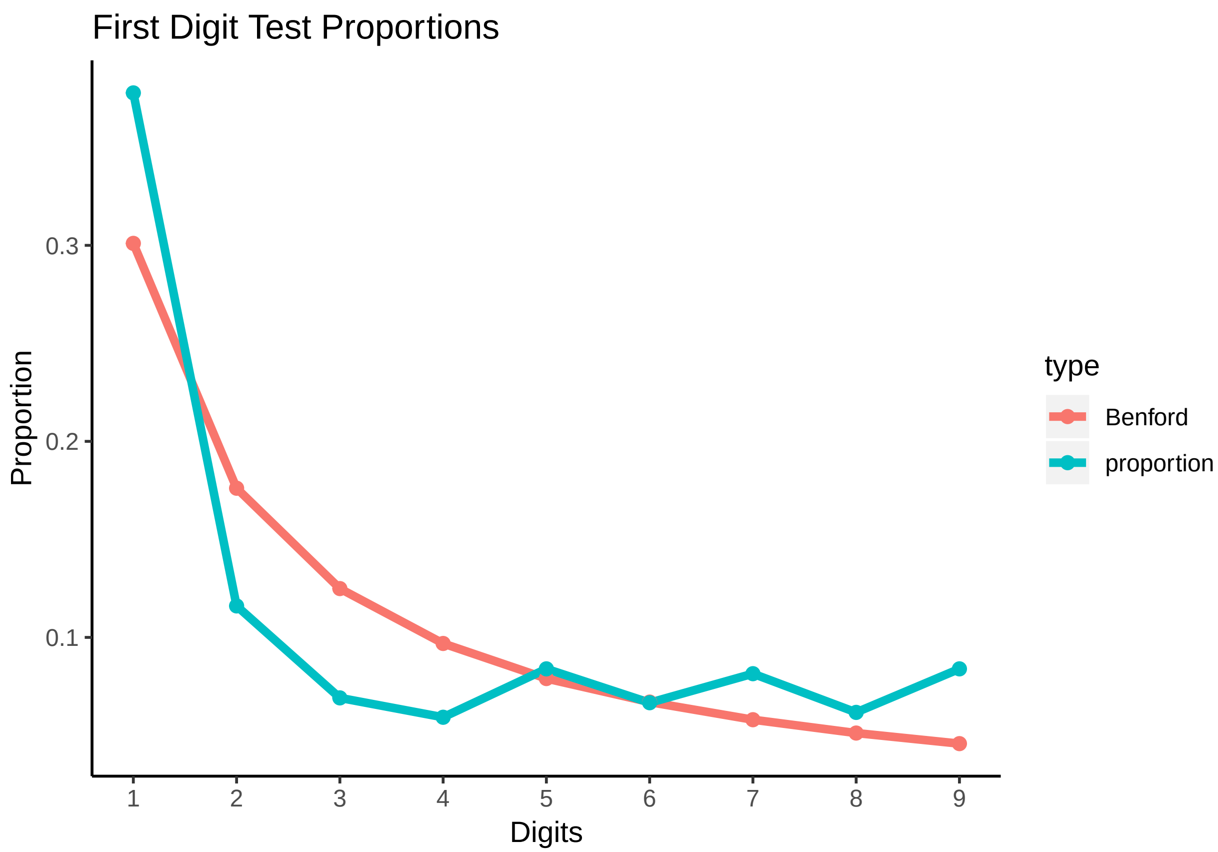
The first step is to count the frequencies of the first leading digits from 1 to 9 and calculate the proportion of each digit in the data from the frequencies. Then we compare the actual

Table 1: The actual first digit frequency distribution of struggling companies(5th year)

Digit	freq	proportion	Benford
1	153	37.78%	30.10%
2	47	11.60%	17.61%
3	28	6.91%	12.49%
4	24	5.93%	9.69%
5	34	8.40%	7.90%
6	27	6.67%	6.70%
7	33	8.15%	5.80%
8	25	6.17%	5.12%
9	34	8.40%	4.58%

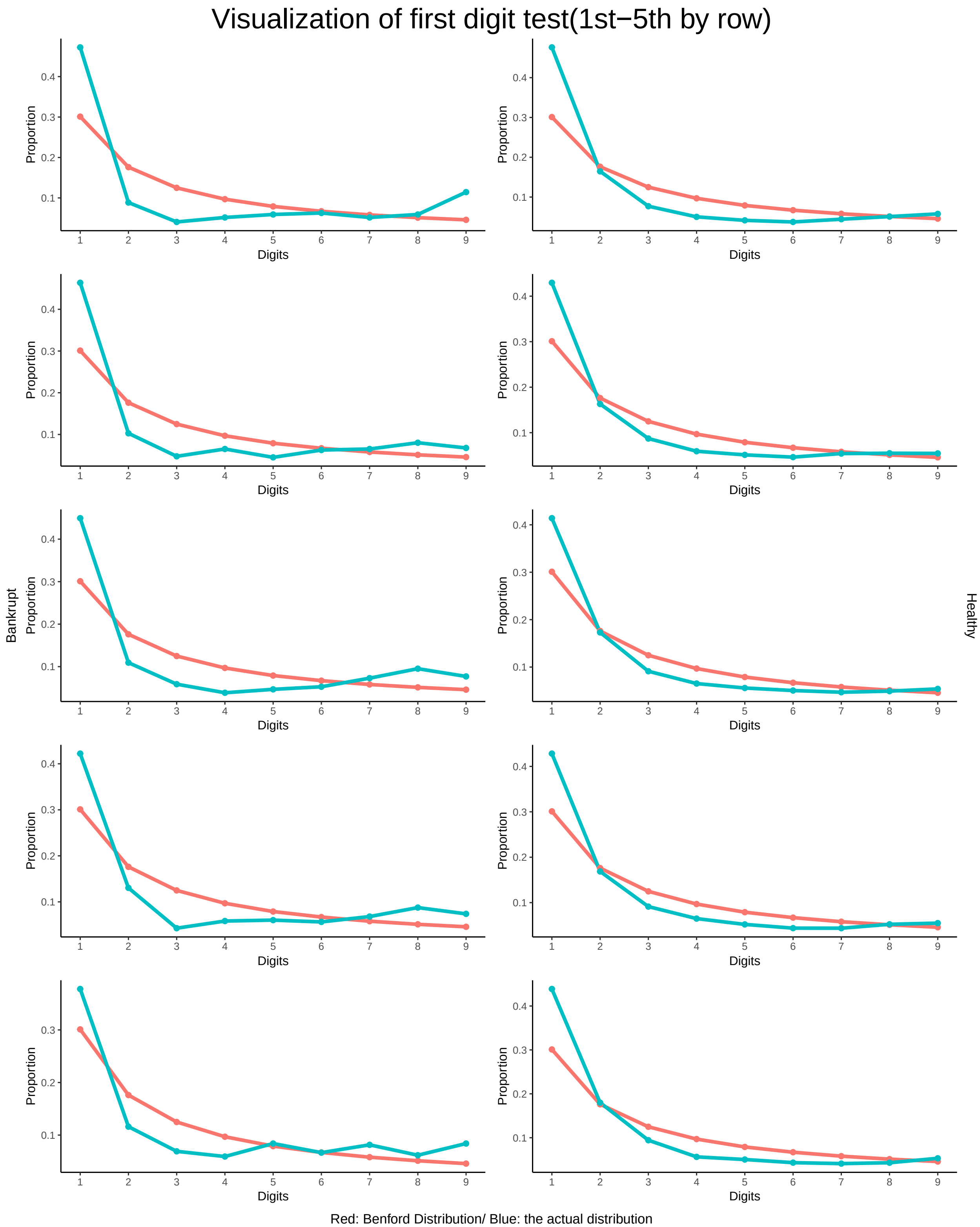
first digit frequency distribution with the theoretical Benford's one by chi square test and visualize it.

For instance, belows are the results of the first digit test on bad companies in the last year of the forcasting period.



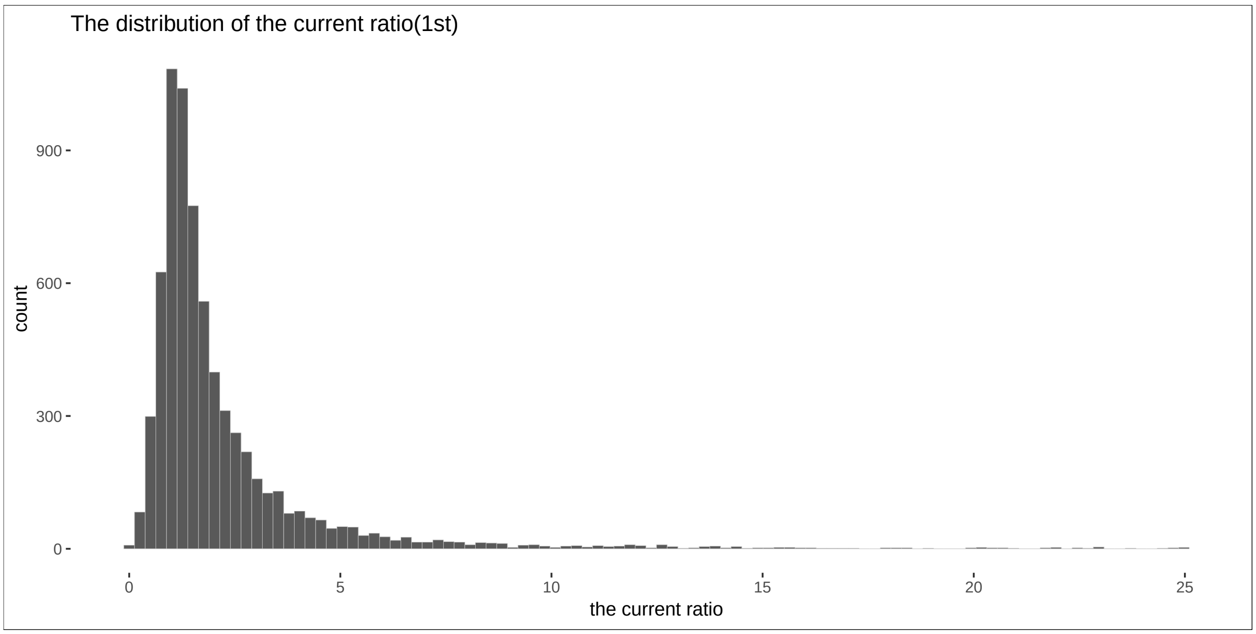
The chi square test confirms that its distribution deviates from the Benford's law.

Similarly, we can perform the test on other sub data sets.



5 Results and Discussions

The chi square test failed to prove that any data sets here are generally conformed to the Benford's law. It is acceptable, since numbers representing the current ratio were not never randomly generated. For instance, by the following plot we can clearly see that the first year's current ratios from all companies gathered around 1 and the distribution is heavily right skewed. As we discussed before, neither a single chi square test nor a first digit test on one specific data set provides any insights.



However, if one carefully examines the integrated results on page 7, it is obvious that significant anomalies occur more with the digits from problematic companies than from the non-problematic. Since companies in healthy financial status do not have a strong motivation to manipulate the ratio, the shape of the real distribution is similar to that of theoretical Benford's law and does not vary much by year. The digits 2, 8 and 9 conform to the Benford's law. The digits 3-7 exhibit less frequency than the expected, and there may be some reasons behind it which I welcome any reader to explore in the future.

The bankrupt companies are interesting for that the tail of the real frequency distribution becomes more and more abnormal as time went closer to the point where it went bankrupt. Companies in a dangerous financial status usually have a current ratio below one, which means that the leading digits extracted from the data would be possibly 7 to 9. If managers tried hard to adjust the ratio up to 1, we expect to see a higher and higher real frequency in the tail by year. And our observation satisfies the hypothesis.

6 Conclusion

To summarize it, bankrupt companies display a predictable pattern in the manipulation of their current ratio before the year they went bankrupt. Compared to the frequency of leading digits from companies in good financial status, that of the bankrupt has a heavy tail which can be used as a benchmark in predicting bankruptcy.

7 Reference

*1.Benford's Law. (n.d.). In Wikipedia. Retrieved December 10, 2018, from https://en.wikipedia.org/wiki/Benford%27s_law

*2.UCI Machine Learning Repository, Polish companies bankruptcy data Data Set.https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data

*3.Zieba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction. Expert Systems with Applications.

8 Appendix

8.1 Original Attribute Information:

Attribute Information:
X1 net profit / total assets
X2 total liabilities / total assets
X3 working capital / total assets
X4 current assets / short-term liabilities
X5 [(cash + short-term securities + receivables - short-term liabilities) / (operating expenses - depreciat
X6 retained earnings / total assets
X7 EBIT / total assets
X8 book value of equity / total liabilities
X9 sales / total assets
X10 equity / total assets
X11 (gross profit + extraordinary items + financial expenses) / total assets
X12 gross profit / short-term liabilities
X13 (gross profit + depreciation) / sales
X14 (gross profit + interest) / total assets
X15 (total liabilities * 365) / (gross profit + depreciation)
X16 (gross profit + depreciation) / total liabilities
X17 total assets / total liabilities
X18 gross profit / total assets
X19 gross profit / sales
X20 (inventory * 365) / sales
X21 sales (n) / sales (n-1)
X22 profit on operating activities / total assets
X23 net profit / sales
X24 gross profit (in 3 years) / total assets
X25 (equity - share capital) / total assets
X26 (net profit + depreciation) / total liabilities
X27 profit on operating activities / financial expenses
X28 working capital / fixed assets
X29 logarithm of total assets
X30 (total liabilities - cash) / sales
X31 (gross profit + interest) / sales
X32 (current liabilities * 365) / cost of products sold
X33 operating expenses / short-term liabilities
X34 operating expenses / total liabilities
X35 profit on sales / total assets