Knowledge-Based Systems 89 (2015) 459–470

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments

Chi-Chen Lin [a,1], An-An Chiu [b,2], Shaio Yan Huang [c,3], David C. Yen [d,*]

[a] Department of Accounting Information, National Taipei University of Business, No. 321, Sec. 1, Jinan Rd., Zhongzheng, Taipei, Taiwan
[b] Department of International Trade, Feng Chia University, No. 100, Wenhwa Rd., Seatwen, Taichung 40724, Taiwan
[c] Department of Accounting and Information Technology, National Chung Cheng University, 168 University Rd., Min-Hsiung, Chia-Yi 62102, Taiwan
[d] School of Economics and Business, 226 Netzer Administration Bldg., SUNY College at Oneonta, Oneonta, NY 13820, United States

ARTICLE INFO

Article history:
Received 4 November 2014
Received in revised form 14 August 2015
Accepted 18 August 2015
Available online 24 August 2015

Keywords:
Fraud factor
Fraud triangle
Data mining

ABSTRACT

The objective of this study is to examine all aspects of fraud triangle using the data mining techniques and employ the available and public information to proxy variables to evaluate such attributes as pressure/incentive, opportunity, and attitude/rationalization, based on the findings from prior studies in this subject field and also the Statement on Auditing Standards. The second objective is to discuss whether or not the suggestion of the experts agrees with the results obtained from adopting those novel techniques. In specific, this study uses both expert questionnaires and data mining techniques to sort out the different fraud factors and then rank the importance of them. The data mining methods employed in this research include Logistic Regression, Decision Trees (CART), and Artificial Neural Networks (ANNs). Empirically, the ANNs and CART approaches work with the training and testing samples in a correct classification rate of 91.2% (ANNs) & 90.4% (CART) and 92.8% (ANNs) & 90.3% (CART), respectively, which is more accurate than the logistic model that only reaches 83.7% and 88.5% of the correct classification in assessing the fraud presence. In addition, type II error of ANNs drops significantly to 23.9% from 43.3% and 27.8% compared to the ones using CART and logistic models. Finally, the differences between different data mining tools and expert judgments are also compared to provide more insights as a research contribution.

1. Introduction

After the occurrence of several major scandals (e.g., Enron Corp., Tyco, and WorldCom Inc.), the loss of market capitalization resulting from the reported financial statement fraud is estimated to be about $460 billion [39]. In 2014, Association of Certified Fraud Examiners (ACFE) reported that the U.S. organizations lose almost 5 percent of their revenue due to fraud, and the Gross Domestic Product (GDP) based annual fraud estimate for U.S. alone is around $3.7 trillion (ACFE, 2014). Sorkin [41] reported that there are 343 criminals and 189 civil defendants involved with fraudulent activities which have harmed more than 120,000 victims with a value of more than $8 billion in recent years in the United States. Financial fraud is becoming an increasingly serious problem and as a result, effective detecting accounting fraud has always been an important but rather complex task for accounting professionals [29,13,37,34]. Examining the financial fraud is in fact one of the hot issues given that the economic and social fallouts from the fraud can be massive [22]. After AICPA issued SAS No. 82, a greater responsibility has been imposed onto the auditors to detect fraud in general, and in dealing with the effective management of fraud in particular. However, this aforementioned act did not provide more specific and objective guidelines. Following the issuance of SAS No. 99 and Sarbanes–Oxley Act, the aim of preventing fraud with a more rigorous internal control oversight is placed as a major focus and it has stimulated and inspired the numerous academic studies [42,33,12,18] in this subject area.

A prolific area of prior research has focused on using different tools and techniques to detect frauds such as analytical procedures, ratio analysis, regression analysis, score propagation over an auction network (SPAN) and checklists to improve the fraud detection [16,19,48]. However, the previous studies may result in too many fraud risk factors to identify the importance of each fraud

* Corresponding author. Tel.: +1 607 436 3458 (office); fax: +1 607 436 2543.
E-mail addresses: c97ve47@yahoo.com.tw (C.-C. Lin), aachiu@fcuoa.fcu.edu.tw (A.-A. Chiu), actsyh@yahoo.com.tw (S.Y. Huang), David.Yen@oneonta.edu (D.C. Yen).
[1] Tel.: +886 2 23226362.
[2] Tel.: +886 4 24517250x4076.
[3] Tel.: +886 5 2720411x34501; fax: +886 5 2721197.

http://dx.doi.org/10.1016/j.knosys.2015.08.011
0950-7051/

factor. Nevertheless, to identify and to rank the importance of the fraud risk factors becomes a critical issue since the limited budgets are always one of the main concerns encountered by today's businesses. This paper tries to rank the importance of frauds to provide the solutions to meet the aforementioned challenges of the limited budget and restricted resources. The rank of the importance of financial frauds may provide a significant advantage to auditors and managers in enhancing the efficiency of fraud detection and critical evaluation.

Nowadays, auditing practices have to be conducted in a timely manner to cope with an increasing number and occurrence of financial statement fraud cases. The novel techniques such as data mining, claims that it has advanced classification and prediction capabilities and can be employed to facilitate auditors' role in terms of successfully accomplishing the task of fraud detection. There has been a limited use of data mining techniques for the detection of financial statement frauds [38]. Data mining plays an important role in financial fraud detection, as it is often applied to extract and uncover the hidden truths behind the very large quantities of data [29]. Lin et al. [23] conducted an experts' questionnaire survey to evaluate the fraud factors using Lawshe's approach. The result of this expert questionnaires shows that 32 factors can be regarded as the suitable measurements for the continuing assessment of fraud detection. Following the study of Lin et al. [23], the first objective of this study is to use different tools and techniques such as logistic regression model and data mining to examine the ranking of the fraud factors and test out the effectiveness of the fraud detection tools by using the published financial data.

Furthermore, Lin et al. [23] listed the top five fraud factors including "Poor performance", "The need for external financing", "Financial distress", "Insufficient board oversight", and "Competition or market saturation" by sequence. However, the judgments of the experts were merely made according to their own experience and specialized knowledge. To resolve this limitation, the second objective of this paper is to discuss whether or not the suggestion of the experts agrees with the result obtained from adopting those novel techniques such as logistic regression model and data mining. It is the authors' hope to use these aforementioned techniques to verify the judgments of the experts to figure out what will be the real financial situation to deal with. In addition, most of previous studies tend to use surveys or subjective measurements to identify the fraud factors and by doing so, the data sets are unavailable to other researches or users so that it is difficult to perform the empirical research to verify the correctness. To bridge this gap, this proposed study uses the public information to proxy variables measurement and consequently, the results can be available for other researches or users for a public scrutiny [24].

2. Literature review

2.1. Fraud triangle

The fraud triangle theory is developed by Cressey [10] and this theory has been widely used by professionals as a useful, theoretical model to explain why most frauds occur. This theory posits that the fraud is likely to occur because of the availability of one or more of the three elements (e.g., pressure, opportunities, or rationalization) of the fraud triangle [1]. Sixty percent of all fraud incidents involved an insider [36]. Srivastava et al. [44] indicated that in the accounting profession, there has been an increased attention on the responsibility of the auditors to adequately assess the risk of fraudulent financial reporting. In fact, the newest fraud standards (e.g., SAS 99, ISA 240, ASA 240, and/or TSAS 43) about fraud risk factors are all based on "the fraud triangle". Understanding the fraud triangle is essential to evaluating financial fraud [38]. The fraud triangle describes the probability of financial reporting fraud which depends on three factors: incentives/pressures, opportunities, and attitudes/rationalization of financial statement fraud. Gozman and Currie [14] suggested that the potential for fraud is increased where there are incentives, often in the form of the need to meet targets or hide losses. The management will face the incentives or pressures to resort to fraudulent practice. Opportunity exists, for example, the absence of controls or ineffective controls that provide an opportunity for fraud to be perpetrated. Rationalization depends on the individuals and the circumstances they are facing and occurs when the perpetrator constructs a justification for the fraud.

2.2. Experts' decision

Because of the limited budgets, how to identify the fraud factors and rank the importance of those fraud factors becomes a critical issue. Prior researches determine the relative importance of fraud factors by using AHP (Analytic Hierarchy Process) in order to determine the relative weightings of each individual item. Apostolou and Hassell [2] used experts' decision such as Big5 auditors, internal auditors, and accounting academics through AHP to determine the relative importance of the 14 fraud risk factors identified in SAS No. 53. Further, Apostolou et al. [3] provided 25 red flags identified in SAS No. 82. They used the experts' decision technique to assess the relative critical fraud factors in three factor group including management characteristics and influence over the control environment, industry conditions, and operation and financial stability characteristics. Mock and Turner [27] also examined the response to SAS No. 82 from three large international audit firms. Of the three audit firms examined, they found that two attempted to reach an assessment through some form of formal scoring system.

Fraud risk factors in the newest fraud standards (SAS 99, ISA 240, ASA 240, and TSAS 43) are all based on "the fraud triangle". Lin et al. [23] used Lawshe's approach and 32 factors are considered by experts to be the measurements suitable for the continuing assessment of fraud detection. The same study further adopts AHP in calculating the weightings of individual measurement items to rank the importance of factors for three aspects of the fraud triangle. Their research indicated that in the fraud triangle dimension, the highest weight is "Pressure/Incentive", and the next is "Opportunity", while the lowest one is "Attitude/rationalization". In terms of the category in each dimension, the top five most important factors are "Poor performance", "The need for external financing", "Financial distress", "Insufficient board oversight", and "Competition or market saturation" by sequence. In specific, 11 of 32 factors belong to the pressure/incentive dimension, the other 15 factors belong to opportunity dimension and the last 6 of 32 factors belong to the attitude/rationalization dimension. In addition, the same study utilizes the Analytic Hierarchy Process (AHP) to calculate the weightings of individual measurement items and then, rank the importance of factors to form the three aspects of fraud triangle.

Experts do make judgments according to their work experience and professional knowledge. The results of experts' decision in relative importance of fraud risk factors might be different from the real situations. To bridge this gap, this research hopes to analyze these differences through data mining technique [24].

2.3. Detecting tools

Traditional analytical review, which mainly involves with the ratio analysis, has yielded a rather limited success in identifying the fraud. One of the problems with using ratio analysis is related to the subjectivity involved in the identification of the ratios that are likely to indicate a fraud [16,17]. The study of Nigrini and Mittermaier [30] discussed various analytical procedures which