JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING
Vol. 4
2007
pp. 183–197

Identifying Qualified Auditors' Opinions: A Data Mining Approach

Efstathios Kirkos
Technological Educational Institution of Thessaloniki

Charalambos Spathis
Aristotle University of Thessaloniki

Alexandros Nanopoulos
Aristotle University of Thessaloniki

Yannis Manolopoulos
Aristotle University of Thessaloniki

ABSTRACT: Data Mining methods can be used in order to facilitate auditors to issue their opinions. Numerous of these methods have not yet been tested on the purpose of discriminating cases of qualified opinions. In this study, we employ three Data Mining classification techniques to develop models capable of identifying qualified auditors' reports. The techniques used are C4.5 Decision Tree, Multilayer Perceptron Neural Network, and Bayesian Belief Network. The sample contains 450 publicly listed, nonfinancial U.K. and Irish firms. The input vector is composed of one qualitative and several quantitative variables. The three developed models are compared in terms of their performance. Additionally, variables that are associated with qualified reports and can be used as indicators are also revealed. The results of this study can be useful to internal and external auditors and company decision-makers.

Keywords: qualified auditors' opinions; auditing; data mining; classification.

INTRODUCTION

In today's modern business era, auditing is becoming a more demanding task. The advancements in the auditing conceptual frameworks, the massive application of information technology in business, and the new knowledge extraction technologies, such as data mining, constitute a field of necessities and capabilities that presents new challenges to the applied auditing methods.

We thank three anonymous reviewers and Prof. M. A. Vasarhelyi (Editor) for valuable comments and suggestions that resulted in significant improvements in the current version. Also, we thank the reviewers of the ADMKD 2006 workshop, Thessaloniki, and the Department of Economics and Department of Informatics research workshops at Aristotle University for their helpful comments and suggestions.

Corresponding author: Charalambos Spathis
Email: hspathis@econ.auth.gr

The relevant research literature repeatedly recognizes the importance of the new technology and of the elaborated knowledge discovery techniques in auditing. The new auditing methodologies embrace the concept of business risk. Risk based auditing (RBA) refers to the risk that a company will not achieve its objectives. RBA adopts a top-down approach starting with the strategic goals of an entity and working downward to business process, controls, and financial statements (Lemon et al. 2000; Salterio 2000). Calderon and Cheh (2002) claim that the RBA approach requires that auditors use advanced technologies like Neural Networks to identify the factors that prevent an organization from achieving its goals.

Koskivaara (2004) notes that sophisticated auditing tools could prevent companies from manipulating account value and help auditors to answer the demands of today's business environment. The increasing number of management fraud cases amplifies the necessity of new elaborated auditing tools (Vasarhelyi 2005; Kirkos et al. 2007). Bell and Tabor (1991), as well as Chen and Church (1992), note that auditors can use the output of models to plan specific auditing procedures that can be applied in order to achieve an acceptable level of audit risk. These models can also be used as a quality-control tool in the review or final stage of an engagement and for contingency analyses on how changes to specific variables could add or detract from the probability of obtaining a qualified opinion (Kleinman and Anandarajan 1999).

A significant number of models used to predict cases, where qualification should be issued, have been developed through several research studies. An examination of the relevant literature reveals that these studies use either some version of regression or some version of neural networks. Although regression probably constitutes the most commonly used method, it is not free of limitations. Several regression versions assume arbitrarily a linear relationship between the dependent variable and the independent variables. Logistic regression, which is particularly suitable for dichotomous classes, overcomes this restriction. However, other assumptions are still valid. Logistic regression assumes a linear relationship between the independents and the log of (logit) of the dependent, while it also assumes that all relevant variables are included in the model and that the irrelevant variables are excluded. Logistic regression also requires independent sampling, a principle that is violated in samples containing matched-pairs observations (Tabachnick and Fidell 2001).

Data mining is an umbrella term that embraces methodologies aiming to extract human knowledge from data. One of the differences between a traditional analysis of data and the data mining is that the former supposes that the hypotheses are already constructed and validated against the data, whereas the latter supposes that the patterns and hypotheses are automatically extracted from the data. Specific fields of data mining are characterization and comparison, classification and prediction, and cluster analysis. Identifying cases of qualified reports can be regarded as a typical classification problem. Data mining includes several classification methods, like Decision Trees, Bayesian Belief Networks, Neural Networks, Rough Sets, Support Vector Machines, and Genetic Algorithms. Until now, most neural networks have been tested for their applicability to detect cases where auditors issue qualified opinions. The fact that alternative classification methods provide significant advantages constitutes a challenge to evaluate these methods in terms of their predicting accuracy and explanatory power and thus propose the method that is most suitable as a decision support tool for auditors.

In this study, we employ three alternative methods i.e., Decision Trees, Neural Networks, and Bayesian Belief Networks. These methods have completely different theoretical foundation and working mechanism. Two of them, the Decision Trees and the Bayesian Belief Networks present the advantage that they represent the decision making process in a comprehendible form. The knowledge of the decision-making mechanism is very important to auditors. That is, the auditor can be assured that the logic of the model is reasonable and that it complies or even does not contradict with recognized auditing principles and practices. All three methods are intensively data driven and extract the decision making process directly from the data.

Neural networks (NNs) are an effective data mining classification method. They provide several advantages over logistic regression, as they are very effective in cases where nonlinear relationships exist between the dependent and the independents. NNs do not impose arbitrary assumptions; they are tolerant to noisy data and are capable of classifying patterns on which they have not been trained. One of their disadvantages is that they act as "black boxes," as the user is provided with a classification decision without having any control or even knowledge on how this decision has been reached. For an auditor, the knowledge of the decision-making mechanism is of a particular importance. Several methods have been proposed to produce comprehendible rules derived from neural networks (Gallant 1988; Fu 1991). However, auditors that are not artificial intelligence experts may find these algorithms very complicated and difficult to use. Another disadvantage of NNs that may be significant to auditors is that the definition of the optimum network's architecture requires expert's experience.

Decision Trees (DTs) have several advantages. They are highly interpretable since they can easily be transformed to a set of meaningful IF-THEN rules. Decision trees are also nonparametric, they make no assumptions about the distribution of the data, and they incorporate a built-in feature selection method which makes them immune to the presence of irrelevant independents. DTs also have a fast learning mechanism and they can handle categorical values.

Bayesian Belief Networks (BBNs) are very suitable in cases where complex interrelationships exist between the dependent and independent variables or even among the dependents themselves. Presumably, this is the case of financial data. BBNs also achieve very good accuracy rates and they can handle categorical variables. They also enable the user to modify the model which was extracted from the data. Irrelevant dependencies can be removed and significant ones added. In this manner, BBNs allow the integration of the extracted model with the expert's domain knowledge.

The motivation of this study is to evaluate the three alternative methods according to their ability to predict the qualified or unqualified cases. Another aim is to reveal the factors found significant by each method. The final goal is to propose the most suitable method. In addition, the purpose of the study is to extend the auditing literature by investigating the efficiency of three methods in the development of classifications models for replicating auditors' opinion.

In our study, the three models are compared in terms of their overall predictive accuracy as well as Type I and Type II error rates. Each model reveals specific input variables that are considered significant for the detection of the qualified cases, and highlights factors associated with qualified reports. A comparative assessment of the outcomes of the three models exhibits the input variables that have been found significant by more than one model. The sample used in the study contains data concerning approximately 450 publicly listed, nonfinancial, U.K. and Irish firms. The input vector consists of several quantitative and one qualitative variables.

This study has implications on internal and external auditors, company decision-makers, investors, financial analysts, and researchers. It helps the company management executives to realize how auditors evaluate their clients and the importance of the financial and nonfinancial factors used in their evaluation. It can also be used to identify the most probable outcome ahead of the external audit. Moreover, with the employment of classification models, auditors can simultaneously screen a large number of firms and direct their attention to the ones that the model will identify as having a high probability of receiving a qualified opinion, hence saving time and money. Researchers can use empirical models to assess the extent to which a qualification could be expected, based on publicly available data (Dopuch et al. 1987).

The paper is organized as follows: The second section reviews relevant former research. The third section constitutes an insight into the research methodology being used. In the fourth section, the developed models are described and their results are being analyzed. Finally, the fifth section comprises the concluding remarks.

FORMER RESEARCH

In accomplishing their task, auditors use Statements on Auditing Standards (SAS), which define factors associated with qualified reports and provide auditors with guidelines to minimize the audit risk. In the U.S.A., the SAS 59 offers guidance to auditors to identify the possibility of qualification problems. It identifies the conditions that an auditor should take into consideration in evaluating a going-concern status. These conditions include financial and operating problems.

Considerable research effort has been directed towards the development of models capable of identifying cases of qualified audit opinions. Dopuch et al. (1987) examined the extent to which a probit model, based on financial and market variables, can be used to predict auditors' decisions when issuing qualified audit reports. Among the three qualification types, the going-concern opinion had the highest accuracy rate in prediction. Coats and Fant (1993), using COMPUSTAT data, developed an NN model to predict going-concern audit opinion. They found that NNs perform better than multiple discriminant analysis. Lenard et al. (1995) developed models to identify cases where firms obtained a modified audit report for going-concern uncertainty. They compared a Generalized Reduced Gradient (GRG) optimizer for NNs model, a backpropagation NN and a Logit model. The data used was drawn from the 1988 Disclosure II database and the input vector consisted of publicly available financial ratios and account values. They concluded that the GRG NN model performed better, achieving an accuracy rate of 95%. Laitinen and Laitinen (1998) used the multivariate logistic regression analysis, based on 17 financial and nonfinancial variables, to explain qualifications in large Finnish companies. Results showed that the likelihood of receiving a qualified audit report is larger, the lower the growth of the firm, the lower the share of equity in the balance sheet, and the smaller the number of employees.

Ramamoorti et al. (1999) examined the potential effectiveness of NNs as a risk assessment tool for internal auditors. According to the authors, logistic regression and stepwise multiple regression results compare favorably with neural network results. Anandarajan and Anandarajan (1999) compared three alternative methods—multiple discriminate analysis, expert systems, and NNs to predict the type of going-concern report that should be issued. Fourteen financial ratios were used as input variables. They report a better NN performance. Three alternative models for predicting the future going-concern status were tested by Etheridge et al. (2000). The three models were different approaches to NNs, i.e., Backpropagation NN, Categorical Learning NN, and Probabilistic NN. They used 57 financial ratios as input variables. According to the defined misclassification cost, different models have been found preferable. The study of O'Leary (1998) is indicative of the extent to which neural networks have been applied to develop predicting models. O'Leary (1998) analyzed 15 articles that applied ANNs to predict corporate failure. He provided information on data, ANN models, software, and architecture.

Spathis (2003), in a logistic regression study, tested various combinations of variables and the ability to correctly predict the audit opinion. He also used OLS regression analysis to find the significant independent variables. The results of the model suggested that there is potential in detecting qualified audit reports through analysis of publicly available financial statements and firm litigation data. Spathis et al. (2003) developed a model using the multicriteria technique UTADIS. The analysis suggested that receivables to sales, net profit to total assets, sales to total assets, and working capital to total assets are useful predictors of audit qualifications. The UTADIS method was found quite effective in predicting qualified/clean reports, providing an estimated classification accuracy of approximately 80 percent. This suggests that there is potential for identifying pre-engagement factors associated with qualified audit reports in the analysis of publicly available financial statements.

Hudaib and Cooke (2005) used Logistic Regression to explore the combined effects of the change of the managing director in situation of financial distress to model audit opinion. They found that companies that are financially distressed and that have their managing director changed are most likely to receive a qualified audit report.

Considering the disadvantages of Backpropagation NNs, Gaganis et al. (2007) investigated the capacity of Probabilistic Neural Networks (PNN) to predict qualified audit opinions. According to the reported results, the PNN model outperforms an Artificial Neural Network model and a logistic regression model.

Bayesian Belief Networks offer additional advantages. Kirkos et al. (2007) in a study referring to the application of BBNs for the detection of fraudulent financial statements found that BBNs outperform NNs in predictive accuracy.

A critical observation of the research literature reveals that developing models in order to predict cases of qualifications is an active research field. However, all studies employ either some version of regression or some version of neural networks. Apart from their advantages, these methods have also specific limitations. Regression imposes arbitrary assumptions and requires the inclusion of all significant variables in the input vector. Neural networks act as "black boxes" and require expertise for the definition of the network's topology. In this study, we test Decision Trees and Bayesian Belief Networks against the widely used neural networks. As described in the introduction section, these two methods offer significant advantages over Logistic Regression and Neural Networks.

RESEARCH METHODOLOGY

Data

We use in this study the FAME (Financial Analysis Made Easy) Database. It contains data on 3,000,000 U.K. and Irish firms. There are several types of qualified opinions in the U.K. that are related to accounting treatment, account disclosure, lack of audit evidence, and going-concern issues. In FAME, the audit information refers to whether the auditors issued a qualified or a nonqualified audit opinion. Thus, there is no way of distinguishing the firms according to the type of qualification.

After excluding the financial companies (i.e., banks and insurance companies), we selected the publicly listed firms that were qualified at least once over the past ten years (1995–2004). Many firms were qualified for successive years. This could relate structural characteristics of specific companies that are present for a certain time period. However, the multiple participation of a firm in the sample, for each year that it obtained a qualified report, could create observations containing repetitive information. Since such observations could bias the sample, we chose to introduce each qualified firm only once. The sample contained 225 qualified firms in total. The qualified firms were matched with an equal number of unqualified firms. The match was performed in terms of activity (four-digit SIC code) and fiscal year to eliminate macroeconomic influences. Further matching criteria like size and turnover could also have been imposed. However, matching observations, according to values of an attribute, eliminates the potential of the attribute to be used as a possible predictor. In order to avoid such an information loss, we did not impose additional matching criteria.

The final sample contained 450 firms. A number of companies belonging to the sample contained missing values (missing value percentage 10.69 percent). We substituted the missing values for the mean per class value (Han and Camber 2000).

Variables

The selection of variables to be used as candidates for participation in the input vector was based upon former research work, linked to the topic of qualified opinion detection. In total, we selected 26 financial ratios. In an attempt to reduce dimensionality, we ran one-way ANOVA to test whether the differences between the two classes (groups) were significant for each variable or not. If the difference was not significant (high p-value), the variable was considered non-informative. Table 1 depicts the means, standard deviations, F-values, and p-values for each variable.

As can be seen in Table 1, sixteen variables presented low p-values ($p < = 0.05$). These variables were chosen to participate in the input vector, while the remaining variables were discarded.

Descriptive statistics provide some initial indicators concerning the characteristics of the firms that obtain qualified opinions. Unqualified firms tend to be considerably bigger in terms of Total Assets, Turnover, and Shareholders Funds. The audit fees paid by the unqualified firms are higher than those paid by the qualified firms, but this can be attributed to their larger size. The ratio Audit Fees to Total Assets reveals that qualified firms pay much higher Audit Fees in relation to their size. Non-Audit Fees do not seem to be significant. All variables related to profitability present lower mean values for the qualified firms. Trends also do not seem to be significant, since all trend variables present high p-values. The Z-Score mean value is much lower for the qualified firms, indicating that financial distress may be associated with qualified reports. Finally, liquidity seems to be irrelevant since all relevant ratios have high p-values and were rejected from the input vector.

Methods

In order to identify qualified reports, we used a two-step procedure. In the first step, a model is trained by using a training sample. The sample is organized in tuples (rows) and attributes (columns). One of the attributes, the class-label attribute, contains values indicating the predefined class to which each tuple belongs. This step is also known as supervised learning. In the second step, the model attempts to classify objects that do not belong to the training sample and that form the validation sample. The methods we used are Decision Trees, Neural Networks, and Bayesian Belief Networks.

Decision Trees

A decision tree (DT) is a tree structure, where each node represents a test on an attribute and each branch represents an outcome of the test. In this way, the tree attempts to divide observations into mutually exclusive subgroups. Decision Trees offer considerable advantages. They are nonparametric and make no assumptions about the independence of the input variables or the distribution of the data. Decision trees produce comprehendible models, since the decision-making process is constituted of a series of simple decisions. Their

TABLE 1
Descriptive Statistics, F, and p-values by One-Way ANOVA

Variables	Qualified Mean	Qualified Std. Dev.	Unqualified Mean	Unqualified Std. Dev.	F	p-value
Turnover	**109981**	**624955**	**490298**	**1665774**	**8.56**	**0.004**
Profit (Loss) before Taxation	**−44020**	**412009**	**58349**	**367942**	**7.70**	**0.006**
Profit before Tax. Margin	0.0294	1.2603	0.0705	0.4356	0.20	0.654
Working Capital	**10059**	**93915**	**54668**	**194939**	**8.94**	**0.003**
Increase (Decrease) Cash and Equivalents	−302	9166	3310	27512	3.39	0.066
Current Ratio	2.408	5.046	2.955	5.331	1.24	0.266
Liquidity Ratio	2.163	4.998	2.690	5.344	1.16	0.282
Solvency Ratio %	**38.97**	**42.90**	**54.30**	**24.81**	**21.00**	**0.000**
Gearing %	**271.4**	**859.3**	**79.5**	**199.0**	**9.59**	**0.002**
Current Assets (Trend)	26.5	149.4	22.8	65.0	0.11	0.735
Total Assets (Trend)	11.34	106.55	24.12	79.95	1.95	0.163
Current Liabilities (Trend)	27.40	91.39	26.69	83.57	0.01	0.933
Return on Shareholders Funds	**−112.7**	**189.6**	**−5.1**	**101.1**	**51.64**	**0.000**
Return on Total Assets %	**−67.49**	**116.15**	**−2.40**	**26.74**	**66.92**	**0.000**
Turnover % Trend	20.9	129.2	36.2	97.4	1.72	0.191
QuiScore	**32.06**	**29.08**	**61.06**	**26.44**	**121.92**	**0.000**
IfBig	**0.4170**	**0.4942**	**0.6116**	**0.4885**	**17.52**	**0.000**
Total Assets	**120227**	**702949**	**687713**	**2776016**	**8.76**	**0.003**
Current Liabilities	**57664**	**425605**	**156541**	**569321**	**4.34**	**0.038**
Long Term Liabilities	**66946**	**366974**	**259546**	**1101855**	**4.39**	**0.037**
Shareholders Funds	**15334**	**199740**	**288748**	**1259448**	**10.34**	**0.001**
Audit Fees	**98.1**	**317.4**	**301.7**	**798.7**	**12.33**	**0.000**
Non Audit Fees	171.2	749.0	354.2	1042.9	3.66	0.057
Zscore	**−4.22**	**21.23**	**1.26**	**1.66**	**14.23**	**0.000**
WC/TA	−0.396	4.843	0.137	0.146	2.67	0.103
Audit Fees/TA	**0.02776**	**0.12978**	**0.00280**	**0.00302**	**8.25**	**0.004**

Bold characters indicate the significant input variables.

learning algorithm is very fast. Decision Trees are immune to the presence of irrelevant input variables or the presence of missing values and outliers. They are able to handle categorical attributes. A major disadvantage of decision trees is that they are sensitive to changes of the sample.

In this study, we use the C4.5 decision tree algorithm. C4.5 is an extension of ID3 that accounts for unavailable values, continuous value attribute ranges, pruning of decision trees, and rule derivation.

Neural Networks

Neural Networks (NN) is a mature technology with an established theory and many recognized application areas. Neural Networks make no assumptions about attributes' independence, are capable of handling noisy or inconsistent data, and are a suitable alternative for problems where an algorithmic solution is not applicable. Another advantage is their predictive performance. Major disadvantages of neural networks are their slow learning algorithm, their poor interpretability, and the experience required for the definition of their topology. Calderon and Cheh (2002) point out that NNs are subject to problems of local minima. NNs can also be very sensitive to specification of learning rates, momentum and other processing elements, and there is no clear guidance on selecting these parameters.

Bayesian Belief Networks

A BBN is a directed acyclic graph, where each node represents an attribute and each arrow represents a probabilistic dependence. The network structure can be defined in advance or can be inferred from the data. For classification purposes, one of the nodes can be defined as the class node. The network can calculate the probability of each alternative class and assign a new observation to the class with the maximum probability.

Bayesian Belief Networks offer significant advantages. BBNs model the probability distribution of the problem domain. In this sense, they are very suitable in cases where complex interrelationships exist between the dependent and independent variables or even among the dependents themselves. BBNs are highly understandable by humans since they present a graph which depicts the relationships between the dependent and independent variables. They also enable the user to modify the model that was extracted from the data. Irrelevant dependencies can be removed and significant ones added. In this manner, BBNs allow the integration of the extracted model with the expert's domain knowledge. Kirkos et al. (2007), in a study referring to the application of BBNs for the detection of fraudulent financial statements, found that BBNs outperform NNs in predictive accuracy. BBNs can handle categorical variables and are tolerant to missing data. Perhaps the most significant disadvantage of an approach involving Bayesian Networks is the fact that there is no universally accepted method for constructing a network from data.

Table 2 summarizes specific characteristics of the three employed methods. Both Neural Networks and Bayesian Belief Networks achieve high accuracy rates. Decision Trees and Bayesian Belief Networks are highly interpretable by humans.

EXPERIMENTS AND RESULTS ANALYSIS

Three models were built. The first model was the C4.5 Decision Tree. The software used was the TANAGRA (Rakotomalala 2005) data mining research software. The tree was built with 0.25 percent confidence level. The whole sample was used as the training set. The constructed decision tree included 55 nodes and 28 leaves.

TABLE 2
Method's Characteristics

	Decision Trees	Neural Networks	Bayesian Networks
Interpretability	high	low	high
Accuracy	medium	high	high
Learning Speed	high	low	medium

The Decision Tree uses, as first level splitter, the variable Z-Score. In the first splitting node, the tree differentiates 210 out of 225 unqualified firms and 132 out of 225 qualified firms according to their Z-Score value. It seems that companies in a rather good financial position manage to obtain clean reports, whereas financially distressed companies tend to obtain qualified reports. One variable associated with leverage (*Gearing*) and one variable associated with profitability (Return on Shareholders Funds—*ROSF*) are used as second-level splitters.

In terms of performance against the training sample, the tree managed to correctly classify 408 cases, achieving a general performance of 90.67 percent. Accuracy per class rate is 92.00 percent for the qualified (207 cases) and 89.33 percent for the unqualified (201 cases) firms.

The second model was the Multilayer Perceptron (MLP). The model was built with Tanagra Software. In neural networks, the definition of the network's topology is a matter of experience. In order to assure the selection of a proper topology, we tested four alternative architectures. All alternative models had one hidden layer and a different number of hidden neurons. According to a ten-fold cross validation test, the model having ten hidden neurons achieved the highest overall performance (81.11 percent). The other performances were 80.47 percent for a network with six hidden nodes, 80.80 percent for a network with eight hidden nodes, and 80.47 percent for a network with twelve hidden nodes. According to the performances, we choose the topology that had ten hidden nodes. The defined learning rate for training was 0.15 and the error-rate threshold was 0.01. The network was trained by using the whole sample and was tested against the training sample. The performance against the training sample was 81.56 percent. More specifically, the model correctly classified 175 qualified firms (77.78 percent) and 192 unqualified firms (85.33 percent).

In order to estimate the attributes' contribution for the multilayer perceptron classifier, Tanagra performs an iterative test by excluding one attribute each time and recalculating the error rate for each case. Although the differences are rather small, it is worth mentioning that the variables *ROSF* and *Z-Score* that are used as high-level splitters in the C4.5 tree present a higher statistical value in the error-rate change test. The variable *Audit Fee to Total Assets* appears also to have a considerable contribution.

The third model was the Bayesian Belief Network (BBN). In order to develop the network, we used the BN Power Predictor software. Although in many BBN software packages the user has to define the structure of the network, the algorithm of BN Power Predictor is capable of extracting the network from the data. The implemented algorithm belongs to the category of conditional independence, test-based algorithms, and does not require node ordering (Cheng and Greiner 2001). The software allows the user to modify the extracted model. Such interference could favor the BBN model and thus we decided not to modify the model. One limitation of the software being used is that it requires discretised data. We performed supervised entropy-based discretisation, because this method uses the class information to define the intervals. Thus, the discretised data is more suitable for the classification task.

The model was built by using the whole sample as a training set. The network achieved a general classification accuracy of 86.44 percent against the training set, managing to correctly classify 184 out of 225 qualified firms (accuracy rate 81.78 percent) and 205 out of 225 unqualified firms (accuracy rate 91.11 percent).

The Bayesian Belief Network directly associates specific input variables with the class attribute by recording dependencies between them. The developed model associates qualifications with the variables *Z-Score*, *ROSF*, *Turnover*, *PLBT* (Profit [Loss] before Taxation), *Gearing*, *ROTA*, Quiscore, and Long-Term Liabilities. Remarkably, the variables *Z-Score* and *ROSF* have also been found significant both by the C4.5 Decision Tree and the Multilayer Perceptron classifier. The variables *Turnover* and *Gearing* have also been found significant by the C4.5 Tree.

A basic aim of this study is to locate the variables that are of significant importance in discriminating the qualified cases from the unqualified ones. All of the three models agree on the fact that financial distress, which is measured by the input variable *Z-Score*, is associated with audit qualifications. Profitability matters are also strongly related to qualifications, since all of the three models associate the variable *ROSF* with qualifications. These variables are proposed as possible indicators. Both C4.5 and BBN models reveal dependencies between audit qualification and the variables *Gearing* and *Turnover*, thus providing clues that leverage and sales performance can be related to qualified opinions. Liquidity seems to be irrelevant, since all of the three variables associated with liquidity were discarded according to the results of ANOVA. The results support the findings of previous studies, which indicate that firms that receive qualified opinions are less profitable (Loebbecke et al. 1989; Laitinen and Laitinen 1998; Spathis 2002) and more likely to default (Bell and Tabor 1991; Laitinen and Laitinen 1998; Reynolds and Francis 2001; Spathis 2002; Spathis et al. 2003; Hudaib and Cooke 2005).

In terms of the possible relation between auditors' characteristics and qualified opinions, our results seem to grant credits to auditors. Only the MLP model associates the level of audit fees with qualified reports, by ranking the variable Audit Fees to Total Assets in the second place in the attributes' contribution table. Moreover, the similar input variable *Audit Fees* is rejected by all of the three models. None of the three models associates the qualitative variable *IfBig* (which indicates whether the auditor is a large auditing firm) with qualified reports. Finally, the variable *Non-Audit Fees* was discarded according to ANOVA.

Models' Performance and Validation

The first three lines of Table 3 depict the performances of the models against the training sample. The Decision Tree model achieves the best performance. Additionally, the Decision Tree achieves a relatively balanced accuracy per class rate. The BBN and the MLP have a lower performance. They also both present a significantly better accuracy rate for the unqualified cases.

Using the training set in order to estimate a model's performance might introduce a bias. In many cases, the models tend to memorize the sample instead of "learning" (data over-fitting). The effect of data over-fitting is that the model describes in detail the training set, but it is unable to classify correctly new observations. Decision Trees achieve accuracy rate 100 percent against the training sample if no pruning is applied. In this sense, the performances against the training sample are just a measure of how well the models describe the specific sample. The true power of a model is its ability to classify previously unseen patterns.

TABLE 3
Training and Validation Performances

Model	Qualified	Unqualified	Type I Error	Type II Error	Total
C4.5 (train. set)	92.00%	89.33%	8.00%	10.67%	90.67%
MLP (train. set)	77.78%	85.33%	22.22%	14.67%	81.56%
BBN (train. set)	81.78%	91.11%	18.22%	8.89%	86.44%
C4.5 (10-f cross val.)	76.62%	78.76%	23.38%	21.24%	77.69%
MLP (10-f cross val.)	78.44%	83.78%	21.56%	16.22%	81.11%
BBN (10-f cross val.)	76.44%	88.00%	23.56%	12.00%	82.22%

There are several methods for validating the model against new observations. The simplest method is to use a part of the sample for training and another part for validation (hold-out sample). Another validation method is the ten-fold cross validation. In ten-fold cross validation, the sample is divided in ten equal, randomly selected folds. The selection of the folds is not significant. Nine folds are used to train the model and one fold is used for validation. This process iterates ten times by using each time a different fold for the validation. In this way, the whole sample is used for validation. Compared to the hold-out sample method, ten-fold cross validation is more reliable because it introduces lower bias and variance.

The two software packages used presented differences in their validation capabilities. BNP allows only the validation against a hold out sample. Tanagra embodies build-in modules for both hold out sample and ten-fold cross validation. In order to assure results that are more reliable, we avoided the convenience of using the built-in hold-out-sample method provided by both software packages, and we applied ten-fold cross validation. Tanagra performs ten-fold cross validation automatically, and the user has no control on the selection of the folds. For the BNP case, we created manually the ten folds and we iterated the hold-out-sample test ten times. The folds used for validating each model were different; yet this is acceptable since the random selection of the folds is a basic supposition of this validation method. The last three lines of Table 3 depict the performances of the models according to the ten-fold cross validation test.

As expected, the accuracy rates are lower, when models try to predict the class label of unknown observations. However, each model presents a different Behavior. The Decision Tree, which achieves the best performance against the training sample, reduces its classification accuracy by a magnitude of 13 percent and manages to correctly classify 77.69 percent of the total cases. The performance of C4.5 is the lowest compared with the other two models. The MLP model correctly classified 81.11 percent of the total cases. Remarkably, this performance is almost equal to its accuracy rate against the training sample. Finally, although the BBN model presents a considerable decrement of its performance, it still achieves the best accuracy rate (82.22 percent of the total cases). In a comparative assessment of the models' performance, we can conclude that the Bayesian Belief Network outperforms the other two models.

In assessing the performance of a model, another important consideration is the Type I and Type II error rates. A Type I error is committed when a qualified company is classified as unqualified. A Type II error is committed when an unqualified company is classified as qualified. Although in most cases Type I and Type II errors have different costs, in this research topic both types of errors have significant costs. The misclassification of a qualified firm may lead to a clean report that does not disclose the true picture of the company, whereas the misclassification of an unqualified firm may lead to an unfairly qualified report that may cause to the company economic problems. Former studies in auditing that deal with the application of different forms of neural network models in fraud detection have achieved rather unbalanced Type I and Type II error rates (Fanning et al. 1995; Fanning and Cogger 1998). According to the results of ten-fold cross validation, our three models have comparable classification accuracy rates for the qualified cases. For the unqualified cases, the BBN model has a significantly lower error rate followed by the MLP and the C4.5 models. A limitation of both software packages used is that they do not inform the user on which observations are misclassified. Thus, it was not possible to estimate the concurrence of the predictions of the three models.

CONCLUSIONS

Modern auditing can be facilitated by the new knowledge extraction techniques of Data Mining. Considerable research effort has already been directed towards the development of models capable of identifying qualified reports and of extracting variables and factors significant in forming the auditors' opinion. A review of the related research studies reveals that researchers mostly use either Logistic Regression Analysis or some version of Neural Networks. However, there are numerous Data Mining classification techniques that have not yet been applied towards the purpose of developing qualified opinion identification models, despite the fact that these techniques offer considerable advantages.

In this study, we employed three Data Mining classification techniques to develop models capable of identifying cases of qualified audit opinions. The input vector contained one qualitative and several quantitative variables. Preliminary feature selection was performed by running a one-way ANOVA. The three developed models have been proven capable of distinguishing the qualified cases. The Decision Tree model achieves the highest accuracy rate against the training set. However, the performance against the training set is just a measure of how well the model describes the specific sample. We estimated the true predictive power of the models by using ten-fold cross validation. According to the ten-fold cross validation results, the Bayesian Belief Network achieves the highest classification accuracy (82.22 percent of the total observations). The Multilayer Perceptron model achieves a marginally lower performance (81.11 percent). Finally, the Decision Tree Model achieved the lowest performance (77.69 percent). The three models have almost similar Type I error rates. The differences in their overall performance arise mainly from their different Type II error rates. In terms of Type II errors, the BBN model outperformed the MLP model by a magnitude of 4 percent and the C4.5 model by a magnitude of 9 percent. The C4.5 model has been proven the most balanced, having comparable Type I and Type II error rates. The fact that the Bayesian Belief Network achieves the highest accuracy rate when tested against unknown patterns, as well as the fact that this model is highly interpretable, provides evidence that Bayesian Belief Networks is the most suitable method.

According to our results, financial distress and profitability are strongly related to qualified opinions, since the corresponding variables Z-Score and Return on Shareholders Funds have been selected by all of the three models. The variables Gearing and Turnover have been found significant by the C4.5 and BBN models. All the liquidity related variables were rejected. Finally, three out of four variables concerning the auditors' characteristics, i.e., the variables Audit Fees, Non-Audit Fees, and the qualitative variable that indicates that the auditor is a large auditing firm, have been found irrelevant. Only the MLP model associated the variable Audit Fees to Total Assets with qualified opinions.

By processing the data, the models manage to predict the auditors' opinions to a percentage of approximately 80 percent. This performance provides evidence that these methods are reliable predictive tools. Typically, an auditor is provided with much more information than common financial statement ratios and values. Providing models with this additional information could further improve their predictive power. Moreover, these methods highlight factors that are associated with qualified opinions and can be used as red flags. These models could be of particular use to auditors in their daily practice regarding the assessment and monitoring of their clients. In spite of their predictive or explanatory power, these models can only be used as assisting tools for auditors. Decision support systems cannot surrogate the human subjective judgment, experience, and initiative, but can substantially support auditors to handle data with many records and several variables that are difficult to analyze otherwise.

The results of our study could be of assistance to internal and external auditors, taxation and other state authorities, credit scoring agencies, financial analysts, individual and institutional investors, and law firms. Such studies that develop predicting models and reveal red-flag indicators could be beneficial for the auditing profession, towards the purpose of identifying cases where qualified opinions should be issued.

This study can also be used as a stepping-stone for further research. One limitation of this study is that we use only three classification methods. Other Data Mining methods, like Support Vector Machines and Rough Sets, are noted as good classifiers but remain to be tested in terms of their performance and explanatory power. Financial matters measured with financial ratios and account values have been well covered in this study. Still, there are numerous qualitative variables that merit further study. Arnold et al. (2001) compare the practices of auditors and insolvency managers and highlight the importance of aspects described by qualitative variables. Enriching the input vector with these qualitative variables could further improve the models' performance. The fact that both DTs and BBNs can handle categorical variables makes them a candidate tool for modeling these aspects. Another consideration is to construct models able to discriminate the cases according to the type of qualifications. Such a research presupposes a sample with a multi-label class attribute describing the specific type of qualification. A topic that merits further study is to investigate the coincidence of the predictions of different models. Such research could contribute to the development of an aggregated classifier that involves individual classifiers in a weighted voting scheme. Finally, industry-specific studies could reveal industry-specific indicators. We hope that the research presented in this paper will, therefore, stimulate additional work regarding these important topics.

REFERENCES

Anandarajan, M., and A. Anandarajan. 1999. A comparison of machine learning techniques with a qualitative response model for auditors' going concern reporting. *Expert Systems with Applications* 16: 385–92.

Arnold, V., A. P. Collier, A. S. Leech, and G. S. Sutton. 2001. The impact of political pressure on novice decision makers: Are auditors qualified to make going concern judgments? *Critical Perspectives in Accounting* 12: 323–338.

Bell, T., and R. Tabor. 1991. Empirical analysis of audit uncertainty qualifications. *Journal of Accounting Research* 29: 350–370.

Calderon, T. G., and J. J., Cheh. 2002. A roadmap for future neural networks research in auditing and risk assessment. *International Journal of Accounting Information Systems* 3 (4): 203–236.

Chen, K., and B. Church. 1992. Default on debt obligations and the insurance of going concern opinions. *Auditing: A Journal of Practice & Theory* (Fall): 30–49.

Cheng, J., and R. Greiner. 2001. Learning Bayesian belief network classifiers: Algorithms and systems. *Proceedings of the 14th Biennial Conference of the Canadian Society on Computational Studies of Intelligence: Advances in Artificial Intelligence*, 141–151.

Coats, P. K., and L. F. Fant. 1993. Recognizing financial distress patterns using a neural network tool. *Financial Management* 22 (3): 142–155.

Dopuch, N., R. Holthausen, and R. Leftwish. 1987. Predicting audit qualifications with financial and market variables. *The Accounting Review* 62 (3): 431–454.

Etheridge, H. L., R. S. Sriram, and H. Y. K. Hsu. 2000. A comparison of selected artificial neural networks that help auditors evaluate client financial viability. *Decision Science* 31 (2): 531–50.

Fanning, K., K. O. Cogger, and R. Srivastava. 1995. Detection of management fraud: A neural network approach. *International Journal of Intelligent Systems in Accounting Finance and Management* 4 (2): 113–126.

———, and ———. 1998. Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting Finance and Management* 7 (1): 21–41.

Fu, L. M. 1991. Rule learning by searching on adapted nets. *Proceedings of the Ninth National Conference on Artificial Intelligence*, 500–595.

Gaganis, C., F. Pasiouras, and M. Doumpos. 2007. Probabilistic neural networks for the identification of qualified audit opinions. *Expert Systems with Applications* 32 (1): 114–124.

Gallant, S. I. 1988. Connectionist expert systems. *Communications of the ACM* 31: 152–169.

Han, J., and M. Camber. 2000. *Data Mining Concepts and Techniques.* San Diego, CA: Morgan Kaufman.

Hudaib, M., and T. E. Cooke. 2005. The impact of managing director changes and financial distress on audit qualification and auditor switching. *Journal of Business Finance & Accounting* 32 (9–10): 1703–1739.

Kirkos, E., C. Spathis, and Y. Manolopoulos. 2007. Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications* 32 (4): 995–1003.

Kleinman, G., and A. Anandarajan. 1999. The usefulness of off-balance sheet variables as predictors of auditors' going concern opinions: An empirical analysis. *Managerial Auditing Journal* 14 (6): 286–293.

Koskivaara, E. 2004. Artificial neural networks in analytical review procedures. *Managerial Auditing Journal* 19 (2): 191–223.

Laitinen, E. K., and T. Laitinen. 1998. Qualified audit reports in Finland: Evidence from large companies. *European Accounting Review* 7 (4): 639–653.

Lemon, W. M., K. W. Tatum, and W. S. Turley. 2000. *Developments in the audit methodologies of large firms.* Canadian Institute of Chartered Accountants and American Institute of Certified Public Accountants (CICA/AICPA).

Lenard, M. J., P. Alam, and G. R. Madey. 1995. The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sciences* 26 (2): 209–227.

Loebbecke, J., M. Eining, and J. Willingham. 1989. Auditors' experience with material irregularities: Frequency, nature, and detectability. *Auditing: A Journal of Practice & Theory* 9: 1–28.

O'Leary, D. E. 1998. Using neural networks to predict corporate failure. *International Journal of Intelligent Systems in Accounting, Finance & Management* 7: 187–97.

Rakotomalala, R. 2005. TANAGRA: Un logiciel gratuit pour l'enseignement et la recherche. In *Actes de EGC'2005, RNTI-E-3* (2): 697–702.

Ramamoorti, S., A. D. Bailey, and R. O. Traver. 1999. Risk assessment in internal auditing: A neural network approach. *International Journal of Intelligent Systems in Accounting, Finance & Management* 8: 159–180.

Reynolds, J., and J. Francis. 2001. Does size matter? The influence of large clients on office-level auditor reporting decisions. *Journal of Accounting and Economics* 30: 375–400.

Salterio, S. E. 2000. A primer on the strategic systems approach to auditing. Working paper, University of Waterloo.

Spathis, C. 2002. Detecting false financial statements using published data: Evidence from Greece. *Managerial Auditing Journal* 17 (4): 179–191.

———. 2003. Audit qualification, firm litigation, and financial information: An empirical analysis in Greece. *International Journal of Auditing* 7 (1): 71–85.

———, M. Doumpos, and C. Zopounidis. 2003. Using client performance measures to identify pre-engagement factors associated with qualified audit reports in Greece. *The International Journal of Accounting* 38 (3): 267–284.

Tabachnick, B., and L. Fidell. 2001. *Using multivariate statistics.* Fourth edition. Boston, MA: Allyn & Bacon.

Vasarhelyi, M. 2005. Editor's letter. *Journal of Emerging Technologies in Accounting* 2 (1).