Journal of Forecasting, *J. Forecast.* **33**, 611–626 (2014)
Published online 1 October 2014 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.2294

Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China

XIN-PING SONG,[1,2] ZHI-HUA HU,[3]* JIAN-GUO DU[1,2] AND ZHAO-HAN SHENG[2]
[1] *College of Business and Management, Jiangsu University, Zhenjiang, China*
[2] *College of Engineering and Management, Nanjing University, Nanjing, China*
[3] *Logistics Research Center, Shanghai Maritime University, Shanghai, China*

ABSTRACT

This study presents a method of assessing financial statement fraud risk. The proposed approach comprises a system of financial and non-financial risk factors, and a hybrid assessment method that combines machine learning methods with a rule-based system. Experiments are performed using data from Chinese companies by four classifiers (logistic regression, back-propagation neural network, C5.0 decision tree and support vector machine) and an ensemble of those classifiers. The proposed ensemble of classifiers outperform each of the four classifiers individually in accuracy and composite error rate. The experimental results indicate that non-financial risk factors and a rule-based system help decrease the error rates. The proposed approach outperforms machine learning methods in assessing the risk of financial statement fraud. Copyright © 2014 John Wiley & Sons, Ltd.

KEY WORDS financial statement fraud; fraud risk assessment; fraud risk factors; machine learning; rule-based system

INTRODUCTION

The problem of financial statement fraud

Enterprise risk management is increasingly important in all aspects of business. Numerous researchers and practitioners have attempted to identify methods to monitor and control enterprise risks. Hazard risk, financial risk, operational risk and strategic risk are the typical risks (Wu and Olson, 2009, 2010a,2010b; Wu *et al.*, 2010). Because of the importance of financial statement fraud, identification and management of the risks associated with financial statement fraud help to reduce enterprise risks. Fraudulent financial statements have become increasingly common (Spathis, 2002). Furthermore, surveys have predicted that, because of the ongoing global economic recession since 2008, financial statement fraud becomes increasingly prevalent. The statement on audit standards (SAS) issued by the American Institute of Certified Public Accountants (AICPA) requires auditors to assess the risk of financial statement fraud and design appropriate audit procedures for such assessments. The failure to detect financial statement fraud imposes adverse legal consequences on auditors, and the related people, companies and governments will face substantial economic penalties (Wells, 1997).

Numerous investigations have contributed to the examination of risk factors and the auditing decision aids for fraud assessment (Beasley, 1996; Bell and Carcello, 2000; Apostolou and John, 2001; Asare and Wright, 2004; Cha and Wu, 2006). However, assessing the risk of financial statement fraud involves complex decision making. Most auditors suffer problems when they assess the risk of financial statement fraud, probably because they lack knowledge of its characteristics (Wells, 2001).

Methods from machine learning have been used for fraud detection, including detection of corporate fraud, credit card fraud, money laundering, mortgage fraud, mass marketing fraud and commodities fraud. However, few studies have examined financial statement fraud detection (Koskivaara, 2004; Dowling and Leach, 2007; Ngai *et al.*, 2011), and those that did so failed to reach a consensus on the usefulness to auditors of related decision tools, such as logistic regression models and expert systems (Eining *et al.*, 1997; Tseng and Chang, 2006). Moreover, Thomas *et al.*, (2002) claimed that the use of neural networks for fraud detection was not accurate when applied to bankruptcy prediction.

The studied problem

The detection of financial statement fraud is of great concern in China because fraud is a significant problem in China's unique transition economy and immature corporate governance. To reduce fraud, the Chinese government introduced commercial laws and attempted to strengthen the regulatory effectiveness and enhance corporate governance (Cha and Wu, 2006). Additionally, assessments of financial statements are increasingly demanded because of a rapid increase in the number of listed companies.

* Correspondence to: Zhi-Hua Hu, Logistics Research Center, Shanghai Maritime University, Shanghai 200135, China. E-mail: zhhu@shmtu.edu.cn

China has recently developed models for financial statement fraud detection. Some studies using Chinese data have examined the relationship between governance structure and fraud risk (Fang, 2003; Liu and Du, 2003). Meanwhile, others have developed fraud detection models using statistical methods and neural networks (Cha and Wu, 2006; Qiao and He, 2007; Gu and Feng, 2009). Moyes *et al.* (2005) found that cross-cultural differences can influence the understanding and use of risk factors. This work uses Chinese company data to identify risk factors and develop models for the assessment of financial statement fraud.

The objective of this work is to examine the effectiveness of the proposed approach in the assessment of the risk of financial statement fraud. The proposed approach integrates a system of financial and non-financial risk factors, and a hybrid assessment method that combines machine learning methods and a rule-based system. This study focused on comparisons among four classifiers, both individually and in combination. It then applies the proposed approach to improve the prediction results obtained from the classifier system.

The remainder of this paper is organized as follows. The next section reviews the relevant literature. The third section then presents an integrated risk assessment approach. The fourth section briefly introduces the machine learning methods used in this study. Next, the fifth section discusses the research methods. Subsequently, the sixth section 6 analyzes the experimental results, and the final section presents conclusions.

RELATED STUDIES

This study reviews three streams of literature on risk management in relation to financial statement fraud comprises: identification of risk factors (red flags), fraud risk assessment methods and fraud risk prevention measures.

Fraud risk factors

Auditing practitioners and researchers have explored the usefulness of fraud risk factors (red flags). AICPA issued several statements on auditing standards (SAS)—SAS 53 in 1998, SAS 82 in 1997 and SAS 99 in 2002—that clarified the responsibility of auditors for fraud detection. Furthermore, in 2002, the International Auditing and Assurance Standards Board (IAASB) of the International Federation of Accountants (IFAC) issued the International Standards on Auditing (ISA) 240. Following ISA 240, the Auditing Standards Boards (ASBs) of other countries issued their own standards. These auditing standards outlined numerous fraud risk factors.

Following the AICPA-issued auditing standards, Loebbecke *et al.* (1989) created a fraud prediction model based on the risk factors identified in SAS 53. Subsequently, most studies used questionnaires and checklists to study fraud risk factors, which were included in SAS 53 and SAS 82 (Beasley, 1996; Pincus, 1997; Apostolou and John, 2001). Although the red flags provide useful information on fraud, Wilks and Zimbelman (2004) showed that they lack management integrity and do not provide an effective basis for subjective judgments.

Subsequently, the usefulness of financial indicators in fraud detection was examined. Persons (1995) studied the likelihood of financial statement fraud, and found financial leverage, capital turnover, asset composition and firm size to be significant factors for fraud detection. Non-financial indicators were also suggested to be used in financial statement fraud detection models. Apostolou and John (2001) found that management characteristics are crucial to fraud indicators, and a computational model using only financial ratios cannot adequately detect financial statement fraud.

Recent investigations have explored the effectiveness of all SAS risk factors in various contexts. Moyes *et al.* (2005) examined the effectiveness of the 42 SAS 99 red flags in the detection of financial statement fraud, and classified them into three categories (very effective, average effectiveness and ineffective) using statistical testing. Of the 42 red flags, 11 relate to financial indicators, of which four were classified as very effective and seven as not effective. This result suggests that financial and non-financial factors offer important clues for fraud detection. This statement holds especially true for listed companies in China, which suffers from weak corporate governance (Chen *et al.*, 2006). Additionally, the effectiveness of the SAS red flags was shown to vary across countries (Moyes *et al.*, 2005).

Fraud risk assessment methods

Because SAS provides few guidelines on how to use the information in red flags for judgment formation, in practice auditors need decision aids to improve their judgments. Generally, fraud detection requires auditors to use analytical review methods that enable estimations of account balances without examining relevant individual transactions. Analytical review methods are classified as non-quantitative, simple quantitative and advanced. Advanced quantitative methods include sophisticated methods derived from statistical and machine learning methods.

Previous studies demonstrated that questionnaires and checklists based on risk factors in SAS help auditors identify risk factors (Pincus, 1997). Asare and Wright (2004) subsequently argued that questionnaires and checklists are less effective because they represent subjective judgments.

Subsequently, statistical methods were applied to assess management fraud and were reported to be generally more consistent and accurate than human judgment, because they could convert red flag cues into assessments

(Eining *et al.*, 1997). Logistic regression models were used to examine the relationship between the likelihood of financial statement fraud and specific fraud risk factors such as composition of boards of directors, insider trading and the existence of an independent audit committee (Beasley, 1996; Summers and Sweeney, 1998; Abbot *et al.*, 2000). Bell and Carcello (2000) estimated the likelihood of financial statement fraud for a sample of 77 fraud and 305 non-fraud cases by using a logistic regression model.

Spathis (2002) proposed a logistic regression model that achieved prediction accuracy exceeding 84%. Spathis *et al.* (2002) later reported that the UTADIS method, which is based on a non-parametric regression framework, outperformed statistical methods. Both studies used a balanced sample of 76 manufacturing firms.

Advanced quantitative methods based on machine learning methods were applied to financial statement fraud detection. Eining and Dorr (1991) developed a knowledge-based system that used red flag cues to assess financial statement fraud risks. Eining *et al.* (1997) found that an expert system could outperform approaches based on checklist and logistic regression models. Neural networks were found to be more effective for fraud detection than standard statistical methods (linear and quadratic discriminate analysis and logistic regression) (Green and Choi, 1997; Fanning and Cogger, 1998). Thomas *et al.* (2002) summarized the studies on auditing and fraud risk assessment, and several notable deficiencies of data mining methods. Jerry *et al.* (2003) applied an integrated fuzzy neural network (FNN) to fraud detection. Furthermore, Kotsiantis *et al.* (2006) used discriminant analysis (DA), logistic regression (LR), C4.5 decision tree (DT), K-nearest neighbor (KNN), Bayesian networks and sequential minimal optimization (SMO) to predict financial statement related fraud.

Glancy and Yadav (2011) designed a computational fraud detection model using SAS Enterprise Miner. The model used a quantitative approach for textual data and document clustering. Humpherys *et al.* (2011) examined the strategic use of deceptive language in risk assessment of financial statement fraud using linguistic cues extracted from 202 publicly available financial disclosures. The model incorporating Naive Bayes and C4.5 DT achieved the highest classification accuracy. Ravisankar *et al.* (2011) used a neural network, SVM, genetic programming and logistic regression to identify financial statement fraud. Furthermore, Yen (2007) used the unsupervised adaptive resonant theory to detect accounting frauds, and proposed an approach to identify potential fraud based on warning signals.

INTEGRATED APPROACH

Risk of financial statement fraud

Financial statements provide a formal record of the financial activities of a business entity. Financial statements typically comprise four basic types: statement of financial position, statement of comprehensive income, statement of changes in equity and statement of cash flows. Statements may also include an extensive set of notes, explanation of financial policies and management analysis. Listed companies typically issue financial statements in three forms, namely quarterly, half-yearly and yearly reports. This experimental study used annual financial statements because they include audit opinions. The data in the financial statements of listed Chinese companies indicate liability, solvency, asset structure, efficiency, growth and profitability.

Methods used for fraud comprise three types: first, misstating margins, inventories and earnings; second, overstating inventory; and third, recording revenues and receivables from vendors as budgeted rather than actual amounts. Unlike unintentional misstatements, financial statement fraud involves the calculated and intentional misstatement of financial data. According to the definition of risk, financial statement fraud risk is defined as the uncertainty concerning the occurrence of financial statement fraud. Financial statement fraud risk represents the likelihood of the occurrence of financial statement fraud. It is difficult to predict whether a company will perform financial statement fraud. Computational models were generally reported to be more accurate than human judgment, because they can mechanically weigh and combine risk factors to create comprehensive assessments. Financial statement fraud is essentially a type of enterprise risk. Although predicting fraud risk is difficult, its occurrence can be controlled and monitored.

Fraud risk factors based on the fraud triangle theory

Loebbecke *et al.* (1989) conceptualized the judgments regarding financial statement fraud risk based on the fraud triangle theory. This theory includes three components: motivation, condition and attitude. This theory subsequently became the core of all fraud auditing standards, including SAS.

Two significant problems exist with the use of SAS standards. First, different versions of SAS involve different fraud risk factors. In 1997, SAS 53 described 14 red flags and SAS 82 added 25 red flags. SAS 99 revised the set of red flags to total 42. Meanwhile, SAS 82 paid more attention to a control environment, whereas SAS adopted a different approach by paying more attention to fraud-related factors. Although the AICPA listed a number of red flags in SAS 53 and SAS 82, only SAS 99 categorized the risk factors based on the fraud triangle theory. Furthermore, SAS is a US-based Statement of Auditing Standard. The perceived effectiveness of each red flag in SAS differs across countries. Thus a proper system of fraud risk factors and corresponding assessment methods suitable for Chinese listed companies must be developed.

Limitations of prior studies

Although SAS standards require auditors to assess the risk of financial statement fraud, they provide few guidelines on how to use these risk factors for detection of fraud risk. Additionally, the previous literature presents limitations concerning risk factors and detection methods. First, one group of studies examined how personal judgment or the logistic regression model can be used to combine the risk factors identified in SAS into an overall assessment (Eining and Dorr, 1991; Eining *et al.*, 1997; Bell and Carcello, 2000). The risk factors are frequently given by the binary values based on questionnaire responses. Moreover, only around 10 red flags in SAS relate to financial indicators, yet in reality several dozen financial ratios can be obtained from financial statements. Therefore, the available financial ratios are not fully used to detect financial statement fraud.

The second group of studies examined the development of fraud prediction models based on financial ratios listed in financial statements and various computational methods (Green and Choi, 1997; Fanning and Cogger, 1998; Glancy and Yadav, 2011). Obviously, the financial indicators used there do not correspond to the red flags in SAS standards. Furthermore, non-financial factors are not used despite being crucial indicators in fraud detection. Designing an integrated framework that includes risk assessment factors and assessment methods thus deserves further exploration.

The proposed method

The proposed new method for risk assessment of financial statement fraud comprises a system of risk assessment factors and a cascaded assessment model.

The system of assessment factors is derived from a set of original risk factors in SAS. SAS 99 provides numerous risk factors grouped into three categories: motivation (16), condition (14) and attitude (12). Within this framework, the original risk factors were further reorganized and reclassified into four groups: financial stability and operating characteristics (F), management characteristics and control environment (M), governance and organization structure (G) and relation and industry conditions (R/I). Figure 1 presents a conversion that maps the original risk factors in SAS into new risk indicators. Table I lists nine 'F' group red flags, 17 'M' group red flags, four 'G' group red flags and six 'R/I' group red flags. Consequently, the initial four groups of red flags were converted into 23 'F' group indicators, 17 'M' group indicators, four 'G' group indicators and six 'R/I' group indicators, respectively.

The proposed system of fraud risk factors has two features. First, it is mainly concerned with financial indicators because they are most likely correlated with fraud risk. Essentially, the symptoms of fraud motivation are inevitably reflected in financial ratios, which are easily accessible from financial statements. Second, the control environment and governance structure are major concerns because China has a relatively underdeveloped capital environment when compared to the USA. According to the audit technology for detection of financial fraud risk (ATW No. 1) issued by the Chinese Public Accountant (CPA) committee, imperfect governance and internal environment must be seriously considered when assessing financial statement fraud risk. This study thus obtained 'M' group indicators from the corresponding original red flags and extended 'G' group indicators based on the research of Cha and Wu (2006).

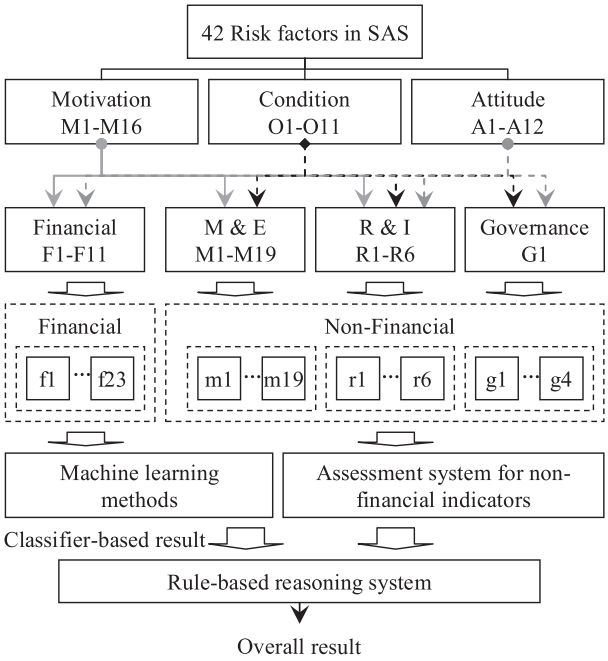

Figure 1. The proposed risk assessment framework for financial statement fraud

Consequently, the four extended 'G' group indicators are 'the number of board meetings' (G1), 'the proportion of outside directors' (G2), 'the duality of the chairman and CEO position' (G3) and 'setting auditing committee' (G4).

The cascade model provides a flexible assessment approach that can convert identified risk factors into an overall assessment. The approach comprises three stages. The first stage applies financial-related factors and machine learning methods to assess each financial component. According to the SAS standards, the auditor should use professional judgments to assess the significance and relevance of fraud risk factors based on consideration of concrete circumstances. Professional knowledge is thus used during assessments in the second and third stages. The second stage utilizes the fraud risk factors related to non-financial status and a rule-based system to assess each non-financial component (i.e. management characteristics, control environments and industrial status). The third stage utilizes four

Table I. A list of fraud risk factors in SAS 99

		Risk factors (red flags)	Category after conversion			
C	1.	Formal or informal restrictions on the auditor that inappropriately limit his access to people or information or limit his ability to communicate effectively with the board of directors or the audit committee			R1	
	2.	Significant related-party transactions not in the ordinary course of business or with related entities that are not audited or audited by another firm			R2	
	3.	Domination of management by a single person or small group in a non-owner-managed business without compensating controls		M1		
	4.	Ineffective accounting and information systems		M2		
	5.	Inadequate monitoring of significant internal controls		M3		
	6.	Ineffective board of directors or audit committee oversight over the financial reporting process and internal control system				G1
	7.	High turnover rates of ineffective accounting, internal audit, information technology staff		M4		
	8.	Significant bank accounts or subsidiary or branch operations in tax haven jurisdictions for which there appears to be no clear business justification			R3	
	9.	Assets, liabilities, revenues or expenses based on significant estimates that involve subjective judgments or uncertainties that are difficult to corroborate	F1			
	10.	High turnover of chief executive officers or board of directors		M5		G1
	11.	Difficulty in determining organization or individuals have controlling interest in entity				
	12.	Overly complex organizational structure involving unusual legal entities or managerial lines of authority				
	13.	A strong financial presence or ability to dominate a certain industry sector that allows the entity to dictate terms or conditions to suppliers or customers that may result in inappropriate or not arm's length transactions			R4	
	14.	Significant operations located or conducted across international borders in jurisdictions where differing business environments and cultures exist				
M	1.	Excessive motivation on management or operating personnel to meet financial targets set up by the board of directors or management		M6		
	2.	Significant portions of managements compensation, represented by bonuses and stock options, being contingent upon achieving aggressive targets for stock price, operating results, financial position or cash flow	F2	M7		
	3.	Rapid growth or unusual profitability, especially compared to that of other companies in the same industry	F3			
	4.	Unrealistic profitability or trend-level expectation by management in overly optimistic press releases or annual report messages		M8		
	5.	Recurring negative cash flows from operations or an inability to generate cash flows while reporting earnings and earnings growth	F4			
	6.	Adverse effects of reporting poor financial results on significant pending transactions, such as business combinations or contract awards				
	7.	Unrealistic profitability or trend-level expectations of external parties (investment analysts, creditors) in overly optimistic press releases or annual report messages	F5			
	8.	Management and/or board directors have personally guaranteed significant debts	F6			
	9.	Operating losses causing threat of imminent bankruptcy or foreclosure, or takeover	F7			
	10.	Management and/or board directors holding significant financial interests in the entity				
	11.	Marginal ability to meet exchange listing requirements or debt repayment	F8			
	12.	High vulnerability to rapid changes in technology, product obsolescence				
	13.	High degree of competition or market saturation accompanied by declining margins	F9		R5	
	14.	Significant declines in increasing business failures in the industry			R6	
	15.	Need to obtain additional debt or equity financing of major research and development or capital expenditures to stay competitive	F10			
	16.	New accounting, statutory, or regulatory requirements				

Table I. *(Continued.)*

A	1.	Significant, unusual, or highly complex transactions, especially occurring close to year end that pose difficult 'substance over form' questions		M9
	2.	Domineering management behavior in dealing with the auditor, especially involving attempts to influence the scope of the auditor's work		M10
	3.	Known history of violations of securities law, or claims against the entity, its senior management or board directors' violations of securities laws		M11
	4.	Ineffective communication, implementation, support or enforcement of the entity's values or ethical standards by management or the communication of inappropriate values or ethical standards		M12
	5.	Frequent disputes with the current or previous auditor on accounting, auditing, or reporting matters		M13
	6.	An interest by management in employing inappropriate means to minimize reported earnings for tax-motivated reasons		M14
	7.	Recurring attempts by management to justify marginal or inappropriate accounting on the basis of materiality		M15
	8.	Management failure to correct known reportable conditions on timely basis		M16
	9.	Non-financial management's excessive participation in the selection of accounting principles or the determination of significant estimates		M17
	10.	A practice used by management of committing analysts, creditors and other third parties to achieve aggressive or unrealistic forecasts		M18
	11.	Unreasonable demands on the auditor, such as unreasonable time constraints regarding the completion of audit or issuance of auditor's report		M19
	12.	Excessive interest by management in maintaining or increasing the entity stock price or earnings trend	F11	

Note: H, high risk; M, medium risk; L, low risk; F, financial stability and operating characteristics; M, management characteristics and its influence over control environment; I, Industry conditions; R, relation; G, governance structure.

Table II. A rule set for non-financial risk factor assessment

No.	Rules
R1-1	If (number of 'M_{high}= yes' is greater than 3) or (number of 'M_{medium}=yes'pulsing number of'M_{low}= yes' is greater than 5) then risk level of M = high risk. Otherwise, risk level of M = low risk
R1-2	If (any one 'R_{high}= yes') or (number of 'R_{medium}= yes'pulsing number of'R_{low}= yes' is greater than 3) then risk level of R = high risk. Otherwise, risk level of R = low risk
R1-3	If (number of 'G1, G2, G3, G4 = yes' is greater than 2) then risk level of G = high risk. Otherwise, risk level of G = low risk.

Table III. A rule set for overall fraud risk level assessment

No.	Rules
R2-1	If F = high risk and (M = low risk and R = low risk and G = low risk) then the overall risk level = medium risk. Otherwise, the overall risk level = high risk.
R2-2	If F = low risk and (M = high risk and R = high risk and G = high risk) then risk level of G = medium risk. Otherwise, the overall risk level = low risk.

components and a rule-based system to obtain an overall assessment. The advantage of the proposed model is that it combines the precision prediction capability of machine learning methods with expert knowledge.

This study derived two sets of fraud risk rules based on discussions and interviews with a group of experienced CPAs (12). Tables II and III list the sets. The first rule set converts the risk factors in groups 'M', 'R' and 'G' into the corresponding non-financial components. The second rule set combines the financial and non-financial components to yield an overall assessment.

THE MACHINE LEARNING METHODS

Intentional fabrication of financial data leads to some accounts or financial ratios being overestimated or underestimated. Thus the financial ratios differ between the fraud and non-fraud classes. Moreover, the inherently balanced and protracted hook relationship among financial ratios is destroyed. From the machine learning perspective an abnormal pattern occurred, which demonstrated that the abnormal model hidden in complicated financial indicators can be captured using machine learning methods. Six methods (outlier detection, clustering, classification, prediction, regression and visualization) can be used for fraud detection by extracting abnormal relationships from the data (Ngai *et al.*, 2011).

Machine learning methods can be applied to detect financial statement fraud because they can capture the changes and relationships between groups of account balances. Classification models were used to obtain prediction results as binary variables that represent the likelihood of fraud. Usually, one indicates a high likelihood of fraud/non-fraud, whereas zero indicates a low likelihood. The prediction result may provide a warning signal to indicate a need to further investigate the account balances for a specific financial statement (Green and Choi, 1997).

The classification methods, decision trees (DT), case-based reasoning (CBR), BPNN and SVM were used to forecast credit scoring and financial distress (Huang *et al.*, 2004; Zhu *et al.*, 2007; Ding *et al.*, 2008; Li and Sun, 2009; Ravisankar *et al.*, 2010; Song *et al.*, 2010). Therefore, this study adopted these classifiers for fraud detection. In particular, an ensemble of these classifiers was developed to improve the prediction performance. A logistic regression model was used to appraise baseline performance of the ensemble.

The logistic regression method

When fraud risk is assessed using the logistic regression method, a key process of fraud classification is to calculate fraud probability. The relationship between the logistic regression model and the set of financial attributes is represented using a cumulative logistic probability function expressed as equation (4), where the binary variable (class labels) takes the value of 1 with a probability of fraud, p_i, and 0 with a probability of non-fraud (1 - p_i). As an advantage of this method, it does not assume multivariate normality and equal covariance matrices as multiple discriminate analysis (MDA):

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} + \text{K}\,\beta_n x_{in} \tag{1}$$

C5.0 DT

This work used Quinlan's C5.0 DT algorithm (Quinlan, 2007) for its good classification accuracy. The decision tree is generated as follows. First, the aforementioned selected input variables are taken as inputs, whereas the output is a decision tree. Second, the decision tree comprises one root, and multiple branch and leaf nodes. The leaf nodes represent class labels (fraud or non-fraud), and the other nodes are associated with the classes being classified. An instance is classified by beginning from the root and moving through it until a leaf node is reached. Third, for each node, information gain and entropy reduction are used as estimation criteria to identify the best attribute. In the post-pruning of the C5.0 DT, error-based pruning (EBP) is used as the pruning algorithm to control the overfitting behavior. Generally, the tree is pruned until the cross-validation error reaches its minimum.

back-propagation neural network (BPNN)

The neural network was applied to capture abnormal patterns associated with fraud cases. Thomas *et al.* (2002) reported that the prediction accuracy of BPNN equals or exceeds other advanced methods, including MDA, logistic regression and DT. However, its performance is not better than these methods. For example, BPNN suffers from overfitting when it is trained with small samples (Thomas *et al.*, 2002). Although BPNN was reported to be the most accurate among numerous types of neural networks, its performance deteriorates considerably compared with SVM. This study applied BPNN to the detection of financial statement fraud.

SVM

SVM has good prediction accuracy and generalization ability, especially for small, nonlinear and high-dimension samples. Previous studies intended to predict financial distress reported that SVM outperformed other classifiers, including artificial neural network (ANN), case-based reasoning (CBR), MDA and logistic regression in terms of generalization performance (Lima *et al.*, 2007; Ding *et al.*, 2008; Song *et al.*, 2010; Maldonado *et al.*, 2011; Yu, 2012). However, these investigations rarely suggested SVM as a means to assess risk of financial statement fraud.

SVM implements the principle of structural risk minimization by constructing an optimal separating hyperplane (Vapnik, 1998). SVM uses a linear model to implement nonlinear class boundaries through nonlinear mapping functions that map input vectors into a high-dimensional feature space. The SVM algorithm is concisely described as follows. Given a training set $D = \{x_i\, y_i\}_{i=1}^{N}$ with input vectors $x_i = (x_i^{(1)}, \ldots, x_i^{(n)}) \in R^n$ and target labels $y_i \in \{-1, +1\}$, a basic model of SVM is built as shown in equation (2), where w is a weight vector and b is a bias value. A nonlinear function ($\phi(\cdot) : R^n \rightarrow R^{n_k}$) maps the input vectors into a high-dimensional feature space. Identifying the optimal separating hyperplane is thus a dual QP-problem. The nonlinear decision function in the primal space for the linearly non-separable case is created using equation (3), where $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is a kernel function that satisfies Mercer's condition:

$$\min_{w,b} \frac{1}{2} w^{\text{T}} w \tag{2}$$

s.t. $y_i(w^T\phi(x_i) + b \quad \geq 1), i = 1, \ldots, N$

$$y(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i k(x, x_i) + b\right) \tag{3}$$

This study adopts a radial basis function (RBF), $k(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2)$, for the SVM-based classification model, because of its better performance compared to other kernels (Hsu *et al.*, 2004). Two parameters (C and γ) are associated with RBF kernels. Improper selection of these two parameters can cause problems of overfitting or underfitting. Hsu *et al.* (2004) suggested a practical guideline to SVM using grid search and cross-validation to identify the values of parameters C and γ. In n-fold cross-validation, the training set is first divided into n equal-sized subsets. One subset is tested sequentially using the classifier trained on the remaining ($n \quad - 1$) subsets. The final classifier performance is thus evaluated using the average classification accuracy of the $n$fold subsets. During the grid search, each parameter pair of (C, γ) is checked, and that with the highest cross-validation accuracy is selected as the optimized pair (C, γ). The SVM model is then trained using the optimal parameters. Finally, the testing sets are predicted using the trained SVM model.

The proposed classifier ensemble

An ensemble of classifiers combines these classifiers to classify a new instance. The integration of the classification results aims to generate more certain, precise and accurate results than that which any single classifier can achieve (Hansen, 1999; Oza and Tumer, 2008; Chen, 2013). Because the classifiers are based on different learning algorithms, they are likely to make errors, and may have unique advantages and deficiencies. Studies in finance-related fields reported that BPNN can achieve high prediction accuracy when the BPNN is trained with a large sample, and the best generation capability of SVM can easily be obtained (Ding *et al.*, 2008). However, the decision tree has the advantage of producing interpretable rules (Nanni and Lumini, 2009). The different classification capabilities of these rules make it necessary to create an ensemble of classifiers to detect financial statement fraud.

The most common method of combining the prediction results of all assembled classifiers is voting. This study uses voting with different weights to the classifiers. The misclassification of the ensemble of classifiers is to be minimized, which is expressed by e as formulated in equation (4), where n denotes the number of records, m represents the number of classifiers, C_i is the class label of the ith instance in the test sample, P_i denotes the prediction result of the ith instance in the test sample using the ensemble of classifiers, R_{ij} represents the prediction result of the ith instance in the test sample using the classifier j, w_j is the weight of the classifier j, t denotes the threshold, and $\oplus$ represents an 'OR' function. The function F is zero when $\sum_{j=1}^{m} w_j R_{ij} \geq t$, and is one when $\sum_{j=1}^{m} w_j R_{ij} < t$:

$$\min_{w,t} e = \frac{1}{n}\sum_{i=1}^{n} P_i \oplus C_i \tag{4}$$

s.t. $P_i = F\left(\sum_{j=1}^{m} w_j R_{ij} - t\right) \quad i = 1, \ldots, N; j = 1, \ldots, M$

The model in equation (4) determines an appropriate threshold of t that minimizes the misclassification degree of the ensemble of classifiers. The prediction accuracy for the training sample set is an effective criterion for evaluating classifier performance. The weight of each classifier represents its influence on the prediction accuracy of the final ensemble. The steps for modeling the ensemble of classifiers are described below:

- Step 1. After training and testing of all classifiers, the prediction accuracy of each classifier is obtained.
- Step 2. The outcome accuracy of classifier j on the training set is denoted by w_j, and the prediction output of classifier j for all instances in the test set is denoted by R_{ij}.
- Step 3. Calculate $F\left(\sum_{j=1}^{m} w_j R_{ij} - t\right)$ for classifier j and implement the OR operation between P_i and C_i.
- Step 4. Calculate the total misclassification e, return to step 3, then adjust the threshold t, and finally repeat the computation until e is minimized.

RESEARCH METHODOLOGY

Sampling

The sample comprises fraud and non-fraud groups. The fraud group was identified using the following criteria: first, observations that reflected verified proofs or indications involving fraudulent financial statements published by the administrative departments (the Chinese Securities Regulatory Commission and the Chinese Ministry of Finance); second, the observations that resulted in dubitable audit opinions, including reserved and rejected audit opinions; and third, the observations that were suspended for reasons associated with financial statement fraud. In contrast, this

study considered observations with qualified audit opinions in the non-fraud class Qualified audit opinions existed for all of the non-fraud cases.

According to these criteria, 110 fraudulent firms not involved in the finance industry and listed on the A-share markets of the Shanghai and Shenzhen stock exchanges in China between 2008 and 2012 were selected. The 110 fraudulent firms matching with 440 non-fraudulent firms of equivalent size in the same industries were used to improve the classification accuracy. The total sample comprised 110 fraudulent companies and 440 non-fraudulent companies.

Variables selection

The variables were selected from the financial indicators associated with three components (conditions, motivation and attitude). Because Moyes *et al.* 2005 demonstrated that 'motivation' group factors were more relevant to fraud than 'condition' and 'attitude', this study focused on extracting financial indicators from the 'motivation' group. In real life, top-level managers have specific motivations to deliberately deceive auditors, to improve their salaries by increasing profits from bad financial conditions or increasing stock prices in advance of capital raising, because their salaries or incomes are related to financial performance (Fanning and Cogger, 1998). Additionally, the symptoms associated with motivation to perform fraud are inevitably reflected by financial ratios, which are easily accessible from financial statements. Thus this study selected financial ratios based on the following three typical motivations for engaging in fraudulent activities.

First, managers deliberately attempt to deceive auditors to increase profits even though their current financial condition may be bad. Because managers understand the limitations of standard auditing procedures, they often fabricate financial data using various tactics. Fanning and Cogger (1998), Persons (1995) and Stice (1991) suggested that certain financial accounts or ratios are easy to manipulate because they are based on subjective estimations. Such easily manipulated indicators include accounts receivable, business incomes, inventories, profits, gross margins and asset quality. Seventy-five percent of firms were found to have overestimated accounts receivable and income from their main business (Persons, 1995). Some companies may overstate inventory and record consigned inventory as owned inventory (Stice, 1991). Others may bolster their balance sheets by exaggerating gross margins, earnings and net incomes, or by diverting assets to affiliated companies through manually written checks (Fanning and Cogger, 1998). Numerous notoriously fraudulent listed companies in China, such as Lantian Share, Ying Guang Xia and Liming Share, sought significant profits by fabricating sales, accounts receivable and incomes from their main businesses. Wells (2001) suggested that fraudulent and non-fraudulent companies differed significantly in gross margin and asset quality.

Second, business failure may be one of the main motivations for managers to engage in fraudulent activities (Loebbecke *et al.*, 1989; Abbot *et al.*, 2000). Chinese studies have indicated that specially treated (ST) companies (meaning companies that are financially distressed) are highly correlated with management fraud (Ma and Wang, 2003). The profitability, efficiency, liquidity, growth and cash flows of financially distressed companies deteriorate more sharply than those of non-distressed companies (Liang and Wu, 2005; Ding *et al.*, 2008; Song *et al.*, 2010) Additionally, a serious debt structure is likely to force a company into financial distress.

Third, attempting to maintain continuous corporate growth may become a motivation for the managers of a company to engage in fraud, and companies that grow rapidly are likely to participate in fraudulent activities (Stice, 1991). This study tested corporate growth using indicators such as growth in accounts receivable, inventory and income from their main businesses.

Additionally, this study considered the components of condition and attitude, and analyzed relevant research reports, papers, journals and business practice studies to incorporate other variables (Beasley, 1996; Pincus, 1997; Abbot *et al.*, 2000; Bell and Carcello, 2000). Finally, this study identified 23 financial ratios. As listed in Table IV, the variables are organized into six groups (namely size, liability, asset structure, growth, efficiency and profitability).

Model validation procedures

This study involved analysis steps as follows:

- Step 1. Data pre-processing: this step renders the ready-to-use data. The financial data pre-processing processes include the manipulation of missing values, filtering of duplicate data and normalization.
- Step 2. Variable extraction: this is important because it can simplify the data description and increase prediction accuracy. ANOVA factor analysis is applied to select variables that differ significantly between the fraud and non-fraud groups. As Table IV shows, 14 variables indicated by an asterisk are obtained as the final input variables, which reveals a significant difference between the fraud and non-fraud groups ($p \leq 0.05$).
- Step 3. Classification model validation: the pre-processed dataset was further randomly divided into training and testing sets via K-fold cross-validation. A typical experiment uses $k = 10$. The sample was divided 10-fold via stratified 10-fold cross-validation. Each fold contained equal numbers of fraud and non-fraud cases. To effectively assess the performance of the different models, an experimental protocol was defined. Each fold of the sample

Table IV. Statistical analysis for input variables of classification methods

	No.	Variable	Non-fraud cases		Fraud cases		ANOVA analysis	
			Mean	SD	Mean	SD	*F*	*p*
S	X1*	log (total assets)	8.931	0.464	8.61	0.401	24.69	0
L	X2	Liabilities/owners' equity	1.319	1.209	−3.471	3.857	0.71	0.412
	X3*	Liabilities/total assets	0.853	0.181	1.393	2.529	7.96	0.007
	X4	long-term liabilities/total assets	0.051	0.098	0.312	0.497	3.31	0.151
	X5*	Current ratio	1.421	0.931	0.821	1.043	5.87	0.008
	X6	Quick ratio	0.969	0.993	0.789	0.983	1.13	0.329
	X7	log (liabilities)	9.219	0.556	8.766	0.443	1.57	0.256
A	X8*	Cash/total assets	0.316	0.149	0.099	0.076	26.83	0
	X9*	Inventory/total assets	0.329	0.21	0.129	0.071	13.85	0.003
	X10	Fixed assets/total assets	0.421	0.181	0.391	0.172	0.02	0.912
	X11*	Operating capital/total assets	0.159	0.318	−1.262	2.437	7.58	0.005
G	X12*	Growth of accounts receivables	0.954	0.483	0.231	2.671	4.56	0.06
	X13*	Growth of Inventory	0.102	0.232	−0.345	1.618	3.99	0.028
	X14	Growth of income from principal business operations	0.187	0.18	−29.982	69.892	2.18	0.315
	X15	Growth of accounts receivables	−0.089	0.853	0.124	2.319	0.23	0.598
E	X16	Cost of principal business operation/ inventory	4.119	3.113	3.765	2.215	0.309	0.591
	X17*	Income from principal business operation/ accounts receivable	82.119	19.382	10.283	16.861	1.39	0.049
	X18*	Log (cost of principal business operation)	8.992	0.612	8.283	0.786	28.16	0
P	X19*	Net earnings/total assets	0.143	0.345	−0.395	0.687	8.13	0.003
	X20*	Net earnings/fixed assets	0.2139	0.325	−2.97	8.131	5.15	0.043
	X21	Net earnings/income from principal business operation	0.197	0.191	−16.763	59.123	3.11	0.422
	X22*	Income from principal business operation/ total assets	1.018	0.436	0.603	0.367	27.54	0.001
	X23*	Margin profit/total assets	0.312	0.336	0.079	0.234	29.17	0

Note: S, L, A, G, E and P denote size, liability, asset structure, growth, efficiency and profitability, respectively; asterisk denotes $*p \leq 0.05$.

was used individually to define parameters and train classifiers, while the remaining nine folds were used as test sets to assess the sample performance. After the parameters were set and the classifiers were trained, the methods were evaluated by applying them to the test sets. Finally, the average classification accuracy of the test sets was calculated. After preparation of the 10-fold cross-validation datasets, these datasets were used by the four classifiers (LR, C5.0 DT, BPNN, and SVM). The proposed ensemble of classifiers was developed and validated based on the classifier results. The BPNN and proposed ensemble of classifiers were implemented using MATLAB 6.5. The algorithms from LIBSVM (Hsu *et al.*, 2004), C5.0 DT (Quinlan, 2007) and SPSS 13.0 are used for SVM, C5.0 DT and LR respectively.

- Step 4. Revision of the classification results: after obtaining the classification results, this study revised the misclassified cases using non-financial risk factors and a rule-based system. First, a group of experienced CPAs (12) were invited to assess non-financial risk factors for the misclassification cases. Table I shows how the non-financial risk factors were converted into questions, to which the CPAs were asked to provide answers (high risk or low risk). For various reasons, some CPAs could not provide answers. The missing values from the Type I or Type II error cases were replaced with 'high risk' or 'low risk'. Second, each non-financial component was assessed using the binary answer and rule set No. 1. Third, the rule set No. 2 was used, and the financial and non-financial components were combined to form an overall assessment. Finally, the misclassification cases were reclassified into three groups, i.e. 'high risk', 'medium risk' and 'low risk'.

EXPERIMENTS

Modeling description

The validation procedure yielded five classification methods for the risk assessment of financial statement fraud. Table V lists the prediction results, where the first-fold dataset illustrates the modeling process of each classification method.

Table V. Performance of tenfold cross validation

Different classifier	Accuracy (%)			Error rate (%)		
	Average	Non-fraud	Fraud	Average	Type I	Type II
LR	77.9	74.5	72.7	22.1	19.5	27.3
BPNN	85.1	88.1	80.9	14.9	11.9	19.1
C5.0 DT	78.6	77.7	89.1	21.4	22.3	10.9
SVM	85.5	90.2	74.5	14.5	9.8	25.5
Proposed ensemble of classifiers	88.9	90.9	89.1	11.1	9.1	10.9

Note: Error rate of Type I: a non-fraud case is misclassified as a fraud class; error rate of Type II: a fraud case is misclassified as a non-fraud class.

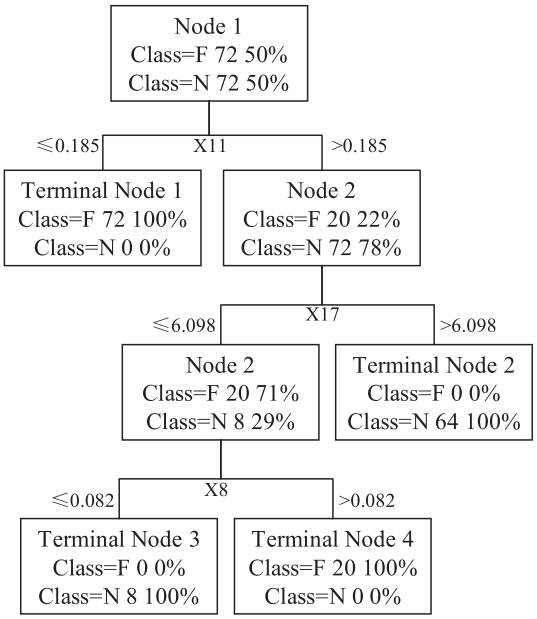

Figure 2. A decision tree returned by the C5.0 DT classification method

Table VI. Analytic rules extracted using C5.0 DT

Terminal node	Rule	Prediction class
1	If X11 <= −0.185 Then class = 'F'	Fraud
2	If X11 > −0.185 and X17 <= 6.098 then class = 'N'	Non-fraud
3	If X11 > −0.185 and X17 <= 6.098 and X8 <= 0.082 then class= 'N'	Non-fraud
4	If X11 > −0.185 and X17 <= 6.098 and X8> 0.082 then class = 'F'	Fraud

The LR model used the procedure for the selection of stepwise variables. The final model used six independent variables (X1, X3, X11, X17, X20 and X23), and the test sets had an average accuracy of 77.9%.

For the BPNN model, trial-and error experiments determined the number of hidden nodes of the BPNN. Because the input layer has 14 input nodes, the initial numbers of hidden nodes to be tested were 12, 14, 16, 18, 20, 22 and 24 (other numbers did not obtain improved results). The learning rate was set to 0.01, and the momentum term was set to 0.7. Two alternative termination criteria were used for training, with mean square errors (MSE) less than 0.0001, and a maximum of 2000 iterations. The hidden nodes and the output nodes used the TANSIG and PURELIN transfer functions, respectively. After comparing the results of all possible hidden nodes, the best prediction accuracy for the test set was found given 18 hidden nodes. The average accuracy of the test sets was 85.1%.

For C5.0 DT, a 10-fold cross-validation of the dataset was used to estimate a best setting for the pruning parameters. The pruning CF was set to a value between 0 to 0.8, with intervals of 0.2. For all the results, the best pruning CF was

0.20 and the 10-fold cross-validation was a minimum of 17.01%. For the final pruned tree, the number of terminal nodes equaled four. As shown in Figure 2 and Table VI, the generated C5.0 DT was transformed into a set of rules. X11, X17 and X8 are clearly important in the rule induction process. Furthermore, the operating capital/total assets (X11) exceeds -0.185, and the income from principal business operation/accounts receivables (X17) does not exceed 6.098. These two conditions fall into terminal nodes 2, whose prediction class is fraud.

For the RBF-kernel SVM model, the grid-search method and 10-fold cross-validation were used to optimize the parameter pair (C, γ). Consider a grid-search space ($C = 2^{-5}, 2^{-3}, \ldots . 2^{15}$; $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$), an optimal pair (0.03125, 0.125) was found in the green zone with a cross-validation rate of 85.84%. The test sets had an average accuracy of 85.5%.

This study presented an ensemble of four classifiers, including logit, BPNN, SVM and C5.0 DT. The values of the prediction accuracy of the four classifiers in the training sample were taken as the weights. The testing sets had an average accuracy of 88.9%.

Experimental results

Besides classification accuracy, this study also used misclassification cost and receiver operating characteristic (ROC) curve analysis to compare the diagnostic performances of the five classifiers. Generally, misclassification cost is associated with two error types. A Type I error occurs when a non-fraud case is classified as a fraud class. Meanwhile, a Type II error is committed when a fraud case is classified as a non-fraud class. The misclassification costs associated with Type II errors are reportedly much higher than those associated with Type I errors (West, 2000). Classifying a fraud case into a non-fraud class may result in incorrect decisions and economic damage. Moreover, classifying a non-fraud case into a fraud class may result in expenses and excess time associated with additional investigation. Based on ROC curve analysis, the sensitivity and specificity for each possible cut-off point are analyzed, and this information is used to discriminate between the two classes. The area under the curve (AUC) served as the evaluation criteria.

The 10-fold cross-validation performances of the five classification methods were calculated and compared. The suggested ensemble of classifiers achieved the highest average accuracy (88.9%), followed by SVM (85.5%), BPNN (85.1%), C5.0 DT (78.6%) and LR (77.9%). For Type I errors, the suggested ensemble of classifiers achieved the lowest error rate (9.1%), followed by SVM (9.8%), BPNN (11.9%), LR (19.5%) and C5.0 DT (22.3%). For Type II errors, the proposed ensemble of classifiers and C5.0 DT achieved the lowest error rate (10.9%), followed by BPNN (19.1%), SVM (25.5%) and LR (27.3%).

When the first-fold dataset is used as an example, Figure 3 depicts the result curve of ROC. The AUC values of LR, BPNN, C5.0 DT, SVM and the ensemble of classifiers were 0.7553, 0.8231, 0.7714, 0.8351 and 0.8470, respectively. As shown in Table VII, the average AUC for the 10 folds of the test set were 0.7641 (LR), 0.8276 (BPNN), 0.7784 (C5.0 DT), 0.8375 (SVM) and 0.8452 (ensemble of classifiers), respectively.

Analyzing experimental results

The above data were used to draw the following conclusions:

1. The machine learning methods (BPNN, C5.0 DT, SVM and the proposed ensemble of classifiers) outperformed the statistical method (LR). Moreover, the performance of these methods can be improved through trial-and-error adjustment of the parameters.

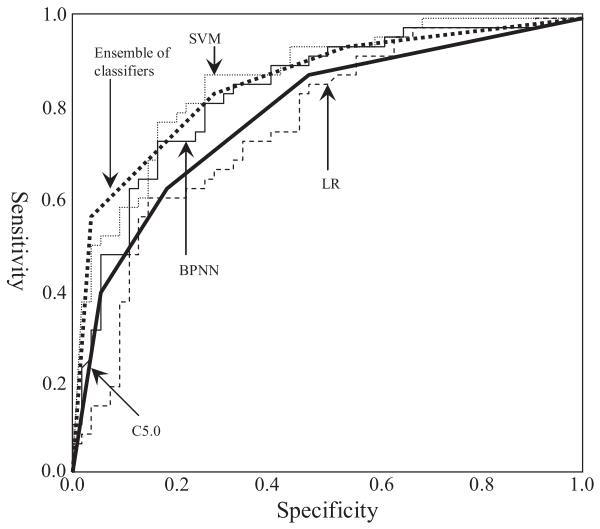

Figure 3. A ROC curve of the first-fold dataset Table I. A list of fraud risk factors in SAS 99

Table VII. Average AUC for datasets using different classifiers

Classifier	Average AUC	Confidence interval
LR	0.7641	±0.0199
BPNN	0.8276	±0.0257
C5.0 DT	0.7784	±0.0187
SVM	0.8375	±0.0050
Proposed ensemble of classifiers	0.8452	±0.0025

2. Regarding average accuracy and average AUC, the proposed ensemble of classifiers significantly outperformed all of the four individual classifiers (BPNN, C5.0 DT, SVM and LR). With respect to error rate, the proposed ensemble of classifiers achieved the lowest composite error rate (sum of Type I and Type II errors), possibly because an ensemble of classifiers combining different assembled learning algorithms can strengthen the best features and minimize misclassification of each assembled classifier. The proposed ensemble of classifiers thus reduced the possible high risks associated with misclassification.
3. SVM and BPNN outperformed C5.0 DT and LR in average accuracy, composite error rate and AUC. However, SVM and BPNN achieved high Type II error rates (25.5% and 19.1%, respectively). This result indicates that SVM and BPNN achieve better classification performance than LR and C5.0 DT, but may suffer weaknesses identifying fraud cases, which may cause risks associated with Type II errors.
4. C5.0 DT achieved the lowest Type II error rate (10.9%) among the four classifiers (BPNN, C5.0 DT, SVM and LR), thus reducing the economic losses from incorrect decisions. C5.0 DT is clearly more accurate than SVM and BPNN when it is used for fraud case identification. The result supports the studies on fraud detection (Green and Choi, 1997; Fanning and Cogger, 1998; Thomas *et al.*, 2002). Thomas *et al.* (2002) and Fanning and Cogger (1998) reported that, although ANN produced the lowest composite error rate, it did not outperform the other methods in prediction accuracy. Moreover, the rules extracted by C5.0 DT are interpretable and useful for audit decision making.
5. Because the classification accuracy is as high as 90.9% for non-fraud cases and 84.2% for fraud cases, the proposed classification methods effectively assessed the risk of financial statement fraud. However, the methods still have high error rates, of 19.5 for Type I errors and 31.7 for Type II errors. The following reasons account for these high error rates.

First, the classification methods suffer inherent limitations in terms of fraud detection capability. The application of these methods to financial data with many dimensions and nonlinear features increases error rates. Whether overlapping data are in multidimensional spaces of the original problems or are in the mapping problems, they must exist between the fraud and non-fraud groups. Statistical analysis in Table IV shows that only 14 variables differed significantly between the fraud and non-fraud groups. Moreover, further analysis of the misclassification confirmed this result. For example, the overlapping financial ratios were X1, X3, X8 and X12.

Second, another limitation associated with classifiers relates to the capture of anomaly patterns. The anomalies are divided into two property classes, namely outliers and violators, both of which occur in financial fraud data (Ngai *et al.*, 2011). Generally, outlier detection methods are more widely used for fraud detection than are violator detection methods (Hodge and Austin, 2004), partly because the majority of violator cases are also detected using classification-based outlier detection methods. Additionally, the violator detection algorithms are less developed than outlier detection algorithms. This study thus adopted classification-based outlier detection methods to assess the risk of financial statement fraud. However, the outlier detection methods cannot always effectively identify violators through anomalous matching between attributes of fraud groups and non-fraud groups, which should not be neglected.

Revising the assessment results

Given the limitations of classifiers, this study adopted the proposed-approach to revise the results from the classifier system. As shown in Figure 1, this study used non-financial indicators and a rule-based system to revise the previous classification results. Considering the ensemble of classifiers as an example, 19 cases of Type II errors and 40 cases of Type I errors exist. After assessing misclassified cases using the rule-based system, Type II error cases (fraud) that were classified into non-fraud class are revised in the final assessment results. Among the 19 misclassified fraud cases, eight were 'high risk', seven 'medium risk' and four 'low risk'. The Type II error rate thus decreased from 17.5% to 10%. A similar improvement occurred with the 40 misclassified non-fraud cases where the Type I error rate decreased from 9.1% to 6%. Because classification errors may cause great losses to stakeholders, auditors should pay further attention to (the real-world 'medium risk' and 'high risk' cases). If a case is identified as 'high risk', then it most likely involves abnormal account balance. Hence the relevant transactions should be further examined to find evidence to identify the fraud cases. Cases classified as 'medium risk' should be closely observed by applying statistical process

control methods such as the Six Sigma monitoring system to evaluate and track firm overall financial performance (Faltin and Faltin, 2003).

As the experiment demonstrated, non-financial risk factors can help reduce two types of error rates. However, assessment of non-financial risk factors is subjective because they are based on questionnaire-based methods.

CONCLUSIONS

This study designed an improved framework for assessing the risk of financial statement fraud. The framework comprises a system of financial and non-financial risk factors, and a hybrid assessment method The study focused on examining the performance of the four classifiers and an ensemble of those classifiers. A statistical method (LR) and four machine learning methods (C5.0 DT, BPNN, SVM and a proposed ensemble of classifiers) were compared. The findings were that the proposed ensemble of classifiers outperformed LR, C5.0 DT, BPNN and SVM in accuracy, composite error rate and AUC. The experimental results suggest that this ensemble is a competitive tool for risk assessment of financial statement fraud. SVM outperformed other classifiers. C5.0 DT can extract interpretable knowledge in the form of rules and achieve the lowest Type II error rate. Moreover, this study further revised the misclassification results yielded by the classifiers. The experimental results indicate that non-financial risk factors and a rule-based system can help decrease the two types of error rates.

The experimental results demonstrate that the proposed classification method can help assess the risk of financial statement fraud. The proposed approach thus can assist stakeholders (such as internal or external auditors, investors, economic analysts, banks and governments) to decrease financial risks. Furthermore, the variables and rules associated with critical financial fraud risk factors are easy to understand and thus are significant for auditing decision making.

The study has several limitations. First, the factors, including sample, period and changes in the Chinese economy, may influence the prediction modes. The authors are trying to improve the accuracy and interpretability of the machine learning methods by exploring problems such as the selection of critical financial ratios and using various methods to refine knowledge into rules. Moreover, future studies can evaluate the studied methods using data from other countries. Finally, future studies will explore machine learning methods for detecting violators.

ACKNOWLEDGEMENTS

This work was supported in part by the National Nature Science Foundation of China (71101088, 71390521, 71471109, 71471076, 71171099, 71373818, 71201071, 713111111), the National Philosophy and Social Science Foundation of China (10CTQ18), the Science Foundation of Ministry of Education of China and Shanghai (20113121120002, 14YZ100, 20123121110004, 13SG48, 20123227110011). This work was also sponsored by Qing Lan Project and 333 Project of Jiangsu Province, and Jiangsu University Top Talents Training Project.

REFERENCES

Abbot JL, Park Y, Parker S. 2000. The effects of audit committee activity and independence on corporate fraud. *Managerial Finance* **26**: 55–67.

Apostolou BA, John M. 2001. The relative importance of management fraud risk factors. *Behavioral Research in Accounting* **13**: 1–24.

Asare SK, Wright AM. 2004. The effectiveness of alternative risk assessment and program planning tools in a fraud setting. *Contemporary Accounting Research* **21**: 325–352.

Beasley M. 1996. An empirical analysis of the relation between board of director composition and financial statement fraud. *Accounting Review* **71**: 443–466.

Bell T, Carcello J. 2000. A decision aid for assessing the likelihood of fraudulent financial report. *Auditing* **9**: 169–178.

Cha Z-Y, Wu S-N. 2006. An early warning study on disclosure fraud detection based on corporate governance (in Chinese). *Journal of financial and Accounting* **19**: 80–90.

Chen G, Firth M, Gao DN, Rui OM. 2006. Ownership structure, corporate governance, and fraud Evidence from China. *Journal of Corporate Finance* **12**: 424–448.

Chen MY. 2013. A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Information Sciences* **220**: 180–195.

Ding Y-S, Song X-P, Zeng Y-M. 2008. Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications* **34**: 3081–3089.

Dowling C, Leech S. 2007. Audit support systems and decision aids: current practice and opportunities for future research. *Journal of Accounting Information Systems* **8**: 92–116.

Eining MM, Dorr PB. 1991. The impact of expert system usage on experiential learning in an auditing setting. *Journal of Information Systems* **5**: 1–16.

Eining MM, Jones DR, Loebbecke JK. 1997. Reliance on decision aids: an examination of auditors' assessment of management fraud. *Auditing* **16**: 1–19.

Faltin DM, Faltin FW. 2003. Toe the line: no more WorldComs. *Quality Progress* **36**: 29–35.
Fang J-X. 2003. An empirical study on financial fraud identification of listed companies in China (in Chinese). *Journal of Listed Companies* **2003**: 40–45.
Fanning K, Cogger K. 1998. Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance and Management* **7**: 21–24.
Glancy FH, Yadav SB. 2011. A computational model for financial reporting fraud detection. *Decision Support Systems* **50**: 595–601.
Green BP, Choi J. 1997. Assessing the risk of management fraud through neural network technology. *Auditing* **16**: 14–28.
Gu N-S, Feng Q-C. 2009. The empirical studying on detecting the fraudulent financial statements based on LVQ neural network (in Chinese). *Journal of Value Engineering* **2009**: 111–118.
Hansen JV. 1999. Combining predictors: comparison of five meta machine learning methods. *Information Sciences* **119**: 91–105.
Hodge V, Austin J. 2004. A survey of outlier detection methodologies. *Artificial Intelligence* **22**: 85–126.
Hsu C-W, Chang C-C, Lin C-J. 2004. A practical guide to support vector classification. *Technical Report*. Department of Computer Science and Information Engineering, National Taiwan University Available: http://www.csie.ntu.edu.tw/~cjlin/papers/guide [accessed on 16 March 2014].
Huang Z, Chen H, Hsu C-J, Chen W-H, Wu S. 2004. Credit rating analysis with support vector machine and neural networks: a market comparative study. *Decision Support Systems* **37**: 543–558.
Humpherys SL, Moffitt KC, Burns MB, Burgoon JK, Felix WF. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* **50**: 585–594.
Jerry WL, Mark IH, Jack DB. 2003. A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal* **18**: 657–665.
Koskivaara E. 2004. Artificial neural networks in analytical review procedures. *Managerial Auditing Journal* **19**: 191–223.
Kotsiantis S, Koumanakos E, Tzelepis D. 2006. Forecasting fraudulent financial statements using data mining. *International Journal of Computational Intelligence* **3**: 1304–2386.
Li H, Sun J. 2009. Gaussian case-based reasoning for business failure prediction with empirical data in China. *Information Sciences* **192**: 89–108.
Liang L, Wu D-S. 2005. An application of pattern recognition on scoring Chinese corporations financial conditions based on back propagation neural network. *Computers and Operations Research* **32**: 1115–1129.
Lima C, Coelho A, Zuben F. 2007. Hybridizing mixtures of experts with support vector machines: investigation into nonlinear dynamic systems identification. *Information Sciences* **177**: 2049–2074.
Liu L-G, Du Y. 2003. An empirical research on the relationship between corporate governance and the quality of accounting information (in Chinese). *Accounting Research* **2003**: 28–33.
Loebbecke J, Eining M, Willingham J. 1989. Auditor's experience with material irregularities: frequency, nature and detectability. *Auditing* **9**: 1–28.
Ma L, Wang Y-M. 2003. Business failure, auditing failure and auditing risk (in Chinese). *Journal of Chinese CPA* **2003**: 23–28.
Maldonado S, Weber R, Basak J. 2011. Simultaneous feature selection and classification using kernel-penalized support vector machines. *Information Sciences* **181**: 115–128.
Moyes GD, Lin P, Raymond ML. 2005. Raise the red flag: a recent study examines which SAS No. 99 indicators are more effective in detecting fraudulent financial reporting. *Internal Auditor* **62**: 47–51.
Nanni L, Lumini A. 2009. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications* **36**: 3028–3033.
Ngai EWT, Hu Y, Wong YH, Chen Y, Sun X. 2011. The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. *Decision Support Systems* **50**: 559–569.
Oza NC, Tumer K. 2008. Classifier ensembles: select real-world applications. *Information Fusion* **9**: 4–20.
Persons O. 1995. Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research* **11**: 38–46.
Pincus KV. 1997. The efficacy of a red flags questionnaire for assessing the possibility of fraud. *Accounting, Organizations and Society* **16**: 153–163.
Qiao H, He C-Z. 2007. GMDH model for identification of fraudulent financial report of listed companies in China (in Chinese). *Journal of Soft Science* **21**: 45–49.
Quinlan JR. 2007. C5.0 online tutorial (2004). Available: http://www.rulequest.com [acccessed on 16 March 2014].
Ravisankar P, Ravi V, Bose I. 2010. Failure prediction of dotcom companies using neural network–genetic programming hybrids. *Information Sciences* **180**: 1257–1267.
Ravisankar P, Ravi V, Rao GR, Bose I. 2011. Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems* **50**: 491–500.
Song X, Ding Y, Huang J, Ge Y. 2010. Feature selection for support vector machine in financial crisis prediction: a case study in China. *Expert Systems* **27**: 299–310.
Spathis C. 2002. Detecting false financial statements using published data: some evidence from Greece. *Managerial Auditing Journal* **17**: 179–191.
Spathis C, Doumpos M, Zopounidis C. 2002. Using client performance measures to identify pre-engagement factors associated with qualified audit reports in Greece. *International Journal of Accounting* **38**: 267–284.
Stice J. 1991. Using financial and market information to identify pre-engagement market factors associated with lawsuits against auditors. *Accounting Review* **66**: 516–533.
Summers SL, Sweeney JT. 1998. Fraudulent misstated financial statements and insider trading: an empirical analysis. *Accounting Review* **71**: 131–146.
Thomas G, Calderon TG, Cheh JJ. 2002. A roadmap for future neural networks research in auditing and risk assessment. *International Journal of Accounting Information Systems* **3**: 203–236.
Tseng Y-C, Chang R-D. 2006. Effects of a decision aid for assessment the fraudulent financial reporting: an application of SAS No. 82. *Journal of Business and Economic Research* **4**: 57–67.
Vapnik V. 1998. *Statistical Learning Theory*. Springer: New York.
Wells JT. 1997. *Occupational Fraud and Abuse*. Obsidian: Austin, TX.

Wells JT. 2001. Irrational ratios. *Journal of Accountancy* **192**: 80–89.
West D. 2000. Neural network credit scoring models. *Computers and Operations Research* **27**: 1131–1141.
Wilks TJ, Zimbelman MF. 2004. Decomposition of fraud-risk assessments and auditors' sensitivity to fraud cues. *Contemporary Accounting Research* **21**: 719–745.
Wu D, Olson DL. 2009. Enterprise risk management: small business scorecard analysis. *Production Planning and Control* **120**: 362–269.
Wu D, Olson DL. 2010a. Enterprise risk management: a DEA VaR approach in vendor selection. *International Journal of Production Research* **48**: 4919–4932.
Wu D, Olson DL. 2010b. Introduction to special section on 'Risk and Technology'. *Technology Forecasting and Social Change* **77**: 837–839.
Wu DD, Kefan X, Hua L, Shi Z, Olson DL. 2010. Modeling technological innovation risks of an entrepreneurial team using system dynamics: an agent-based perspective. *Technology Forecasting and Social Change* **77**: 857–869.
Yen E. 2007. Warning signals for potential accounting frauds in blue chip companies: an application of adaptive resonance theory. *Information Sciences* **177**: 4515–4525.
Yu L. 2012. An evolutionary programming based asymmetric weighted least squares support vector machine ensemble learning methodology for software repository mining. *Information Sciences* **191**: 31–46.
Zhu Z, He H, Starzyk JA, Tseng C. 2007. Self-organizing learning array and its application to economic and financial problems. *Information Sciences* **177**: 1180–1192.

Author's biographies:

Xin-Ping Song is an associate professor in School of Management at Jiangsu University. She got the PhD in management science and engineering at Donghua University. She also had worked as a postdoctor at Nanjing University. Her researches are focused on business intelligence, information science, knowledge management and network marketing.

Zhi-Hua Hu is an associate professor in Logistics Research Center at Shanghai Maritime University. He got the PhD in control theory and engineering at Donghua University. He had worked as a postdoctor at Tongji University. His researches are focused on logistics and supply chain management, intelligent algorithms and systems.

Jian-Guo Du is a professor in School of Management at Jiangsu University. He got the PhD in management science and engineering at Nanjing University. He had also worked as a postdoctor at Nanjing University. His researches are focused on regional economic management, enterprise environment management, risk management, management methodology and applications.

Zhao-Han Sheng is a professor in School of management and engineering at Nanjing University. His researches are focused on complex system, complex network, social and economic system engineering, economic and management systems based on computational experiment.

Author's address:

Xin-Ping Song and **Jian-Guo Du**, College of Business and Management, Jiangsu University, Zhenjiang 212013; and College of Engineering and Management, Nanjing University, Nanjing 211108, China.

Zhi-Hua Hu, Logistics Research Center, Shanghai Maritime University, Shanghai 200135, China.

Zhao-Han Sheng, College of Engineering and Management, Nanjing University, Nanjing 211108, China.

