會計與公司治理
第五卷　第二期
2008 年 12 月
55-78 頁

運用決策樹技術探討會計師選擇之關鍵決定因素

洪嘉聲
南華大學

盧鈺欣
國立中正大學

歐進士
國立中正大學

摘要

本研究利用多元選擇迴歸模型及決策樹技術，從代理問題、股權結構、營運績效、風險等構面，探討影響我國企業會計師選擇決策之關鍵決定因素。本研究不僅探討企業會計師選擇決策之關鍵決定因素，亦研究各決定因素在會計師選擇決策過程中的優先順序。由多元選擇迴歸的實證結果可知，我國企業之會計師選擇決策之關鍵決定因素主要包括現金流量權、最終股東控制權、負債資產比率、資產報酬率、資產周轉率、子公司家數、監督股東及負債新增比率有顯著關連。決策樹的實證結果則顯示影響企業會計師選擇五個關鍵決定因素之先後順序分別為負債資產比率、資產周轉率、負債新增比率、資產報酬率及是監督股東。本文的最主要貢獻之一是不但找出企業會計師選擇決策之關鍵決定因素，更進一步透過決策樹分析找出上述因素之先後順序。本文之研究結果使會計師了解企業選擇會計師決策以進一步研擬對策，對擴展會計師業務有參考價值。

關鍵字：會計師選擇、資料探勘、決策樹、分類與迴歸樹

Journal of Accounting and Corporate Governance
Volume 5 Number 2, Dec. 2008
pp.55-78

The Critical Deciding Factors of Auditor Choice Decision：An Application of the Decision Tree Technique

Chia-Sheng Hung
Nan Hua University

Yu-Hsin Lu
National Chung Cheng University

Chin S. Ou
National Chung Cheng University

Abstract

In this paper, we apply the multinomial logit regression and decision tree approaches to examine the critical deciding factors that affect auditor choice decisions. We first examine the relationship between the auditor choice and firm specific factors. In order to get deeper insights about the association between auditor choice and firm's characteristics, this paper employs the decision tree approach, a data mining technique, to explore the priority of variables affecting audit choices. The decision tree approach is an innovative tool that can help researchers choose between alternative options and investigate the possible outcomes of choosing options. To the best of our knowledge, this study is one of few studies that use the decision tree techniques to investigate the sequential deciding factors of audit choice decision. By applying both the regression and CART approaches, we find that auditor choice is affected first by agency conflict, and then by operating efficiency consideration, funding needs, and operating performance consideration sequentially.

Keywords: *Auditor choice, Data mining, Decision tree, Classification and regression tree.*

I. Introduction

There are growing studies on how firms select their auditors. Wallace (1980) proposed three sources of demand for auditor services: agency demand, information demand and insurance demand. The agency demand is derived from the agency theory developed by Jensen and Meckling (1976). Auditors can reduce agency costs arising from the self-interested behavior of agents. Heterogeneous agency costs across firms may accounts for the differential demands for auditor services (Dopuch and Simunic 1980, 1982). The information demand is derived from the argument that the selection of a credible auditor signals management honesty and quality to all related stakeholders. The insurance demand is originated from the argument that auditors indemnify investors and creditors against financial losses via auditors' professional liability exposure.

According to the previous researches, auditor choice is affected by firm specific factors or characteristics. These factors include the agency cost, ownership structure, complexity and risk of firms. In this paper, we first examine the relationship between the auditor choice and firm specific factors. We then use the decision tree approach to investigate the priority of these factors.

The regression models are the conventional methodologies for many empirical investigations (Francis and Wilson 1988; Firth 1999; Lee, Stokes, Taylor and Walter 2003; Fan and Wong 2005). The regression methods usually require the assumption of linear relationship between dependent variables and independent variables and require the independence among all independent variables. However, the above assumptions can be easily challenged. In order to get deeper insights about the association between auditor choice and firm's characteristics, this paper employs the decision tree approach, a data mining technique, to explore the priority of variables affecting audit choices. The decision tree approach is an innovative tool that can help researchers choose between alternative options and investigate the possible outcomes of choosing options. Decision trees are considered to be a map of reasoning process that helps solve the task of classifying cases into individual categories. To the best of our knowledge, this study is one of few studies that use the decision tree techniques to investigate the sequential deciding factors of audit choice decision.

This paper consists of five sections. Section 1 gives an overall introduction of this paper. Section 2 presents the literature review. Section 3 introduces the research method. The data and empirical results are reported in Section 4. Section 5 concludes this paper.

II. Literature Review

Auditor Choice

Prior studies proposed three primary sources of demand for auditors, including agency demand, information demand, and insurance demand (Wallace 1980; Beattie and Fearnley 1995; Abbott and Parker 2000). When the agency conflicts exist, a firm may hire an external auditor to enhance corporate governance mechanism in order to mitigate agency problems. The more serious the agency problem is, the more intensive the need for the external auditor (Jensen and Meckling 1976; Francis and Wilson 1988; Francis, Maydew and Sparks 1999).

Concentrated ownership induces agency problems. When the degree of ownership concentration is high enough, major shareholders will obtain effective control of the firm. Under this circumstance, the nature of the agency problem shifts from shareholder-manager conflicts to the conflicts between the controlling shareholders and the minority shareholders (Fan and Wong 2002; Lemmon and Lins 2003). Shareholders have the voting rights to deploy corporate assets and cash flow right to share the earnings (Fan and Wong 2002). When major shareholders obtain voting rights in the excess of their cash flow rights, the minority shareholders' interest is easily entrenched.

There are other factors that may influence the auditor choice in addition to the agency problem and the ownership structure. Francis et al. (1999) and Reed, Trombley and Dhaliwal (2000) claimed that the audit firms are valuable in moving up the credibility of firms. The contribution of audit firms is especially valuable for firms which have urgent need of rising external funding by issuing debt or equity. Abbott and Parker (2000) and Chaney and Jeter (2004) proposed that the complexity of enterprise and the probability of bankruptcy may also influence the audit choice decisions. Therefore, number of business segment, firm size, profitability, asset utilization efficiency and capital structure are important deciding factors for auditor choice.

The international big audit firms have brand-name reputations and are viewed as producing higher quality audits than non-big firms. DeFond (1992) and Francis and Wilson (1988) proposed that the demand of Big 6 audits is an increasing function of agency costs. Craswell, Francis and Taylor (1995) argued that Big 6 auditors earn significant premiums over non-Big 6 auditors due to their quality differentiation.

The quality of audit depends on the expertise of auditors. Prior studies suggested that the brand name and industry specialization can be used as the proxies of audit quality (Velury, Reisch and O'Reilly 2003; Hay and Davis 2004; Dunn and Mayhew 2004; Fan and Wong, 2005). Abdolmohammadi, Searfoss and Shanteau (2004) examined the differences in expertise attributes by industry specialization. The experience of industry specialization can help auditor to effectively solve the problems of specific industries (Chan, Ferguson, Simunic and Stockes 2001).

Decision Trees

The decision tree approach is one of data mining methods. Data mining is a systematic approach to find underlying patterns, trends, and relationships buried in data and is regarded as a knowledge discovery method. Roiger and Geatz (2003) defined data mining as a process of employing one or more computer learning techniques to automatically analyze and extract knowledge out of data contained within a database. Data mining can be an automatic or semi-automatic process to discover and analyze volumes of data and find meaningful patterns or rules for many decision making problems (Berry and Linoff, 1997).

The researches regarding data mining can be classified into two categories: methodologies and techniques. The methodologies researches consist of data visualization, machine learning, statistical technique, and deductive database (Curt, 1995). The relevant applications of these methodologies include classification, prediction, clustering, summarization, linkage analysis, and sequential analysis (Fayyad, Piatetsky-Sharpiro and Smyth 1996). The techniques of data mining include statistical methods, neural networks, decision trees, genetic algorithms, and non-parametric methods.

The classification analysis is a process for building a systematic classification model that establishes relationships between decision outcome and input variables. Several classification techniques have been proposed, including decision tree, neural network, nearest-neighbor classification, decision-rule induction, and Bayesian networks (Wei, Piramuthu and Shaw 2003).

The decision tree approach has been applied to many issues. Beynon, Peel and Tang (2004) proposed a decision tree model to examine the relationship between firm characteristics and audit fees. Lee, Chiu, Chou and Lu (2006) used the Classification And Regression Tree (CART) approach to explore the performance of credit scoring. They presented four reasons of using CART as a research methodology. First, CART exhibits the capability of modeling complex relationships between variables without strong model assumption. Second, CART can identify "important" independent variables through the built tree when many variables are considered. Third, CART does not require much time for modeling and training process. Finally, the results of CART can be easily interpreted.

The CART has been applied to many research issues. Li (2006) applied CART to stainless steel production and found that the CART could produce insight to materials usage behaviors. Waheed, Bonnell, Prasher and Paulet (2006) utilized CART to investigate the hyper-spectral remote sensing data to extract better crop information. Lee et al. (2006) employed CART to examine the customer credit of banks. Chang and Chen (2005) and Chang and Wang (2006) used CART to examine the risk factors associated with freeway traffic accidents and found that daily traffic volume is the most important determinant for freeway accidents.

Although decision tree and CART have been applied to empirical researches in many fields, application of decision tree or CART to auditor choice has been very limited. In the auditor choice literature, the logistic regression model and the ordered choice model are the conventional methods for empirical investigations (Francis and Wilson 1988; Firth 1999; Lee et al. 2003; Fan and Wong 2005). The regression models require some specific model assumptions and pre-defined underlying relationships between dependent and independent variables. For example, the regression models assume linear relationships between dependent variable and independent variables and require the independence for each independent variable. If these assumptions are violated, the estimated results could be biased. On the other hand, CART is a non-parametric model without pre-defined relationships between dependent variable and independent variables, thus is more flexible in model specification.

III. Sample, Variables and Research Method

Sample

The objective of this paper is to examine the auditor choice decision of business groups in Taiwan. Business groups are popular in the developing or developed world, including Taiwan and China. A business group is a gathering of formally independent firms under a single common administrative and financial control center. Ownership structures of firms in business groups are usually pyramidal and more concentrated (La Porta, Lopez-de-Silanes and Shleifer 1999; Claessens, Djankov and Lang 2000; Chang and Hong, 2000; Morck and Yeung, 2003) and thus makes agency conflicts worsen. Since audit is mechanism to reduce agency costs, it is worthy to investigate how business groups make their auditor choice decisions.

Many Taiwan enterprises, such as the Taiwan Semiconductor Manufacturing Company (TSMC), Foxconn Technology Group, Formosa Plastic Group, and Uni-President Group, have invested a lot of capital and resources in China. Since joining the WTO on December 11, 2002, China has removed many geographic limitations and business restrictions. These actions have created many investment opportunities and thus create more and more demand for audit firms' services. For international audit firms, they need to have a good grasp of Chinese cultural factors and local business practices before penetrating this huge auditing market. China had implemented a centrally planned economy for many years before the start of its reform program. As a result, there is little information about Chinese business practices. Since the business practices and value judgement are similar between Taiwan and China, this paper provides the lessons learned from Taiwanese firms to facilitate the understanding of the auditor choices of China businesses which are useful for international audit firms to penetrate the Chinese audit market.

In order to avoid the impact of the merge between KPMG and Coopers & Lybrand in 1999 and the merge of Arthur Andersen and Deloitte Touche Tohmatsu in 2003, we choose the research period from 2000 to 2002. During this period, there were three big audit firms in Taiwan, Arthur Andersen, Pricewaterhouse Coopers, and KPMG. The data about auditor and firm characteristics are drawn from datasets of Taiwan Economic Journal (TEJ), Taiwan Stock Exchange, Smart Net[1], and Joint Credit Information Center. We exclude financial institutions since they are highly regulated. Overall, there are 874 observations used in this paper.

Variables

This paper examines the critical deciding factors affecting the auditor choice decisions. We classify the audit firms into three categories. The first category is the non-Big 3, the second category is the Big 3, and the last one is Big 3 with industry specialization. The big audit firms carry names that can be viewed as "brand" (Hay and Davis 2004; Kane and Velury 2004; Fan and Wong 2005). The experience of industry specialization can help auditor to solve auditing problems specific to industries effectively (Chan et al., 2001). Prior studies suggested that both brand name and industry specialization are indicators of audit quality (Velury et al. 2003; Hay and Davis 2004; Dunn and Mayhew 2004; Fan and Wong 2005).

Prior researches used audit market share within a specific industry as a proxy of industry specialization (Palmrose 1986; Craswell et al. 1995; Ferguson and Stokes 2002; Velury et al. 2003). Following the definition of Simunic (1980) and Palmrose (1986), we adopt two conditions to define audit firms with industry specialization within a specific industry. First, the audit firm's market share within a industry should exceed 24%. Second, the number of companies within the industry should be more than 20. The audit firm may not be viewed as with industry specialization just because the number of companies within a specific industry is less than the requirement of definition.

From the perspective of agency cost, DeFond, Francis and Wong (2000) indicated when current liability ratio (the ratio of current liabilities to current assets) is larger, agency problem is more serious. Jung and Kwon (2002) and Velury et al. (2003) proposed that the second largest shareholder can play a supervising role to help outside shareholders monitor

[1] Smart Net: http://www.smartnet.com.tw.

a corporation's management. We use the ratio of long term debt to total assets and the existence of the second largest shareholder investor as proxies of agency conflicts. These two variables are denoted as *LEV* and *BLOCK*, respectively. When the second largest shareholder (if the share of second largest shareholder is more than 5%) exists, the variable, *BLOCK*, equals to one; otherwise, zero. When the ratio of long term debt to total assets is higher or the second largest shareholder exists, the firm will require higher audit quality. Thus, we predict the coefficients of *LEV* and *BLOCK* to be positive.

As large shareholders' control increases, their abilities to expropriate minority shareholders' interest increase. We use the controlling shareholder's voting rights and cash flow rights as the proxies of control and ownership (Fan and Wong, 2005). These two variables are denoted as *VOTE* and *CASH*, respectively. We use ownership of 10% or more of shares to define controlling shareholder (La Porta et al., 1999). If there is no controlling shareholder, we define the voting rights (*VOTE*) and cash flow rights (*CASH*) to be zero. When the controlling shareholder's control increased or their ownership decreased, the possibility for the controlling shareholders to entrench small investors will rise. In this case, the agency conflict between large shareholders and small shareholders becomes more serious and the demand for higher audit quality services increases. Hence, we predict the coefficients of *VOTE* and *CASH* o positive and negative, respectively.

The auditor with high audit quality can rise up the credibility for the auditees, especially for those planning to issue debt or equity to raise funds (Francis et al. 1999; Reed et al. 2000). If the ratio of the increment of long debts and equities to total assets is higher than the average, the firm's demand for the endorsement of an audit firm is more urgent than other companies. We use a dummy variable (denoted as *NEW*) to indicate whether the demand for the endorsement of an audit firm is urgent. The variable *NEW is* equal to one if the ratio of the increment of long debts and equities to total assets is higher than the average. Otherwise, the dummy variable *NEW is* given zero. If the demand for the endorsement is relative intensive, the firm will require higher auditing quality. Thus, the predicted sign of the coefficient of variable *NEW is* positive.

Besides, the complexity of firms can influence auditor choice decision as well. We use size and number of subsidiary companies as the proxies of organizational complexity of firms. These two variables are denoted as *SEGNUM* and *SIZE is*. The variable *SEGNUM* is the square root of the number of subsidiary companies. *SIZE* is a dummy variable. If a firm's total assets belong to top 25%, *SIZE* equals to 1, otherwise 0. Larger audit firms are expected to have higher capabilities to audit large and complex companies, thus the predicted signs of coefficients of *SEGNUM* and *SIZE* are both positive.

Abbott and Parker (2000) and Chaney and Jeter (2004) found that operating performance and efficiency were related to auditor choice decisions. In this paper, we use the return of assets (denoted as *ROA*) and the ratio of sales to total assets (denoted as

ATURN) as the proxies of operating performance and operating efficiency, respectively. The variable *ROA* is a conventional accounting performance index and the variable *ATURN* represents a firm's assets utilization efficiency. In addition, we control for the impact of negative income on *ROA* and thus include the variable *ROA *LOSS* as an additional independent variable. The *LOSS* variable is a dummy variable, which equals to 1 when a firms' net income is negative and 0 otherwise.

In order to avoid business failure, a firm has to be conscious about managing credit, liquidity, and risk. We use the ratio of current assets to total assets (denoted as *CURR*) as a risk proxy. The variable *CURR* represents the liquidity of assets. If the liquidity of assets is low, the risk of bankruptcy will be high. The audit firms provide management consulting service and risk management service to help firms enhance their operating performance and control operating risk to an acceptable level. If a firm is not very efficient in operation or their risk is relatively high, its demand for the services from audit firms will be more intensive. On the other hand, a firm with higher operating efficiency or lower risk could be the consequence of following an audit firm's consultant services. It is hard to identify the causality between audit choice and operating efficiency and risk, therefore we don't make prediction of the signs of the coefficients of *ROA*, *ROA *LOSS*, *ATURN*, and *CURR*.

In short, in this paper we include the variables of agency cost, ownership structure, and needs of raising funds, firm complexity, operating performance and efficiency when examining the auditor choice decision.

In order to avoid business failure, a firm has to be conscious about managing credit, liquidity, and risk. The audit services can help firms (auditees) enhance their credibility and transparency in financial reporting. Pittman and Fortin (2004) examined the impact of auditor choice on debt pricing. They found that the auditor of higher quality can reduce debt-monitoring costs by enhancing the credibility of financial statements.

Research Method

The main propose of this paper is using the decision tree approach to find the critical deciding factors of the auditor choice in addition to logit regression model. Previous studies argued that the combination of multiple techniques of data mining can improve the accuracy of estimation (Kim, Kim and Lee 2003; Kim, Min and Han 2006). This paper combines the decision tree approach and multinomial logit regression model to estimate the critical deciding factors of the auditor choice. We first perform multinomial logit regression to examine the relationship between auditor choice and the proposed explanatory factors. The empirical results of this stage will show the factors which are significantly related with auditor choice decisions. Next, we precede the decision tree analysis only on the significant variables of the first stage. By doing so, we can simply the decision tree and delete the non-significant variables out of the decision tree analysis. The results of decision tree will indicate the sequential order of these deciding factors when firms make their auditors choice decisions.

Multinomial Logit Regression

This paper classifies auditors into three classes: non-Big 3, Big 3 and Big 3 with industry specialization. The audit quality of non-Big 3 is presumed to be worse than Big 3 and Big 3 with industry specialization.

The multinomial logit regression model is specified as the following equations:

$$AUD\_CH = \beta 0 + \beta 1\ CASH + \beta 2\ VOTE + \beta 3\ LEV + \beta 4\ SIZE + \beta 5\ BLOCK + \beta 6\ NEW + \beta 7\ SEGNUM + \beta 8\ ROA + \beta 9\ ATURN + \beta 10\ CURR + \beta 11\ LOSS*ROA + \varepsilon \quad (1)$$

The definitions of variables are presented in table 1.

Table 1 Definition of variables

AUD_CH	If an auditor is non-Big 3, AUD_CH is given 0; if an auditor is Big 3 without industry specialization, AUD_CH is given 1; if an auditor is Big 3 with industry specialization, AUD_CH is given 2.
CASH	Cash flow right. The percentage of cash flow right of controlling shareholders.
VOTE	Voting right. The percentage of voting right of controlling shareholders.
LEV	The ratio of long debts to total assets.
SIZE	Dummy variable. When the log of total assets belongs to top 25% of the sample, SIZE is given 1 and 0 otherwise.
BLOCK	Dummy variable. If percentage of shares of the second largest shareholder is more than 5%, BLOCK is given 1 and 0 otherwise.
NEW	Dummy variable. If the growth ratio of long debts and equities to total assets is higher than the average, NEW is given 1, otherwise 0.
SEGNUM	The square root of the number of subsidiary companies.
ROA	The ratio of income before tax and interest to total assets.
ATURN	The ratio of sales to total assets.
CURR	The ratio of current assets to total assets.
LOSS*ROA	LOSS is a dummy variable, 1 when a firm had net loss last year and 0 otherwise.

Decision Tree Approach

The decision tree approach is a powerful tool for making prediction and classification. Several algorithms of decision trees are developed in recent years, for example C4.5, C5.0 and CART (classification and regression tree). Proposed by Breiamn, Friedman, Olshen and Stone (1984), CART is a non-parametric method that can find the rules for classifying the dependent variable. CART has been found to produce more accurate predictions than other statistical approaches in many cases (Kattan, Hess and Beck 1998). CART is also found to be better than conventional discriminate analysis (for example Chang and Chen 2005; Chang and Wang 2006). One of the distinguishing features of CART is that it cab be applied to both categorical and continuous dependent variable. To the best of our knowledge, this paper is one of very few studies that use CART to investigate the auditor choice.

CART has the capacity of automatically searching of the best predictors and the best threshold values for all predictors to classify the dependent variable. It uses a stepwise method to establish splitting rules. The procedure is binary and recursive. The parent nodes are always split into two child nodes and the process is repeated by treating each child node as a parent node. The process is repeated until further partitioning is impossible. The node is called a terminal node if the node data cannot be split into additional child nodes. Once the first terminal node has been created, the algorithm repeats the procedure for each set of data until all data are categorized as terminal nodes.

Beginning with the first parent node, CART finds the best possible variable to split the node into child nodes. The algorithm checks all possible splitting variables and all possible threshold values of the variable to be used to split the node. The algorithm seeks to maximize the average "purity" of the two child nodes. The "Gini method" is the default splitting criterion (Han, 2007).

CART attempts to pull out the largest category in the child nodes. For example, in our case, we classify the data into three categories, non-Big 3, Big 3 without industry specialization, and Big 3 with industry specialization. The proportions of these categories are 15.22%, 75.97%, and 8.81%, respectively. The CART would attempt to pull out the "Big 3 without industry specialization" category into a node first. The procedure is recursively repeated until the terminal node has been created.

If we have n-category classification problem, the impurity of node t can be calculated by the following equation: $i(t) = 1 - \sum_{j=1}^{n} p(j|t)^2$, where $i(t)$ is Gini impurity, $p(j|t)$ is the proportions of category j in the node t. In this study, n is 3 and the proportion of each categories is 15.22%, 75.97%, and 8.81 %, respectively.

IV. Empirical Results

Descriptive Statistics

The descriptive statistics are presented in Table 3. The average cash flow right of controlling shareholder (*CASH*) is 19.880 and the voting right (*VOTE*) is 27.691. The *BLOCK* value of 0.344 suggests that about one-third of sample companies have the second largest shareholders with more than 5% of shares. About 25% companies issue large long debts and equities. The mean value of *ROA*LOSS* is -0.382 indicating that the average *ROA* is -0.382 for firms suffering net losses.

The correlate analysis of independent variables is shown in Table 2. The figures above the diagonal are the Spearman coefficients, and the figures below the diagonal are the Pearson coefficients. The correlation coefficients between variables are low except for *CASH* and *VOTE* (correlation coefficient between *CASH* and *VOTE* is 0.637). Although the correlation coefficient between *CASH* and *VOTE* is 0.637, the VIFs of all independent variables are smaller than 4, suggesting that no co-linearity exists in our regression model.

Results of Multinomial Logit Regression Model

The auditor choice is a polychromous-response variable. In this paper we use multinomial logit regression model to analyze the critical deciding factors of auditor choice. We use non-Big 3 as the baseline auditor choice case. The empirical results of multinomial logit regression are shown in Table 4.

With regard to the proxies of agency conflicts, the coefficients of *LEV* and *BLOCK* are significantly different from zero for Big 3 and Big 3 with industry specialization models. The coefficients of *LEV* of both models are statistically significant, indicating that the auditor choice is significantly affected by agency conflicts. The coefficient of *LEV* for Big 3 being negative and the coefficient for Big 3 with industry specialization being positive indicate that firms with higher *LEV* have higher probability of choosing Big 3 with industry specialization and have lower probability of choosing Big 3. The above results reveal that firms with higher ratio of long term debt to total assets prefer non-Big 3 or Big 3 with industry specialization to deal with this problem.

Table 2 Correlation coefficients of independent variables

	CASH	VOTE	LEV	SEGNUM	ROA	ATURN	CURR	SIZE	BLOCK	NEW	LOSS* ROA
CASH1	**1.000**	.561	.007	.020	.094	.056	.082	-.162	-.047	-.040	-.039
VOTE1	.637	**1.000**	.097	-.091	-.036	-.078	-.059	-.048	-.086	-.047	-.034
LEV	.007	.125	**1.000**	.057	-.383	-.043	.004	.176	-.015	.067	-.118
SEGNUM	-.016	-.075	.078	**1.000**	.077	.003	-.218	.286	-.151	-.062	.000
ROA	.064	-.046	-.383	.068	**1.000**	.456	.280	-.083	.123	.084	.328
ATURN	.003	-.089	.037	.005	.316	**1.000**	.525	-.229	.035	.041	.146
CURR	.045	-.065	.047	-.229	.221	.504	**1.000**	-.286	.133	.055	.010
SIZE	-.153	-.034	.157	.288	-.051	-.166	-.252	**1.000**	-.038	.046	.035
BLOCK	-.040	-.083	-.018	-.138	.077	.077	.132	-.038	**1.000**	.076	-.001
NEW	-.025	-.029	.040	-.055	.037	.090	.059	.046	.076	**1.000**	.009
LOSS* ROA	-.018	-.016	-.150	.023	.553	.104	.016	.075	-.010	-.010	**1.000**
	CASH	VOTE	LEV	SEGNUM	ROA	ATURN	CURR	SIZE	BLOCK	NEW	LOSS* ROA

The figures above the diagonal are Spearman coefficients, and the figures below the diagonal are Pearson coefficients.

Table 3 Basic descriptive statistics

	Mean	Std. Dev.	Minimum	Maximum
AUD_CH	0.936	0.486	0	2
CASH	19.880	13.808	0	87.73
VOTE	27.691	13.554	10.06	88.97
LEV	41.823	15.686	3.080	94.52
SEGNUM	1.927	0.976	0	6.403
ROA	4.890	9.353	-73.400	42.19
ATURN	77.639	61.076	0	446.830
CURR	41.110	21.947	3.101	96.860
SIZE	0.391	0.4883	0	1
BLOCK	0.344	0.475	0	1
NEW	0.247	0.432	0	1
LOSS*ROA	-0.382	4.747	0	1

Table 4 Results of multinomial logit regression

$$AUD\_CH = \beta_0 + \beta_1\ CASH + \beta_2\ VOTE + \beta_3\ LEV + \beta_4\ SIZE + \beta_5\ BLOCK + \beta_6\ NEW + \beta_7\ SEGNUM + \beta_8\ ROA + \beta_9\ ATURN + \beta_{10}\ CURR + \beta_{11}\ LOSS*ROA + \varepsilon$$

		AUD_CH =1			AUD_CH =2		
	Sign	Parameter	p value	VIF	Parameter	p value	VIF
Constant		-1.131	0.002 ***		-6.185	0.000 ***	
CASH	−	-0.035	0.000 ***	1.826	-0.055	0.000 ***	1.826
VOTE	+	0.021	0.014 ***	1.819	0.059	0.000 ***	1.819
LEV	+	-0.014	0.013 ***	1.773	0.044	0.000 ***	1.773
SIZE	+	0.052	0.763	1.285	0.242	0.424	1.285
BLOCK	+	0.439	0.009 ***	1.053	0.689	0.018 **	1.053
NEW	+	0.634	0.001 ***	1.223	-1.220	0.009 ***	1.223
SEGNUM	+	0.327	0.000 ***	1.213	0.579	0.000 ***	1.213
ROA	?	-0.020	0.080 *	1.721	0.009	0.681	1.721
ATURN	?	0.009	0.000 ***	1.470	0.008	0.003 ***	1.470
CURR	?	0.014	0.001 ***	1.746	0.000	0.987 ***	1.746
LOSS*ROA	?	-0.013	0.540	1.501	-0.004	0.910	1.501
Pseudo R^2	0.127						
Predict Correct Percentage	60.870						
Chi Square	204.838						
Degree of freedom	22						

a. * P< 0.1, ** P< 0.05, *** P< 0.01.

As to the ownership structure, the coefficients of *CASH* and *VOTE* are both significantly different from zero. The coefficients of *CASH* are negative and those of *VOTE* are positive[2] in the Big 3 or Big 3 with industry specialization models. These results show that the more the controlling shareholders own cash flow right, the less likely the firms would select Big 3 or Big 3 with industry specialization. The more the controlling shareholders own the voting right, the more likely the firms would select Big 3 or Big 3 with industry specialization. These evidence are consistent with prior studies (for example, Claessens, Djankov, Fan and Lang 2002; Fan and Wong 2002, 2005) which argued that cash flow right and voting right can be used as proxies for agency conflicts. Our results suggest that firms are more likely to select Big 3 or Big 3 with industry specialization when the agency conflicts are more serious.

The variable *NEW* is the proxy for firms' funding needs. We find that the auditor choice is significantly affected by *NEW*. Firms with higher funding needs choose Big 3 for their endorsements for funds rising. The proxies of firm complexity are *SEGNUM* and *SIZE*. *SEGNUM* has significant impact on auditor choice, but *SIZE* doesn't. The coefficients of *SEGNUM* are positive, indicating that firms will be more likely to choose Big 3 with or without industry specialization when their organizations are more complex.

The variable *CURR* represents the liquidity of assets. If the liquidity of assets is low, the risk of company will be high. The coefficients of *CURR* are significantly positive, suggesting that firms with higher risks tend to hire an auditor of Big 3 with or without industry specialization.

The variables *ROA* and *ATURN* represent the return of assets and the ratio of sales to total assets, proxies of operating performance or efficiency. The coefficients of *ATURN* are significantly positive in both models suggesting that firms with higher operating efficiency tend to hire an auditor of Big 3 with or without industry specialization. The coefficient of *ROA* is significantly negative for the Big 3 model, indicating that firms with lower operating performance tend to choose Big 3. The coefficients of *ROA *LOSS* are not significantly different from zero, indicating that firms with negative net income have no obvious impact on auditor choice.

From the above empirical results of multinomial logit regression, we find that only two independent variables, *SIZE* and *LOSS*ROA*, have no significant associations with audit choice decisions. The variables of *LEV*,*BLOCK*, *CASH* and *VOTE* all have significant impacts on auditor choice decisions. The variable *NEW* is associated with the demand for the financial advisory service. The variable *ROA* , *ATURN*, *SEBNUM* are related with the demands for management consulting service and the variable *CURR* is corresponding to the risk management service. In short, we find that auditor choice is associated with firms'

[2] Some researches take the deviation of cash flow right and voting right as the proxies of agency conflicts. In our sample, the deviation of cash flow right and voting right does not significantly affect the auditor choice.

needs for financial advisory service, management consulting service and risk management service.

Results of CART

In order to avoid the interferences of irrelevant variables and simplify the decision tree, we exclude the variables of *SIZE* and *LOSS*ROA* in the second stage CART analysis. The CART uses a stepwise method to establish splitting rules recursively. The “Gini method” is the default splitting criterion and attempt to pull out the largest category in the node. In our sample, there are 664 observations choosing Big 3 audit firms without industry specialization. Therefore, the CART pulls out the observations which chose Big 3 audit firms without industry specialization recursively. In order to simplify the splitting rules, the CART sets the maximum tree depth of 5 as the stopping rule.

The results of CART are shown in Figure 1. The sample is first split by *LEV* , and then by the order of *ATURN*,*NEW*,*ROA*, and *BLOCK*.

The splitting rule for nodes of second layer depends on the value of the variable *LEV*. Observations with $LEV \geq 0.619$ are gathered in node 2. In this node, there are 53.12% of observations choosing Big 3 audit firms. This result shows that when the agency conflicts are serious, more than 50% of firms will choose Big 3 audit firms. There is no enough information in node 2 to reduce the impurity of node, so no node is extended from node 2.

Observations with $LEV < 0.619$ are gathered in node 1. In the node 1, the class of Big 3 is the largest category. There are 75.97% of observations choosing Big 3 audit firms. The node 1 is splitted into node 3 and 4 according the value of *ATURN*. In node 1, observations with $ATURN \geq 1.326$ are grouped into node 3; and the other observations are grouped into node 4. In node 4, 93.27% of observations chose Big 3, 6.73% of observations chose Big 3 with industry specialization, and no observation chose non-Big 3 audit firms. Since the variable *ATURN* represents operating efficiency, the above results show that when agency conflicts are not very serious, operating efficiencies will domain the auditor choice decision.

In following layers of the nodes, the splitting criterions depend on the values of *NEW*, *ROA*, and *BLOCK*. There are 87.65%, 86.05%, and 78.57% of observations choosing Big 3 in nodes 6, 8, and 10, respectively. These nodes show that funding needs, operating performance consideration, and agency conflict sequentially domain the auditor choice. There is no node developed from node 2, 4, 6, and 8, because there is no additional information in these nodes to reduce the Gini impurity by further splitting.

From the above CART analysis, auditor choice is affected first by *LEV* and *ATURN*, and then by *NEW*, *ROA*, and *BLOCK*. The variable *LEV* and variable *ATURN* are the proxies for agency conflicts and operating efficiency and play key roles in auditor choice decisions.

Although the variables of *CASH*, *VOTE*, *SEGNUM*, and *CURR* are significantly associated with auditor choice in multinomial logit regression, these variables have no high priority in the decision tree analysis, they do not show up in Figure 1. Our results demonstrate the contribution of using the traditional regression model first and then apply the CART approach to find the sequential order of factors that have significant impacts on auditor choice decision. From the first stage of multinomial logit regression, we find that two independent variables, *SIZE* and *LOSS*ROA*, have no significant associations with audit choice. In the second stage of CART analysis, we find that although variables of *CASH*, *VOTE*, *SEGNUM*, and *CURR* are significantly in multinomial logit regression, these variables have low priority in the decision tree. By applying both the regression and CART approaches, we find that auditor choice is affected by *LEV*, *ATURN*, *NEW*, *ROA*, and *BLOCK* sequentially.

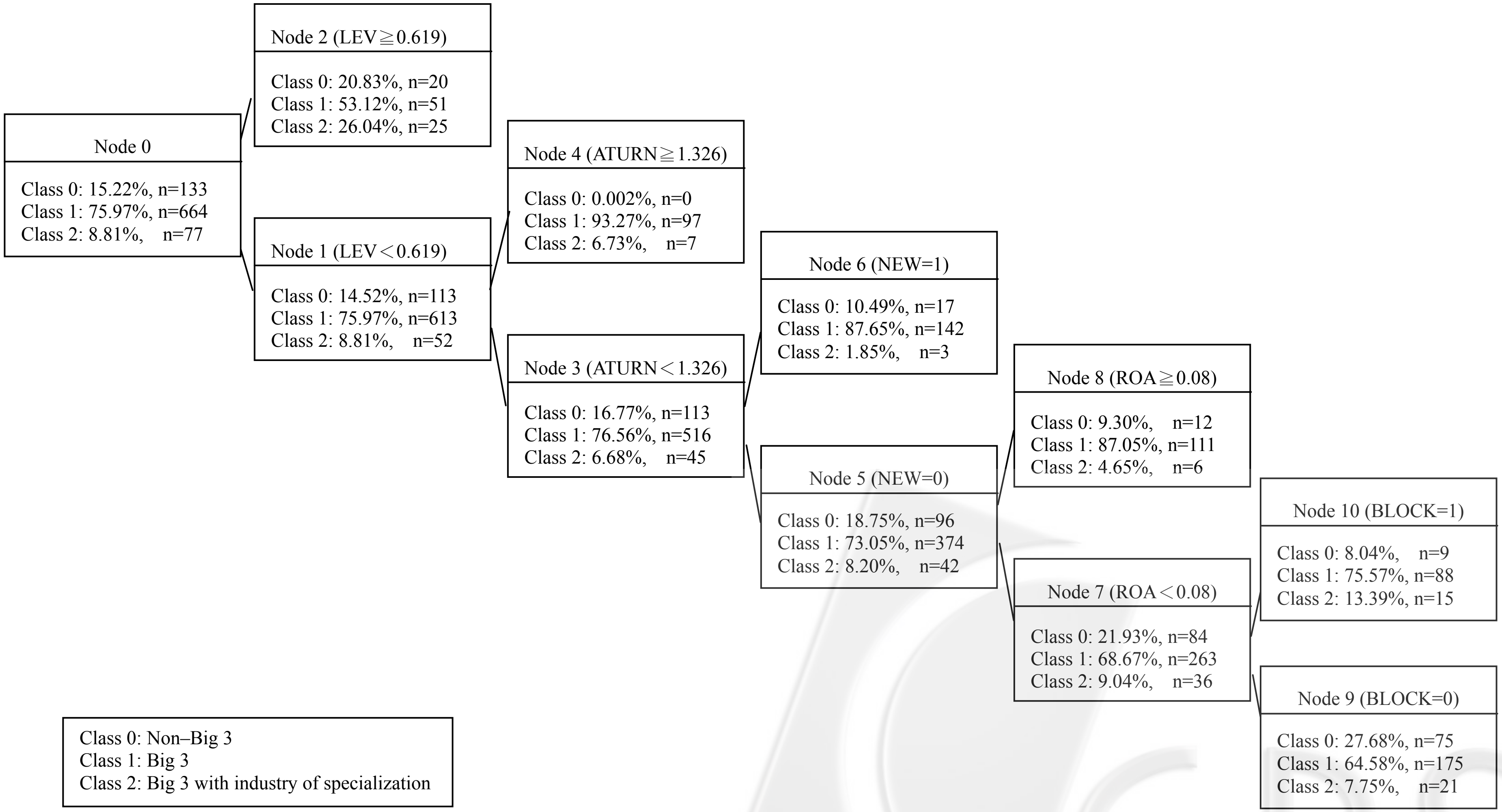
V. Conclusion

This study is an exploratory application of data mining to examine the sequentially critical deciding factors of auditor choice decision. We examine the relationship between auditor choice and the deciding factors from the perspectives of agency conflicts, ownership structure, funding needs, organizational complexity, operating performance and efficiency, and liquidity of assets. In the first stage, our regression results indicate that auditor choice is associated with agency conflicts, ownership structure, funding needs, organizational complexity, operating performance and efficiency, and liquidity of assets. In the second stage, the CART results show that agency conflict and operating efficiency have high priority in the auditor decision process. By applying both the regression and CART approaches, we find that auditor choice is affected first by agency conflict, and then by operating efficiency consideration, funding needs, and operating performance consideration sequentially.

There are two limitations in this paper. DeAngelo (1981) claimed that the low rate at which firms change auditors is an evidence of switching costs. The auditor choice may not a response to firm characteristics over time. Prior researches examined the auditor choice as the consequences of changes in circumstances and focused on the auditor change (Francis and Wilson 1988; Beattie and Fearnley 1995; Firth 1999; Branson and Breesch 2004) or the auditor choice initial public offering (IPO) firms (Lee et al. 2003; Pittman and Fortin 2004). If we follow prior researches and study the audit change or the audit choice of IPO firms, the number of sample will be small and the empirical results may loose the generality. Thus, we assume firm characteristics hold out over time. In the future, the effects of changes in firm characteristics on auditor changes would be an interesting issue to be explored.

The second limitation is that we do not investigate the role of corporate governance in auditor choice decisions in this paper. Since the enactment of Sarbanes-Oxley act, there are more and more firms establishing audit committee and giving the power of choosing auditor to the committee. It would be interesting to examine the effects of audit committee on auditor choice decisions in addition to the effects of firm characteristics.

Figure 1: Decision tree analysis



References

Abbott, L. J., and S. Parker. 2000. Auditor selection and audit committee characteristics. *Auditing: A Journal of Practice & Theory*, 19(Fall): 47-66.

Abdolmohammadi, M.J., D. G. Searfoss, and J. Shanteau. 2004. An investigation of the attributes of top industry audit specialists. *Behavioral Research in Accounting*, 16(Fall): 1-17.

Beattie, V., and S. Feranley. 1995. The importance of audit firm characteristics and the driver of auditor change in UK listed companies. *Accounting and Business Research*, 25(Autumn): 227-239.

Berry, M.J., and G. Linoff. 1997. *Data mining techniques: For marketing sale and customer support*. John Wiley and Sons, Inc., Canada.

Beynon, M. J., M. J. Peel and Yu-Cheng Tang. 2004. The application of fuzzy decision tree analysis in an exposition of the antecedents of audit fees. *Omega*, 32(June): 231-244.

Branson, J. and D. Breesch. 2004. Referral as a determining factor for changing auditors in the Belgian auditing market: An empirical study. *The International Journal of Accounting*, 39: 307-326.

Breiman, L., J. Friedman, R. Olshen, and C. Stone. 1984. *Classification and regression trees.* Wadsworth.

Chan, D. K., A. Ferguson, D. A. Simunic, and D. Stockes. 2001. A spatial analysis and test of oligopolistic competition in the market for assurance services. Working paper, University of British Columbia.

Chang, S. J., and J. Hong. 2000. Economic performance of group-affiliated companies in Korea: Intra-group resource sharing and internal business transactions. *Academy of management Journal*, 43(June): 429-48.

Chaney, P. K., and D. C. Jeter. 2004. Self-selection of auditors and audit pricing in private firms. *The Accounting Review*, 79 (January): 51-72.

Chang, L.Y. and H. W. Wang. 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis and Prevention*, 38(September): 1019-1027.

Chang, L.Y. and W. C. Chen. 2005. Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, 36: 365-375.

Claessens, S., S. Djankov, J. P.H. Fan, and L. H. P. Lang. 2002. Disentangling the incentive and entrenchment effects of large shareholdings. *Journal of Finance*, 57(December): 2741-2771.

Claessens, S., S. Djankov, and L. H. P. Lang, 2000. The separation of ownership and control in East Asian corporation. *Journal of Financial Economics*, 58: 81-112.

Craswell, A. T., J. R. Francis, and S. L. Taylor. 1995. Auditor brand name reputations and industry specialization. *Journal of Accounting and Economics*, 20(December): 297-322.

Curt, H. 1995. The devil's in the detail: techniques, tools, and applications for database mining and knowledge discovery-Part I. *Intell. Software Strategies*, 6: 1-15.

DeAngelo, L. E. 1981. Auditor size and audit quality. *Journal of Accounting and Economics*, 3(December): 183-199.

DeFond, M., J. Francis, and T. J. Wong. 2000. Auditor industry specialization and market segmentation: Evidence from Hong Kong. *Auditing: A Journal of Practice and Theory*, 19(Spring): 49-66.

DeFond, M., T. J. Wong, and S. Li. 1999. The impact of improved auditor independence on audit market concentration in China. *Journal of Accounting and Economics*, 28(December): 269-305.

DeFond, M. 1992. The association between changes in client firm agency cost and auditor switching. *Auditing: A Journal of Practice and Theory*, 11(Spring): 16-31.

Dopuch, N. and D. Simunic. 1980. *The nature of competition in the auditing profession: A descriptive and normative view.* In Buckley, J. W. and Weston, J. F. (eds.), Regulation and the Accounting Profession (Lifetime Learning Publications).

Dopuch, N. and D. Simunic. 1982. *Competition in auditing: An assessment.* Fourth Symposium on Auditing Research (University of Illinois).

Dunn, K. A., and B. W. Mayhew. 2004. Audit firm industry specialization and client disclosure quality. *Review of Accounting Studies*, 9(March): 35-58.

Fan, J. P. H., and T. J. Wong. 2005. Do external auditors perform a corporate governance role in emerging markets? Evidence form East Asia. *Journal of Accounting Research*, 43(March): 35-72.

Fan, J. P. H., and T. J. Wong. 2002. Corporate ownership structure and the informativeness of accounting earnings in East Asia. *Journal of Accounting and Economics*, 33(August): 401-425.

Ferguson, A., and D. Stokes. 2002. Brand name, audit pricing, industry specialization, and leadership premiums post-Big 8 and Big 6 mergers. *Contemporary Accounting Research*, 19(Spring): 77-110.

Fayyad, U., G. Piatetsky-Sharpiro, and P. Smyth. 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(November): 27-34.

Firth, M. 1999. Company takeovers and the auditor choice decision. *Journal of International Accounting, Auditing & Taxation*, 8: 197-214.

Francis, J. R., E. L. Maydew, and H. C. Sparks. 1999. The role of Big 6 auditors in the credible reporting of accruals. *Auditing: A Journal of Practice & Theory*, 18(Fall): 17-34.

Francis, J. R., and E. R. Wilson. 1988. Auditor changes: A joint test of theories relating to agency costs and auditor differentiation. *The Accounting Review*, 63(October): 663-682.

Han, J. and M. Kamber. 2007. *Data mining-concepts and techniques.* Morgan Kaufmann Pub. San Francisco, CA. USA.

Hay, D., and D. Davis. 2004. The voluntary choice of an auditor of any level of quality. *Auditing: A Journal of Practice & Theory*, 23(September): 37-53.

Jensen, M. C., and W. H. Meckling. 1976. Theory of the firm：Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(October): 305-360.

Jung, K., and S. Y. Kwon. 2002. Ownership structure and earnings informativeness: Evidence from Korea. *The International Journal of Accounting*, 37: 301-325.

Kane, G. D., and U. Velury. 2004. The role of institutional ownership in the market for auditing services: An empirical investigation. *Journal of Business Research*, 57(September): 976-983.

Kattan, M.W., K.R. Hess, J. R. Beck. 1998. Experiments to determine whether recursive partitioning (CART) or an artificial neural network overcomes theoretical limitations of Cox proportional hazards regression. *Computers and biomedical research*, 31(October): 363-373.

Kim, E., W. Kim, and Y. Lee. 2003. Combination of multiple classifiers for the customers' purchase behavior prediction. *Decision Support Systems*, 34(January): 167-175.

Kim, M.J., S.H. Min, and I. Han. 2006. An evolutionary approach to the combination of multiple classifiers to predict a stock price index. *Expert Systems with Applications*, 31(August): 241-247.

La Porta, R., F. Lopez-de-Silanes, and A. Shleifer. 1999. Corporate ownership around the world. *Journal of Finance*, 54(April): 471-517.

Lee, P., D. Stokes, S. Taylor, and T. Walter. 2003. The association between audit quality, accounting disclosures and firm-specific risk: Evidence form initial public offerings.

Journal of Accounting and Public Policy, 22(September-October): 377-400.

Lee, T.S., C.C. Chiu, Y.C. Chou and C.J. Lu. 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50(February): 1113-1130.

Lemmon, M. L., and K. V. Lins. 2003. Ownership structure, corporate governance, and firm value: Evidence from the East Asian financial crisis. *Journal of Finance*, 58(August): 1445-1468.

Li, Y. 2006. Predicting materials properties and behavior using classification and regression trees. *Materials Science and Engineering: A*, 433(October): 261-268.

Morck, R., and B. Yeung. 2003. Agency problems in large family business group. *Entrepreneurship Theory and Practice*, 27(Summer): 367-382.

Palmrose, Z. V. 1986. Audit fees and auditor size: Further evidence. *Journal of Accounting Research*, 24(Spring): 97-110.

Pittman, J.A. and S. Fortin. 2004. Auditor choice and the cost of debt capital for newly public firms. *Journal of Accounting and Economics*, 37(February): 113-136.

Reed, B. J., M. A. Trombley, and D. S. Dhaliwal. 2000. Demand for audit quality: The case of Laventhol and Horwath's Auditees. *Journal of Accounting, Auditing and Finance*, 15(Spring): 183-198.

Roiger, R. J. and M. W. Geatz. 2003. *Data mining: A tutorial–based primer.* Boston MA: Addision Wesley.

Simunic, D. 1980. The pricing of audit services: Theory and evidence. *Journal of Accounting Research*, 18(Spring): 161-190.

Velury, U., J. T. Reisch, and D. M. O'Reilly. 2003. Institutional ownership and the selection of industry specialist auditors. *Review of Quantitative Finance and Accounting*, 21 (July): 35-48.

Waheed, T., R.B. Bonnell, S.O. Prasher and E. Paulet. 2006. Measuring performance in precision agriculture: CART– A decision tree approach. *Agricultural Water Management*, 84(July): 173-185.

Wallace, W. A. 1980. *The economic role of the audit in free and regulated markets.* New York: The Touch Ross and Co. Aid to Education Program, reprinted in Auditing Monograhps (Macmillan Publishing Co., 1985).

Wei, C. P., S. Piramuthu, M. J. Shaw. 2003. Knowledge discovery and data mining, *Handbook of Knowledge Management*, vol.2, in C.W. Holsapple (Ed.), Germany, Springer-Verlag, Berlin, Ch. 41: 157-189.