2016

Improving forensic software tool performance in detecting fraud for financial statements

Brian Cusack
*Auckland University of Technology*, brian.cusack@aut.ac.nz

Tau’aho Ahokov
*Christ's University in Pacific*, tahokovi@bigpond.com

Follow this and additional works at: [https://ro.ecu.edu.au/adf](https://ro.ecu.edu.au/adf)

Part of the Accounting Commons, Forensic Science and Technology Commons, and the Information Security Commons

**Recommended Citation**

Cusack, B., & Ahokov, T. (2016). Improving forensic software tool performance in detecting fraud for financial statements. DOI: [https://doi.org/10.4225/75/58a4f9b604efd](https://doi.org/10.4225/75/58a4f9b604efd)

DOI: 10.4225/75/58a4f9b604efd


This Conference Proceeding is posted at Research Online. [https://ro.ecu.edu.au/adf/163](https://ro.ecu.edu.au/adf/163)
Abstract

The use of computer forensics is important for forensic accounting practice because most accounting information is in digital forms today. The access to evidence is increasingly more complex and in far greater volumes than in previous decades. The effective and efficient means of detecting fraud are required for the public to maintain their confidence in the reliability of accounting audit and the reputation of accounting firms. The software tools used by forensic accounting can be called into question. Many appear inadequate when faced with the complexity of fraud and there needs to be the development of automated and specialist problem-solving forensic software. In this paper we review the context of forensic accounting and the potential to develop improved support tools. The recommendation is for adopting financial ratio analysis as the basis for an improved fraud detection software.

Keywords

Financial Ratio Analysis, Forensic accounting, Forensic Software Tools, Fraud

INTRODUCTION

Before a potential fraud can be investigated, it must be detected. The process of fraud detection involves searching for symptoms that may indicate that fraud exists (Kovalerchuk et al., 2007, p.2). One of the most popular mathematical techniques which are currently used in forensic accounting, is that Benford law. The main issue with this technique is that it takes a relatively broad, rather than narrow approach to detecting fraud. As a result a lot of false signals can occur with the impact of time consumption and increase costs in investigations. Another technique known as relative size factor (RSF), is designed to detect any outliers or unusual data. However, any outlier can be just errors in data entry but not financial fraud. Panighahi (2006, p.3) stated that “RSF is simple to calculate but not an effective and efficient tool”. An abundance of data creates both challenges and opportunities for the forensic accountant. A digital forensics software tool is employed to help deal with the big data problem and to speed up both the accuracy and the completion of investigations. The problem with computer based fraud detection in the field of forensic accounting is that there are significant differences in task performance and knowledge requirements for the completion of investigation. For example, computer forensics requires knowledge of computing systems, log files, graphics and other formats, and many other non-accounting knowledges. Similarly, data theft prevention and investigation requires database knowledge, computer security knowledge, encryption and computer systems (Albright, 2008, p.5).

In the field of forensic accounting, there are categories of investigation such as:

- **Data mining for fraud** Techniques and methodologies for discovering fraud in corporate databases;
- **Financial statement fraud**: Ratio analysis and other methods of finding financial statement manipulation
- **External information sources**: Information about perpetrator finances and other data, usually found in websites; and,
- **Computer forensics**: Investigating by sifting through computer hard drives and other information devices.

Each context has its own challenges and ways of investigation. As a result, digital forensic software tools have to declare capability before use. Otherwise a null or an erroneous result may be obtained from the tool as it is not capable in a particular area, and yet fraud exists within the dataset. Therefore, till testing is a critical issue that impacts potential evidential outcomes. Verification is required from approved by independent bodies and not just proprietary vendor’s (Al Mutawa, et al., 2012, p.26). Researchers in the field suggested that it is vital for investigators to compare the suitability of forensic tools in relation to various application environments (Guo & Slay, 2010, p.297). Some have common views in analysis of large data (e.g. several terabytes) where tools should be able to efficiently and effectively handle the volumes (Yannikos et al., 2011, p.200). Kimmel et al., (2012) recommended that an in – depth assessment of the tool based on requirements and an evaluation of different packages and their functions within all available types of fraud pattern. What is missing from this
advice are statements about how software forensic tools will stay up-to-date and adapt themselves to new patterns of fraudulent activity (p.765).

In this paper we briefly review the problem area of relevancy for current digital forensic and accounting forensic software tools. We then propose a focus for the development of accounting forensic tools that takes ratio analysis and pattern recognition to be the foundational building block for the analysis. We then propose a fraud detection model that is based on the similarity metric for determining the similarity of patterns and the proximity of data to events. We conclude by recommending care in the use of current forensics software tools and emphasise the necessity of tools being able to not only cope with big data but also complexity.

THE PROBLEM AREA

Forensic accounting is the application of accounting knowledge and investigative skills to identify and document potential matters with legal implication for fraud and other financial crime (Houck et al., 2006, p.68). Forensic accounting is a developing area of specialisation in the field of accounting. Its main concerned with the detection and prevention of financial frauds and other forms of economic crime (Dhar and Sarkar, 2010, p.94). In the accounting field, there are a number of ongoing activities that, collectively, allow business owners or managers to access the information when they need it in order to make well informed decisions. In terms of business operation, this includes basic transactions such as purchases and sales, and marketing and strategic planning as well as summative information. Accounting refers to financial record keeping and data reporting that businesses engage in to meet legal requirements and keep the organisation’s stakeholders informed of the organization's financial position at any point of the year. Accounting is defined as a systematic process of identifying, recording, measuring, classifying, verifying, summarizing, interpreting and communicating financial information. It reveals profit or loss for a given period, and the value and nature of a firm's assets, liabilities and owners' equity (Business Dictionary, 2016, p.1). However, the financial statement will only be as good as the journal entries (Haber, 2004, p.7). Missing or fraudulent journal entries will produce financial statement that is fraudulent (Basuhaul, 2010, p.97).

Financial statements are prepared to present fair information about the financial position, operating performance of the organisation or the business. The international standards on auditing 240 stated that, fraud and error must be considered when auditing financial statements (Smith et al., 2008, p.18). Furthermore, the auditor must perform procedures to assist the entity in the detection of fraud (Council, 2013, p.13). As a result of the Enron and WorldCom failures (Reinstein and McMillan, 2004, p.956), the accounting arena had undergone fundamental changes to redress the shortcomings found in previous audit requirements. Therefore, a new market with a new class of accountants known as forensic accountants has emerged. The white collar crimes is in focus and occupational fraud. The Association of Certified Fraud Examiners (ACFE) estimates that occupational fraud losses cost organizations $994 billion annually (Davis et al., 2010, p.5; Crumbley et al., 2005, p.400). Forensic accounting and fraud examination are different but related. Forensic accounting work is done by accountants in anticipation of litigation and can include fraud, valuation, bankruptcy and a host of other professional services. Fraud examinations can be conducted by either accountants or non-accountants and refer only to anti-fraud matters (Wells, 2003, p.76).

DIGITAL FORENSIC SOFTWARE TOOLS

The amount of data stored in accounting files requires digital management. This means that relevant and appropriate software must be available to process the data and to extract the relevant information. Forensic investigation and the analysis of accounting data for fraudulent patterns has become more and more complex (Albano, et al., 2011, p.381). Digital forensics integrates the fields of computer science and law to investigate crime (Dezfouli, et al., 2012, p.186). For any digital evidence to be used in court, investigators must follow a proper set of procedures when collecting and analysing data from computer systems (Jansen & Ayers, 2007, p.6). Hence the interface between the investigator and the data is mediated by software but due to the differences in terms of the technologies, investigators will have to engage different methods and tools depending on the category of information involved (Albano, et al., 2011, p.381). Therefore, prior to acquiring data from a comprised device, extra cautions must be taken, standard procedures and base practises must be followed carefully. This is to avoid altering data in the process because digital data can be easily corrupted (Jansen & Ayers, 2007, p.45). The first challenge for forensic software tools is the bridging of the structures, protocols, and designs in which the data resides. The second challenge is then to be able to identify accurately patterns of conformance, compliance, and aberrations that may occur within the data (Ayers, 2007, p.1). Each tool has a different core set of features that are designed to deliver specific outcomes (NIST, 2013, p.4). Reliable digital forensic techniques are therefore important for prevention, detection, and investigation of electronic crime (Nissan, 2012, p.843). For example, in table 1 and analysis has been completed of different accounting forensic
and digital forensic tools that may be employed in an investigation. The analysis shows that each tool has a different capability and that some do overlap. However, an investigator has to be fully aware of these limitations before a tool is selected or used for forensic accounting purposes (Mohtasebi & Dehghantanha, 2013, 353). For most the integrity of the evidences and its admissibility in the court of law has to be preserved. As a result, it is critical for an investigator to know the reliability and accuracy of the tool (Kubi et al., 2011, p.2) and to select one’s that are to be effective and efficient in the particular situation.

### Table 1. Experiment results (Grispos, et al., 2011, p.30).

<table>
<thead>
<tr>
<th>Item</th>
<th>Type</th>
<th>Logical Acquisition Examination Analyzer (configured)</th>
<th>Physical Scalpel (default)</th>
<th>Foremost (configured)</th>
<th>Simple File Carver</th>
<th>Phone Image Carver</th>
<th>WinHex (modified image)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 docx</td>
<td>N</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>P</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>2 docx</td>
<td>N</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>P</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>3 txt</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>4 log</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>5 txt</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>6 pcap</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>7 pdf</td>
<td>N</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>8 jpg</td>
<td>N</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>N</td>
<td>D</td>
</tr>
<tr>
<td>9 jpg</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>10 jpg</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>11 jpg</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>12 jpg</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>13 jpg</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>14 mtl</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>15 wav</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>P</td>
<td>F</td>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>16 avi</td>
<td>N</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>17 wmv</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>18 sip</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>19-23 Appointments</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>24-28 Contacts</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>29-30 Email Sent</td>
<td>N</td>
<td>F</td>
<td>F</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>F</td>
</tr>
<tr>
<td>31-32 Email Received</td>
<td>N</td>
<td>F</td>
<td>F</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>F</td>
</tr>
</tbody>
</table>

Keys that the authors used in Table 3.2 are F = Full, P = Partial, D = Detected and N = Not.

In spite of everything, all logical data can be acquired and analysed yet, tool developers seem to over claim their tool’s support while the tool can only obtain some of the requirements. As a result, forensic tools should be evaluated based on their abilities and not the costs. For example, a tool may have better support for a particular brand of operating system or device (Morrissey, 2010, p.130). Many of the current software tools for forensic investigation may not be used in financial related investigation beyond the extraction of data because they do not discriminate sufficiently at the higher complexity levels. A forensic accountant requires detection of fraud in financial statement reports which is more than just extracting the report data. As a result, the abstraction of models is required to detect fraud in a financial report.

### ACCOUNTING RATIOS

The initial design of the proposed model was developed based on the learning from the analysis of the literature, formulated to fill the gap, and avoid repetition. The reading and analysis gain sufficient abstraction that the modelling system would rest above the data level. The result is a tentative detection model for financial fraud that requires testing and validation. The initial design of the financial fraud detection model comprises of five financial ratios - **Return on Assets (ROA)**, **Accounts Receivable (A/R) to sales ratio**, **Current Ratio**, **Total Asset**.
**Turnover and Inventory Turnover.** The implication is that for a financial report - which brings together many segregated areas – financial ratios from each of the implicated areas are required comparison against the industry standard or a benchmark from the trend analysis ratios (Albano, et al., 2011). Models are an abstraction of a process to examine potential evidence, irrespective of the originality of the evidence (Peisert, et al., 2008, p.116). Forensic experts also believe that forensic investigation process models generalise an informal procedure to deliver a framework. That framework provides a detailed understanding of what each process is to do and not do. Jankun-Kelly, et al. (2007) explained that the model and framework provides an effective means to acquire information within the process. These processes are used to capture relevant aspects of the investigation (p.357). A breakdown of each ratio for building into the tool is as follows:

**Return on Assets (ROA), Accounts Receivable (A/R) to sales ratio, Current Ratio, Total Asset Turnover and Inventory Turnover:** If an organisation or business is subject to an audit, the auditors will review its accounts receivable in some detail. Accounts receivable is frequently the largest asset that a company has, so auditors tend to spend a considerable amount of time gaining assurance that the amount of the stated asset is reasonable.

Return on Assets: \[
\frac{\text{Profits after taxes}}{\text{Total assets}}
\]

A measure of the return on total investment; It is sometimes desirable to add interest to after tax profits to form the numerator of the ratio since total assets are financed by creditors as well as by stockholders; hence, it is accurate to measure the productivity of assets by the returns provided to both classes of investors.

Current ratio: \[
\frac{\text{Current assets}}{\text{Current liabilities}}
\]

Indicates the extent to which the claims of short-term creditors are covered by assets that are expected to be converted to cash in a period roughly corresponding to the maturity of the liabilities.

Total assets turnover: \[
\frac{\text{Sales}}{\text{Total assets}}
\]

A measure of the utilization of all the firm's assets; a ratio below the industry average indicates the company is not generating a sufficient volume of business, given the size of its asset investment.

Inventory turnover: \[
\frac{\text{Sales}}{\text{Inventory of finished goods}}
\]

When compared to industry averages, it provides an indication of whether a company has excessive or perhaps inadequate finished goods inventory.

**Accounts Receivable (A/R) to sales ratio:** Shows the relationship between unpaid sales and the total sales revenue. It is considered high if it is near to 1.0, because that means a significant amount of cash is tied up with the slow paying customers. Formula: Total accounts receivable (outstanding in an accounting period) ÷ sales revenue (in the same period).

**Accounts Receivable (A/R) turnover =** Average credit sale/Accounts Receivable: Ratio that shows the relationship between unpaid credit sales to total credit sales. It indicates, in general, the effectiveness (or lack of it) of a firm's credit policies and cash collection efforts. Formula: Outstanding accounts receivable (in an accounting period) ÷ credit sales revenue (in the same period), also called receivable turnover.

**FINANCIAL RATIO ANALYSIS**

Financial analysis techniques can help investigators discover and examine unexpected relationships in financial information (Davis, et al., 2010). These analytical procedures are based on the premise that relatively stable relationships exist among economic events in the absence of conditions to the contrary. Known contrary conditions which cause unstable relationships to exist might include unusual or nonrecurring transactions or events, and accounting, environmental, or technological changes. Public companies experiencing these events must disclose and explain the facts in their financial statements. Increasingly, private and not-for-profit companies follow best practices and do the same. Financial ratios are a great way to analyse a company’s strengths and weaknesses. Ratios convert financial information to standardised format that can be used to compare with other companies and industry expectations. Unexpected deviations in relationships most likely indicate errors, but also might indicate illegal acts or fraud (Kovalerchuk et al., 2007). Therefore, deviations in
expected relationships warrant further investigation to determine the exact cause. Several methods of analysis assist the reader of financial reports in highlighting the areas that most likely represent fraudulent accounting methods (Dhar et al., 2010). An understanding of general relationships between certain financial statement balances is necessary to identify relationships that appear unusual. If sales increase, how should the cost of sales respond? If commission expense decreases, what would be expected of sales? Answers to questions such as these are the foundation of financial analysis. The following relationships are general, and traditionally occur between financial accounts; however, unique circumstances may render different results. The following is a worked example:

For the current year, ABC Ltd reported the following ratios:

- ROA = 20%
- Asset Turnover (being Sales/Average Total Assets) = 2.5 times
- Net Profit Margin = 15%

**QUESTION:** Why would we suspect at least one of these ratios is not correct?

We must be able to identify a relationship between the given ratios:

- ROA = Net Income/Average Total Assets
- ASSET TURNOVER = Sales/Average Total Assets
- NET PROFIT MARGIN = Net Income/Sales

Now, we can look at the common factors:

- ROA and ASSET TURNOVER both use Average Total Assets
- ROA and NET PROFIT MARGIN both use Net Income
- ASSET TURNOVER and NET PROFIT MARGIN both use Sales

Now, we can look at the values provided:

- ROA = 0.20
- ASSET TURNOVER = 2.5
- NPM = 0.15

We can then prove these values are not consistent. This can be done either by using algebra, or by simply assuming a value for one of the common factors. In this case, let’s assume average total assets equal $100 (although any value can be chosen to prove if the ratios are consistent).

If that is the case; Net income/100 = 0.20 (so Net Income = 20)

Sale / 100 = 2.50 (so Sales = 250)

Then: Net Income / Sales would have to be: 20 / 250 = 0.08

Conclusion: NPM of 15% is not consistent with the values of ROA = 20% and Asset Turnover = 2.5 times.

These ratios may also reveal frauds other than accounting frauds. If an employee is embezzling from the company’s accounts, for instance, the amount of cash will decrease disproportionately and the current ratio will decline. Liability concealment will cause a more favourable ratio. Similarly, a cheque-tampering scheme will usually result in a decrease in current assets, namely cash, which will, in turn, decrease the current ratio. In fact, these frauds may be more easily detected with ratio analysis because employees other than management would not have access to accounting cover-ups of non-accounting frauds. Anomalies in ratios could point directly to the existence of fraudulent actions. Accounting frauds can be much more subtle and demand extensive investigation beyond the signal that something is out of the norm.

**THE PROPOSED FRAUD DETECTION MODEL**

The Euclidean Distance based similarity metric is to evaluate the similarity of the standard industry ratios (Return on Assets (ROA), Accounts Receivable (A/R) to sales ratio, Current Ratio, Total Asset Turnover and Inventory Turnover) against the ratios of the current financial statement in order to determine if a fraud has occurred. Bajcsy and Kovačič (1989) argued that defining the problem will be the best way to understand the nature of the problem in its entirety. To evaluate the similarity between two different objects, x and y, a distance metric known as Euclidean Distance (EU) is used, this defines as follows: $EU(x,y) = \sqrt{(x - y)^2}$ (1)
This metric can be generalized into n-dimensions points, such that \( a=\{x_1, x_2, \ldots, x_n\} \) and \( b=\{y_1, y_2, \ldots, y_n\} \). In this case, n-dimensions EU metric is defined as:

\[
EU(a, b) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}
\]  

(2)

Let \( L_1 \) and \( L_2 \) be the existing standard industry ratios and the current financial statement ratios, respectively. Let \( x_i \) represent each ratio from the standard industry ratios and \( y_i \) represent each ratio from the financial statement, where \( i=\{1, 2, \ldots, n\} \) and \( n \) is the total number of ratios. In this case, \( L_1=\{x_1, x_2, \ldots, x_n\} \) and \( L_2=\{y_1, y_2, \ldots, y_n\} \).

Euclidean distance can be normalized into a distance based similarity as follow:

\[
S = \frac{1}{1 + EU(L_1, L_2)}
\]  

(3)

Similarity normalized \( EU \) into a value in between 0 and 1, where a value of 1 means that the two objects are identical and a value of 0 means that the two objects are not identical. This study focuses on detecting financial fraud and identifying the means of the fraud. In order to detect the fraud, the similarity between the industry standard ratios and the ratios of the current financial report is ranked. In doing so, the Euclidean distance between \( L_1 \) and \( L_2 \) is calculated first by using equation (1) and then the similarity can be ranked based on equation (3).

For example:
Using equation 1, if the industry standard for Return on assets: profit after tax/total assets = 1.5 and in the current financial report its 2.5 therefore it is calculated as follows:

\[
EU(x, y) = \sqrt{(1.5 - 2.5)^2} = \sqrt{1} = 1
\]

In order to normalise this into a value in between 0 to 1, use the similarity metric as showed in equation 3

\[
S = \frac{1}{1 + EU(x, y)} = 0.5
\]

The result is interpreted from predefined tabulated charts or user constructed tables based on empirical sample data constructed from use cases. The default setting is a result that falls between 0.5 and 1; and it indicates acceptance, but if it falls below 0.5 then it signals a red flag. A red flag means that the data requires further investigation. The metric acts as an indicator that eliminates possibilities rather than a deterministic measure that nominates an outcome. It cannot be used in isolation from expert knowledge and practitioner experience but in large datasets it can heighten the awareness of variations that are most likely indicate errors, and illegal acts or fraud. Therefore, deviations in expected relationships warrant further investigation to determine the exact cause.

CONCLUSION

Current digital forensics tools when applied to forensic accounting came up inadequate in our testing because they lack models that can sufficiently abstract from the data concerned. In this paper, the distance based similarity metric was proposed as the method for detection and identification of fraud, and a solution to the complexity problem. The result from the simple case developed to validate the method showed that the distance based similarity metric can detect financial statement variations that are an advancement on the simple ratio model tests and these variations can be mapped onto patterns of different activities, including fraud. It is also evident in this study that the metric effectively improved the performance, effectiveness and efficiency of the examination and analysis of large datasets of financial statements. This is helpful given the big data issues surrounding accounting and audit practice. For future work, a use case database of reference tables is to be developed from the application of this tool in industry so that the normalisation measure may be better matched onto scenario and context based situations. In its current form the metric is an improvement on the current ratio detection systems and has potential for coding into other digital forensic tools or into its own accounting forensic tool.
REFERENCES


