**Forensic Accounting Analysis on Walmart**

Abe Archer, Coleman Dawkins, Isabella McCoy, Lauren Petty

University of Arkansas

DASC 1223/H – Role of Data Science in Today’s World

Dr. Schubert

03/03/2024

**Executive Summary**

Forensic accounting is a series of techniques used to detect discrepancies in financial data. This can be extremely helpful in uncovering financial fraud and finding issues with corporate business reports. Using forensic accounting techniques, this project is an analysis of Walmart financial data (Mergent Online (Walmart Inc)). There were four goals for this analysis, the first of which was to examine the data by columns, view trends, and see how different columns interact with each other. The second was to explore Walmart’s stock prices and see if and when any large jumps or dips occur (*Walmart Inc. (WMT))*. The third goal was to compare Walmart’s data to that of a similar competitor, in this case being Costco. Finally, the fourth goal was to perform statistical analysis on the data. For this project, it was decided to use Walmart quarterly data from 2009 – 2019, as this would give a long enough range to see trends develop, while avoiding possible discrepancies resulting from COVID-19.

The majority of financial fraud occurs in reporting revenue, assets, and liabilities (Bloomenthal, 2019). This can be done by overreporting revenue, creating sales that never occurred, or overestimating the value of assets. Because of this, the assets and liabilities recorded in the dataset were examined closely to determine trends and view any discrepancies. In the analysis of the individual columns, these had special attention paid to them. The current ratio was calculated by dividing the current assets by the current liabilities, and was then analyzed alongside the individual elements. All the analysis of the individual columns was done in R, and the data was also cleaned in R.

In examining the stock data, it was just as important to look at the dividends paid out by Walmart to stockholders, and to determine what the effects of these might be on the company’s financial information. These were examined using principal component analysis, which is a machine learning method that condenses large datasets into more manageable smaller sets while still maintaining the important patterns from the original data (Jaadi, 2021). Principal component analysis requires standardizing the data, which eliminates possible issues from variables with larger or smaller ranges, and then generating a covariance matrix, which shows all possible pairs of the variables being used. The sign of the covariance of each of these pairs indicates whether there is a correlation or not, with positive meaning there is and negative meaning there is none. The eigenvectors and eigenvalues of this matrix are then computed, and these are used to create feature vectors, which are then finally used to replot the data.

To compare to the Walmart results, similar data was pulled for Costco and analyzed following the same methods (Mergent Online (Costco)). Comparing these can demonstrate if there may be issues occurring with Walmart’s data. This analysis measured and compared growth patterns for assets, liabilities, and stock prices. It also looked for outliers in a similar fashion to the Walmart analysis. Investigating Costco’s data was also helpful to explain some of what occurs in Walmart’s data, such as learning that they pay dividends differently. Similarly to the Walmart data, the Costco data was cleaned and analyzed using R.

Statistical analysis is used as a major form of forensic accounting. Following along with this method, Benford’s Law was used to examine the Walmart data. Benford’s Law measures the frequency of the first integer of each number in the dataset. Most sample sets, regardless of the subject of the data, will follow a trend established by Benford’s Law. If a leading integer occurs more or less frequently than predicted by Benford’s Law, it can be an indicator that the data is incorrect. The Kolmogorov-Smirnov and Chi Square tests were both also applied to the data, being used to determine goodness of fit.

After all of the analyses were performed on the data, the results were examined to determine whether there may be discrepancies indicative of fraud. While none was explicitly found, there were several interesting points to explore where the data diverges from expected results. These included a certain quarter having significant outliers for multiple variables, as well as the data not conforming properly to Benford’s Law. The statistical analysis raises questions as to why Walmart’s data behaves unexpectedly. These issues will be explained in further detail in the following sections.

**Literature Review**

When examining past financial accounting analysis, one study in particular was helpful when planning the statistical analysis. This was Juan Chang’s 2017 dissertation at the University of Nebraska, which concerned using Benford’s Law, particularly to analyze the fraud committed by Bernie Madoff in various balance sheets from 1990 – 2008. The study also examined two other companies, Toshiba Corporation and Valeant Pharmaceuticals International, Inc. To test the conclusions drawn from the Benford’s Law analysis, Chang also used the Kolmogorov-Smirnov and Chi Square tests. Chang’s dissertation was also helpful because it includes the R code used, which made it easier to see how the analysis was performed step-by-step.

In addition to this study, many articles have been written about how to prevent corporate fraud, and about instances where it should have been noticed sooner. Some of these were written due to the incidents Chang analyzed, such as the Japanese company Toshiba Corporation, found in 2015 to have doctored their financial information (Du, 2015). Toshiba had massively overstated their profits by over $1 billion from 2008 to 2014, and as a result the Japanese government introduced stricter restrictions on corporations to prevent similar incidents from happening in the future.

A 2019 study on the Toshiba Corporation’s accounting fraud helps explain some of what occurred (Caplan et al., 2019). This study focuses less on statistical analysis than Chang’s research, and more on how the fraud occurred in the first place and what exactly Toshiba was doing behind the scenes. It explains how Toshiba artificially inflated their reported income by using their business structure of having contract manufacturers, where they would sell these manufacturers the parts to produce their products, and then buy back the finished products. If the contract manufacturers sold all their finished products back to the company and still had leftover parts they had not used, there would be a credit balance on the production cost account. Using this knowledge, Toshiba would intentionally send far more production materials than were needed, which increased the credit balance on the account, indicating that Toshiba had higher earnings than they actually did.

Looking at the background of forensic accounting, it becomes apparent why it is so important to companies. It can help uncover fraud and prevent it in the future, and as a result is also helpful for overseeing bodies that regulate corporations. Due to their crimes being uncovered, Toshiba had to pay considerable fines, and several of their high-ranking executives were forced to resign amidst the scandal. Bernie Madoff was sentenced to 150 years in prison due to the massive amount of fraud he committed. Both outcomes occurred because of the work of people analyzing their falsified data and represent the importance forensic accounting can have. While nothing as extreme as either of these examples was found in this research, it is important to understand the background of forensic accounting when performing analyses used in these cases.

**Process**

In order to be able to use the Walmart data, it had to be cleaned significantly. The original dataset was formatted with each different category organized by rows instead of columns, which made it significantly more difficult to analyze. To fix this, the data was read into R, transposed using the t() function, and then outputted into a new Excel file, which then displayed the data with the different categories organized by column. Following this, there were spaces between each column that were deleted to make reading the data easier. The dates of each quarterly report were also reformatted, as they were originally in MM-DD-YYYY format and were then changed to YYYY-MM-DD, which was easier to perform analysis on in R.

To analyze the individual columns of data, several boxplots and line graphs were made to examine distributions and correlations. Boxplots can be useful to determine what an expected range for a set of data is, and to indicate outliers. One such outlier occurs in the category Cash and Cash Equivalents (Fig. 1). In the second quarter of 2018, this had a reported value of $15,840,000,000, which was significantly higher than the main range of the data. Looking further into this quarter, it also contains one other large discrepancy in the category Short-term Borrowings with a value of $444,000,000, which is significantly lower than the reported value for any other quarter (Fig. 2). These are two of the most significant outliers in the entire dataset.

A graph with a bar

Description automatically generated

*Fig. 1*

*A graph of a graph showing the number of short-term sales

Description automatically generated with medium confidence*

*Fig. 2*

Using the knowledge that a significant portion of financial fraud occurs in the reporting of assets and liabilities, both the total assets and the current assets were examined next. The difference between these is that current assets are those that will provide financial value for the company within one year, while total assets include current assets but also all other assets that may not provide value within that timeframe (Tamplin, 2022). We can see that there is a large range for the total assets, with many of the datapoints appearing as outliers in the boxplot (Fig. 3). We can also see that there is fairly consistent growth for the total assets, with it generally following an up-and-down trend, and then rising more in the last years of the dataset (Fig. 4).

A graph with a line

Description automatically generated

*Fig. 3*

A graph showing the growth of assets

Description automatically generated

*Fig. 4*

Comparing the current assets and current liabilities, we can see that they follow similar growth trends (Fig. 5). The liabilities are higher than the assets for every quarter measured, and the difference between them is fairly consistent. They follow very similar patterns quarter by quarter until about 2016, where they invert and high assets correspond to low liabilities, and vice versa. While this does not imply that there is fraud happening, it is curious why it happens from a corporate perspective. Using the current asset and liability data, the current ratio was then calculated by dividing the assets by the liabilities and was graphed (Fig. 6). This does not follow any visible trends, despite the current assets and liabilities themselves following trends.

A graph showing the value of a quarter report

Description automatically generated with medium confidence

*Fig. 5*

A graph showing the growth of a stock market

Description automatically generated

*Fig. 6*

To analyze Walmart’s stock data, a graph was made of their stock price over time in R. Ggplot was used in R to create a visualization of the rises and falls in Walmart’s stock price over time (Fig. 7). A principal component analysis was also conducted on the balance sheet items in R. The Excel spreadsheet was simplified by using the total short-term assets, total long-term assets, and total short-term liability categories instead of using all the individual variables for those categories. After simplifying the Excel spreadsheet, this was converted over to a .csv file. The corrr, FactoMineR, and ggcorrplot packages were installed and used in R to conduct principal component analysis on the .csv file.

A graph of a stock market

Description automatically generated

*Fig. 7*

A graph was then generated displaying correlations between the variables (Fig. 7). It was found that the retained earnings deficit variable had the highest correlation with the stockholders’ equity deficit variable. This makes sense because retained earnings are a component of stockholders’ equity. Further research was then conducted to explain these findings, and it was learned that a retained earnings deficit can occur when the company pays out a significant number of dividends. It was determined that Walmart pays out significant dividends four times each year, higher than some of their competitors (*Full Dividend History)*. Some of Walmart’s key financial ratios were calculated and analyzed, and it was found that in 2019 the return on equity ratio was 12% and the debt-to-equity ratio was 78%, which are both standard for the industry (St, 2019).

A close-up of a computer screen

Description automatically generated

*Fig. 8*

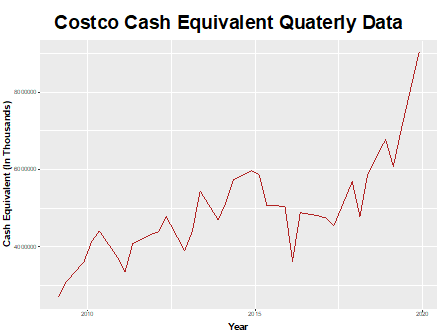
The next objective was to compare and find differences between the quarterly balances/stock data found for Walmart from 2009-2019 and quarterly balances/stock data for Costco from 2009 -2019. To do this, it was necessary to find a dataset that had enough information to serve as a sufficient comparison for the data gathered for Walmart. After exploring several possible datasets, it was found that on the original website there was a “Competitors” tab that had Costco data. This proved perfect because now all the variable names from both the datasets were the same, and both dataset had their information logged in the same way, with the dollar values being kept in thousands on both quarterly datasets. For the stock data for Costco, Yahoo Finance was used because it is easily accessible, and it is easily formattable for R (Costco Wholesale Corporation (COST)). From there, R was used to analyze and visualize both datasets.

Similar to the Walmart data, the Costco data needed to be cleaned. The most prominent issue with the original data was that it was in row form instead of columns. While the Walmart data was successfully transposed in R, this method was less successful for the Costco data, and it ended up being simplest to transpose it in Excel. Then, the different variables were modeled in similar fashion to the Walmart data. The code for plotting the Cash and Cash Equivalencies variable can be viewed below (Fig. 9). The as.Date() function is used on the column ReportDate to modify every date into a new column called Datenew, which recognizes every date as a continuous variable instead of each of them being recognized as a unique string. This was used to generate the following graphs using ggplot (Fig. 10) (Fig. 11). The balances increased overall and had some spikes and dips throughout the time as well. Even though there is one outlier on the boxplot for Costco’s quarterly cash equivalent data, it is evident that this is due to the quarterly cash equivalents increasing over time.

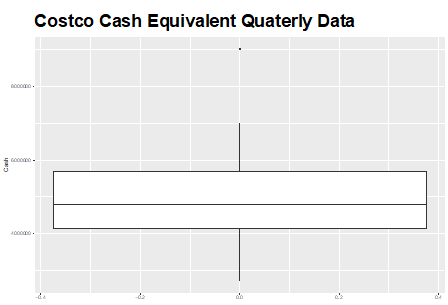
A screenshot of a computer program

Description automatically generated

*Fig. 9*

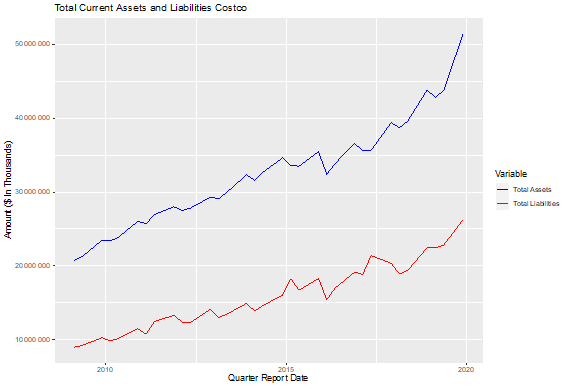


*Fig. 10*

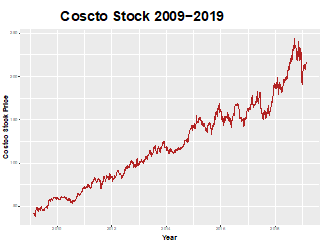


*Fig. 11*

Next, the total current assets and total current liabilities were analyzed and used to generate a line graph (Fig. 12). This graph demonstrates that unlike Walmart, Costco’s total current assets are higher than its total current liabilities. This simply demonstrates that Costco, unlike Walmart, has prioritized minimizing its debts on the books unlike Walmart. However, this does not have as dramatic of an effect as could be expected, as demonstrated in Costco’s stock history (Fig. 13). This figure demonstrates the stock price of Costco from 2009 to 2019. One interesting distinction between this stock and Walmart’s is that Costco’s has generally fewer extreme peaks and valleys than Walmart’s stock history. Also, the general trend of the graph seems to have a very strong correlation to the trend for both the assets and liabilities, and below is shown a comparison between them (Fig. 14).



*Fig. 12*



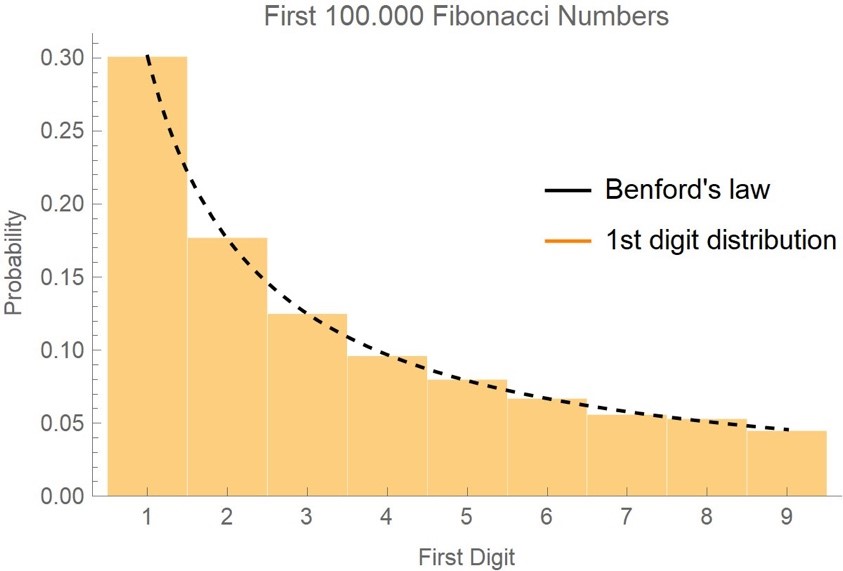
*Fig. 13*

A graph of growth and loss of a company

Description automatically generated with medium confidence

*Fig. 14*

After all of this, it was time to perform the statistical analysis. First, it was necessary to learn about Benford’s Law and its applications. Juan Chang’s dissertation is where most of the knowledge used in this research was gained. Chang describes Benford’s law as a “numerical phenomenon [...] which asserts numerical values for the empirical probabilities of first digits appearing in many lists of numbers.” This means the smaller the number is, the more common it becomes in all the lead digits in a data set (Fig, 15). The proportions of leading digits in a dataset are modeled by P(d) = Log10(1+1/d) where P(d) is the numerical probability based on the d digit.



*Fig. 15*

While the scope of this dissertation is much beyond what was required for this analysis, it gave a complex and deep understanding of Benford’s law. So how is it used in forensic accounting? In the dissertation, Change retroactively applied Benford’s law analysis on multiple different real-world examples. For the purposes of this project, it was most helpful to focus on Bernie Madoff and his financial statements. In short, Bernie Madoff used his legitimate stock broking company to hide a 65-billion-dollar Ponzi scheme. He went to private investors to get money that he would pretend to put into stocks. Due to his immense experience in the stock market, he was trusted to get an above average return on investment compared to index funds like the S&P 500. He then took that money and put it into a personal account that he used to fuel a lavish lifestyle. To cover this up, he fabricated balance sheets and showed them to private investors to show the profit they were making. He then used the money new investors gave him to pay out some of the funds to his first investors. This lasted for over 20 years. Chang applied a Benford’s law analysis to Madoff’s balance sheets and potential returns for his private investors. As expected, Madoff’s numbers deviated significantly from the expected proportions of Benford’s law. The standard practice of Benford’s law is analysis of the first, first and second, and second digits of a dataset. Here are the P-values calculated by Chang’s statistical analysis (Fig. 16).

A table with numbers and text

Description automatically generated

*Fig. 16*

Standard statistical analysis analyzes the P value with an alpha of 0.05, meaning 95 percent significance. Chang performed two different tests as seen above. The reason for there being two different tests is that when performing analysis on Benford’s law, Chi Square tests generally are weaker than the Kolmogorov-Smirnov test. With a standard alpha of .05, you could reach different conclusions about Madoff’s potential fraud.

With all this knowledge, it was time to see how Walmart’s data fared when faced with the same tests. Using R studio and the Benford.analysis package in R, P-values were calculated for both a Chi Square and Kolmogorov-Smirnov tests. The data was read into R, and all non-numeric rows were removed to better read the values into a numeric vector. Then the following statistics were calculated for Chi Square (Fig. 17).

A close-up of a list of different times

Description automatically generated with medium confidence*Fig. 17*

The Benford.analysis package provides a mean absolute deviation of the data when performing Chi Square to better show the data’s conformity to Benford’s law due to the weakness of Chi Square tests. When using the mean absolute deviation and P value for both first and first and second digits, the claim that the data follows Benford's law can be rejected. Next, a Kolmogorov-Smirnov test was performed on the data using the same package in R as above. Using the default 10,000 iterations, R calculated a D statistic of 23.001 and a P value of 2.2 E^-16. Using a null hypothesis, H0 = Walmart’s data from 2009 to 2019 conforms to Benford’s law, the H0 is rejected. There is significant statistical evidence that Walmart’s data deviates from Benford’s law with a 99 percent level of significance.

What does this all mean? While a test for Benford’s law can show forensic accountants that there is a possibility of fraud, more inspection of the data is required to find out if there is fraud. It can be easily concluded from the evidence, that Walmart’s data deviates significantly from Benford’s law. Upon further inspection of the data, a graph was constructed of the first- and second-digit frequency (Fig, 18). As shown, the frequency of digit pairs ending in zero is significantly higher in Walmart’s data. This most likely is a rounding issue or an effort on Walmart’s part to simplify their data for the use of consumers. These could be some reasons why Walmart’s data does not conform to Benford’s law.

A group of graphs with different numbers

Description automatically generated with medium confidence

*Fig. 18*

**Retrospective**

While this project does not prove or even necessarily imply fraud, it does indicate that there may be some concerns with the data Walmart has released. It is unusual that the data would fail testing via Benford’s Law in the way that it does, suggesting that the numbers are wildly different than what is expected, and could be investigated further to determine why this issue occurs. According to the Kolmogorov-Smirnov test, there is also a 99% significance level to suggest that Walmart’s data does not conform to Benford’s Law, which supports this conclusion. However, as mentioned, there could be other reasons for this, such as rounding or simplifying the data with no malintent.

This project required significant statistical analysis. Many of these techniques are commonly used in forensic accounting to uncover the truth in cases such as Madoff’s, or as mentioned earlier, in Toshiba’s. Much of this analysis is greatly simplified by using R, which has several packages specifically designed for such purposes. It was also helpful to see how similar analyses had been run in the past in Chang’s dissertation for similar purposes.

This project also required some knowledge of business operations and financial reporting in order to be able to accurately interpret the data. For example, while Walmart’s current liabilities are consistently higher than their current assets, this is not inherently a bad thing and merely represents where Walmart’s corporate views lie on what amount of debt they should take on. A modest amount of debt can be healthy for a company, and is not necessarily indicative of fraud or corporate failure. This knowledge was also used to compute business ratios, such as the current ratio, return-on-equity ratio, and debt-to-equity ratio.

In this forensic accounting analysis, no fraud was uncovered. However, some discrepancies were found that may merit further investigation to determine if there is a simple explanation or if they are indicative of a corporate issue. Unexpectedly, Walmart’s data does not conform to Benford’s Law, and contains some curious outlier variables in the individual columns. Further statistical analysis could be done to determine if there are other tests the data does not conform to, or to try to find possible explanations for the results of the Benford’s Law analysis. All in all, this study on Walmart’s data required significant research and the use of unique forms of analysis, and was a fascinating investigation into the world of forensic accounting.

**Bibliography**

Bloomenthal, A. (2011, July 20). *Detecting financial statement fraud*.

Investopedia. <https://www.investopedia.com/articles/financial-theory/11/detecting-financial-fraud.asp> ​

Caplan, D. H., Dutta, S. K., & Marcinko, D. J. (2019). Unmasking the fraud at Toshiba. *Issues in Accounting Education*, *34*(3), 41–57. <https://doi.org/10.2308/iace-52429>

​

Chang, J. C. (2017). A Study Of Benford’s Law, With Applications To The Analysis of

Corporate Financial Statements. University of Nebraska. <https://etda.libraries.psu.edu/files/final_submissions/14504>.​

​

Costco Wholesale Corporation (COST) stock historical prices & data - Yahoo Finance. (n.d.).  ​

 Retrieved February 19, 2023, from <https://finance.yahoo.com/quote/COST/history/>​

Du, L. (2015, July 21). *5 Things to Know about Toshiba’s Accounting Scandal*. Wsj.com.

<https://www.wsj.com/articles/BL-263B-5336>

Jaadi, Z. (2021, April 1). *A step-by-step explanation of principal Component Analysis (PCA)*. Built In. <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>

Mergent Online - As reported : Costco Wholesale

Corp. (n.d.). <https://www.mergentonline.com/companyfinancials.php>?      pagetype=asreported&compnumber=16720 &period=Quarters&dataarea=BS&range=50¤cy=AsRep&scale=AsRep&Submit=Refresh&csrftokenmol=a60a01f7ec ​

​

Mergent Online – As reported : Walmart Inc. (n.d.) (2023, October).  ​

 Retrieved February 11, 2023, from <https://www.mergentonline.com/companyfinancials.php?compnumber=8865>

St, S. W. (2019, July 1). *A note on Walmart inc.’s (NYSE:WMT) ROE and debt to equity*. Yahoo

Finance. <https://finance.yahoo.com/news/note-walmart-inc-nyse-wmt-095520157.html>​

​

Tamplin, T. (2022, December 26). *Is current assets the same as total assets?* Finance

Strategists. <https://www.financestrategists.com/accounting/financial-statements/balance-sheet/current-assets/is-current-assets-the-same-as-total-assets/>

*Walmart Inc. (WMT)*. (n.d.). Yahoo.com. Retrieved March 2, 2024,

from <https://finance.yahoo.com/quote/WMT?.tsrc=fin-srch>​

​

*Full dividend history*. (n.d.). Walmart.com. Retrieved March 2, 2024,

from <https://stock.walmart.com/stock-information/dividend-history/full-dividend-history/>​