

AUDIT-FIRM GROUP APPOINTMENT: AN ARTIFICIAL INTELLIGENCE APPROACH

Efstathios Kirkos,^a Charalambos Spathis^{b*} and Yannis Manolopoulos^c

^a *Department of Accounting, Technological Educational Institution of Thessaloniki, PO Box 141, 57400, Thessaloniki, Greece*

^b *Division of Business Administration, Department of Economics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece*

^c *Department of Informatics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece*

SUMMARY

Auditor appointment can be regarded as a matter of pursued audit quality and is driven by several factors. The adoption of an effective auditor procurement process increases the likelihood that a company will engage the right auditor at a fair price. In this study, three techniques derived from artificial intelligence (AI) are used to propose models capable of discriminating between cases where companies appoint a Big 4 or a Non-Big 4 auditor. These three AI methods are then compared with the broadly used method of logistic regression. The results indicate that two of the AI techniques outperform logistic regression. In addition, one method further improves its performance by applying bagging. Finally, significant factors associated with auditor appointment are revealed. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: auditor appointment; artificial intelligence; audit quality; data mining

1. INTRODUCTION

Today's economy is facing several unsettling phenomena. The collapse of major corporations like Enron, WorldCom and Tyco, the increasing number of management fraud cases (Kirkos *et al.*, 2007a) and the recent financial crisis have led to the losses of billions of dollars. Such phenomena call for improved controlling mechanisms. Governments are currently attempting to address the situation by establishing rules and regulations, such as the Sarbanes Oxley Act in the USA and the 8th Directive in the EU.

One of the main controlling mechanisms is auditing. The primary objective of auditing is to ensure the proper disclosure of a company's financial status and, thus, to reduce the asymmetry of the information flow among managers, shareholders and creditors. Auditing can have a variety of beneficial effects. According to agency theory, the separation of ownership and control in modern corporations creates incentives for managers, who are not owners, to act in their own interests at the expense of stakeholders and creditors (Kane and Velury, 2004). Auditing can prevent management's malfeasance, by adding credibility to statutory financial reporting, thus reducing the risk of misinformation and, consequently, investment risk. In this sense, reliable auditing can drive share prices higher and reduce the cost of capital (Houghton and Jubb, 2003). Auditing may also lead to an improvement in the

* Correspondence to: Charalambos Spathis, Division of Business Administration, Department of Economics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece. E-mail: hspathis@econ.auth.gr

efficiency of business processes and assist in regulatory compliance (Knechel *et al.*, 2008). Such phenomena can gain momentum through the implementation of an audit with a higher level of quality (Broye and Weill, 2008).

Despite the abovementioned beneficial effects, however, modern auditing suffers from an internal contradiction. This contradiction arises from the fact that the auditor must remain independent and protect investors and creditors, when it is the auditee that determines the auditor's appointment, retention and fees. This contradiction may jeopardize the auditor's objectivity. Thus, the question of quality auditing remains unresolved. Nevertheless, it is a fact that when a higher quality level of auditing is selected, the phenomenon of asymmetry is reduced even further (Broye and Weill, 2008).

It is generally acknowledged that audit markets are segmented into at least two categories: the Big 4 auditors and the Non-Big 4 auditors. Big auditors are regarded as higher quality auditors in the relevant literature (DeAngelo, 1981; Mutchler, 1986; Palmrose, 1988; Bartov *et al.*, 2001; Craswell *et al.*, 2002). Large auditing firms, owing to their size, are in a better position to withstand client pressure than smaller firms are; they invest more in technology, training and facilities and they have increased incentives to maintain their reputation and professional standards. Earlier studies have shown that reported discretionary accruals that increase income are significantly lower for clients of Big Five auditors (Becker *et al.*, 1998; Francis *et al.*, 1999; Gul *et al.*, 2002; Kim *et al.*, 2003). For these reasons, a number of prior research studies have used audit firm size as a proxy for audit quality (Teoh and Wong, 1993; Becker *et al.*, 1998).

The appointment of an audit firm is a complex process. Shareholders are keen to hire high-quality auditors in order to minimize the opportunistic behaviour of management and confirm the credibility of statutory reports; in this way they reduce the cost of capital and facilitate share trading (Houghton and Jubb, 2003). Managers who wish to express their alignment with the shareholders' interests also wish to hire high-quality auditors. However, high-quality auditors invest more in technology and training and, therefore, charge higher fees. Another concern is that, in the case of business failure and subsequent audit failure, the appointment of a particular auditor may have to be justified. The appointment of a big auditor ensures the reliability of the accounting information which is used to price debt and evaluate the risk of covenant violation (Broye and Weill, 2008). The adoption of an effective auditor procurement process increases the likelihood that the company will engage the right auditor at a fair price. Decision patterns, in the form of models which link the demand for quality-differentiated audits with company-specific factors, may contribute to this end.

An examination of the relevant literature reveals that researchers reject the null hypothesis that clients are randomly allocated to big and non-big auditors. Research efforts have been directed towards the question of appointing auditors. However, related studies have only made use of traditional statistical techniques, like logistic regression. In most cases, their aim is to confirm that specific managerial or financial characteristics are linked to the auditor's appointment. Although regression is the method most commonly used, it is not free of limitations. Several regression versions arbitrarily assume that a linear relation exists between the dependent variable and the independent variables. Logistic regression, which is particularly suited to dichotomous classes, assumes that the logarithm of odds is a linear function of the independent variables; furthermore, irrelevant variables are excluded.

The auditor appointment question can be regarded as a typical classification problem. The methodological framework of artificial intelligence provides methods, techniques and concepts that can be used for classification purposes. As opposed to other well-examined fields, like bankruptcy prediction, credit risk estimation and fraud detection, artificial intelligence (AI) classification methodologies have not yet been tested vis-à-vis the auditor appointment question. AI techniques have a theoretical

advantage that lies in the fact that they do not impose any arbitrary assumptions, like the linear relation between the logarithm of the dependent variable and the independent variables.

The aim of this study is to contribute to the auditor procurement process by putting forward models that are capable of distinguishing cases where firms choose an audit-firm group (i.e. Big 4 or Non-Big 4 auditor). Three methods derived from AI will be used here for the first time, namely decision trees (DTs), neural networks (NNs) and k -nearest neighbours. These methods were selected on the basis of their previous successful application in auditing-related topics (Fanning and Cogger, 1998; Lin and McClean, 2001; Gaganis *et al.*, 2007b; Kirkos *et al.*, 2007a,b). The models developed are compared in terms of their accuracy rate. Techniques for improving model performances are applied and evaluated. The performances of the models are compared with the accuracy rate of the broadly used logistic regression method. Our assumption is that AI methods can outperform traditional statistical techniques, such as logistic regression. A series of important factors associated with the auditor appointment process is revealed. The sample consists of 338 UK and Irish company-year observations during the period 2003–2005. The input vector contains both quantitative (financial) and qualitative variables. The data used are publicly available and easily found in a standard financial database.

The decision of the appointment of an audit-firm group and the decision to switch auditors occasionally are complex choices. Our study contributes to the existing literature to increase our understanding in the appointment behaviour of the management according to audit-firm group, using AI classification models. Audit firms can use these models in designing their competition strategy, providing their managers with indications on what characteristics of client firms they should target. Also, institutional investors have influence in the market for auditing services (Kane and Velury, 2004). Audit firms may increase their attractiveness to institutional investors when the firms provide quality audit services.

The paper proceeds as follows. Section 2 reviews the existing research on this subject. Section 3 provides an insight into the research methodology used. Section 4 describes the models developed and analyses the results. Finally, Section 5 presents the concluding remarks.

2. PRIOR RESEARCH

In the past, research studies have used auditor type in several ways. Waren (1980) investigated uniformity surrogates for auditing standards among the Big 8 auditing firms and other auditors. He concluded that further research was required in the area of auditor differences. DeFond (1992) used the auditing companies' size, brand name (Big 6 auditor), expertise and independence as a combined measure of their ability to alleviate agency conflicts. Citron and Taffler (1992) examined audit firm size and other factors as variables that affect the value of the audit report in a going-concern qualification situation. Krishnan *et al.* (1996) used auditor type (Big 6 versus Non-Big 6) as an independent variable to model audit qualifications and auditor switching.

Recent research studies have attempted to model the auditor appointment process and reveal significant factors. Chaney *et al.* (1996) conducted a survey in privately owned UK firms to investigate the pricing of audit fees and to classify auditor appointment. The method applied was ordinary least-squares regression. In order to model auditor appointment, they used financial ratios and two qualitative variables that indicated whether the firm made sales outside the UK and whether it had incurred a loss in the previous year. Their findings suggest that, when clients are not compelled by market pressure to appoint a big auditor, they opt for the auditor that is available for the lowest cost. The researchers classified auditor appointment and then used this information in the fee analysis. The latter suggests that auditors structure their business in a manner that befits specific client segments.

Citron and Manalis (2001) investigated auditor nomination in Greece after the liberalization of the audit market. Their sample contained 205 companies listed on the Athens Stock Exchange. They developed a model to classify the appointment of a big auditor by using binomial logistic regression. Six financial and non-financial variables were used as input variables. According to their results, the level of shareholdings owned by foreign shareholders is positively associated with the appointment of a big auditor. Beasley and Petroni (2001) examined the US property-liability insurance industry and reported a positive association between insurers who employ specialized big auditors and the percentage of external members on a company's board of directors. The method applied was logistic regression and the input vector mainly contained non-financial variables. Their findings suggest that external board members tend to monitor management more closely and, therefore, influence the board of directors in appointing higher quality audit providers.

Velury *et al.* (2003) provide evidence linking corporate governance mechanisms to auditor nomination. They employed a two-stage least-squares regression method to test the relation between audit quality and the level of institutional investment. Their findings indicate that firms which have relatively higher levels of institutional ownership tend to employ industry specialist auditors. Kane and Velury (2004) investigated the relation between the appointment of a big auditing firm and the level of institutional ownership. They used logistic regression to develop a model that would classify the appointment of a big auditor. Their sample came from Compustat and the Compact Disclosure database. The input vector contained financial ratios and a variable defining the proportion of shares held by institutions. They found that firms with a high level of institutional ownership are more likely to engage a big auditor. They also found a positive association between size and debt and the appointment of a big auditor.

A recent study by Knechel *et al.* (2008) analysed auditor nomination by Finnish firms using logistic regression analysis. They found that the need for a higher quality auditor is first driven by complexity, and then, as the firm grows, it is supplemented by the use of debt financing and ultimately by the need to raise equity as well as debt financing. Broeye and Weill (2008) investigated the impact of the legal environment on the relation between leverage and auditor appointment in 10 European countries using logit models. They found that the need for improved protection of creditor rights and a high disclosure requirement are factors which encourage the demand for a high-quality auditor by leveraged firms. Also, leveraged firms from countries with less stringent auditor liability exposures are more likely to appoint a Big 5 auditor.

A critical observation of the collected literature indicates that all related studies are based on some form of regression. From the research that examines similar classification problems, like bankruptcy prediction or fraud detection, it is observed that methodologies derived from AI perform at least to the same level as statistical techniques (Fanning and Cogger, 1998; O'Leary, 1998; Lin and McLean, 2001). In the field of auditing, relevant studies have only recently employed AI techniques, such as NNs (e.g. Lenard *et al.*, 1995; Gaganis *et al.*, 2007a), support vector machines (Doumpos *et al.*, 2005), nearest neighbours (Gaganis *et al.*, 2007b), DTs and Bayesian belief networks (Kirkos *et al.*, 2007a,b).

Moreover, data mining classification literature provides additional methodologies concerning the improvement of model performance, including the validation of the models against previously unseen patterns. These methodologies have not yet been applied as regards the question of audit-firm group nomination. Thus, the purpose of this study is to apply and compare AI classification techniques in order to validate the audit-firm group nomination and evaluate a performance improvement technique. The derived models are then validated against unknown patterns.

3. RESEARCH METHODOLOGY

3.1. Methods

The issue of modelling auditor nomination can be regarded as a typical classification problem. Classification includes 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 (big4 versus non-big4 auditor). The step of model training is also known as supervised learning. In the second step, the model attempts to classify objects that do not belong to the training sample; they form the validation sample.

AI proposes several classification methods. Three methods that enjoy a good reputation for their classification capabilities are employed in this research study. These methods are DTs, NNs and k -nearest neighbours (k -NNs).

Decision Trees

A 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. The goodness of a split is based on the selection of the attribute that best separates the sample (Han and Kamber, 2000). There are several proposed DT algorithms. The best known is the ID3 algorithm, where the selection of the splitting variable is based on an entropy-based measure, called information gain. In a binary class example, entropy $E(S)$ is defined as

$$E(S) = -p_+ \log_2(p_+) - p_- \log_2(p_-) \quad (1)$$

where S is the set of observations, p_+ is the proportion of positive observations in S and p_- is the proportion of negative observations in S . If the class attribute can obtain c alternative values, then entropy is defined as

$$E(S) = -\sum_{i=1}^c p_i \log_2(p_i) \quad (2)$$

where p_i is the proportion of observations which belong to class i . The notion of information gain $G(S, A)$ represents the reduction in entropy in S if the attribute A is selected as a splitter. Information gain is defined as

$$G((S, A)) = E(S) - \sum_{u \in \text{Values}(A)} \frac{|S_u|}{|S|} E(S_u) \quad (3)$$

where $E(S)$ is the entropy of S , A is the splitting attribute, $\text{Values}(A)$ are the values of A , S_u is the proportion of observations with $A = u$ and $E(S_u)$ is the entropy of the observations with $A = u$.

According to these principles the sample is successively divided into subsets until either no further splitting can produce statistically significant differences or the subgroups are too small to undergo similar meaningful division. The successive division of the sample may produce a large tree. Some of the tree's branches may reflect anomalies in the training set, like false values or outliers. For that reason, tree pruning is required. Tree pruning involves the removal of splitting nodes in a way that does not significantly affect the model's accuracy rate.

DTs offer considerable advantages. They make no assumptions about the independence of the input variables or the distribution of the data. They produce comprehensible models which can be easily converted into a set of If–Then rules. DTs are immune to the presence of irrelevant input variables or the presence of missing values and outliers. Their learning algorithm is very fast. A major disadvantage of DTs, however, is that they are sensitive to sample changes.

In this study, the C4.5 DT algorithm was used. C4.5 is an extension of ID3 that accounts for unavailable values, continuous value attribute ranges, pruning of DTs and rule derivation.

Neural Networks

An NN consists of a number of neurons, i.e. interconnected processing units. Each neuron is connected to other neurons. Associated with each connection is a numerical value, called ‘weight’. Each neuron receives signals from connected neurons and the combined input signal is calculated. The total input signal for neuron j is $u_j = \sum w_{ij}x_i$, where x_i is the input signal from neuron i and w_{ij} is the weight of the connection between neuron i and neuron j . If the combined input signal strength exceeds a threshold, then the input value is transformed by the transfer function of the neuron and finally the neuron fires (Han and Kamber, 2000).

The neurons are arranged into layers. A layered network consists of at least an input (first) and an output (last) layer. Between the input and output layers there may exist one or more hidden layers. Different kinds of NN have different numbers of layers. Self-organizing maps have only an input layer and an output layer, whereas a backpropagation NN has additionally one or more hidden layers.

After the network architecture is defined, the network must be trained. In backpropagation networks, a pattern is applied to the input layer and a final output is calculated at the output layer. The output is compared with the desired result and the errors are propagated backwards in the NN by tuning the weights of the connections. This process iterates until an acceptable error rate is reached.

One of the major advantages of NNs is that they do not assume a linear relation between the independent variable (or its logarithm) and the dependent variables. NNs are also capable of handling noisy or inconsistent data and they are a suitable alternative for problems where an algorithmic solution is not applicable. Another advantage is their predictive performance. Major disadvantages of NNs are their slow learning algorithm, their poor interpretability and the experience required for the definition of their topology. In this study, the multilayer perceptron (MLP) model is used.

k-Nearest Neighbours

According to the k -NN method, a sample containing objects with n attributes is considered to be an n -dimensional space. Each object is one point in the n -dimensional space. k -NN introduces a similarity metric for the objects. When an unknown observation is given, a k -NN classifier searches the sample space for the k training cases that are closest to the unknown observation. These k cases are the k -NNs of the new observation. The classifier assigns the new observation to the most common class among the k -NNs.

In its simplest version, the k -NN method requires arithmetic input variables. The similarity of two objects is calculated as their Euclidean distance. For two objects $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$ their Euclidean distance is defined as

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

One well-known problem with the Euclidean distance is that attributes that have a big value range contribute more than attributes with a small value range. In this study, the heterogeneous Euclidean-overlap metric is used, introduced by Wilson and Martinez (1997), which permits the normalization of the attributes.

The main advantages of the k -NN classifier are its simplicity and performance. The disadvantages are its sensitivity to the participation of irrelevant attributes and the need for efficient indexing techniques. A more detailed, but nevertheless brief, presentation of DTs, backpropagation NNs and k -NNs can be found in Han and Kamber (2000).

3.2. Data

The data used in this study come from the FAME (Financial Analysis Made Easy) database. FAME contains data relating to British and Irish firms. In order to construct the sample, publicly listed companies belonging to the manufacturing, construction, mining and computer sector (UK SIC codes 10–45 and 72) were selected, which had changed auditors during the years 2003–2005. Some companies in this initial sample contained many missing values. These observations were thus considered non-informative and were removed. Companies which mistakenly appeared to have changed auditor due to auditor merges or whose auditor in 2002 was Arthur Andersen, and were thus forced to change auditor in the year 2003, were also removed from the initial sample. Finally, two companies that changed auditors during 2003–2005, namely Royal Dutch Shell and British Petroleum, were considered outliers and were also removed, based on the assumption that they could bias the sample due to their size.

The remaining observations were matched with an equal number of companies that did not switch auditors for at least three consecutive years. The set of samples was stratified according to the auditor change cases. The matching was performed in terms of industry (four-digit SIC code) and fiscal year to eliminate macroeconomic influences. The final sample contained 338 company-year observations. The distribution between Big 4 and Non-Big 4 cases was almost balanced, since 181 companies had a Big 4 auditor (i.e. KPMG, PricewaterhouseCoopers, Ernst & Young and Deloitte & Touche), and 157 companies had a Non-Big 4 auditor.

3.3. Variables

For the purposes of the study, 338 company-year observations were chosen. Auditor category and a number of independent variables were selected for each observation. Both the auditor category and the independent variables refer to the same fiscal year.

In previous studies dealing with the question of auditor nomination, independent variables were found to be relevant. The research literature provides insights regarding factors that affect auditor appointment. For example, Krishnan *et al.* (1996) found that smaller companies in the USA are less likely to be audited by big auditors. The variables Turnover, Total Assets, Fixed Assets and Shareholders' Funds were thus used as proxies for auditee size. Companies with many subsidiaries are more complex and are more likely to be audited by a large auditing firm. Subsequently, the variable Number of Subsidiaries was also included in our sample.

Studies conducted both in the past and also quite recently associate debt with auditor engagement (DeFond, 1992; Knechel *et al.*, 2008; Broye and Weil, 2008). The debt-related variables Long Term Debt, Total Debt and Gearing were tested. Inventory and accounts receivable require audit adjustments (Icerman and Hillison, 1991). A firm with a high proportion of receivables and inventories is more

likely to employ a large audit firm. The variables Accounts Receivable, Stock & WIP and Inventory to Total Assets (INVTA) were included in the test.

A previous study noted that non-financial companies that appoint a big auditor are more profitable (Citron and Manalis, 2001). Several profitability-related accounts and ratios were thus tested, namely Gross Profit, Operating Profit, Retained Profit, Profit Margin, Return on Shareholders' Funds and Return on Total Assets. A significant association was found to exist between auditor size and the ability to issue a qualified opinion (Waren, 1980; Chow and Rice, 1982). Other studies suggest that the receipt of a qualified opinion has a significant effect on a company's propensity to switch auditors (Citron and Taffler, 1992; Krishnan *et al.*, 1996) and that the direction of the switch is from a larger to a smaller auditor (Johnson and Lys, 1990). In order to relate qualifications to auditor nomination, four dummy variables were created, each of which expresses a possible combination between the qualifications of the switch year and the qualifications of the previous year. These variables are Unqualified–Unqualified, Unqualified–Qualified, Qualified–Unqualified and Qualified–Qualified. For each observation, the variable which depicts the real situation obtains the value 1 and the other three dummy variables obtain the value 0.

Firms that are developing rapidly have a relatively greater need for additional external financing. In anticipation of this need, managers of such companies may employ higher quality auditors in order to signal to capital suppliers that their financial statements are of a higher quality (Velury *et al.*, 2003). Such trends were tested by using the variables Total Assets Trend, Current Assets Trend, Current Liabilities Trend and Long Term Liabilities Trend.

Big auditors are assumed to have a high cost structure because the personnel they hire are more expensive and because their training and other costs are higher. These costs are passed on to their clients through their audit fees. Another reason for a positive relation between size and audit fees is that big auditors are more exposed to litigation risk and, therefore, charge higher fees to compensate for this risk (Firth, 2002). The variables Audit Fees and Audit Fees to Total Assets (AFTA) were included for this reason. In the UK, there is no legal restriction concerning the provision of both audit and non-audit services. In many cases auditors provide audit and consultancy services. The variable Non-Audit Fees was used as a possible predictor.

Kane and Velury (2004) use the variable Market Value of Equity to model auditor nomination. In this case, the variable Market Capitalization was used. Finally, some characteristic accounts and financial ratios were tested as possible candidates for inclusion in the final input vector. These are: Current Assets, Current Liabilities, Working Capital, Current Ratio, Liquidity Ratio, Solvency Ratio, Quick Ratio, Sales to Total Assets (SALTA), Working Capital to Total Assets (WCTA), Price to Book Value and, finally, Altman's *Z* score as a proxy for financial distress.

Thirty-five financial ratios and accounts were selected in total. Before the classification was carried out, feature selection was required. Feature selection is a standard data preprocessing step in classification and is defined as the task of selecting the input variables that are useful for classification. This step is required because while some methods, like DTs, are immune to the presence of irrelevant input variables, other methods, like *k*-NNs, are sensitive to the presence of irrelevant attributes. Feature selection may speed up the learning step and increase classification accuracy (Han and Kamber, 2000). The significance of the selected input variables is then further assessed by methods which provide interpretable classification models. In an attempt to reduce dimensionality, a one-way analysis of variance (ANOVA) was used to test whether the differences between the two classes were significant for each variable. Descriptive statistics and the results of the ANOVA are given in Table I. According to Table I, numerous variables present low *p* values. Eighteen variables were selected, which had the lowest *p* values ($p \leq 0.005$). These variables, together with the

Table I. Descriptive statistics, *F* and *p* values by one-way ANOVA

Variable	Big		Non-Big		<i>F</i>	<i>p</i>
	Mean	SD	Mean	SD		
<i>Significant input variables (p ≤ 0.005)</i>						
Audit Fees	566	1362	89	492	17.31	0.000
Non-Audit Fees	506	1237	92	585	12.64	0.000
Current Assets	359,633	1,022,334	46,337	412,695	12.91	0.000
Current Ratio	1.910	1.828	3.898	6.619	15.03	0.000
Liquidity ratio	1.604	1.851	3.550	6.540	14.68	0.000
Profit Margin	3.26	19.41	-6.93	26.50	14.15	0.000
Number of Subsidiaries	33.85	53.39	9.83	34.63	23.26	0.000
Accounts Receivable	273,888	792,423	25,108	291,499	13.84	0.000
Solvency Ratio	42.63	27.51	54.33	35.68	11.44	0.001
Market Capitalization	985	2690	156	1480	11.01	0.001
Turnover	877,229	2,779,232	109,893	894,534	9.60	0.002
Working Capital	101,180	307,032	16,878	150,894	9.41	0.002
Audit Fees to Total Assets (AFTA)	3.314	5.969	5.701	8.187	9.53	0.002
Operating Profit	100,188	390,501	6,241	72,018	8.82	0.003
Total Debt	424,392	1,439,138	67,000	478,444	8.83	0.003
Current Liabilities	303,331	1,082,573	42,439	415,358	8.08	0.005
Total Assets	1,115,192	3,750,319	184,974	1,948,673	7.82	0.005
Gearing	107.5	206.9	51.1	90.6	7.97	0.005
<i>Non-significant input variables</i>						
Sales to Total Assets (SALTA)	1.63	4.61	14.47	62.31	7.64	0.006
Quick Ratio	-7.2	43.9	-43.8	171.6	7.65	0.006
Return on Total Assets	-3.17	48.26	-18.01	53.38	7.18	0.008
Current Liabilities Trend	17.2	54.9	49.2	158.7	6.35	0.012
Retained Profit	28,945	131,084	2,285	37,181	6.07	0.014
Fixed Assets	755,559	2,798,495	140,425	1,546,409	5.94	0.015
Inventory to Total Assets (INVTA)	20.9	86.6	49.3	132.0	5.60	0.019
Shareholders' Funds	373,305	1,314,165	88,565	909,180	5.21	0.023
Gross Profit	327,944	1,479,756	31,310	234,231	4.80	0.029
Long Term Debt	339,191	1,259,099	45,161	387,991	4.72	0.031
Return on Shareholders' Funds	15.9	60.8	-8.5	135.7	4.68	0.031
Stock & WIP	85,745	297,810	21,230	156,637	3.83	0.051
Working Capital to Total Assets (WCTA)	1.74	1.44	6.06	32.34	3.23	0.073
Price/Book Value	3.678	4.810	2.931	3.769	2.19	0.140
Total Assets Trend	352	4394	106	495	0.47	0.494
Current Assets Trend	237	2934	76	311	0.45	0.503
Long Term Liabilities Trend	159.5	1008.0	91.6	638.8	0.40	0.528
Z score	12.74	75.11	17.36	63.93	0.36	0.547

four dummy variables, were included in the final input vector. The variables selected are given in bold in Table I.

Descriptive statistics and the *p* values provide some initial indications concerning the features relevant to auditor nomination. Auditee size is a significant factor, since two out of the three size-related variables (Turnover and Total Assets) belong to the selected group with the lowest *p* values. The mean values of these variables reveal that considerably larger companies tend to appoint big auditors. Fee-related issues are also significant, since all the relevant variables were selected to participate in the final input vector.

Liquidity also seems to be associated with auditor nomination, since the variables Liquidity Ratio, Current Ratio and Working Capital have $p \leq 0.005$, whereas the variable Quick Ratio has a marginally higher *p* value. The mean values of Current Ratio and Liquidity Ratio reveal that companies which appoint a non-big auditor tend to have a better 'Current Assets to Current Liabilities' proportion.

Two out of the three variables associated with debt (Gearing and Total Debt) were selected to participate in the input vector, thus proving that debt structure significantly affects auditor nomination. The mean value of the ratio Gearing reveals that companies with a high debt tend to appoint a big auditor.

The fact that the p value of the Z score is remarkably high provides some indication that financial distress is not associated with auditor engagement. Trends also seem to be irrelevant, since all variables recording trends were rejected due to their p values.

4. EXPERIMENTS AND RESULT ANALYSIS

4.1. The Models' Development and Interpretation

Three alternative classification techniques were employed to model the auditor appointment decision. The techniques employed were C4.5 DT, MLP and k -NNs. These AI techniques were compared with logistic regression. All models were built using Tanagra (Rakotomalala, 2005) data-mining research software.

In the first experiment, the C4.5 model was developed. The tree was built with a 0.25% confidence level and was developed by using the whole sample. The tree produced had 25 nodes and 13 leaves. The tree managed to classify 92.90% of the total cases correctly. In particular, it correctly classified 92.82% of the big auditor cases and 92.99% of the non-big auditor cases.

Table II exhibits the high-level splitters of the C4.5 DT. The tree uses the variable Total Debt as first-level splitter. In order to select the splitting nodes, the C4.5 algorithm performs a univariate analysis of all input variables and selects the input variable that best separates the samples according to their class label (big auditor versus non-big auditor in the present study). In this way, the tree uses the most significant variable, according to the Information Gain measure, as first-level splitter. By defining a cutoff value of £72,865,000, the algorithm identifies the companies which have the highest values for Total Debt. The vast majority of these companies (95 out of 98 observations—96.94%) have appointed a big auditor. Thus, the DT model indicates that Total Debt is the most significant variable and that companies with a high debt pursue audit quality. The tree uses the variables Audit Fees and Accounts Receivable as second- and third-level splitters respectively. Companies with a Total Debt < £72,865,000 and Audit Fees < £53,000 but whose Accounts Receivable \geq £45,464,500 appoint a non-big auditor at a percentage of 94.44% (119 out of 126 observations).

In the second experiment, the MLP model was developed. After testing a number of alternative designs, a topology with one hidden layer containing 11 hidden nodes was selected. The network achieved a general classification accuracy of 83.43% against the training set and managed to correctly classify 79.56% of the big auditor cases and 87.90% of the non-big auditor cases.

Table II. The C4.5 variables

Variable	Splitters' level
Total Debt	1st
Audit Fees	2nd
Account Receivable	3rd

In order to estimate the attributes' contributions to the MLP classifier, Tanagra performed an iterative test by excluding one attribute each time and recalculating the error rate for each repetition. Table III exhibits the most significant input variables together with the differences in the corresponding error rates (DER).

The attribute contribution test recognized the variable Total Debt as the most significant attribute. This variable was also used as first-level splitter in the C4.5 DT model. The variable Audit Fees, which was used as third-level splitter by the C4.5 model, was also found to be significant according to the MLP model. Other significant variables were Non-Audit Fees and Current Ratio.

In the third experiment, the k -NNs method was applied. The number of nearest neighbours was defined as being equal to five. The algorithm utilizes a heterogeneous Euclidean-overlap metric as the distance function (Wilson and Martinez, 1997). The model correctly classified 268 cases (performance 79.29%). More specifically, it classified 130 big auditor cases (71.82%) and 138 non-big auditor cases (87.90%) correctly. Unfortunately, Tanagra does not provide an attribute contribution metric for the k -NN classifier; thus, it was impossible to estimate the input variables' significance for the k -NN model.

The fourth method employed was binary logistic regression. The model successfully classified 127 big auditor cases (70.17%) and 139 non-big auditor cases (88.54%) and achieved an overall performance of 78.70%. The logistic regression method provides the Wald measure in order to estimate the independent variables' contributions. Table IV gives the most significant variables together with the corresponding Wald values at $p < 0.01$. According to these results, logistic regression recognized the variable Total Debt as the most significant variable. Logistic regression also ranked the variable Audit Fees in fifth place in the table. Thus, remarkably enough, three different methods, i.e. DT, NNs and logistic regression, all agree that the most significant input variable is the variable Total Debt, and Audit Fees is also included in the list of significant variables. This conclusion complies with the findings of DeFond (1992); Firth (2002); Broye and Weill (2008) and Knechel *et al.* (2008).

Table III. The MLP variables

Variable	DER ^a
Total Debt	0.1154
Non Audit Fees	0.1036
Audit Fees	0.0858
Current Ratio	0.0414

^a DER: differences in the corresponding errors rates.

Table IV. The logistic regression variables^a

Variable	Wald
Total Debt	60.1388
Total Assets	45.8971
Operating Profit	18.5470
Working Capital	10.4826
Audit Fees	9.3564

^a Statistically significant at 1% level.

4.2. The Models' Validation

Using the training set in order to estimate a model's performance might introduce a bias. In many cases the models, especially those derived from AI, tend to memorize the samples instead of 'learning' (data overfitting). This means that a model embodies decision mechanisms for outliers and exceptions of the training set. These decision mechanisms confuse the model when dealing with new observations. As a result, the model achieves a very high performance against the training set, but a low performance against new, out-of-the-training-set observations. However, the true value of a model is its successful application in the real world, where it will face new, unknown, out-of-the-training-set cases. For this reason the true performances of models must be estimated against previously unseen patterns. In order to cope with this problem, AI introduces the validation of the models. The basic idea of validation is to divide the samples into two subsets. One subset is used to train the model. After the model has been trained, it attempts to guess the class label of the observations of the second subset. The classification results of the model are compared with the actual class labels. The accuracy achieved is the percentage of correct classifications. Validation ensures that the model can correctly classify new observations, rather than simply analyse a specific set of observations.

There are several alternatives to validation, like splitting the samples into a training set and a validation set, x -fold cross-validation, leave-one-out, etc. In this study, two different validation techniques are used. The first technique is the 10-fold cross-validation approach. This validation method is the one proposed in the relevant literature (e.g. Han and Kamber, 2000). In 10-fold cross-validation, the sample is divided into 10 folds. For each fold, the model is trained by using the remaining nine folds and tested by using the hold-out fold. Finally, the average performance is calculated. Table V summarizes the 10-fold cross-validation performances of the three models.

As can be seen in Table V, the C4.5 model outperforms all the other models. The DT model correctly classifies 82.12% of the total cases, 83.43% of the big auditor cases and 80.65% of the non-big auditor cases. The MLP model follows, by correctly classifying 77.27% of the total cases, 70.86% of the big auditor cases and 84.52% of the non-big auditor cases. Logistic regression achieves an overall performance of 76.66% and successfully classifies 68.57% of the big auditor cases and 85.81% of the non-big auditor cases. Finally, the k -NNs model, despite the improved distance function, lags behind in terms of accuracy rate achieved. The k -NNs model correctly classified 69.09% of the total cases, 61.71% of the big auditor cases and 77.42% of the non-big auditor cases. In terms of Type I and Type II error rates, it is observed that the C4.5 model achieves balanced performances. On the contrary, the MLP, the logistic regression and the k -NNs models present a considerably higher Type I error rate. Moreover, it is noted that the differences in the overall performance of the three models are mainly attributed to the differences in the corresponding Type I error rates.

The second validation method is the splitting of the samples into the training set and the validation set. Additional to this technique, a time lag notion is also introduced. In this validation procedure we test the 'out of time' performance of the models. For this purpose, the samples are divided into two subsets. The first subset contains the observations of 2003 and 2004 (220 observations), while the second subset contains the observations of 2005 (118 observations). The models are trained by using the first subset. Afterwards, the models attempt to find the auditor's category as regards the cases belonging to the second subset. The corresponding performances are summarized in Table V. As can be seen, the C4.5 model outperforms all the other models. It correctly classifies 82.09% of the big auditor cases, 82.35% of the non-big auditor cases and 82.20% of the total cases. The MLP, the k -NNs and the logistic regression models follow, by achieving overall performances of 72.88%, 71.19% and 63.56% respectively. According to this test, all the AI methods outperform logistic regression.

Table V. Models' performances

Model	Big auditor (%)	Non-Big auditor (%)	Total (%)
<i>Training set</i>			
C4.5	92.82	92.99	92.90
MLP	79.56	87.90	83.43
<i>k</i> -NNs	71.82	87.90	79.29
Logistic regression	70.17	88.54	78.70
<i>10-fold cross validation</i>			
C4.5	83.43	80.65	82.12
MLP	70.86	84.52	77.27
<i>k</i> -NNs	61.71	77.42	69.09
Logistic regression	68.57	85.81	76.66
<i>Validation on year 2005</i>			
C4.5	82.09	82.35	82.20
MLP	74.63	70.59	72.88
<i>k</i> -NNs	59.70	86.27	71.19
Logistic regression	52.24	78.43	63.56
<i>Bagging on training set</i>			
C4.5	97.24	96.18	96.75
MLP	77.90	89.17	83.14
<i>k</i> -NNs	70.17	92.36	80.47
Logistic regression	74.59	91.08	82.25
<i>Bagging on 10-fold cross validation</i>			
C4.5	88.00	82.58	85.45
MLP	75.43	83.23	79.09
<i>k</i> -NNs	60.57	78.71	69.09
Logistic regression	71.43	85.16	77.88

Moreover, the C4.5 and the MLP models achieve a balanced performance in terms of Type I and Type II errors, whereas the *k*-NNs and logistic regression models present a much higher rate of Type I errors.

4.3. Bagging the Classifiers

Although the performances of the models (especially the DT model) against the training set are high, what is important from the view of AI is the ability of the model to cope with new, real-world problems. For that reason the true performance of the model must be estimated against unknown, out-of-the-training-set observations by employing a validation method. The 10-fold cross-validation performances of our models are satisfactory but can be improved further. Increasing the classifiers' accuracy rates is a permanent concern in classification problems. AI proposes techniques for improving a model's performance. Such techniques have not yet been applied and evaluated in relation to the issue of auditor nomination. However, the successful application of these techniques in specific problematic areas may lead to improved models. These models may achieve higher accuracy rates when they are applied in the real world and address the question of auditor appointment in a real business environment. Any improvement to the models' performances in the real world can be experimentally evaluated by employing a validation technique.

In this study, an attempt is made to improve performances through bagging. Bagging is a general technique for improving a classifier's accuracy (Breiman, 1996). The bagging algorithm creates multiple training sets. The training sets are created from the original training set by random sampling

with replacement. Since replacement is used, some observations may participate repeatedly in a training set, whereas some others may not participate at all. For each training set, one classifier is learned. In order to classify an unknown sample, the algorithm utilizes all the learned classifiers. Each classifier's decision is regarded as one vote. The aggregated classifier assigns the sample to the class with the majority of votes. Since multiple versions of the classifier are produced, the interpretable structure is lost. What one gains is increased accuracy. Breiman (1996) showed that bagging is effective on 'unstable' learning algorithms, where small changes in the training set result in large classification changes. He also claimed that NNs and DTs are examples of unstable learning algorithms.

The bagging technique was applied in relation to the four classification methods used. Then the accuracy rates were tested against the training set, and the bagged classifiers were also tested by using 10-fold cross-validation. Table V gives the accuracy rates against the training set and the results of the 10-fold cross-validation.

By comparing the accuracy rates against the training set, it is observed that the bagged C4.5 classifier increases its total accuracy rate by almost 4%. The improvement in the performance of the logistic regression method is comparable. However, a comparison of the 10-fold cross-validation results reveals that only the C4.5 model considerably increases its performance and achieves a total classification accuracy of 85.45%. The results of the 10-fold cross-validation for the bagged classifiers also reveal that the differences in the models' performances mainly arise from the large differences in the corresponding Type I error rates.

5. CONCLUSIONS

Audit quality is an open question due to the contradictory nature of the relation between auditor and auditee. Big auditing firms are identified in the relevant literature as higher quality auditors. In considering auditor size as a proxy for audit quality, researchers have developed models to classify the appointment of auditors. However, all these studies employ some version of regression analysis. Several research studies have employed AI methods to cope with financial classification problems, such as bankruptcy prediction or fraud detection. These studies report that the results obtained by using AI methods are similar to or better than those provided by traditional statistical methods. AI methods have not been used to date in order to address the question of auditor nomination.

This study is the first to use three AI techniques to develop models capable of classifying auditor nomination. The methods used are C4.5 DT, MLP NN and k -NNs, with improved distance function. These methods are then compared with logistic regression. The sample contains 338 company-year observations related to British and Irish firms. The input vector is composed of financial accounts and ratios, as well as variables associated with qualifications and auditor switching. A preliminary feature selection is performed by running a one-way ANOVA.

The four models manage to classify the training set and achieve satisfactory accuracy rates. In terms of the models' interpretations, the C4.5 model, the MLP model and the logistic regression model agree that there are dependencies between the auditor nomination and debt level. The three models also agree that the former is also related to audit fees.

In assessing the models' performances, it is observed that the C4.5 model outperforms the other three models. According to the 10-fold classification results, the total accuracy rates for the C4.5, MLP, logistic regression and k -NNs models are 82.12%, 77.27%, 76.66% and 69.09% respectively. Two AI methods outperform logistic regression. Only the DT presents a balanced

performance in terms of Type I and Type II error rates. In classifying auditor category for future cases, all the AI classifiers outperform logistic regression. By bagging the classifiers, the C4.5 increases its performance significantly and reaches an accuracy rate of 85.45%, according to a 10-fold cross-validation evaluation.

The input variables used in this study are publicly available financial ratios and account values. These ratios and values depict common aspects of a firm's financial status. Moreover, these ratios and values can be found in the financial statements issued by companies in all developed countries. This goes to prove that the AI methodologies employed in this study can also be applied to develop models capable of classifying auditor nomination for non-UK firms. The study's results concerning the significance of the debt level and audit fees comply with the results of other studies (DeFond, 1992; Firth, 2002; Broye and Weill, 2008; Knechel *et al.*, 2008), despite the fact that they used different samples and methods.

One limitation of this study involves the use of the *k*-NNs method, which does not reveal which variables contribute to the decision that affects auditor nomination. The main purpose of this paper is not to provide evidence of the association between the appointment of an audit-firm group and a company's financial or non-financial characteristics. Instead, the paper focuses on the development of AI models capable of classifying the category of the nominated auditor.

As is usually the case, this study can also be used as a stepping stone for further research. Numerous studies associate auditor nomination with managerial issues, such as the percentage of external board members or the level of institutional ownership. Owing to the unavailability of any relevant data, such managerial characteristics were not addressed. If the input vector were to be enriched with qualitative, management-related variables, then this could potentially improve the models, both in terms of their explanatory power and their classification accuracy.

Another consideration is the combination of different classifiers. Bagging is a relatively simple technique, where alternative instances of the same classifier are produced. Aggregating different classifiers by involving them in a voting scheme and, thus, constructing a new assembled classifier could further improve classification accuracy. We hope that the research presented in this paper will, therefore, stimulate additional work regarding these important topics.

ACKNOWLEDGEMENTS

We would like to thank Daniel O'Leary (editor) and Bob Berry (former editor) for the encouragement and three anonymous referees for their helpful comments.

REFERENCES

- Bartov E, Gul FA, Tsui JSL. 2001. Discretionary-accruals models and audit qualifications. *Journal of Accounting and Economics* **30**: 421–452.
- Beasley M, Petroni K. 2001. Board independence and audit firm type. *Auditing: A Journal of Practice and Theory* **20**: 97–114.
- Becker CL, DeFond ML, Jiambalvo J, Subramanyam KR. 1998. The effect of audit quality on earnings management. *Contemporary Accounting Research* **15**: 1–24.
- Breiman L. 1996. Bagging predictors. *Machine Learning* **24**(2): 123–140.
- Broye G, Weill L. 2008. Does leverage influences auditor choice? A cross-country analysis. *Applied Financial Economics* **18**(9): 715–731.

- Chaney P, Jeter D, Shivakumar L. 1996. Self selection of auditors and audit pricing in private firms. *The Accounting Review* **79**: 51–72.
- Chow C, Rice S. 1982. Note: qualified audit opinions and auditor switching. *The Accounting Review* **37**: 326–335.
- Citron D, Manalis G. 2001. The international firms as new entrants to the statutory audit market: an empirical analysis of auditor selection in Greece, 1993 to 1997. *The European Accounting Review* **10**: 439–459.
- Citron D, Taffler R. 1992. The audit report under going concern uncertainties: an empirical analysis. *Accounting and Business Research* **22**: 337–345.
- Craswell A, Stokes DJ, Laughton J. 2002. Auditor independence and fee dependence. *Journal of Accounting and Economics* **33**: 253–275.
- DeAngelo L. 1981. Auditor size and auditor quality. *Journal of Accounting and Economics* **1**: 113–127.
- DeFond M. 1992. The association between changes in client firm agency costs and auditor switching. *Auditing: A Journal of Practice and Theory* **11**(1): 16–31.
- Doumpos M, Gaganis C, Pasiouras F. 2005. Explain qualifications in audit reports using a support vector machine methodology. *Intelligent Systems in Accounting, Finance and Management* **13**: 197–215.
- 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.
- Firth M. 2002. Auditor-provided consultancy services and their associations with audit fees and audit opinions. *Journal of Business Finance and Accounting* **29**: 661–693.
- Francis J, Maydew E, Sparks HC. 1999. The role of big 6 auditors in the credible reporting of accruals. *Auditing: A Journal of Practice and Theory* **18**(2): 17–34.
- Gaganis C, Pasiouras F, Doumpos M. 2007a. Probabilistic neural networks for the identification of qualified audit opinions. *Expert Systems with Applications* **32**(1): 114–124.
- Gaganis C, Pasiouras F, Spathis C, Zopounidis C. 2007b. A comparison of nearest neighbours, discriminant and logit models for auditing decisions. *Intelligent Systems in Accounting, Finance and Management* **15**: 23–40.
- Gul F, Lynn S, Tsui J. 2002. Audit quality, management ownership, and the informativeness of accounting earnings. *Journal of Accounting, Auditing and Finance* **17**: 25–49.
- Han J, Kamber M. 2000. *Data Mining Concepts and Techniques*. Morgan Kaufman: San Diego, CA.
- Houghton K, Jubb C. 2003. Auditor selection: what influences decisions by listed companies? *Australian Accounting Review* **13**(3): 67–72.
- Icerman R, Hillison W. 1991. Disposition of auditor-detected errors: some evidence on evaluative materiality. *Auditing: A Journal of Practice and Theory* **10**: 22–34.
- Johnson B, Lys T. 1990. The market of auditor services—evidence from voluntary auditor changes. *Journal of Accounting and Economics* **12**: 281–308.
- Kane G, Velury U. 2004. The role of institutional ownership in the market for auditing services: an empirical investigation. *Journal of Business Research* **57**: 976–983.
- Kim JB, Chung R, Firth M. 2003. Auditor conservatism, asymmetric monitoring, and earnings management. *Contemporary Accounting Research* **20**: 323–359.
- Kirkos E, Spathis C, Manolopoulos Y. 2007a. Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications* **32**(4): 995–1003.
- Kirkos E, Spathis C, Nanopoulos A, Manolopoulos Y. 2007b. Identifying qualified auditors' opinions: a data mining approach. *Journal of Emerging Technologies in Accounting* **4**: 183–197.
- Knechel R, Niemi L, Sundgren S. 2008. Determinants of auditor choice: evidence from a small client market. *International Journal of Auditing* **12**: 65–88.
- Krishnan J, Krishnan J, Stephens R. 1996. The simultaneous relation between auditor switching and audit opinion: an empirical analysis. *Accounting and Business Research* **26**: 224–236.
- Lenard M, Alam P, Madey G. 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.
- Lin FY, McLean S. 2001. A data mining approach to the prediction of corporate failure. *Knowledge-Based Systems* **14**(3–4): 189–195.
- Mutchler J. 1986. Empirical evidence regarding the auditor's going-concern opinion decision. *Auditing: A Journal of Practice and Theory* **6**: 148–163.
- O'Leary DE. 1998. Using neural networks to predict corporate failure. *Intelligent Systems in Accounting, Finance and Management* **7**: 187–197.

- Palmrose Z. 1988. An analysis of auditor litigation and audit service quality. *The Accounting Review* **63**: 55–73.
- Rakotomalala R. 2005. TANAGRA: un logiciel gratuit pour l'enseignement et la recherche. In *Actes de EGC'2005*, RNTI-E-3, vol. 2, pp. 697–702.
- Teoh SH, Wong TJ. 1993. Perceived auditor quality and the earnings response coefficient. *The Accounting Review* **68**: 346–366.
- Velury U, Reish J, O'Reilly D. 2003. Institutional ownership and the selection of industry specialist auditors. *Review of Quantitative Finance and Accounting* **21**: 35–48.
- Waren C. 1980. Uniformity of auditing standards: a replication. *Journal of Accounting Research* **18**: 312–324.
- Wilson R, Martinez T. 1997. Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research* **6**: 1–34.