Predicting Auditor Changes Using Financial Distress Variables And The Multiple Criteria Linear Programming (MCLP) And Other Data Mining Approaches

Wikil Kwak, University of Nebraska at Omaha, USA
Susan Eldridge, University of Nebraska at Omaha, USA
Yong Shi, Graduate University of the Chinese Academy of Science & University of Nebraska at Omaha, USA
Gang Kou, University of Electronic Science and Technology of China, China

ABSTRACT

Our study evaluates a multiple criteria linear programming (MCLP) and other data mining approaches to predict auditor changes using a portfolio of financial statement measures to capture financial distress. The results of the MCLP approach and the other data mining approaches show that these methods perform reasonably well to predict auditor changes using financial distress variables. Overall accuracy rates are more than 60 percent, and true positive rates exceed 80 percent. Our study is designed to establish a starting point for auditor-change prediction using financial distress variables. Further research should incorporate additional explanatory variables and a longer study period to improve prediction rates.

Keywords: auditor change; data mining; multiple criteria linear programming

1. INTRODUCTION

Auditor change issues are not new topics, yet auditor changes still occur unexpectedly and frequently without explanation. Hackenbrack and Hogan (2002) document that 43.7 percent of the 4,066 Securities and Exchange Commission (SEC) registrants that had auditor changes from late 1991 through late 1997 provided no reason for the change. More recently, almost 63 percent of the 2,514 companies that changed auditors during 2003 and 2004 gave no reason for the change (Turner et al. 2005).

Financial distress is one possible reason for an auditor change that a company may not want to disclose. As a company's financial health declines, it may decide to change auditors to reduce its audit fees burden or to engage auditors who may be more flexible or less conservative in applying accounting standards. Bankruptcy is the extreme form of financial distress. Schwartz and Menon (1985) conclude based on results of Chi-square tests with bankrupt and non-bankrupt firms that "failing firms have a greater tendency to switch auditors than do healthy firms" (p. 248), and they suggest the "need to control for the presence of financial distress in studies on auditor switching" (p. 248). Chen et al. (2004) and Chen et al. (2009) include an auditor change variable in their bankruptcy prediction studies where the auditor change precedes bankruptcy. However, auditor changes studies, discussed in Section 2 of this paper, have typically not emphasized or controlled for financial distress as a reason for auditor changes. An exception is Hudaib and Cooke (2005). In their study, Hudaib and Cooke (2005) diagram (in Figure 1 on page 1715) and test the influence of managing director changes, financial distress, and other-than-clean audit opinions on auditor switching. The significant negative coefficient on their financial distress explanatory variable in their multivariate logistical regression analysis (Table 10 on page 1733) indicates that the probability of switching auditors increases as financial health declines.

Many of the bankruptcy and auditor switching studies use logistical regression models to predict bankruptcy, auditor changes, or the direction of auditor changes (between large and small audit firms). Data mining classification methods such as multiple criteria linear program (MCLP) and classification and regression trees (CART) may provide better predictive power than traditional logistic regression, but this is an empirical question.

The primary purpose of this study is to analyze the predictive nature of financial distress variables for predicting auditor changes using MCLP and other data mining methods. Our study is designed to be a starting point to establish the usefulness of a portfolio of financial distress variables in predicting auditor changes before any additional predictor variables are considered. Because auditor changes are often viewed negatively by investors (Fusco 2006) and because companies often do not disclose the reason for an auditor change decision (Hackenbrack and Hogan 2002, Turner et al. 2005), our results may provide investors, regulators, and others with better understanding of the influence of financial distress on auditor changes. Our comparison of results using MCLP and other data mining classification methods may also provide insights into the relative usefulness of these statistical tools.

Our paper proceeds in the following order. The next section reviews prior research and presents the variables we use in our data mining models. The third section describes the multiple criteria linear programming (MCLP) model, identifies the comparison models we use, and explains the performance metrics for evaluating the results of the data mining classifications. Section four presents sample selection procedures, data, and empirical results. The last section presents contributions and limitations of our study and further research avenues.

2. PRIOR RESEARCH AND FINANCIAL DISTRESS VARIABLES

A firm's deteriorating financial condition may trigger an auditor change, and bankruptcy (extreme financial distress) may occur subsequent to an auditor change. Our discussion of prior research starts with bankruptcy studies that also incorporate auditor changes. Then we discuss auditor change or switching studies. This section of our paper concludes with the identification of the financial distress variables that we include in our analysis.

Schwartz and Menon (1985) provide early evidence of the association between bankruptcy and auditor changes in their study of 132 companies that filed bankruptcy between 1974 and 1982. As is the case with this and other bankruptcy studies, the sample firms include bankruptcy firms and matched firms that were not bankruptcy firms. Only 48 of the 264 sample firms changed auditors within the four years prior to the event date (bankruptcy filing for the bankruptcy firms), and 35 of these 48 were bankruptcy firms. Chi-square tests document a significant association between bankruptcy and auditor changes.

Chen et al. (2004) document the statistical significance of an auditor change variable in addition to all except one (Current assets-to-Sales) of their six financial statement ratio variables in a logistic regression model using a sample of 472 firms in bankruptcy proceedings during 1990 through 1998 and 424 matched non-bankruptcy firms. Of the 458 sample firms that changed auditors, 283 (62 percent) were bankruptcy firms, and Chi-square tests also confirm the significant association between bankruptcy and auditor changes previously documented by Schwartz and Menon (1985). Chen et al. (2004) included the following six financial statement ratio variables in their bankruptcy prediction logistic regression model: Cash-to-Total assets, Current assets-to-Current liabilities, Current assets-to-Sales, Current assets-to-Total assets, Long-term debt-to-Total assets, and Net income-to-Total assets. Chen et al. (2009) is similar to Chen et al. (2004) because they include an auditor change variable with a financial distress index variable in a logistic regression model for prediction of financial distress (bankruptcy) subsequent to the auditor change and both of these variables have statistically significant coefficients. There are two key differences between these two studies. First, Chen et al. (2009) has a small sample of firms (29 bankrupt and 58 healthy) listed on the Taiwan Securities Exchange during 1996 through 2001. The second, and possibly more important, difference is the use of a single financial distress index variable instead of using a portfolio of financial statement ratio variables. Chen et al.'s (2009) index variable uses coefficients derived from estimating Zmijewski's (1984) bankruptcy prediction model for their bankrupt firms in previous years to calculate a Z-score used in their logistic regression. The three financial ratios included in the Z-score estimation are Net income-to-Total assets, Total debt-to-Assets, and Current assets-to-Current liabilities. The use of an index instead of a portfolio of financial ratios may result in loss of information, and the index could be biased based on the time period used to estimate the coefficients. Current data mining techniques such as MCLP allow researchers to use more variables compared with traditional regression models.

The problem with the bankruptcy prediction models is that they are not designed to evaluate the effect of financial distress prior to auditor changes on the auditor change decision. Also, not all companies that experience financial distress end up in bankruptcy. Hudaib and Cooke (2005) investigate the possible influence of financial distress, managing director changes, and qualified audit opinions on auditor switching using a stratified sample of non-financial UK companies listed on the London Stock Exchange during the period from 1987 through 2001. Their Figure 1 (page 1715) illustrates their hypotheses that financial distress, managing director changes, and control variables may influence the auditor's opinion on financial statements and that the auditor's opinion in addition to these other variables may influence a company's decision to change auditors. Of the 4,176 firm-year observations in their sample, 223 involved auditor changes. They use a Z-score variable to capture financial distress (or health), and this variable is based on a model that incorporates financial ratios to estimate financial solvency (see page 1719). The results of their multivariate logistic regression analyses of auditor changes provide evidence that as financial distress increases (the Z-score variable decreases) companies are more likely to switch auditors. As with Chen et al. (2009), the use of a Z-score single variable for financial distress may not capture as much information about financial condition as a portfolio of financial statement ratios. So, we choose to include a portfolio of thirteen financial statement measures in our MCLP and other data mining methods to evaluate the predictive ability of financial distress measures for auditor change decisions.

Extant literature includes a variety of papers that examine aspects of auditor changes, but many of these use samples that include only auditor changes and do not incorporate a portfolio of financial condition variables. Calderon and Ofobike (2008) use CART methodology to evaluate factors (none of which are financial statement ratios) that influence whether auditor changes are client-initiated or auditor-initiated. Francis and Wilson (1988) test whether agency costs influence companies to change from a non-Big Eight to a Big Eight audit firm or vice versa, and Debt-to-Total assets is the only financial statement ratio they include in their explanatory variables. Davidson et al. (2006) test for effects of earnings management on the direction of auditor changes (Big-to-Small, Big-to-Big, etc.) and control for financial distress using the Altman Z-score; the coefficients for the Altman Z-score in the full models are not statistically significant. Landsman et al. (2009) focus on client risk (both financial and audit) and client misalignment characteristics of audit client portfolio management decisions by the top-tier (Big N) accounting firms in pre- and post-Enron periods. They include five financial statement ratios (Return on assets (ROA), Loss (equals one if ROA is negative), Debt-to-Assets, Cash-to-Assets, and (Inventory plus Receivables)-to-Assets) in the set of client risk measures for their multinomial logistic regression model, and the coefficients on all of these variables except for Debt-to-Assets are statistically significant in at least one of the four scenarios (combinations of pre- and post-Enron and lateral/upward and downward switches).

In our study, financial distress variables are focused on as a base-line or starting point for predicting auditor changes in a firm. Because bankruptcy is the extreme form of financial distress, we include bankruptcy prediction variables in our study to capture financial distress. We use Altman's (1968) and Ohlson's (1980) bankruptcy prediction variables and variables from studies mentioned above. We also include a dummy variable (DIV) to capture Lau's (1987) State 1 – Dividend Omission, an early state of financial distress. Lau (1987) uses five categories, from financial stability (state 0) to bankruptcy and liquidation (state 4), to classify the financial state of a firm. Our thirteen financial statement variables are as follows:

TL/TA	= Total Liabilities ÷ Total Assets (Ohlson (1980), Francis and Wilson (1988), Chen et al. (2004), and Landsman et al. (2009))
WCA/TA	= Working Capital ÷ Total Assets (Altman (1968) and Ohlson (1980))
CL/CA	= Total Current Liabilities ÷ Total Current Assets (Ohlson (1980) and Chen et al. (2004))
NI/TA	= Net Income ÷ Total Assets (Ohlson (1980), Chen et al. (2004), and Landsman et al. (2009))
FU/TL	= Funds from Operations ÷ Total Liabilities (Ohlson (1980))
LOSS	= 1 if a firm has loss in previous years; else LOSS=**0** (similar to Ohlson (1980) and Landsman et al. (2009))
DIV	= 1 if a firm did not pay dividend in a previous year; else DIV=0 (Lau (1987))
CREIN/TA	= Change in the ratio of receivables plus inventories to total assets (similar to Landsman et al. (2009))

RE/TA = Retained Earnings ÷ Total Assets (Altman (1968))
EBIT/TA = Earnings before Interest and Taxes ÷ Total Assets (Altman (1968))
MKV/TD = Market Value of Equity ÷ Book Value of Total Debt (Altman (1968))
SALE/TA = Sales ÷ Total Assets (Altman (1968))
SIZE = Log of Total Assets (similar to Ohlson (1980))

3. MODELS OF MULTIPLE CRITERIA LINEAR PROGRAMMING CLASSIFICATION

Since our study's focus is auditor change prediction, a general data mining approach can be used for our study. Kou et al. (2003) developed a multiple criteria linear programming (MCLP) data mining approach. We summarize the basic concept of the formulation by the following models.

Given a set of k variables about the auditor change or non-auditor change sample firms in database $\boldsymbol{a} = (a_1, \ldots, a_k)$, let $\boldsymbol{A}_i = (A_{i1}, \ldots, A_{ik})\ \boldsymbol{R}^k$ be the sample observations of data for the variables, where $i = 1, \ldots, n$ and n is the sample size. We want to determine the coefficients of the variables, denoted by $\boldsymbol{X} = (x_1, \ldots, x_k)^T$, and a boundary value of b to separate two classes: AC (Auditor change) and NAC (Non-auditor change); that is,

(a) $\boldsymbol{A}_i \boldsymbol{X} \le b, \boldsymbol{A}_i \in$ AC (Auditor change) and $\boldsymbol{A}_i \boldsymbol{X} > b, \boldsymbol{A}_i \in$ NAC (Non-auditor change).

Consider now two kinds of measurements for better separation of AC (Auditor change) and NAC (Non-auditor change) firms. Let α_i be the overlapping degree with respect to $\boldsymbol{A}_i$, and β_i be the distance from $\boldsymbol{A}_i$ to its adjusted boundary. In addition, we define α to be the maximum overlapping of two-class boundary for all cases $\boldsymbol{A}_i$ ($\alpha_i < \alpha$) and β to be the minimum distance for all cases $\boldsymbol{A}_i$ from its adjusted boundary ($\beta_i > \beta$). Our goal is to minimize the sum of α_i and maximize the sum of β_i simultaneously. By adding α_i into (a) above, we have:

(b) $\boldsymbol{A}_i \boldsymbol{X} \le b + \alpha_i, \boldsymbol{A}_i \in$ AC and $\boldsymbol{A}_i \boldsymbol{X} > b - \alpha_i, \boldsymbol{A}_i \in$ NAC.

However, by considering β_i, we can rewrite (b) as

(c) $\boldsymbol{A}_i \boldsymbol{X} = b + \alpha_i - \beta_i, \boldsymbol{A}_i \in$ AC and $\boldsymbol{A}_i \boldsymbol{X} = b - \alpha_i + \beta_i, \boldsymbol{A}_i \in$ NAC.

Our two-criteria linear programming model is stated as

(d) Minimize $\sum_i \alpha_i$ and Maximize $\sum_i \beta_i$

Subject to

$\boldsymbol{A}_i \boldsymbol{X} = b + \alpha_i - \beta_i, \boldsymbol{A}_i \in$ AC,

$\boldsymbol{A}_i \boldsymbol{X} = b - \alpha_i + \beta_i, \boldsymbol{A}_i \in$ NAC,

where $\boldsymbol{A}_i$ are given, $\boldsymbol{X}$ and b are unrestricted, α_i and $\beta_i \ge 0$ (Shi et al., 2001, p. 429).

Our solution method is to use the compromise solution approach (Shi and Yu, 1989; Shi, 2001) to reform the model (d) by systematically identifying the best trade-offs between $-\sum_i \alpha_i$ and $\sum_i \beta_i$. We assume the ideal value of $-\sum_i \alpha_i$ is $\alpha^* > 0$ and the ideal value of $\sum_i \beta_i$ is $\beta^* > 0$. Then, if $-\sum_i \alpha_i > \alpha^*$, we define the regret measure as $-d_\alpha^+ = \sum_i \alpha_i + \alpha^*$; otherwise, it is 0. If $-\sum_i \alpha_i < \alpha^*$, the regret measure is defined as $d_\alpha^- = \alpha^* + \sum_i \alpha_i$; otherwise, it is 0. Thus, we have (i) $\alpha^* + \sum_i \alpha_i = d_\alpha^- - d_\alpha^+$, (ii) $| \alpha^* + \sum_i \alpha_i | = d_\alpha^- - d_\alpha^+$, and (iii) $d_\alpha^-, -d_\alpha^+ \ge 0$. Similarly, we derive $\beta^* - \sum_i \beta_i = d_\beta^- - d_\beta^+$, $| \beta^* - \sum_i \beta_i | = d_\beta^- + d_\beta^+$, and $d_\beta^-, d_\beta^+ \ge 0$. Our MCLP model has been gradually evolved as

(e) Minimize $d_\alpha + d_\alpha^{+} + d_\beta^{-} + d_\beta^{+}$

Subject to

$\alpha^* + \sum_i \alpha_i = d_\alpha^{-} - d_\alpha^{+}$,

$\beta^* - \sum_i \beta_i = d_\beta^{-} - d_\beta^{+}$,

$\boldsymbol{A}_i \boldsymbol{X} = b + \alpha_i - \beta_i$, $\boldsymbol{A}_i \in$ AC,

$\boldsymbol{A}_i \boldsymbol{X} = b - \alpha_i + \beta_i$, $\boldsymbol{A}_i \in$ NAC,

where $\boldsymbol{A}_i$, α^*, and β^* are given, $\boldsymbol{X}$ and b are unrestricted, $\alpha_i, \beta_i, d_\alpha^{-}, d_\alpha^{+}, d_\beta^{-}, d_\beta^{+} \geq 0$.

The data separation of the MCLP classification method is determined by solving the above problem. Our version of the software to implement the MCLP method uses C++ language running on Linux platform (Kou and Shi, 2002). The reason for developing the Linux version of the MCLP classification software is that the majority of database vendors, such as IBM, are aggressively moving to Linux-based system development. This Linux version goes along with the trend of information technology.

Other classification methods

A variety of different classification methods are applied to our data to compare with our MCLP data mining approach. From methods included in Witten and Frank’s (2005) Weka Data Mining Workbench, we have chosen the following three classifiers: classification and regression tree (CART), linear logistic regression, and Bayesian network. CART can predict both continuous and categorical dependent attributes by building regression trees and discrete classes, respectively (Breiman et al. 1984). Linear logistic regression models the probability of occurrence of an event as a linear function of a set of predictor variables (Cessie and Houwelingen, 1992). Bayesian network is a probabilistic graphic model and can represent conditional independencies between variables (Weiss and Kulikowski, 1991).

Performance metrics

We use the five performance metrics described below to evaluate the quality of the four selected classification methods. A “positive” classification means that a company is classified as an Auditor Change company. A “negative” classification means that a company is classified as a Non-Auditor Change company. Actual Auditor Change companies are either correctly classified (True Positive) or incorrectly classified (False Negative), and actual Non-Auditor Change companies also are either correctly classified (True Negative) or incorrectly classified (False Positive).

- Overall accuracy: Accuracy is the percentage of correctly classified companies (Han and Kamber, 2006). It is one of the most widely used classification performance metric.

$$\boxed{\text{Overall Accuracy} = \frac{TN + TP}{TP + FP + FN + TN}.}$$

- True Positive (TP): TP is the number of correctly classified Auditor Change companies. The TP Rate measures how well a classifier can recognize Auditor Change companies. This rate is also called the sensitivity measure.

$$\boxed{\text{True Positive Rate (Sensitivity)} = \frac{TP}{TP + FN}.}$$

- False Positive (FP): FP is the number of Non-Auditor Change companies that are misclassified as Auditor Change companies. The FP Rate measures the percentage of Non- Auditor Change companies that were incorrectly classified.

$$\text{False Positive Rate} = \frac{FP}{FP+TN}.$$

- True Negative (TN): TN is the number of correctly classified Non-Auditor Change companies. The TN Rate measures how well a classifier can recognize Non-Auditor Change companies. This rate is also called the specificity measure.

$$\text{True Negative Rate (Specificity)} = \frac{TN}{TN+FP}.$$

- False Negative (FN): FN is the number of Auditor Change companies that are misclassified as Non-Auditor Change companies. The FN Rate measures the percentage of Auditor Change companies that were incorrectly classified.

$$\text{False Negative Rate} = \frac{FN}{FN+TP}.$$

4. SAMPLE DATA, VARIABLES, AND EMPIRICAL RESULTS

Because this is a classification study, we have a sample of companies that changed auditors in 2007 and 2008 and a sample of companies that did not change auditors, matched with auditor-change companies based on size (using total assets) and industry (using two-digit SIC codes). We identified over 790 firms that changed auditors in 2007 and 2008 using CompuStat's "Auditor" (AU) variable. Our study period of 2007 and 2008 is based on the following discussion of Sarbanes-Oxley Act of 2002 (SOX) implementation. As a result of SOX, the SEC (2003) amended Rule 2-02 of Regulation S-X to require the accountants auditing the annual financial statements to also attest to and report on management's assessment of its internal control effectiveness. The SEC (2004) required this initial attestation report be included with audited financial statements for fiscal years ending on or after November 15, 2004 for accelerated filers and for fiscal years ending on or after July 15, 2005 for non-accelerated filers. In 2005, the SEC (2005) extended the initial compliance date for non-accelerated filers to fiscal years ending on or after July 15, 2006. Fusco (2006) discusses the impact of SOX on trends in auditor changes and reports the following numbers of auditor changes each year during the period from 2002 through 2005: 1,224 in 2002, 1,467 in 2003, 1,736 in 2004, and 1,673 in 2005. To exclude the potential effects of the initial implementation of the SOX attestation requirement on auditor change decisions, our study period includes the post-SOX implementation years of 2007 and 2008. We eliminated auditor-change firms that had multiple auditor changes within the test period. Our final sample is composed of 790 auditor-change (experimental) and 1,132 non-auditor change (matching control) firm-year observations.

As discussed in Section 2 of this paper, our empirical tests are designed to evaluate the comparative quality of MCLP with three other classification methods to predict auditor changes using financial variables shown in prior research to be associated with financial distress or bankruptcy. These predictor variables are defined in Section 2, and descriptive statistics of these predictor variables are presented in Table 1. Panels A1, A2, and A3 show descriptive statistics for auditor change firms for 2007, 2008, and both years, respectively. Panels B1, B2, and B3 show descriptive statistics for matching control (non-auditor change) firms for 2007, 2008, and both years, respectively, and include the t-test statistics for tests of differences in means between the two groups of firms. The t-test statistics in all three panels indicate that means for auditor change firms are significantly different from means for non-auditor change firms for the following five (of thirteen) variables: WCA/TA, CL/CA, DIV, EBIT/TA, and SALES/TA. These results indicate that auditor change firms maintain significantly lower ratios of working capital to total asset and higher current liabilities relative to current assets, generate more negative earnings before interest and taxes as a percent of total assets and lower ratios of sales to total assets, and more often pay no dividends as compared with similar non-auditor change firms.

Table 1: Descriptive statistics of predictor variables for auditor change (Panels A1, A2, and A3) and non-auditor change (Panels B1, B2, and B3) sample firms for 2007, 2008, and both years combined

Panel A1: Auditor Change Firms - 2007

Variables	N	Mean	Std Dev	Min	Max
SIZE	521	2.26	1.09	-1.43	6.11
TL/TA	521	0.25	0.48	0.00	6.85
WCA/TA	521	0.16	0.70	-12.82	1.00
CL/CA	414	0.84	2.11	0.00	34.81
NI/TA	521	-0.13	2.04	-13.27	40.03
FU/TL	406	65.06	731.64	-160.78	13293.20
LOSS	522	0.37	0.48	0	1
DIV	522	0.69	0.47	0	1
CREIN/TA	521	-0.003	0.09	-0.46	0.55
RE/TA	521	-10.03	169.04	-3847.5	1.27
EBIT/TA	521	-0.15	0.88	-12.36	0.65
MKV/TD	374	277.13	2787.36	0.004	50113.10
SALE/TA	521	0.73	0.76	0.00	4.17

Panel B1: Non-Auditor Change Firms - 2007

Variables	N	Mean	Std Dev	Min	Max	t-value [(1)]
SIZE	729	2.27	0.80	-0.82	4.55	-0.301
TL/TA	699	0.23	0.34	0.00	3.64	0.900
WCA/TA	699	0.25	0.36	-3.65	0.98	-3.146***
CL/CA	688	0.65	0.77	0.01	8.96	1.771*
NI/TA	699	-0.08	0.45	-6.82	0.52	-0.555
FU/TL	545	106.87	721.59	-717.67	12972.20	-0.879
LOSS	757	0.38	0.48	0	1	-0.058
DIV	757	0.78	0.42	0	1	-3.531***
CREIN/TA	690	0.003	0.07	-0.57	0.41	-1.184
RE/TA	699	-1.34	5.45	-64.66	1.16	-1.360
EBIT/TA	698	-0.02	0.33	-4.18	0.50	-3.071***
MKV/TD	504	1015.29	11193.22	0.03	241785.40	-1.247
SALE/TA	699	0.91	0.76	0.00	5.83	-4.286***

[(1)] t-value for testing mean differences between auditor change and non-auditor change firms
*: $p < 0.10$
**: $p < 0.05$
***: $p < 0.01$

Panel A2: Auditor Change Firms - 2008

Variables	N	Mean	Std Dev	Min	Max
SIZE	269	2.46	1.07	-1.15	5.82
TL/TA	269	0.25	0.31	0.00	3.14
WCA/TA	269	0.13	0.44	-3.92	0.99
CL/CA	196	1.01	2.46	0.004	29.05
NI/TA	269	-0.40	2.57	-39.85	0.44
FU/TL	217	49.16	394.69	-567.95	4818.57
LOSS	269	0.32	0.47	0	1
DIV	269	0.62	0.49	0	1
CREIN/TA	269	-0.004	0.09	-0.45	0.62
RE/TA	269	-2.69	12.41	-142.84	0.84
EBIT/TA	269	-0.16	0.73	-7.57	0.35
MKV/TD	209	204.27	1640.78	0.00	21838.14
SALE/TA	269	0.72	0.93	0.00	8.61

Table 1 continued:

Panel B2: Non-Auditor Change Firms - 2008

Variables	N	Mean	Std Dev	Min	Max	t-value (1)
SIZE	363	2.48	0.84	-0.92	4.59	-0.311
TL/TA	339	0.25	0.31	0.00	3.23	-0.045
WCA/TA	339	0.23	0.29	-0.83	0.99	-3.338***
CL/CA	336	0.69	0.88	0.00	12.72	1.755*
NI/TA	339	-0.21	2.18	-39.85	0.96	-0.967
FU/TL	262	13.17	108.52	-686.67	1463.49	1.303
LOSS	375	0.35	0.48	0	1	-0.754
DIV	375	0.76	0.43	0	1	-3.862***
CREIN/TA	338	-0.004	0.07	-0.21	0.76	-0.115
RE/TA	339	-0.99	4.80	-74.38	1.53	-2.113**
EBIT/TA	339	-0.02	0.28	-2.38	0.45	-3.048***
MKV/TD	250	147.45	1292.48	0.00	19173.00	0.415
SALE/TA	339	0.90	0.78	0.00	5.88	-2.691***

(1) t-value for testing mean differences between auditor change and non-auditor change firms
*: $p < 0.10$
**: $p < 0.05$
***: $p < 0.01$

Panel A3: Auditor Change Firms – both 2007 and 2008 combined

Variables	N	Mean	Std Dev	Min	Max
SIZE	790	2.33	1.09	-1.43	6.11
TL/TA	790	0.25	0.43	0.00	6.85
WCA/TA	790	0.15	0.62	-12.82	1.00
CL/CA	610	0.90	2.23	0.00	34.81
NI/TA	790	-0.22	2.23	-39.85	40.03
FU/TL	623	59.52	634.59	-567.95	13293.20
LOSS	790	0.36	0.48	0.00	1.00
DIV	790	0.66	0.47	0.00	1.00
CREIN/TA	790	-0.003	0.09	-0.46	0.62
RE/TA	790	-7.53	137.47	-3747.60	1.27
EBIT/TA	790	-0.15	0.83	-12.36	0.65
MKV/TD	583	251.01	2437.77	0.004	50113.10
SALE/TA	790	0.72	0.82	0.00	8.61

Panel B3: Non-Auditor Change Firms – both 2007 and 2008 combined

Variables	N	Mean	Std Dev	Min	Max	t-value (1)
SIZE	1092	2.34	0.82	-0.92	4.59	-0.39
TL/TA	1038	0.24	0.33	0.00	3.64	0.76
WCA/TA	1038	0.25	0.34	-3.65	0.99	-4.00***
CL/CA	1024	0.66	0.80	0.00	12.72	2.48**
NI/TA	1038	-0.12	1.30	-39.85	0.96	-1.11
FU/TL	807	76.45	597.64	-717.70	12972.20	-1.59
LOSS	1132	0.37	0.48	0.00	1.00	-0.49
DIV	1132	0.77	0.42	0.00	1.00	-5.15***
CREIN/TA	1029	0.0007	0.07	-0.57	0.76	-1.01
RE/TA	1038	-1.22	5.25	-74.38	1.53	-1.29
EBIT/TA	1037	-0.02	0.31	-4.18	0.50	-4.20***
MKV/TD	754	727.54	9187.57	0.00	241785.40	-1.22
SALE/TA	1038	0.91	0.77	0.00	5.88	-4.99***

(1) t-value for testing mean differences between auditor change and non-auditor change firms
*: $p < 0.10$
**: $p < 0.05$
***: $p < 0.01$

Table 1 continued:

Variable Descriptions:

SIZE	= Log of Total Assets
TL/TA	= Total Liabilities ÷ Total Assets
WCA/TA	= Working Capital ÷ Total Assets
CL/CA	= Total Current Liabilities ÷ Total Current Assets
NI/TA	= Net Income ÷ Total Assets
FU/TL	= Funds from Operations ÷ Total Liabilities
LOSS	= 1 if a firm has loss in previous years; else LOSS=**0**
DIV	= 1 if a firm did not pay dividend in a previous year; else DIV=**0**
CREIN/TA	= Change in the ratio of receivables plus inventories to total assets
RE/TA	= Retained Earnings ÷ Total Assets
EBIT/TA	= Earnings before Interest and Taxes ÷ Total Assets
MKV/TD	= Market Value of Equity ÷ Book Value of Total Debt
SALE/TA	= Sales ÷ Total Assets

Table 2 presents our results using our MCLP and other data mining models. The Overall Accuracy rates in Panel C (both years combined) are very similar for the four models and range from 63.50 percent using the CART method to 60.17 percent using the Logistic method with 61.85 percent for the MCLP method. When Overall Accuracy rates in Panel A (2007) and Panel B (2008) are compared, MCLP's rate is the most stable (60.16 percent for 2007 and 60.88 percent for 2008), and the rates from the Logistic method are consistently the lowest (53.13 percent and 47.97 percent in 2007 and 2008, respectively). The True Positive and True Negative Rates in Panel C of Table 2 range from 80.28 percent (MCLP) to 95.26 percent (CART) and from 1.43 percent (BayesNet) to 27.76 percent (MCLP), respectively, and these rates indicate, as expected, that the MCLP and other data mining models are all much better at correctly classifying auditor change firms than non-auditor change firms using the financial-distress related classification variables. Because our primary interest in this study is to evaluate the usefulness of financial distress variables in classifying or predicting auditor changes, we focus on the False Negative Rates since these rates reflect auditor change firms incorrectly classified as non-auditor change firms. Although the MCLP method has the highest False Negative Rate of 19.72 percent in Panel C compared to the other data mining methods, the False Negative Rates for the MCLP are the most stable across the two years (21.43 percent in 2007 and 23.19 percent in 2008) compared to the rates in Panels A and B for the other three methods. These results support the usefulness of financial distress variables in identifying auditor change firms but also suggest that additional predictor variables should be considered to capture other characteristics that influence the auditor change decision and thereby to improve our prediction model accuracy.

Table 2: Classification Accuracy Results for 2007 (Panel A), 2008 (Panel B), and both years combined (Panel C)

Panel A - 2007	**Prediction Accuracy Rates**				
Method	Overall	True Positive	False Positive	True Negative	False Negative
MCLP	60.16	78.57	75.56	24.44	21.43
BayesNet	68.29	99.44	91.89	8.11	0.56
CART	65.84	97.87	95.89	4.11	2.13
Logistic	53.13	71.20	81.78	18.22	28.80
Panel B - 2008	**Prediction Accuracy Rates**				
Method	Overall	True Positive	False Positive	True Negative	False Negative
MCLP	60.88	76.81	66.67	33.33	23.19
BayesNet	60.16	94.03	98.33	1.67	5.97
CART	61.59	94.44	94.00	6.00	5.56
Logistic	47.97	56.11	66.00	34.00	43.89
Panel C - Both years	**Prediction Accuracy Rates**				
Method	Overall	True Positive	False Positive	True Negative	False Negative
MCLP	61.85	80.28	72.24	27.76	19.72
BayesNet	60.64	92.72	98.57	1.43	7.28
CART	63.50	95.26	95.33	4.67	4.74
Logistic	60.17	82.83	82.00	18.00	17.17

5. SUMMARY AND CONCLUSIONS

In this paper, we used the MCLP and other data mining approaches to evaluate the usefulness of financial distress variables in predicting auditor changes. Our study is designed to be a starting point for establishing a set of predictor variables to improve prediction of auditor changes. Our univariate statistics indicate that the auditor-change firms in our sample did exhibit lower liquidity (WCA/TA and CL/CA), lower profitability (EBIT/TA), and lower asset turnover (SALE/TA) than was exhibited by the non-auditor change firms and had more instances of not paying dividends in the preceding year. This is consistent with weaker financial conditions for auditor-change firms than for non-auditor change firms. Overall accuracy rates for all four data mining approaches are similar at around 60 percent for both years combined, and true positive rates range from 80 percent to 95 percent for both years combined. Because the true positive rate reflects the percentage of audit-change firms correctly classified, these high rates indicate the usefulness of financial distress variables in identifying audit-change firms. The logistic model has the lowest overall accuracy in each year and when both years are combined, and the MCLP model has the most stable overall accuracy rate and true positive rate. Investors and regulators may find our results helpful in their assessment of auditor changes for which no reason is disclosed.

Although our results provide evidence that a portfolio of financial distress variables is useful in predicting auditor changes, further research is needed to improve overall accuracy, true positive and true negative rates. The auditor change studies discussed in Section 2 of this paper include variables, such as qualified audit opinions, agency-cost measures, internal control weaknesses, and audit fees, that may also motivate firms to dismiss auditors or auditors to resign. A longer study period would also provide a better understanding of the stability of the various data mining methods over time.

AUTHOR INFORMATION

Wikil Kwak is a Professor of Accounting at the University of Nebraska at Omaha. He received his Ph.D. in Accounting from the University of Nebraska in Lincoln. Dr. Kwak's research interests include the areas of mathematical programming approaches in capital budgeting, transfer pricing, performance evaluation and Japanese capital market studies. He has published in the *Engineering Economist*, *Abacus*, *Contemporary Accounting Research*, *Review of Quantitative Finance and Accounting*, *Management Accountant*, *Journal of Petroleum Accounting and Financial Management*, *Business Intelligence and Data Mining, Review of Pacific Basin Financial Markets and Policies*, and *Multinational Business Review*. E-mail: wkwak@mail.unomaha.edu

Susan W. Eldridge is an Associate Professor of Accounting and Accounting Department Chair at the University of Nebraska at Omaha. She received her PhD from the University of North Carolina at Chapel Hill. Her research and teaching interests are in the financial accounting area. Her research has been published in the *Journal of Applied Business Research*, *Journal of Accounting, Ethics and Public Policy*, *Bank Accounting & Finance,* and *Review of Pacific Basin Financial Markets and Policies*. She is a Certified Public Accountant and a member of the American Accounting Association and the American Institute of CPAs.

Yong Shi is the director of Chinese Academy of Sciences Research Center on Data Technology & Knowledge Economy, President's Assistant of the Graduate University of Chinese Academy of Sciences. He had been the Charles W. and Margret H. Durham Distinguished Professor of Information Technology, College of Information Science and Technology, University of Nebraska at Omaha since1999. Dr. Shi's research interests covers data mining, information overload, multiple criteria decision making and telecommunication management. He has published seven books and more than sixty papers in various journals He is the Editor-in-Chief of *International Journal of Information Technology and Decision Making (SCI)* and an Area Editor of *International Journal of Operations and Quantitative Management.* Dr. Shi has received many distinguished awards including Outstanding Young Scientist Award, National Natural Science Foundation of China, 2001; Member of Overseas Assessor for the Chinese Academy of Sciences, May 2000; and Speaker of Distinguished Visitors Program (DVP) for 1997-2000, IEEE Computer Society. He has consulted in a number of famous companies.

Gang Kou is a professor of School of Management and Economics, University of Electric Science and Technology of China and managing editor of *International Journal of Information Technology & Decision Making*. Previously, he was a research scientist in Thomson Co., R&D. He received his Ph.D. in Information Technology from the College of Information Science & Technology, Univ. of Nebraska at Omaha. He got his master's degree in Computer science, University of Nebraska at Omaha, B.S. in Physics, Tsinghua University, and Beijing, China.

REFERENCES

1. Altman E., Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance* (September 1968), 589-609.
2. Breiman L., J. H. Friedman, R.A. Olshen, and C.J. Stone, *Classification and Regression Trees*. Wadsworth International Group, Belmont, California, 1984.
3. Calderon, T. and E. Ofobike, “Determinants of client-initiated and auditor-initiated auditor changes,” *Managerial Auditing Journal*, 23(1), 2008, 4 – 25.
4. Cessie S., and J. C. Houwelingen, Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191-201, 1992.
5. Chen, C., G. Yen, and F. Chang, “Strategic auditor switch and financial distress prediction – empirical findings from the TSE-listed firms,” *Applied Financial Economics* (2009), 59 – 72.
6. Chen, Y., A. Gupta, and D. Senteney, “Predicting Impending Bankruptcy Using Audit Firm Changes,” *The Journal of American Academy of Business* (March 2004), 423 – 433.
7. Davidson, W., P. Jiraporn, and P. DaDalt, “Causes and Consequences of Audit Shopping: An Analysis of Auditor Opinions, Earnings Management, and Auditor Changes,” *Quarterly Journal of Business and Economics* (Winter 2006), Vol. 45, Nos. 1/2, 69-87.
8. Francis, J. and E. Wilson, “Auditor Changes: A Joint Test of Theories Relating to Agency Costs and Auditor Differentiation,” *The Accounting Review* (October 1988). Vol. 63, No. 4, 663-682.
9. Fusco, C., “Is It Time To Revise 8-K Rules on Auditor Changes?,” *Financial Executive* (March 2006), 49 – 51.
10. Hackenbrack, K. and C. Hogan, “Market Response to Earnings Surprises Conditional on Reasons for an Auditor Change,” *Contemporary Accounting Research*, Vol. 19 No. 2 (Summer 2002), 195 – 223.
11. Han J. and M. Kamber, *Data Mining: Concepts and Techniques*. 2nd edition, Morgan Kaufmann, 2006.
12. Hudaib, M. and T. Cooke, “The Impact of Managing Director Changes and Financial Distress on Audit Qualification and Auditor Switching,” *Journal of Business Finance & Accounting*, 32(9) & (10), November/December 2005, 1703 – 1739.
13. Kou, G., X. Liu, Y. Peng, Y. Shi, M. Wise, and W. Xu, Multiple Criteria Linear Programming Approach to Data Mining: Models, Algorithm Designs and Software Development, *Optimization Methods and Software* 18, 453-473, 2003.
14. Kou, G. and Y. Shi, Linux based Multiple Linear Programming Classification Program: version 1.0, College of Information Science and Technology, University of Nebraska-Omaha, Omaha, NE 68182, USA, 2002.
15. Landsman, W., K. Nelson, and B. Rountree, “Auditor Switches in the Pre- and Post-Enron Eras: Risk or Realignment?,” *The Accounting Review*, Vol. 84, No. 2 (March 2009), 531 – 558.
16. Lau, A., “A Five-State Financial Distress Prediction Model,” *Journal of Accounting Research*, Vol. 25, No. 1 (Spring 1987), 127 – 138.
17. Ohlson J., Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research* 18(1), 109-131, 1980.
18. Sarbanes-Oxley Act (SOX). 2002. Public Law No. 107-204. Washington, D.C.: Government Printing Office.
19. Schwartz, K. and K. Menon, “Auditor Switches by Failing Firms,” *The Accounting Review*, Vol. LX, No. 2 (April 1985), 248-261.
20. Securities and Exchange Commission (SEC). Final Rule: Management’s Reports on Internal Control Over Financial Reporting and Certification of Disclosure in Exchange Act Periodic Reports. Release No. 33-8238. Washington, D.C.: SEC, 2003. Available at: http://www.sec.gov/rules/final/33-8238.htm.

21. Securities and Exchange Commission (SEC). Final Rule: Management's Reports on Internal Control Over Financial Reporting and Certification of Disclosure in Exchange Act Periodic Reports. Release No. 33-8392. Washington, D.C.: SEC, 2004. Available at: http://www.sec.gov/rules/final/33-8392.htm.
22. Securities and Exchange Commission (SEC). Final Rule: Management's Reports on Internal Control Over Financial Reporting and Certification of Disclosure in Exchange Act Periodic Reports of Non-accelerated Filers and Foreign Private Issuers. Release No. 33-8545. Washington, D.C.: SEC, 2005. Available at: http://www.sec.gov/rules/final/33-8545.htm.
23. Shi, Y., *Multiple Criteria multiple Constraint-Levels Linear Programming: Concepts, Techniques and Applications*, World Scientific Publishing, River Edge, New Jersey, 2001.
24. Shi, Y., and P. L. Yu, Goal Setting and Compromise Solutions, In B. Karpak and S. Zionts (Eds.), *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers*, 165-204, Springer-Verlag, Berlin, 1989.
25. Turner, L., J. Williams, and T. Weirich, "An Inside Look at Auditor Changes," *The CPA Journal* (November 2005), 12 – 21.
26. Witten I. H., and E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
27. Weiss S. M., and C. A. Kulikowski, *Computer Systems that Learn: Classification and Predication Methods from Statistics, Neural Nets, Machine Learning and Expert Systems*, Morgan Kaufmann, 1991

