Applying Data Mining Methodologies for Auditor Selection

Efstathios Kirkos[1], Charalambos Spathis[2], Yannis Manolopoulos[3]

[1]Department of Accounting
Technological Educational Institution of Thessaloniki
57400, Thessaloniki, Greece
Email: stkirk@acc.teithe.gr

[2]Department of Economics
Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
Email: hspathis@econ.auth.gr

[3]Department of Informatics
Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
Email: manolopo@csd.auth.gr

Abstract

Auditor selection can be regarded as a matter of audit quality. Research studies aiming to model the auditor choice employ statistical techniques. Here we employ three techniques derived from the Data Mining domain to build models capable of discriminating cases where companies choose a Big 4 or a Non-Big 4 auditor. Significant factors associated with the auditor choice are revealed. The three models are compared in terms of their performances. According to 10-fold cross validation bagging increases significantly the performance of one classifier.

Keywords: Auditor choice, audit quality, Big 4 auditors, data mining, bagging

1. Introduction

External auditors are said to add value to financial reports by assuring the proper disclosure of a company's financial status. However, the relationship between auditor and auditee is complex and contradictory. The contradiction arises from the fact that the auditor must remain independent when the auditee determines the appointment, retention and audit fees. This contradiction may jeopardize the auditor's objectivity. Thus, the question for quality auditing remains open.

It is generally acknowledged that the audit markets are segmented into at least two categories, the Big 4 auditors and the Non-Big 4 auditors. Big auditors are big international auditing firms. Some years ago there were 8 big auditors. After auditing firms' merges and the collapse of Arthur Andersen remain 4 Big Auditing firms i.e. KPMG, PricewaterhouseCoopers, Ernst & Young and Deloitte & Touche. All other auditors which have a national or local reputation are the Non-Big Auditors. Big auditors are identified in the literature as higher quality auditors [DeAngelo (1981)], [Mutchler (1986)], [Palmrose (1988)], [Bartov et al. (2001)], [Craswell et al. (2002)]. The big auditing firms due to their size are in better position to withstand client pressure, they invest more in technology, training and facilities and have increased incentive to maintain their reputation and professional standards. For these reasons a number of prior research studies have used audit firm size as a proxy for audit quality [Teoh & Wong (1993)], [Becker et al. (1998)].

From the research literature it is revealed that researchers reject the null hypotheses that clients are randomly allocated across big and non-big auditors. Research effort has been directed towards the auditor selection problem. However, these studies use solely some version of Logistic Regression. In most cases the aim is to prove an association between specific managerial characteristics and the choice of auditor. The auditor selection problem can be regarded as a typical classification problem. The methodological framework of Data Mining (DM) provides methods, techniques and concepts suitable for classification purposes [Han & Camber (2006)]. AI methods have the theoretical advantage that they do not impose arbitrary assumptions about attributes' independence. As opposed to other well-examined fields like bankruptcy prediction, credit risk estimation or fraud detection [Kirkos et al. (2007)], DM classification methodologies remain to be tested against the auditor selection problem.

The aim of this study is to employ three DM methods to develop models capable of predicting the auditor choice. The methods used are *K*-Nearest Neighbor (*k*-NN), Decision Trees and Neural Networks. Significant factors associated with the selection of auditor are revealed. The developed models are compared in terms of their predictive accuracy. Techniques for improving the models' performance are applied and evaluated. The sample is composed of 338 UK and Irish company-year observations. The input vector contains both quantitative financial and qualitative variables. The data used are publicly available and easily found in a typical financial data base. This study has implications for internal and external auditors, company decision makers, investors and researchers. It can also be used to predict the most probable outcome for the selection of auditor.

The paper proceeds as follows: Section 2 reviews prior research related to auditor selection. Section 3 describes the construction of our sample, the initial input variables' selection and the feature selection. Section 4 refers to models' performances, interpretation, validation and classifiers' bagging. Finally, Section 5 presents the concluding remarks.

2. Prior Research

Recent research studies attempted to model the auditor selection and to reveal significant factors. [Citron and Manalis (2001)] investigated the choice of auditor in Greece after the liberalization of the audit market. They developed a model to predict the selection of a big auditor by using Binomial Logistic Regression. According to their results the level of shareholdings by foreign shareholders is positively associated with the choice of a big auditor.

[Velury et al. (2003)] provides evidence linking corporate governance mechanisms to the choice of auditors. They employ a two-stage least square regression method to test the relationship between audit quality and the level of institutional investment. Their findings indicate that firms having relatively greater levels of institutional ownership tend to employ industry specialist auditors.

[Kane and Velury (2004)] investigated the relationship between the selection of a big auditing firm and the level of institutional ownership. They used Logistic Regression to develop a model to predict the selection of a Big Auditor. The input vector contained financial ratios and a variable defining the proportion of shares held by institutions. They found that firms with high level of institutional ownership are more likely to select a Big auditor. They also found a positive association between Size and Debt and the selection of a Big auditor.

[Chaney et al. (1996)] conducted a survey in privately held UK firms to investigate the pricing in audit fees and to predict the auditor choice. The method applied was the OLS Regression. To model the auditor selection they used financial ratios and two qualitative variables. Their findings suggest that when not compelled by market pressures to choose a Big auditor, clients choose the lower cost auditor available. They predicted the auditor choice and then they used this information in the fee analysis. Fee analysis suggests that auditors structure their business in a manner appropriate for specific client segments. The model had 68.5% accuracy rate against the training sample.

A critical observation of the collected literature indicates that all studies employ some version of Logistic Regression. The novelty of this study is that it applies, interprets and evaluates methods derived from DM. These methodologies have not yet been applied for the purpose of predicting the auditor choice.

3. Research Methodology

3.1 Data

The data used in this study come from the FAME (Financial Analysis Made Easy) database. FAME contains data about 3,000,000 UK and Irish firms. Our sample comes from a previous research study concerning auditor switching. However the sample contains fairly equal observations of Big 4 and non Big 4 cases.

To construct the sample we selected the publicly listed companies activated in the sectors of Manufacturing, Construction, Mining and Computers (UK SIC Codes 10-45 & 72), which changed their auditor during the years 2003-2005. Some companies in this initial sample contained many missing values. This observations were considered non informative and were removed. From the initial sample were also removed the companies which, due to auditor merges, appeared mistakenly to change auditor. Finally, the companies Royal Dutch Shell and British Petroleum changed their auditor during 2003-2005. However, these companies could bias the sample due to their size. These two companies were considered outliers and were also removed.

The remaining observations were matched with equal number of companies which did not change auditor for at least three successive years. The matched has been performed in terms of industry (four digits 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 is almost balanced since 181 companies had a Big 4 auditor where 157 companies had a non Big 4 auditor.

3.2 Variables

The research literature provides insights relevant to factors associated with the choice of auditor. [Krishnan et al. (1996)] found that smaller companies in the US are less likely to be audited by Big auditors. We use the variables Turnover, Total Assets, Fixed Assets and Shareholders' Funds as proxies for auditee's size. Companies with many subsidiaries are more complex and are more likely to be audited by a big auditing firm. We include the variable Number of Subsidiaries in our sample.

[DeFond (1992)] associates debt with the selection of auditor. We test the debt related variables Long Term Debt, Total Debt and Gearing. Inventory and accounts receivable need audit adjustments [Icerman & Hillison (1991)]. We check the variables Accounts Receivable, Stock & WIP and Inventory to Total Assets (INVTA).

[Citron & Manalis (2001)] found that that the non-financial companies which choose a Big auditor are more profitable. We test numerous profitability related accounts and ratios. These are Gross Profit, Operating Profit, Retained Profit, Profit Margin, Return on Shareholders' Funds and Return on Total Assets.

[Chow and Rice (1982)] found a significant association between the size of the auditor and the possibility to issue a qualified opinion. Other studies suggest that the receipt of a qualified opinion has a significant effect on a company's propensity to switch auditors [Citron & Taffler (1992)], [Krishnan et al. 1996] and that the direction of the switch is from a larger to a smaller auditor [Johnson & Lys (1990)]. To associate qualifications with the auditor choice we created four dummy variables each of which expresses a possible combination between the qualification of the year of change and the qualification of the previous year. These variables are Unqualified-Unqualified, Unqualified-Qualified, Qualified-Unqualified and Qualified-Qualified. For each observation the variable which depicts the qualifications cases obtains the value 1 where the other three dummy variables obtain the value 0.

Firms that are growing quickly have a relative greater need for additional external financing. Anticipating the need for financing managers of such companies may employ higher quality auditors to signal to capital suppliers that their financial statements are of higher quality [Velury et al. (2003)]. We test trends 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 as the personnel they hire are more expensive and training and other costs are higher. Another reason for a positive relationship between big and audit fees is that the big auditors are more exposed to litigation risk and, thus, they charge higher fees to compensate for the risk [Firth (2002)]. We include the variables Audit Fees and Audit Fees to Total Assets (AFTA). In 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. We check the variable Non-Audit Fees as a possible predictor.

[Kane & Velury (2004)] use the variable Market Value of Equity to model the selection of auditor. We also use the variable Market Capitalization. The binary variable Change Auditor indicates auditor switching. Finally we tested some typical accounts and financial ratios. 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 the Altman's ZScore as a proxy for financial distress.

In total, we selected 35 financial ratios and accounts. In an attempt to reduce dimensionality, we ran one way ANOVA to test whether the differences between the two classes were significant for each variable. Descriptive statistic and the results of ANOVA are depicted in Table 1. As can be seen in Table 1 numerous variables present low p values. We selected 18 variables which had the lowest p values ($p \leq 0.005$). These variables together with the four dummy variables and the binary variable Auditor Change participate in the final input vector.

Descriptive statistics and the p values provide initial indications concerning features relevant to auditor selection. Auditee's size is a significant factor since two out of three size related variables (the variables Turnover and Total Assets) belong to the selected group with the lowest p values. The mean values of these variables reveal that companies which select a Big Auditor are considerably bigger. Fee matters are also significant since all fee related variables were selected to participate in the final input vector.

Table 1. Descriptive statistics, *F* and *P* values by one way ANOVA. Asterisks indicate the significant input variables

	BIG		NON-BIG			
Variable	Mean	StDev	Mean	StDev	*F*	*P*
*Turnover	877229	2779232	109893	894534	9.60	0,002
Gross Profit	327944	1479756	31310	234231	4.80	0,029
*Operating Profit	100188	390501	6241	72018	8.82	0,003
Retained Profit	28945	131084	2285	37181	6.07	0,014
*Audit Fees	566	1362	89	492	17.31	0,000
*Non-Audit Fees	506	1237	92	585	12.64	0,000
Fixed Assets	755559	2798495	140425	1546409	5.94	0,015
*Current Assets	359633	1022334	46337	412695	12.91	0,000
*Current Liabilities	303331	1082573	42439	415358	8.08	0,005
*Working Capital	101180	307032	16878	150894	9.41	0,002
*Total Assets	1115192	3750319	184974	1948673	7.82	0,005
Long Term Debt	339191	1259099	45161	387991	4.72	0,031
Shareholders Funds	373305	1314165	88565	909180	5.21	0,023
*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
*Solvency Ratio	42,63	27,51	54,33	35,68	11.44	0,001
*Gearing	107,5	206,9	51,1	90,6	7.97	0,005
Current Assets Trend	237	2934	76	311	0.45	0,503
Total Assets Trend	352	4394	106	495	0.47	0,494
Current Liabilities Trend	17,2	54,9	49,2	158,7	6.35	0,012
Long Term Liabilities Trend	159,5	1008,0	91,6	638,8	0.40	0,528
*Profit Margin	3,26	19,41	-6,93	26,50	14.15	0,000
Return on shareholders Funds	15,9	60,8	-8,5	135,7	4.68	0,031
Return on Total Assets	-3,17	48,26	-18,01	53,38	7.18	0,008
Stock & WIP	85745	297810	21230	156637	3.83	0,051
*Market Capitalization	985	2690	156	1480	11.01	0,001
Price / Book value	3,678	4,810	2,931	3,769	2.19	0,140
*Number of Subsidiaries	33,85	53,39	9,83	34,63	23.26	0,000
*Total Debt	424392	1439138	67000	478444	8.83	0,003
Z Score	12,74	75,11	17,36	63,93	0.36	0,547
*AFTA	3,314	5,969	5,701	8,187	9.53	0,002
*Accounts Receivable	273888	792423	25108	291499	13.84	0,000
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
INVTA	20,9	86,6	49,3	132,0	5.60	0,019
WCTA	1,74	1,44	6,06	32,34	3.23	0,073

Two out of three variables associated with debt (Gearing and Total Debt) were selected to participate in the input vector providing evidence that a high debt structure is significant in the auditor selection decision. The mean value of the ratio Gearing reveals that companies with high debt tend to choose a Big Auditor.

Liquidity seems also to be associated with auditor choice since the variables Liquidity Ratio, Current Ratio and Working Capital have *p* value ≤ 0.005, where the variable Quick Ratio has marginally higher *p* value. Mean values of Current Ratio and Liquidity Ratio reveal that companies which choose a Non-Big Auditor tend to have a better Current Assets to Current Liabilities proportion.

The *p* value of Z Score which is remarkably high provides indication that financial distress is not associated with the selection of auditor. Trends seem also to be irrelevant since all trends recording variables were rejected according to their p values.

4. Experiments and Results Analysis

Three alternative classification techniques were employed to model the auditor choice decision. The employed techniques are *k*-NN, C4.5 Decision Tree and Multilayer Perceptron (MLP). All models were built using the Tanagra [Rakotomalala (2005)] DM research software.

In the first experiment we developed the C4.5 model. The tree was built with confidence level 0.25%. The whole sample was used as training set. The produced tree had 25 nodes and 13 leaves. The tree managed to classify correctly 93.49% of the total cases. In particular, it classified correctly 95.58% of the Big Auditor cases and 91.08% of the Non-Big Auditor cases.

Table 2. C4.5 significant input variables

VARIABLE
Total Debt
Audit Fees
Accounts Receivable

Table 2 exhibits the high level splitters of the C4.5 Decision Tree. The tree uses as first level splitter the variable Total Debt. By defining a cut-off value of 72,865,000£ the algorithm separates the companies which have the highest values for Total Debt. The vast majority of these companies (95 out of 98 observations – 96.94%) choose a Big Auditor. Thus the decision tree model provides evidence that companies with high debt seek after audit quality.

The tree uses as second and third level splitters the variables Audit Fees and Accounts Receivable respectively. Companies with Total Debt < 72,865,000£ and Audit Fees < 53,000£ but having Accounts Receivable ≥ 45,464,500£ choose a Non-Big Auditor in a percentage of 94.44% (119 out of 126 observations).

In the second experiment we developed the Multilayer Perceptron model. After testing a number of alternative designs we choose a topology with one hidden layer containing 11 hidden nodes. 80% of the sample was used for training and the rest 20% for testing. The network achieved a general classification accuracy of 79.88% against the training set managing to classify correctly 77.35% of the Big Auditor cases and 82.80% of the Non-Big Auditor cases.

To estimate the attributes' contribution for the multilayer perceptron classifier, Tanagra performs an iterative test by excluding each time an attribute and recalculating the error rate for each repetition. Table 3 exhibits the most significant input variables together with the differences in the corresponding error rates (D.E.R.).

Table 3. M.L.P. significant input variables

VARIABLE	D.E.R
Total Debt	0.0592
Non Audit Fees	0.0414
Auditor Change	0.0207
Audit Fees	0.0148

The attribute contribution test recognizes as the most significant attribute the variable Total Debt. This variable was also used as first level splitter in the C4.5 Decision Tree model. The variable Audit Fees which was used as third level splitter by the C4.5 model has also been found significant by the Multilayer Perceptron model.

In the third experiment we developed the *k*-NN model. The algorithm utilizes a Heterogeneous Euclidean-Overlap Metric as the distance function [Wilson & Martinez (1997)]. The whole sample was used for training. After the training the model was tested against the training set.

The model correctly classified 280 cases (performance 82.84%). In particular, it correctly classified 142 Big Auditor cases (78.45%) and 138 Non-Big Auditor cases (87.90%). Unfortunately, Tanagra does not provide an attributes' contribution metric for the *k*-NN classifier and, thus, was impossible to estimate the input variables' significance for the *k*-NN model. Table 4 summarizes the models' accuracy rates against the training set.

Table 4. Models' performance against the training set

Model	Big Auditor%	Non-Big Auditor%	Total%
C4.5	95.58	91.08	93.49
MLP	77.35	82.80	79.88
k-NN	78.45	87.90	82.84

4.1 The Models' Validation

Using the training set to estimate a model's performance might introduce a bias. In many cases the models, especially those derived from artificial intelligence, tend to memorize the sample instead of "learning" (data overfitting). To eliminate such a bias the performance of the models must be estimated against previously unseen patterns. TANAGRA embodies several learning assessment methods. We choose a 10- fold cross validation approach. In 10-fold cross validation, the sample is divided in ten 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 5 summarizes the 10-fold cross validation performances of the three models

Table 5. 10-fold cross validation performance

Model	Big Auditor %	Non-Big Auditor %	Total %
C4.5	82.93	80.87	81.97
MLP	77.91	79.08	78.45
k-NN	68.02	79.22	73.21

As can be seen in Table 5 the C4.5 model outperforms the other two models. The decision tree model classifies correctly 81.97% of the total cases, 82.93% of the Big Auditor cases and 80.87% of the Non-Big Auditor cases. The MLP model follows by classifying correctly 78.45% of the total cases, 77.91% of the Big Auditor cases and 79.08% of the Non-Big Auditor cases. Finally, the *k*-NN model, although the improved distance function comes behind in terms of the achieved accuracy rate. The *k*-NN model classified correctly 73.21% of the total cases, 68.02% of the Big Auditor cases and 79.22% of the Non-Big Auditor cases. In terms of Type I and Type II error rates we observe that the C4.5 model and the MLP model achieve balanced performances. On the contrary, the *k*-NN model presents a considerably bigger Type I error rate. Moreover, we observe that the differences in the total performance of the three models are attributed mainly to differences in the corresponding Type I error rates.

To ensure that our results give a true and fair picture of the classifiers' performances we repeated the 10-fold validation procedure by using subsets of the sample. We built four samples by using 20%, 40%, 60% and 80% of the total sample. The observations participating in these samples were randomly selected. In all cases the C4.5 classifier achieved marginally lower performance and outperformed the other two classifiers followed by the MLP classifier.

4.2 Bagging the Classifiers

Bagging [Breiman (1996)] is a general technique for improving a classifier's accuracy. 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 multiple times in one training set where some others may not participate at all. For each training set one classifier is learned. 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 showed that bagging is effective on "unstable" learning algorithms where small changes in the training set result in large changes in predictions. He also claimed that neural networks and decision trees are examples of unstable learning algorithms

We applied the bagging technique to our three employed classification methods. We tested the accuracy rates against the training set. We also tested the bagged classifiers by using 10-fold cross validation. Table 6 depicts the accuracy rates against the training set. Table 7 depicts the results of the 10-fold cross validation.

Table 6. Bagged classifiers against the training set

Model	Big Auditor%	Non-Big Auditor%	Total%
C4.5	97.23	96.82	97.04
MLP	83.98	85.99	84.91
k-NN	80.11	86.62	83.14

By comparing the accuracy rates against the training sample we observe that the bagged MLP classifier increases its total accuracy rate by almost 5%. The improvement of the performance for C4.5 is lower where bagging seems not to have a considerable effect on the *k*-NN classifier. Thus bagging seems to be beneficial mainly for the MLP model.

Table 7. 10-fold cross validation for bagged classifiers

Model	Big Auditor%	Non-Big Auditor%	Total%
C4.5	88.78	83.48	86.33
MLP	78.26	80.43	79.27
k-NN	67.72	80.27	73.55

However, a comparison of the 10-fold cross validation results reveals that only the C4.5 model increases considerably its performance and achieves a total classification accuracy of 86.33%. The results of the 10-fold cross validation for the bagged classifiers also reveal that the differences in models' performances arise mainly from big differences in the corresponding Type I error rates.

5. Conclusions

Audit quality is an open question due to the contradictory nature of the relationship between auditor and auditee. Big auditing firms are identified in the literature as higher quality auditors. Regarding the auditor's size as a proxy for audit quality researchers developed models to classify the auditor choice. However, all these studies employ some version of the Logistic Regression method. As apposed to other finance related topics like bankruptcy prediction or fraud detection, DM classification concepts and techniques have not been applied for the purpose of predicting the choice of auditor.

The contribution of this study is that it applies, evaluates and interprets Data Mining methodologies to address the question of auditor selection. We employ three alternative classifiers to develop models. The methods used are C4.5 Decision Tree, Multilayer Perceptron Neural Network and k-NN with improved distance function. Our sample contains 338 company-year observations of UK and Irish firms. The input vector is composed of financial accounts and ratios and variables associated with qualifications and auditor changes. Preliminary feature selection has been performed by running one way ANOVA.

The three models manage to classify the training set and achieve satisfactory accuracy rates. In terms of models' interpretation the C4.5 model and the MLP model reveal dependencies between the choice of auditor and the debt level. The two models also agree that the choice of auditor is related to audit fees.

In assessing the models' performance we observe that the C4.5 model outperforms the other two models. According to the 10-fold classification results the total accuracy rates for the C4.5, the MLP and the *k*-NN models are 81.97%, 78.45% and 73.21% respectively. Only the Decision Tree and the MLP present balanced performance in terms of Type I and Type II error rates. By bagging the classifiers both the C4.5 and the MLP models increase their classification accuracy against the training set. However, according to a 10 fold cross validation evaluation only the bagged C4.5 classifier increases significantly its performance and succeeds an accuracy rate of 86.33%.

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 European countries. This provides evidence that the DM methodologies employed in this study can also be applied to develop models capable of predicting the choice of auditor for non-UK firms. Our results concerning the significance of the debt level and the audit fees comply with the results of other studies [DeFond (1992)], [Firth (2002)] although these studies used different samples and different methods.

As usually happens, this study can be used as a stepping stone for further research. Numerous studies associate the auditor choice with managerial issues like the percentage of external board members or the level of institutional ownership. Due to data unavailability we did not address management characteristics. Enriching the input vector with qualitative management related variables could improve models both in terms of explanatory power and in terms of classification accuracy. Another consideration is the combination of different classifiers. Bagging is a relatively simply 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.

References

Bartov E., Gul F.A., Tsui J.S.L. (2001), *Discretionary-accruals models and audit qualifications,* Journal of Accounting and Economics, vol. 30, pp. 421-52.

Becker C.L., DeFond M.L., Jiambalvo J., Subramanyam K.R. (1998), *The effect of audit quality on earnings management,* Contemporary Accounting Research, vol. 15, pp. 1-24.

Breiman L. (1996), *Bagging Predictors*, Machine Learning, vol. 24, pp. 123-140.

Chaney P., Jeter D., Shivakumar L. (1996), *Self selection of auditors and audit pricing in private firms,* The Accounting Review, vol. 79, pp. 51-72.

Chow C., Rice S. (1982*), Notes: Qualified audit opinions and auditor switching*, The Accounting Review, vol. 27, pp. 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, vol. 10 pp. 439-459.

Citron D., Taffler R. (1992), *The audit report under going concern uncertainties: an empirical analysis,* Accounting and Business Research, vol. 22, pp. 337-345.

Craswell A., Stokes D.J., Laughton J. (2002), *Auditor independence and fee dependence*, Journal of Accounting and Economics, vol. 33, pp. 253-275.

DeAngelo L. (1981), *Auditor size and auditor quality*, Journal of Accounting and Economics, vol. 1, pp. 113-127.

DeFond M. (1992), *The association between changes in client firm agency costs and auditor switching,* Audit: a Journal of Practice and Theory, vol. 11, pp. 16-31.

Firth M. (2002), *Auditor-provided consultancy services and their associations with audit fees and audit opinions,* Journal of Business Finance and Accounting, vol. 29, pp. 661-693.

Han J., Camber M, (2006), *Data Mining Concepts and Techniques*, Morgan Kaufman, 2nd edition, San Diego, USA.

Icerman R., Hillison W. (1991), *Disposition of auditor–detected errors: some evidence on evaluative materiality,* Audit a Journal of Practice and Theory, vol. 10, pp. 22-34.

Johnson B., Lys T. (1990), *The market of auditor services – Evidence from voluntary auditor changes,* Journal of Accounting and Economics, vol. 12, pp. 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, vol. 57, pp. 976-983.

Kirkos E., Spathis C., Manolopoulos Y. (2007), *Data mining techniques for the detection of fraudulent financial statements*, Expert Systems with Applications, vol. 32, pp. 995-1003.

Krishnan J., Krishnan J. and Stephens R. (1996), *The simultaneous relation between auditor switching and audit opinion: An empirical analysis*, Accounting and Business Research, vol. 26, pp. 224-236.

Mutchler J. (1986), *Empirical evidence regarding the auditor's going-concern opinion decision*, Auditing: a Journal of Practice & Theory, vol. 6, pp. 148-163.

Palmrose Z. (1988), *An analysis of auditors litigation and audit service quality*, The Accounting Review, vol. 63, pp. 55-73.

Rakotomalala R. (2005), *TANAGRA un logiciel gratuit pour l'enseignement et la recherché*, in Actes de EGC'2005, RNTI-E-3 2, pp. 697-702.

Teoh SH, Wong TJ. (1993), *Perceived auditor quality and the earnings response coefficient*, The Accounting Review, vol. 68, pp. 346–366.

Velury U., Reish J., O'Reilly D. (2003), *Institutional ownership and the selection of industry specialist auditors*, Review of Qualitative Finance and Accounting, vol. 21, pp. 35-48.

Wilson R., Martinez T. (1997), *Improved heterogeneous distance functions*, Journal of Artificial Intelligence Research, vol. 6, pp. 1-34.