Identifying Material Weakness of Internal Control: An Empirical Study for a Multi-year Period

John J. Cheh, The University of Akron, USA

Il-woon Kim*, The University of Akron, USA

Juliana Reifsnyder, The University of Akron, USA

ABSTRACT

According to the Sarbanes-Oxley Act of 2002, the management of publicly-traded companies is required to assess their companies' internal control material weaknesses and provide assertions of effective internal control over financial reporting. Each company's independent audit firm is required to attest these assertions. Even though the Act helped improve the quality and transparency of financial reports, many publicly-traded companies have been concerned about the increase in audit fees charged by external auditors, questioning the practical value of the Act. The main objective of this paper is to provide evidence that data mining can be used as a powerful tool to identify those companies with material weaknesses in internal control. The decision tree model, in particular, reveals decision rules that can be used by auditors to assess whether a company under audit is likely to have weak internal controls. This information will help auditors not only enhance the effectiveness of the audit but also reduce its cost so that the value of the Act can be enhanced.

Key words: internal control, material weakness, Sarbanes-Oxley Act

JEL: 41

* Corresponding author. Email:ikim@uakron.edu

1. Introduction

Publicly listed companies have been required to maintain an effective internal control system since the introduction of the Securities Exchange Act of 1934, but the area of risk assessment did not come under scrutiny until the passage of Section 404 of the Sarbanes-Oxley Act (SOX) of 2002 (Knechtel et al., 2006). The Act not only requires management of publicly traded companies to assess the material weakness (MW) of their internal control over financial reporting and to provide an annual report to stockholders and regulators, but also mandates external auditors to attest to the management's assessment and to report directly on its effectiveness. SOX has increased the accountability of both management and external auditors by requiring all public companies to maintain accurate records and an adequate system of internal accounting control and by dramatically increasing the penalties for false financial reporting.

Since its passage, however, SOX has brought about both positive and negative implications for companies and investors. By helping companies restore a healthy balance between performance and understand the risks and controls, the Act has helped companies gain confidence from the investors on financial reporting (Goodburn, 2011). Furthermore, it has provided unexpected benefits within the organization by improving processes, broadening employee's job responsibilities, eliminating duplicate activities, and automating manual controls (Bradford et al., 2010). These benefits have been accompanied by an increasing cost of compliance, a major concern for companies. According to O'Sullivan, the Act spurred a 58% increase in the fees charged by external auditors (2006). In addition, the Financial Executive International (FEI) surveyed public companies and found that compliance with section 404 of SOX resulted in massive cost increase. The FEI also estimated that the total cost of compliance would cost companies over $20 billion over the four years following the Act's passage in 2002. For every $1 billion in revenues, AMR Research estimates that companies are spending $1 million on SOX compliance. Bradford, et al. summarize the consequences clearly: "Because SOX section 404 significantly increased the depth and breadth of internal control evident that auditors must evaluate, the Act quickly became one of the most costly accounting-related laws in history" (2010). In his recent article evaluating the success of SOX for last 10 years, Verschoor (2012, p. 14) argues that, even though SOX is widely credited for strengthening at least two major areas of investor protection (i.e., accountability and professionalism), some continue to question its overall value, citing, as an example, its failure to prevent the situations that led to the financial crisis of 2008.

The major objective of this study is to use key financial ratios and a data mining algorithm to find those companies that have a material weakness. Once identified, these financial characteristics of these companies can be used by external auditors as an initial screening device in identifying companies with MW in internal control. If the financial characteristics of a company under audit exhibit similar characteristics of the companies as demonstrated in this study, it may be an indication that the company has a higher probability of internal control risk and therefore should be subject to a detailed audit. A less-comprehensive audit procedure may be warranted for the other companies that do not demonstrate those financial characteristics. To auditors, this type of statistical pattern recognition of data mining can provide valuable insights in designing cost-effective and high-quality audits.

Previous literature will be reviewed in the following section. This section will be followed by a brief description of data mining, and then research methodology will be presented. These sections are followed by the analyses of data and discussions on the findings with a concluding remark.

2. Previous Research

During the last few years, several research papers have been published on identifying variables affecting Sarbanes-Oxley Act internal control MW using different variables and different methodologies. Studies by Doyle et al. (2007) and Ogneva et al. (2007) used a traditional method of theorizing the situations and bringing theory-based variables into the research. Doyle et al. in particular, made an attempt to identify the determinants of weaknesses in internal control by examining 779 firms that disclosed material weaknesses from August 2002 to 2005. Eleven variables were used in their study based on two major categories: entity-wide vs. accounting-specific. The study by Ogneva et al. focused on the cost of equity that would be impacted by the deterioration of information quality from the weaknesses in internal control. Ten variables which represented the firm characteristics associated with internal control weaknesses were used in their study. Cheh et al. (2009) used a trial and error approach with different combinations of many variables, and 23 financial variables were finally identified which gave the highest predictive rate for internal control MW. A subsequent study by Cheh et al. (2010) examined a number of the independent variables used in three previous studies and constructed a highly predictive model for internal control weaknesses.

These four studies dealt with a similar topic, but each study had a slightly different research objective. Doyle at al. (2007), for example, strived to find some characteristics of MW companies. Ogneva et al. (2007) were more interested in examining the association between cost of capital and internal control weaknesses. On the other hand, the aim of the study of Cheh et al. (2009) was to find data mining rules that facilitate the identification of MW companies. Cheh at al. (2010) focused on identifying the financial and nonfinancial variables which would be useful in predicting SOX internal control weakness.

Several papers examined some issues related to Sarbanes-Oxley Act and internal control. None of them, however, appeared to deal with ex-ante factors that might have determined the weakness in internal control. In general, they have addressed the effects after the disclosure of internal control weakness. For example, Kim (2008) tested how the disclosure of the weakness in internal control would affect the loan contract, and Li et al. (2008) examined what role remediation of internal control weakness would play in internal and external corporate governance. Tang and Xu (2008) investigated how institutional ownership would affect the disclosure of internal control weakness, and Xu and Tang (2008) studied the relationship between financial analysts' earnings forecasts and internal control weakness. Patterson and Smith (2007) did a theoretical investigation on the effects of the Sarbanes-Oxley Act of 2002 on auditing intensity and internal control strength, but did not address the issues on the determinants of internal control weakness.

3. Research Methodology

3.1. Data Mining and Decision Tree Model

Data mining is an integral part of knowledge discovery from large databases. Data mining techniques can be used to find useful patterns and trends that might otherwise remain unknown within the data. Hence, data mining has been widely used in business to find useful patterns and trends that might otherwise remain unknown within the business data. It has been applied for a long time in the marketing area of business, such as targeted marketing, store layout, customer profitability, and customer relationships. In the financial community, banks are among the earliest adopters of data mining for fraud detection, and credit card issuers use data mining software to detect credit fraud. Data mining has been extensively used to model the stock market using neural networks for financial gain (Groth 2000). Altman's bankruptcy prediction model—Z score model—has been widely used in industry, and became a benchmark for assessing whether a company may go into bankruptcy (1968; 2007).

Data mining, which is often called analytics, has also often been used for academic research. There are numerous conceptual models in data mining, such as decision trees, classification and rules, clustering, instances-based learning, linear regression, and Bayesian networks (Whitten and Frank, 2005). Since one of the main objectives of our research was to find tree-like rules that would help audit practitioners make informed decisions about MW, we decided to use the decision tree model. As a graph of decisions, the model uses a decision tree as a predictive model which maps the variables about an item to the conclusion about the item's target value.

3.2. Data Mining Process

Data mining is the overall process of converting raw data into useful information. Even though there are many different ways to implement a data mining process, two approaches have been widely used in the business world; supervised learning approach and unsupervised learning approach. Unsupervised learning (or clustering) is the process of determining patterns among random variables by fitting the pattern to the data; thus, no hypotheses or independent variables are required. This process uses the fitted pattern to accumulate similar observations into groups or clusters. This approach has been widely used in marketing and retailing to determine relationships between products that are purchased together. Supervised learning or classification is the process of dividing a data set into the groups that are already defined. Hence, the target outcomes or target variables should be clearly defined first. The result can help to not only understand the existing data but also to find predictive patterns. This approach has been widely used in the financial community; for example, in Altman's bankruptcy prediction model previously discussed, the target variable becomes whether or not a company will go bankrupt. The supervised learning approach will be used in this study because the target variable of the study is known: material weakness.

In this study, the following commonly used data mining steps in business will be used for transforming financial data into actionable guidelines, as suggested by Berry and Linoff (2000, p. 48):

- Identify the financial variables relevant for this study;
- Collect the data for the variables for MW companies and non-MW companies;
- Clean the data for data mining;
- Build and train SAS EM's decision tree model; and
- Test the decision model and analyze the findings

3.3. Variable Identification and Data Collection

The data on the target variable for this study (i.e., material weakness) were collected from Audit Analytics and Research Insight's Compustat for the study period from 2004 to 2010. Audit Analytics produced 932 companies from this study period with weaknesses in internal control. For non-MW companies, we obtained 19,933 companies during the same period from Research Insight's Compustat. Hence, the total number of companies in the entire data set was 20,865, including both MW and non-MW companies. Initially, Compustat provided a far larger data set of 49,545 companies; however, this data set included values such as @CF, @IF, @NA, or 0 that could not be processed, so this dirty data set needed to be processed for cleaning. Companies with any values that could not be processed were eliminated from the data set. All companies in our data set have the fiscal year ending December 31 of each year. The data cleaning process produced a final, clean data set of 20,865 companies.

Since Audit Analytics does not provide financial data for a large number of variables required in this study, we obtained the data for financial variables of these MW companies and non-MW companies from the Compustat database of Research Insight. Basically, all 23 financial ratios available in Compustat were used. No attempt was made to identify relevant ratios since this was previously done by Ge and McVay (2005), who used a multivariate regression model in their study on material weakness. This type of prediction modeling based on a prior notion of relationships between MW and independent variables can cause important variables to be overlooked. Some variables may not be theoretically relevant, but they could be significant in actual relationships among the variables. In this study, therefore, we used typical financial ratios available in Compustat that would affect the efficiency, profitability, liquidity, leverage, and solvency of the companies under study. These ratios along with their Compustat mnemonics are shown in Exhibit 1.

Exhibit 1

Financial Ratios and Compustat Mnemonics

Financial Ratios	**Column Name with Compustat Mnemonics**
Net Income/Sales	NI/SALE
Net Income/Total Assets	NI/AT
Net Income/Net Worth	NI/(AT-LT)
Earnings Before Income Tax/Total Assets	EBIT/AT
Sales/Total Assets	SALE/AT
Current Assets/Total Assets	ACT/AT
Quick Assets/Total Assets	(ACT-INVT)/AT
Sales/Quick Assets	SALE/(ACT-INVT)
Current Assets/Sales	ACT/SALE
Inventory/Sales	INVT/SALE
Cost of Goods Sold/Inventory	COGS/INVT
Total Liabilities/Total Assets	LT/AT
Quick Assets/Sales	(ACT-INVT)/SALE
Retained Earnings/Inventory	RE/INVT
Current Assets/Current Liabilities	ACT/LCT
Quick Assets/Current Liabilities	(ACT-INVT)/LCT
Current Liabilities/Total Assets	LCT/AT
Sales/Current Liabilities	SALE/LCT
Cash/Current Liabilities	CH/LCT
Working Capital/Sales	WCAP/SALE

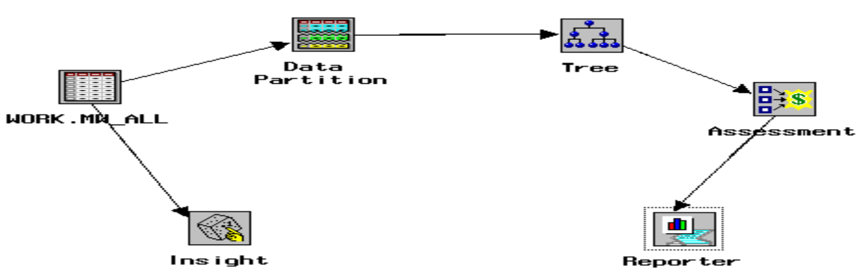
Retained Earnings/Total Assets	RE/AT
Cash/Total Assets	CH/AT
Total Liabilities/Net Worth	LT/(AT-LT)

3.4. Data Cleaning and Preparation

To reduce a biased result, all rows with null values were deleted. For data cleaning and preparation, SAS Enterprise Miner (EM) decision making process was used. SAS EM is a well-recognized industrial-strength decision making tool available in the market. Exhibit 2 shows the graphical representation of SAS EM, and the process is explained in detail in the subsequent section.

Exhibit 2

Graphical Representation of SAS Enterprise Miner's DM Process for All Years



3.4.Building and Training Decision Tree Model

As shown in Exhibit 2, the decision making process started with Insight icon and Input Data Source (i.e., WORK.DATA) icon. The source data file was selected using WORK.DATA. The next step was to designate MW as the target variable, which indicates whether or not each sample company had a material weakness using a binomial. Once this process was completed, the Insight icon was used to indicate that the entire data set would constitute our study sample. In the Data Partition icon, we decided to allocate the sample data, using the partition ratio of 40 percent to training, 30 percent to validation, and 30 percent to testing. SAS EM uses the validation set to refine the trained model for better classification or prediction; thus, validation process can be used as part of a typical training process in other data mining software. Therefore, the training data set in SAS EM can be considered as a pre-training data set, and the validation data set as a data set for fine tuning to refine the model developed in the pre-training stage.

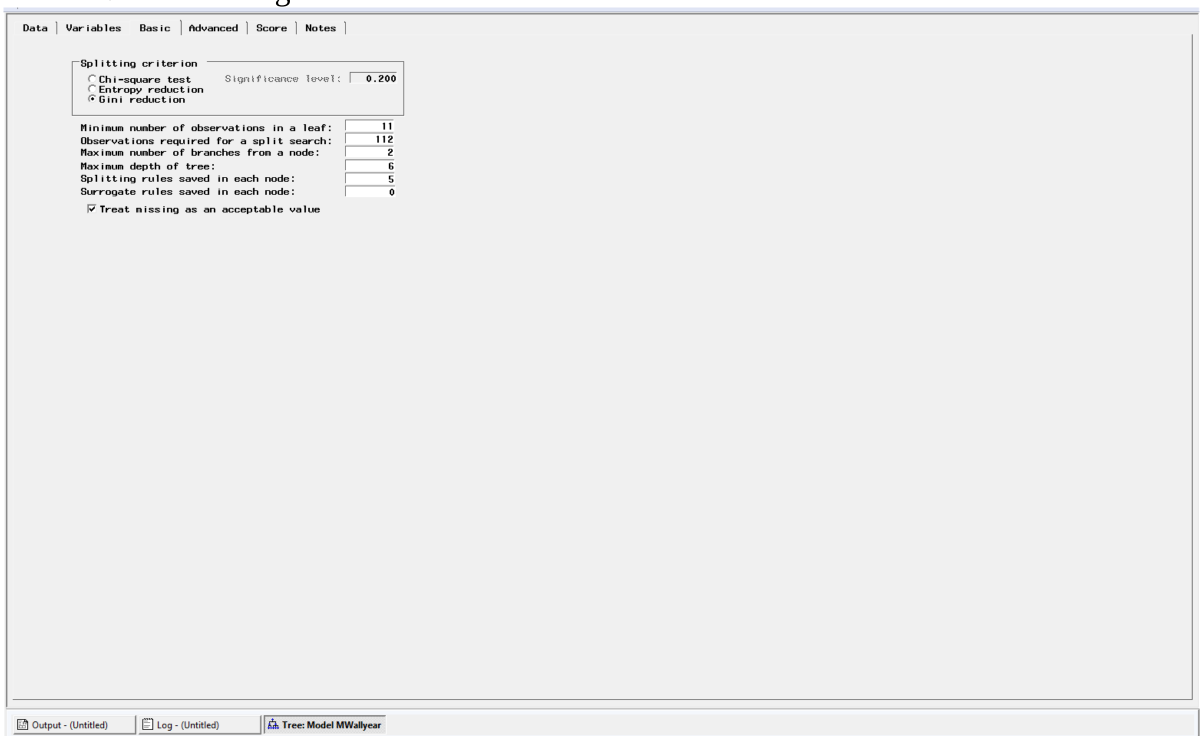
This process of testing in two additional stages, called validation and testing, reduces the errors that may result from using data from the same year for both training and testing in a typical case of data mining. Partitioning the data into three groups consisting of 40 percent training, 30 percent validation, and 30 percent testing is a reasonable allocation, since more than 50 percent of the data were used for validating and testing, while less than a half of the data were used for training. Different combinations of percentages among these three groups will not significantly affect the findings; the results will still show that the financial ratios employed in this paper are useful in recognizing companies with material weaknesses.

There are many different types of data mining algorithms. Many of these machine learning algorithms were originated from artificial intelligence in computer science. Among this variety of algorithms available in SAS EM, we decided to use the decision tree model. This model, often called Classification And Regression Trees or CART, is simple to use and easy to understand (Breiman et al. 1998). The rationale behind the selection of the decision tree model was the model's ability to produce simple, IF-THEN-type decision rules that proved useful for profiling material weakness companies.

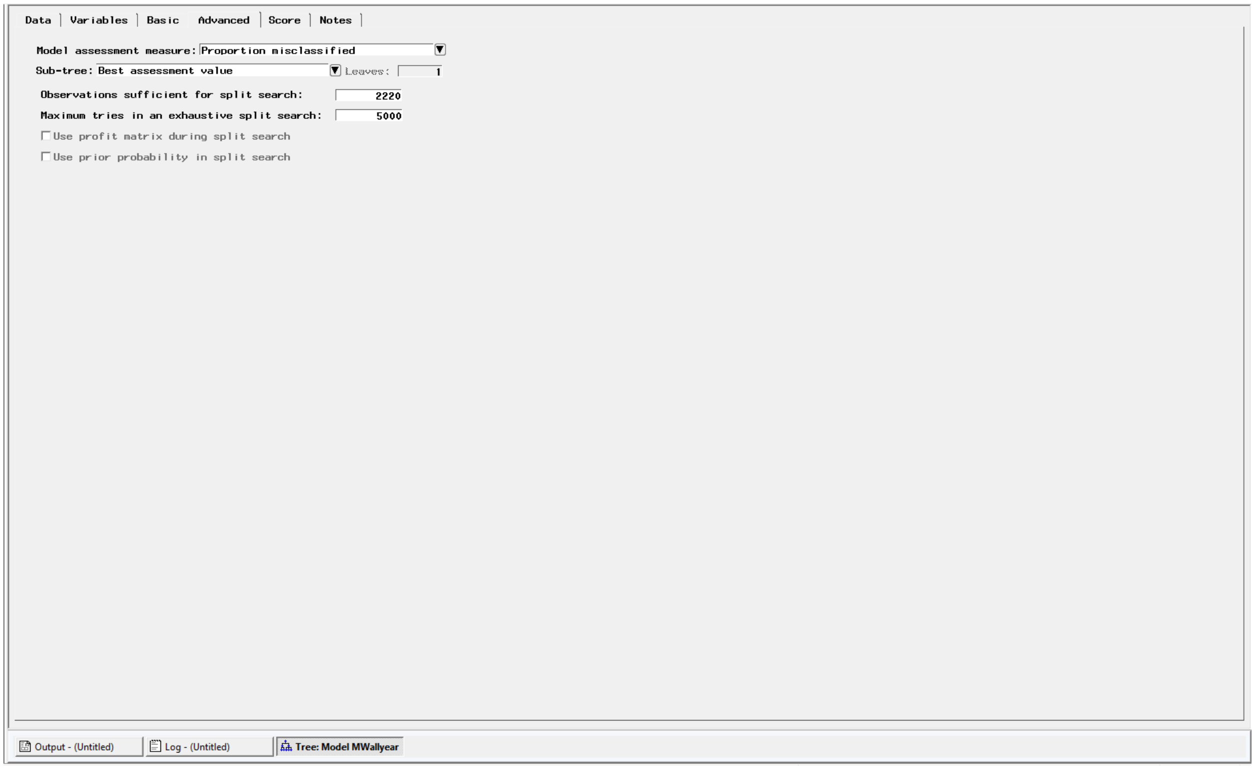
Exhibit 3

A Graphical Representation of Options Used for Building MW Profiling Decision Model in SAS Enterprise Miner

Panel A: Basic Configuration



Panel B: Advanced Configuration



As observed in Panels A and B of Exhibit 3, we chose to use the Gini reduction method in the SAS EM decision tree model and used the numerical configurations shown in the graph for various decision tree options. Once the configuration process was completed, the next step was to run the model using the Tree icon. Later, the Assessment icon and Reporter icon were run to complete the finding report; although these two additional icons were useful, it turned out that the results necessary for this study came mostly from the Tree icon.

4. Findings and Discussions

Exhibit 4 provides our findings on the profiling accuracy and misclassification errors, and Exhibit 5 a graphical representation of data mining results. The Exhibits show the results for two opposite target events each year: MW and non-MW. The order of MW variable was set ascending in the Class Variables of WORK.DATA; this produced the results for when the target event equals zero in SAS EM Assessment. When the target event equals zero, SAS EM is looking for profiling non-MW companies, and the misclassification error rate tells the auditors that SAM EM failed to

Exhibit 4

Data Mining Model Comparison for Profiling Accuracy and Misclassification Errors

Panel A: For the Entire Study Period

Data Mining Model: Decision Tree	Training		Validation		Testing	
	Profiling Accuracy	Misclassification Error*	Profiling Accuracy	Misclassification Error	Profiling Accuracy	Misclassification Error
Target Event: Non-MW Companies	95.410%	4.590%	95.469%	4.531%	95.068%	4.932%
Target Event: MW Companies	95.410%	4.590%	95.469%	4.531%	95.068%	4.932%

* Type I error occurs when SAS EM makes an error in rejecting the null hypothesis that there is no difference in two sets of samples: MW and non-MW. In other words, if SAS EM recognized non-MW companies as MW companies, then the error rate becomes Type I error rate. Thus, misclassification error rate becomes Type I error when target event is non-MW, because SAS EM could not recognize non-MW companies when in fact they were MW companies. This applies to three processes of training, validation, and testing; in this case, SAS EM failed to recognize MW companies when in fact they were MW companies. On the other hand, Type II error occurs when SAS EM failed to reject the null hypothesis. In other words, SAS EM recognized MW companies as non-MW companies, despite the fact that they were in fact MW companies.

Panel B: Each Year Separately

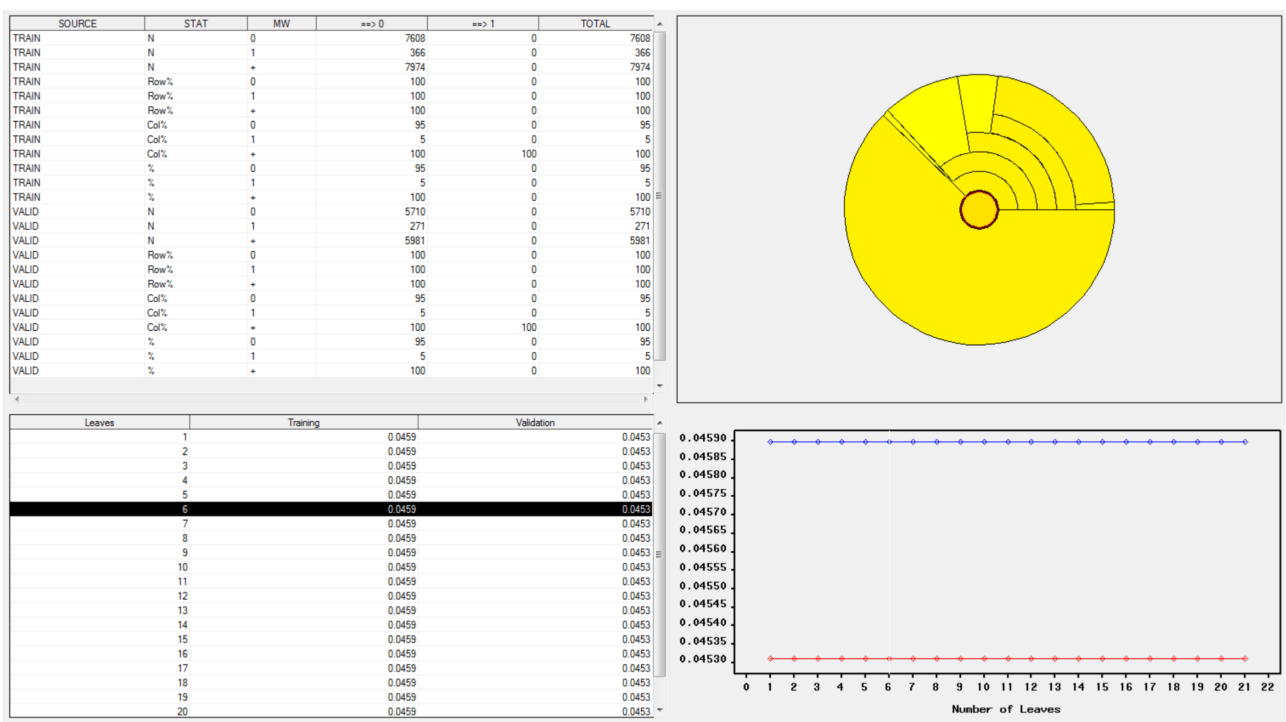
Data Mining Model: Decision Tree	Fiscal Year	Training		Validation		Testing	
		Profiling Accuracy	Misclassification Error	Profiling Accuracy	Misclassification Error	Profiling Accuracy	Misclassification Error
Target Event: Non-MW Companies	2004	93.60%	6.40%	95.54%	4.46%	93.83%	6.17%
Target Event: MW Companies		93.60%	6.40%	95.54%	4.46%	93.83%	6.17%
Target Event: Non-MW Companies	2005	91.93%	8.07%	91.61%	8.39%	94.49%	5.51%
Target Event: MW Companies		91.93%	8.07%	91.61%	8.39%	94.49%	5.51%
Target Event: Non-MW Companies	2006	93.45%	6.55%	93.78%	6.22%	94.02%	5.98%
Target Event: MW Companies		93.45%	6.55%	93.78%	6.22%	94.02%	5.98%
Target Event: Non-MW Companies	2007	94.88%	5.12%	94.10%	5.90%	95.37%	4.63%
Target Event: MW Companies		94.88%	5.12%	94.10%	5.90%	95.37%	4.63%
Target Event: Non-MW Companies	2008	96.24%	3.76%	95.53%	4.47%	96.41%	3.59%

Target Event: MW Companies		96.24%	3.76%	95.53%	4.47%	96.41%	3.59%
Target Event: Non-MW Companies	2009	96.95%	3.05%	97.32%	2.68%	97.64%	2.36%
Target Event: MW Companies		96.95%	3.05%	97.32%	2.68%	97.64%	2.36%
Target Event: Non-MW Companies	2010	97.84%	2.16%	98.50%	1.50%	97.58%	2.42%
Target Event: MW Companies		97.84%	2.16%	98.50%	1.50%	97.58%	2.42%

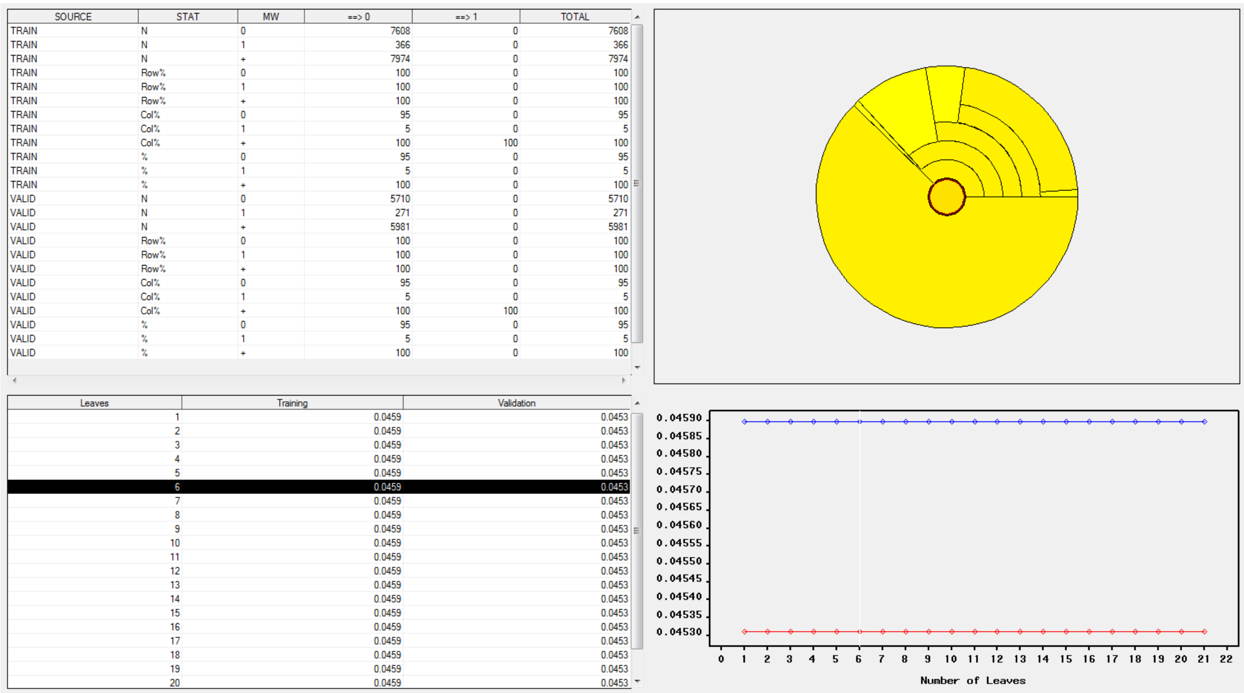
Exhibit 5

A Graphical Representation of Data Mining Results

Panel A: When Target Event is MW for Years 2004-2010



Panel B: When Target Event is Non-MW for Years 2004-2010



SOURCE	STAT	MW	==> 0	==> 1	TOTAL
TRAIN	N	0	7608	0	7608
TRAIN	N	1	366	0	366
TRAIN	N	+	7974	0	7974
TRAIN	Row%	0	100	0	100
TRAIN	Row%	1	100	0	100
TRAIN	Row%	+	100	0	100
TRAIN	Col%	0	95	0	95
TRAIN	Col%	1	5	0	5
TRAIN	Col%	+	100	100	100
TRAIN	%	0	95	0	95
TRAIN	%	1	5	0	5
TRAIN	%	+	100	0	100
VALID	N	0	5710	0	5710
VALID	N	1	271	0	271
VALID	N	+	5981	0	5981
VALID	Row%	0	100	0	100
VALID	Row%	1	100	0	100
VALID	Row%	+	100	0	100
VALID	Col%	0	95	0	95
VALID	Col%	1	5	0	5
VALID	Col%	+	100	100	100
VALID	%	0	95	0	95
VALID	%	1	5	0	5
VALID	%	+	100	0	100

Leaves	Training	Validation
1	0.0459	0.0453
2	0.0459	0.0453
3	0.0459	0.0453
4	0.0459	0.0453
5	0.0459	0.0453
6	0.0459	0.0453
7	0.0459	0.0453
8	0.0459	0.0453
9	0.0459	0.0453
10	0.0459	0.0453
11	0.0459	0.0453
12	0.0459	0.0453
13	0.0459	0.0453
14	0.0459	0.0453
15	0.0459	0.0453
16	0.0459	0.0453
17	0.0459	0.0453
18	0.0459	0.0453
19	0.0459	0.0453
20	0.0459	0.0453

recognize non-MW companies. Depending on how auditors set the order of the target variable, SAS EM produces a different set of results. In this study, the order of MW variable was set descending in the Class Variables of WORK.DATA; this produced the results for when the target event equals one in SAS EM Assessment. When the target event equals one, SAS EM is looking for profiling MW companies and the misclassification error rate tells the auditors that SAM EM failed to recognize MW companies.

The overall results as seen in Panel A of Exhibit 4 are, when the target event is non-MW, SAS EM analysis demonstrates that the profiling accuracy is 95.41 percent in the training process and 95.47 percent in the validation process, and the misclassification rate is reduced from 4.59 percent to 4.53 percent in the testing process. The additional reduction might have been resulted from using the same year's data. That is, the data mining software had probably memorized them, reducing the error rate by a small fractional percentage. It is surprising to see that, when the target event is MW, the results are identical. The breakdown of the results for each year is presented in Panel B, and the results are also identical each year during the study period for both MW and non-MW events. The identical results are explained in the following paragraph.

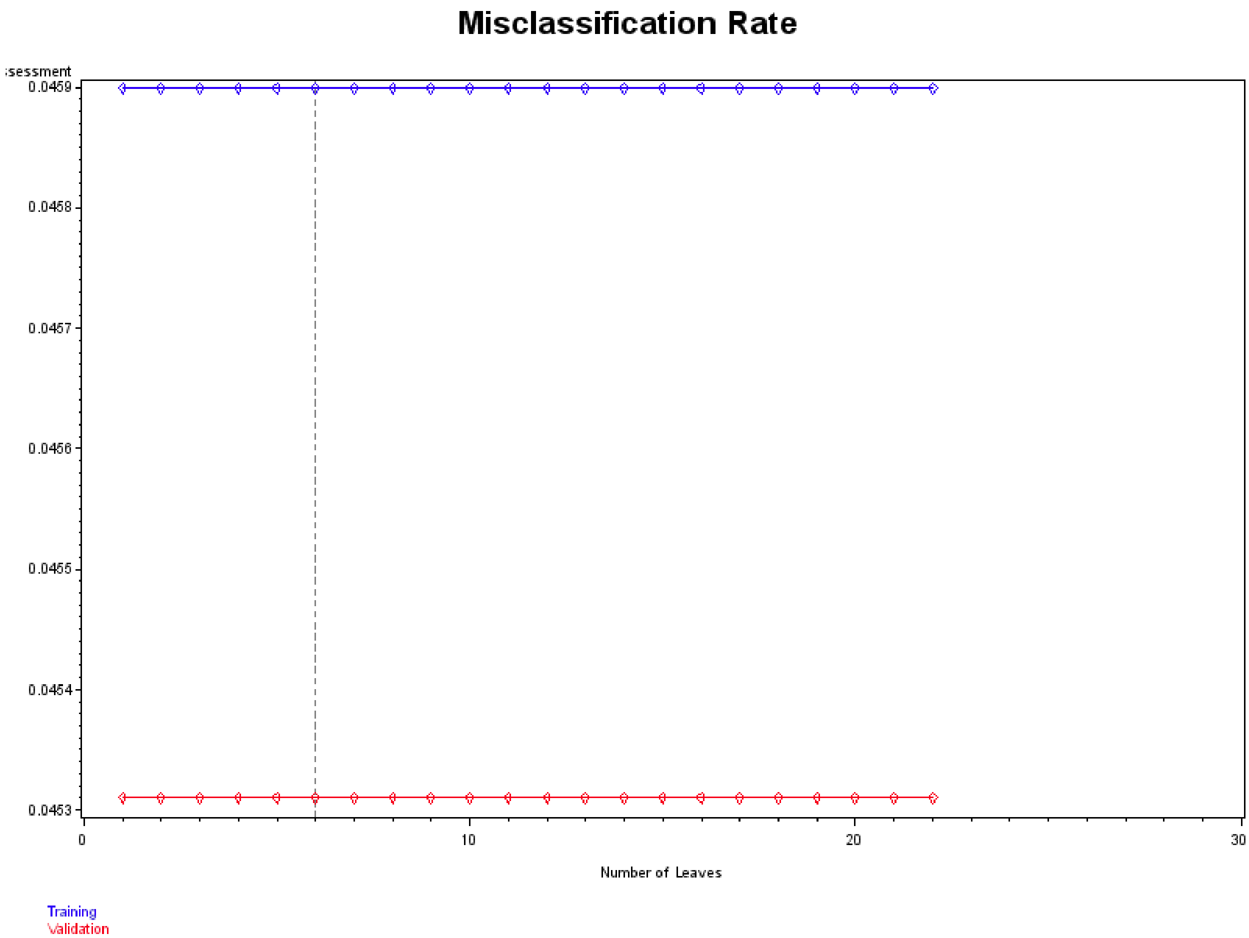
These results are quite promising. They show convincing evidence that data mining can be an effective way of aiding auditors. As auditors are now interested in reducing audit risk and, at the same time, saving the cost of audit, data mining can be an effective tool to achieve the audit objectives. Practitioners, however, may be somewhat cautious of the reported findings when non-MW companies are the target event. When creating a new data set in Enterprise Miner, SAS gives the option to set the level of prior probabilities for the target event. These values represent the proportions of observations in the data set for which the target event is equal to 1 or 0. The default setting for these prior probabilities is 0.5, or exactly half of the data. In this setting, we are assuming that the prior probabilities of Type I and Type II errors as explained in the footnote of Exhibit 4 Panel A are identical. Therefore, SAS Enterprise Miner calculates misclassification errors of not being able to recognize MW companies and non-MW companies on an equal proportion, which may have caused to produce the identical results. In addition, one assumption that we have made in this segment of our study is that the companies not flagged in Audit Analytics as having material weaknesses do not have material weaknesses for a given year. Yet, the material weaknesses latent in that particular year may surface or be uncovered in later years. Making this assumption in studying our sample companies may be another reason that the results for both target events are found to be identical.

An improvement for the profiling accuracy can be achieved by stratifying the data into various segments of industries and different sizes of the sample companies. For example, Altman (1968) examined the sample companies only in the manufacturing sector and eliminated both small and very large companies from the sample data set. This kind of trimming process is often known to improve the classification accuracy. Another way to improve the results is to extend the study period into multiple years. Additional tests with more data can provide a solid foundation for confirming the stability of the results of this study. As shown in Panels A and B of Exhibit 5, the validation process does not change with the addition of new leaves into the decision tree. No matter how many leaves are included, the misclassification rate remains the same.

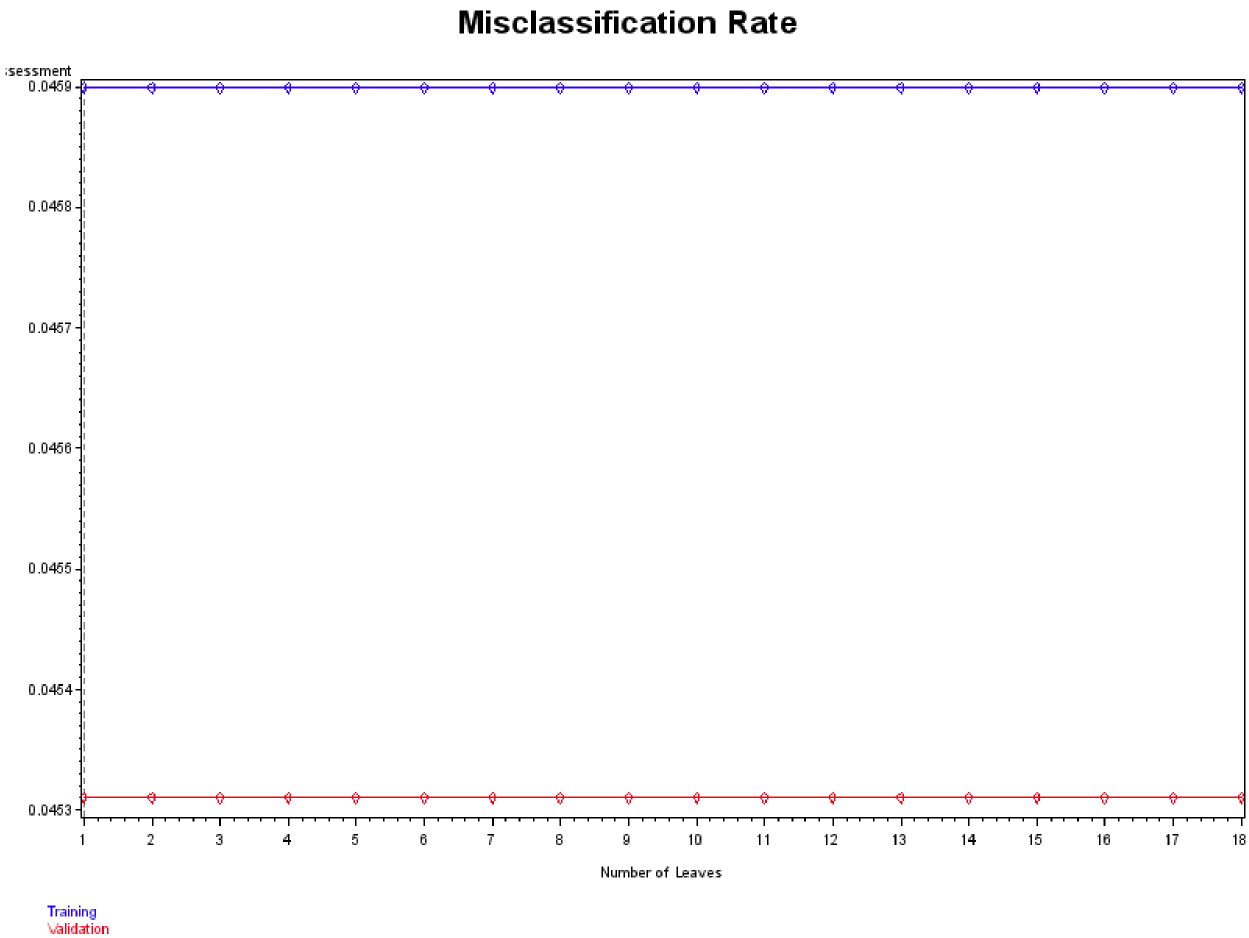
Exhibit 6

A Graphical Representation of Misclassification Rate with Varying Number of Leaves

Panel A: When Target Event is MW for Years 2004-2010



Panel B: When Target Event is Non-MW for Years 2004-2010

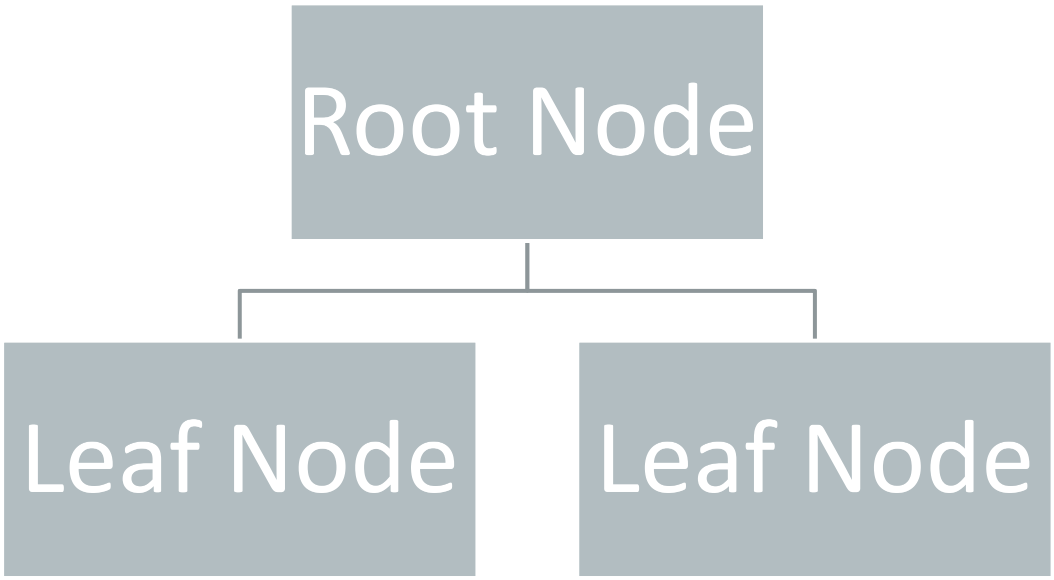


Nevertheless, as noted below, the tree structure reveals additional information as information related to more leaves are carefully examined. Therefore, in Exhibit 7, we show the tree structure with 6 leaves.

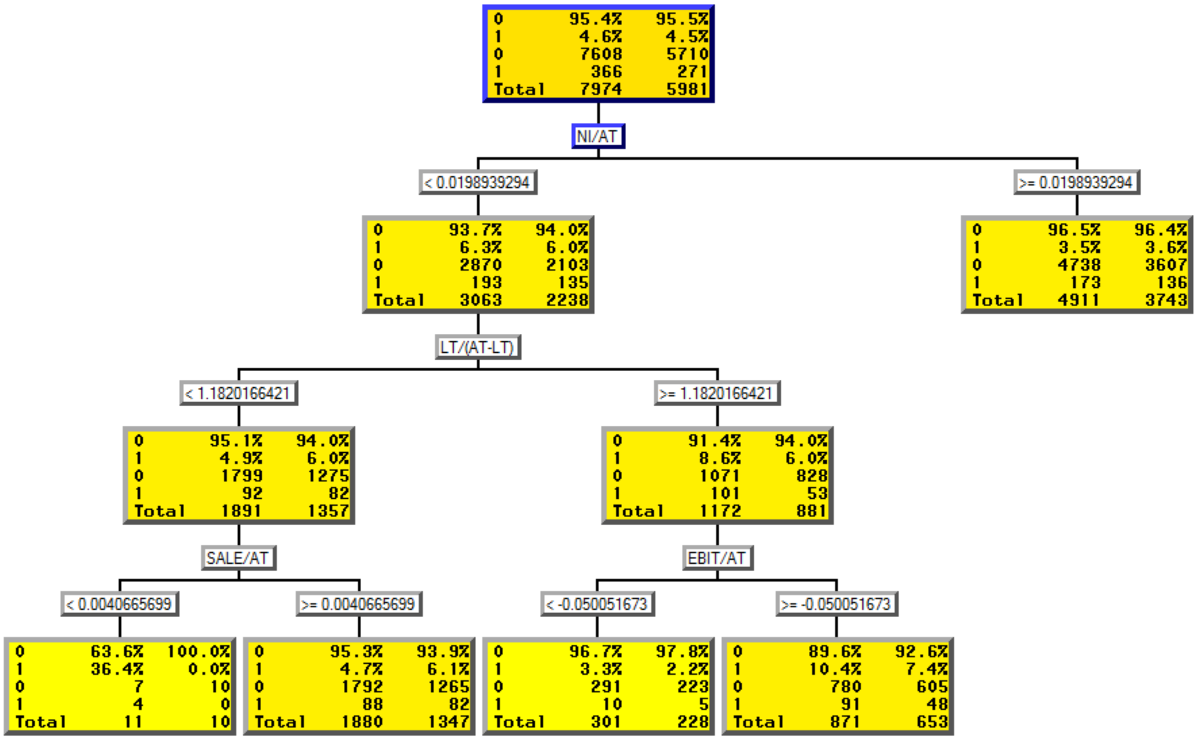
Exhibit 7

Graphical Representation of Decision Tree Rules Diagram with Two Leaves

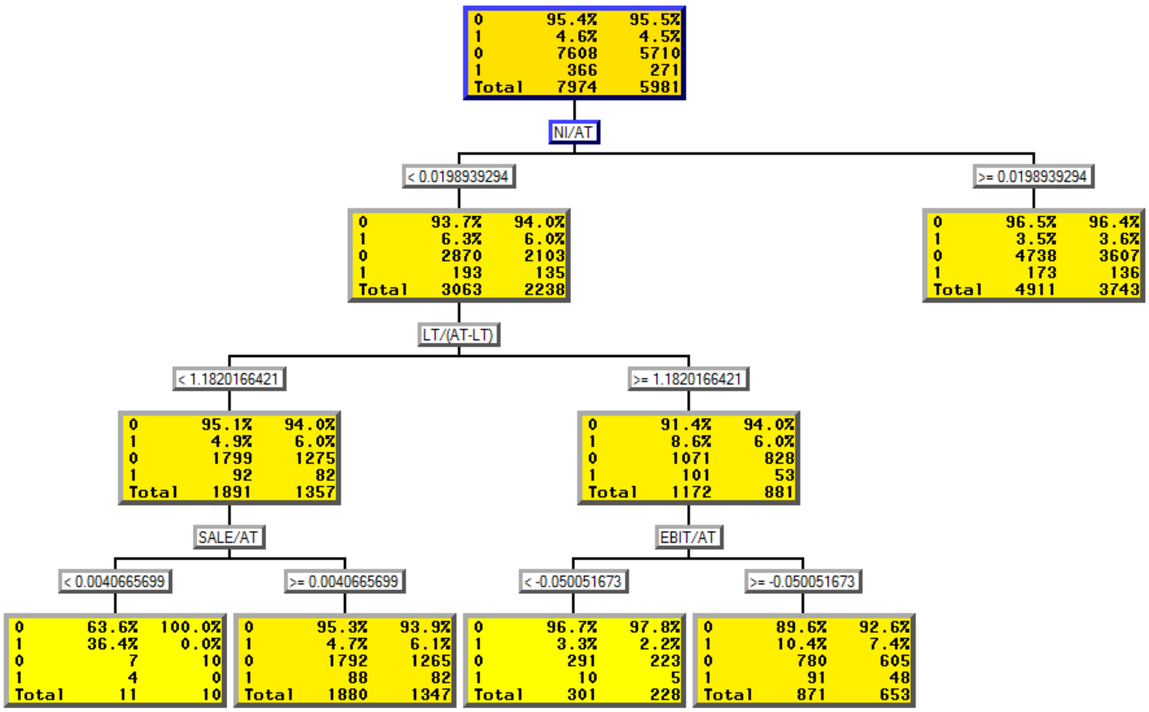
Panel A: Skeleton Representation of Decision Tree Rules Diagram



Panel B: Detailed Graphical Representation of Decision Tree Rules Diagram When Target Event is MW for Years 2004-2010



Panel C: Detailed Graphical Representation of Decision Tree Rules Diagram When Target Event is Non-MW for Years 2004-2010



As shown in Panels A and B of Exhibit 7, the root node—a nonleaf node—is connected to leaf nodes; during the process of splitting the root node into leaf nodes, a test is needed to split the root node into smaller groups of leaf nodes. The root node simply shows how total sample is divided between the training set (40 percent) and the validation set (30 percent); the node also provides information about how many companies in the training and validation sets have non-MW or MW companies. Therefore, data mining needs more than one leaf to show the decision or production rules. A production rule is a test with an IF statement. These additional leaves provide branches that carry the production rule tests. In the decision tree modeling, branches carry tests which will be connected to the lower levels of leave nodes.

Careful observation of Panel B reveals that because the composition of the data set heavily tilted toward non-MW companies might have resulted in more nodes with a higher percentage of non-MW companies than MW companies. Therefore, in the future, paired-matching of non-MW and MW companies—using sales or assets as a matching variable—may reveal additional insights and useful indicators in identifying MW companies.

Nevertheless, this path analysis of tree structure reveals useful insights. For example, in Panel C of Exhibit 7, if Net Income over Total Assets is less than 0.019 (this is a test branched out of the root note), then SAS EM's validation process reveals that 6.3% of the sample with such financial characteristics turns out to be MW companies. As auditors add another criterion such as when Total Liabilities over Net Worth is greater than or equal to1.18, then there is an increase in the MW companies during the testing process, making the percentage of MW companies 8.6%. With still another criteria added, when Earnings Before Interest and Taxes over Total Assets is greater than or equal to -0.05, the percentage of MW companies increases yet again to 10.4%. The validation process exhibits the same pattern of increase in the percentage of MW companies as new criteria are added.

Therefore, auditors may find useful information by analyzing all paths of how each node or criterion is branched out. What the decision tree model does is to show the users what characteristics each group possesses. Therefore, by documenting information from different paths based on each configuration in SAS Enterprise Miner's Decision Tree Model, the auditors would develop a comprehensive map where MW companies are more likely congregate in regard to different criteria or financial ratios. In Panel B of Exhibit 7, auditors may recognize the following promising path: NET INCOME/TOTAL ASSETS < 0.019 → TOTAL LIABILITIES/NET WORTH >= 1.182 → EARNINGS BEFORE INTEREST AND TAXES/TOTAL ASSETS < -0.050. This path leads to a highly concentrated group of non-MW companies. The validation process of this path reveals that 97.8% of companies in the group are non-MW companies, whereas only 2.2% in the group are MW companies. As future researchers store more data in their research data warehouses, more promising criteria may emerge from this type of research, and from such studies, we expect that more useful insight will emerge.

When auditors are interested in finding non-MW companies, the auditors could use Panel C by using a similar analytical process that we provided in the previous paragraphs. For example, auditors may get insights on examining the decision trees by comparing companies with certain contrasting characteristics: (1) NI/AT < 1.98%, LT/(AT-LT) < 118.2%, and SALE/AT < .4%, and (2) NI/AT < 1.98%, LT/(AT-LT) < 118.2%, and SALE/AT > .4%. In the first set, companies have low utilization rate of assets from both NI/AT and SALE/AT. In the second set, however, companies, NI/AT and SALE/AT tell us somewhat conflicting stories on asset utilization, and we witness the increase in material weakness companies from zero in the first set to 6.1%, which were 82 companies of 1,347 companies in the second set.

5. Concluding Remarks

It has been demonstrated in this study that the decision tree model can provide a useful way to identify MW companies with a high accuracy. Considering the increasing cost of quality audits in the post-SOX period, intelligent usage of data mining will help auditors save financial resources and facilitate them to engage in cost-effective audit planning. Further refinements could be made before this type of approach is applied to actual audit cases. For example, different industries may demonstrate different levels of prediction accuracy, and an extension of the study period can also help enhance the validity of the positive result of this study

References

Altman, E. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance* (Volume 23, No. 4), 589-609.

Altman, E. 2007. Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA® Models. Available at: http://pages.stern.nyu.edu/~ealtman/Zscores.pdf.

Berry, M., and Linoff, G. 2000. *Mastering Data Mining.* John Wiley & Sons, Inc., New York, NY.

Bradford, M., Taylor, E., and Brazel, J. 2010. Beyond Compliance: The Value of SOX. *Strategic Finance* (May), 48-53.

Cheh, J., Kim, I., and Moon, C. 2009. Profiling Companies with Internal Control Material Weaknesses Using the Data Mining Decision Tree Model. *Korean Journal of Management Accounting Research* (June), 75-94.

Cheh, J., Lee, J. and Kim, I. 2010. Determinants of Internal Control Weaknesses. *Contemporary Management Research* (June), 159-176.

Doyle, J., Ge, W., and McVay, S. 2007. Determinants of Weaknesses in Internal Control over Financial Reporting. *Journal of Accounting & Economics* (Volume 44), 193-223.

Ge, W., and McVay, S. 2005. The Disclosure of Material Weaknesses in Internal Control After Sarbanes-Oxley Act. *Accounting Horizons* (Volume No. 3), 137-158.

Goodburn, M. 2011. The CFO is the Closest Confident to the CEO. *CFO Magazine* (September), 38-40.

Groth, R. 2000. *Data Mining: Building Competitive Advantage*. Prentice Hall PTR, Upper Saddle River, NJ.

Kim, Y. H. 2008. A Study on Early Adoption on Internal Accounting Control System (in Korean). Deloitte Anjin LLC. Deloitte Consulting Korea. Available at: http://www.deloitte.com/dtt/article/0,1002,cid%253D146291,00.htm1.

Knechel, R., Salterio, S., and Ballou, B. 2006. *Auditing: Assurance and Risk*, South-Western College Pub., 231.

Li, C., Rupley, K.H., and Johnstone, K. 2008. Internal Governance, External Governance, and Internal Control Material Weakness Remediation. Working Paper, University of Pittsburg.

O'Sullivan, K. 2006. The Chase for Clarity. *CFO Magazine* (September), 65-69.

Ogneva, M., Subramanaam, K. R., and Raghunandan, K. 2007. Internal Control Weakness and Cost of Equity: Evidence from SOX Section 404 Disclosures. *The Accounting Review* (Volume 82, No. 5), 1255- 1298.

Patterson, E. R. and R. Smith. 2007. The Effects of Sarbanes-Oxley on Auditing and Internal Control Strength. *The Accounting Review* (Volume 82, No.2), 427-455.

Tang, A.P. and Xu, L. 2008. Institutional Ownership and Internal Control Material Weakness. *Quarterly Journal of Finance and Accounting*, March 2012.

Verschoor, Curtis. 2012. Has SOX Been Successful? *Strategic Finance* (September), 14-17.

Whitten, I. H., and Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition, Morgan Kaufman Publishers, San Francisco, CA.