

Determinants of Internal Control Weaknesses

John J. Cheh
The University of Akron
E-Mail: cheh@uakron.edu

Jang-hyung Lee
Daegu University
E-Mail: goodljh@daegu.ac.kr

Il-woon Kim
The University of Akron
E-Mail: ikim@uakron.edu

ABSTRACT

The major purpose of this paper is to find the financial and nonfinancial variables which would be useful in identifying companies with Sarbanes-Oxley Act internal control material weaknesses (MW). Recently, three research papers have been published on this topic based on different variables. Using the decision tree model of data mining, this paper examines the predictive power of the variables used in each paper in identifying MW companies and compares the predictive rates of the three sets of the variables used in these studies. In addition, an attempt is made to find the optimal set of the variables which gives the highest predictive rate. Our results have shown that each set of the variables is complementary to each other and strengthens the prediction accuracy. The findings from this study can provide valuable insights to external auditors in designing a cost-effective and high-quality audit decision support system to comply with the Sarbanes-Oxley Act.

Keywords: Sarbanes-Oxley Act, Internal Control Material Weakness, Data Mining

INTRODUCTION

The Sarbanes-Oxley Act (SOX) of 2002, also known as the Public Company Accounting Reform and Investor Protection Act, was enacted on July 30, 2002 in response to a number of major corporate and accounting scandals of large companies in the U.S., such as Enron, Tyco International, Adelphia, Peregrine Systems, and WorldCom. Section 404 of the Sarbanes-Oxley Act requires management of publicly traded companies to assess their company's internal control material weakness (MW) and to provide an internal control report as part of their periodic report to stockholders and regulators. The Act requires all public companies to maintain accurate records and an adequate system of internal accounting control. SOX also dramatically increased the penalties for false financial reporting on both management and external auditors.

Even though the Act helped improve the quality and transparency of financial reports and give investors more confidence in financial reporting through added focus on internal control, increasing costs of compliance has been a major concern of many companies. A recent survey by Financial Executive International (O'Sullivan, 2006) found that public companies have incurred greater than expected costs to comply with section 404 of SOX: 58% increase in the fees charged by external auditors. It was estimated that U.S. companies would have spent \$20 billion by the end of 2006 to comply with the law since it was passed in 2002. AMR Research also estimates that companies are spending about \$1 million on SOX compliance for every \$1 billion in revenues.

The major purpose of this paper is to find the financial and nonfinancial variables which would be useful in predicting Sarbanes-Oxley Act internal control material weaknesses (MW). In addition, an attempt will be made using all the variables from these published studies to find an optimal set of the variables which gives the highest predictive rate. Once identified, these variables can be used by external auditors as a screening device in identifying companies with MW in internal control, which can help external auditors save the costs of audit significantly.

PREVIOUS RESEARCH

Recently, three research papers have been published on identifying variables affecting Sarbanes-Oxley Act internal control MW using different variables and different methodologies: Cheh et al. (2006), Doyle et al. (2007) and Ogneva et al. (2007). In the study by Cheh et al., a trial and error approach was used with different combinations of many variables, and 23 financial variables were finally identified

which gave the highest predictive rate for internal control MW. Studies by Doyle et al. and Ogneva et al. used a traditional method of theorizing the situations and bringing theory-based variables into the research. Doyle et al. (2007), 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. (2007) 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.

Although three groups of these researchers worked on a similar topic, each research group had slightly different research objectives. The aim of Cheh et al.'s work (2006) was to find data mining rules that facilitate the identification of MW companies. Doyle et al. (2007) strived to find characteristics of MW companies. On the other hand, Ogneva et al. (2007) were more interested in examining the association between cost of capital and internal control weaknesses. The variables used in each study were not necessarily selected to find MW companies, but they were certainly related to MW. Since the variables in each study are quite different from each other, it would be interesting to see if the predictability can be improved using all the variables from the three studies. Eventually, a certain set of the financial and nonfinancial variables may be able to be identified that can be powerful enough, yet most cost effective, in predicting MW companies.

Besides the three papers mentioned, 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 et al. (2009) 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.

RESEARCH METHODOLOGY

Decision Trees Model

As an integral part of knowledge discovery from large databases, data mining is the overall process of converting raw data into useful information. Data mining techniques tend to allow data to speak for themselves rather than coming into the data analysis with any preconceived notion of hypothetical relationships of the variables under study. Hence, data mining has been widely used in business to find useful patterns and trends that might otherwise remain unknown within the business data.

Data mining, which is often called analytics, has been widely used not only for business practices but also 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 (Witten 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.

Independent Variables

There is conceivably a large number of varying combinatorial sets of independent variables on the dependent variable of material weakness on a company's internal control. In these conceivably huge sets of the independent variables, we decided to use the variables which were used in three published studies: Cheh et al. (2006), Doyle et al. (2007) and Ogneva et al. (2007). In total, 46 variables were used on these studies: 23 by Cheh et al., 10 by Doyle et al., and 13 by Ogneva et al. There were several overlapping variables between studies by Doyle et al. and Ogneva et al. In addition, two of the variables used by Doyle et al. (2007) were excluded in this study because they were not available in our data bases, such as Audit Analytics, CRSP, and Research Insight. These two variables were the number of special purpose entities (SPEs) associated with each firm and governance score. Consequently, 35 variables were used in this study. Also, we excluded in our study three variables from the work by Ogneva et al.; these variables were predicted forecast errors and were known to be associated with systematic biases in analyst forecasts. The variables utilized forecast data from I/B/E/S and Value Line; but these data were not available in our aforementioned data bases.

Panel A of Table 1 provides detailed descriptive explanations of the 35 independent variables, and the descriptive statistics of each variable are presented in Panel B. It can be seen in Panel B that four variables have very large standard deviations and variances. These variables are Net Income/(Total Assets – Total Liabilities), Cost of Goods Sold/Inventory, Retained Earnings/Inventory, and Total Liabilities/(Total Assets – Total Liabilities). Although we have normalized net income, cost of goods sold, retained earnings, and total assets using various measures, these variables apparently have had very large fluctuations among the sample firms. For net income, however, it is interesting to note that Net Income/Sales and Net Income/Total Assets have very small variances while Net Income/(Total Assets – Total Liabilities) has a significant variance. This result indicates that net income of sample firms fluctuated significantly with respect to net equity and was relatively stable with respect to sales and total assets.

Sample Firms

The initial sample companies came from Research Insight which has almost 10,000 companies listed in its Compustat database. These companies are listed in three major U.S. stock exchanges (i.e., New York Stock Exchange, American Stock Exchange and NASDAQ), regional U.S. stock exchanges, Canadian stock exchanges, and over-the-counter markets. For these companies, the data required for this study were retrieved from three research databases: Audit Analytics, CRSP, and Research Insight. The companies with SOX MW for the period of 2002 to 2006 were obtained from Audit Analytics. The data for the 35 variables were retrieved from Compustat database in Research Insight, except for information about the ages of the sample firms. The information about ages was available in the CRSP tape. To remove any companies with incomplete data for any variables, we used an in-house developed software application. This process resulted in 869 companies in the final sample, a considerably smaller subset of the initial sample.

Table 1 Descriptive statistics for 35 variables for 2004-2006

Panel A: Variable Descriptions

Variables	Description	Variables	Description
NI/SALE	Net Income/Net Sales	CH/LCT	Net Sales/Current Liabilities
NI/AT	Net Income/Total Assets	WCAP/SALE	Cash/Current Liabilities
NI/(AT-LT)	Net Income/Net Worth	RE/AT	Working Capital/Net Sales
EBIT/AT	Earnings Before Income Tax/Total Assets	CH/AT	Retained Earnings/Total Assets
SALE/AT	Net Sales/Total Assets	LT/(AT-LT)	Cash/Total Assets
ACT/AT	Current Assets/Total Assets	LOGMARKETCAP	Total Liabilities/Net Worth
(ACT-INVT)/AT	Quick Assets/Total Assets	LOGAGE	LOG of share price × number of shares outstanding(CSHO*PRCCF)
SALE/(ACT-INVT)	Net Sales/Quick Assets	AGGREGATELOSS	LOG Age of company
ACT/SALE	Current Assets/Net Sales	RZSCORE	Indicator variable [1,0] if \sum earning before extraordinary items in years t and t-1 < 0, otherwise(IB)
INVT/SALE	Inventory/Net Sales	LOGSEGGEOMUM	LOG \sum number of operating and geographic segments (Bus Segment-Actual number+Geo Seg Areas-Actual number)
COGS/INVT	Cost of Goods Sold/Inventory	FOREIGN	Indicator variable [1,0] if non-zero foreign currency translation , otherwise(FCA)
LT/AT	Total Liabilities/Total Assets	ACQUISITION VALUE	\sum 50% Ownership of the acquired firm(AQA/MKVAL)
(ACT-INVT)/SALE	Quick Assets/Sales	EXTREME SALESGROWTH	Indicator variable [1,0] if year-over-year industry-adjusted Net Sales growth=top quintile, otherwise
RE/INVT	Retained Earnings/Inventory	RESTRUCTURINGCHARGE	\sum restructuring charges (RCA/MKVAL)
ACT/LCT	Current Assets/Current Liabilities	LOGSEGMUM	LOG Number of business segments
(ACT-INVT)/LCT	Quick Assets/Current Liabilities	INVENTORY	Inventory/Total assets
LCT/AT	Current Liabilities/Total Assets	M&A	Indicator variable [1,0] if company was involved in M&A over 3 years, otherwise
SALE/LCT	Sales/Current Liabilities		

Panel B: Descriptive statistics of the variables

variables	N	Mean	Std. Deviation	Variance	Minimum	Maximum
ICMW *	869	0.520138	0.499882	0.249882	0	1
NI/SALE	869	-0.04617	0.585724	0.343073	-11.4922	2.506701
NI/AT	869	-0.00369	0.258762	0.066958	-2.81327	4.832774
NI/(AT-LT)	869	1.102761	33.98825	1155.201	-14.0419	1001.306
EBIT/AT	869	0.032835	0.158506	0.025124	-2.40982	0.480082
SALE/AT	869	1.098502	0.852066	0.726016	0.026889	8.613351
ACT/AT	869	0.507784	0.235754	0.05558	0.024723	0.991249
(ACT-INVT)/AT	869	0.381526	0.212887	0.045321	0.023738	0.956625
SALE/(ACT-INVT)	869	4.355389	5.787747	33.49802	0.035184	66.39423
ACT/SALE	869	0.750516	1.444692	2.087135	0.033669	28.66403
INVT/SALE	869	0.12852	0.120808	0.014595	0.000373	1.631579
COGS/INVT	869	19.77532	73.00375	5329.548	0.107527	1874.644
LT/AT	869	0.528867	0.418582	0.17521	0.028224	6.811726
(ACT-INVT)/SALE	869	0.621996	1.410005	1.988113	0.015062	28.42211
RE/INVT	869	-17.4043	144.4622	20869.32	-2476.96	568.4035
ACT/LCT	869	2.813127	2.783589	7.748366	0.097934	35.69801
(ACT-INVT)/LCT	869	2.227998	2.66166	7.084434	0.079661	35.41416
LCT/AT	869	0.250558	0.169365	0.028684	0.024158	1.388766
SALE/LCT	869	4.914681	2.970393	8.823232	0.250556	33.59769
CH/LCT	869	0.822242	1.541899	2.377452	0	20.04329
WCAP/SALE	869	0.452549	1.332053	1.774364	-1.74452	26.57621
RE/AT	869	-0.43403	2.365501	5.595594	-42.8029	1.041362
CH/AT	869	0.127273	0.131966	0.017415	0	0.879367
LT/(AT-LT)	869	-4.58159	181.5619	32964.74	-5340.58	226.1027
LOGMARKETCAP	869	2.663385	0.719972	0.51836	-0.28377	4.889938
LOGAGE	869	1.117768	0.18178	0.033044	0.30103	1.30103
AGGREGATELOSS	869	0.219793	0.414345	0.171682	0	1
RZSCORE	869	4.52359	2.906921	8.450192	0	9
LOGSEGGEOMUM	869	0.738405	0.174211	0.030349	0	1.176091
FOREIGN	869	0.218642	0.413563	0.171035	0	1
ACQUISITIONVALUE	869	-0.00023	0.002234	4.99E-06	-0.03232	0.037896
EXTREMESALES GRO WTH	869	0.243959	0.429715	0.184655	0	1
RESTRUCTURING CH ARGE	869	-0.19341	5.104154	26.05239	-149.554	3.574735
LOGSEGMUM	869	0.286145	0.290228	0.084233	0	1
INVENTORY	869	0.126258	0.120126	0.01443	0.000327	0.754949
M&A	869	0.116226	0.32068	0.102836	0	1

* Note that ICMW is the dependent variable of the study while all others are independent variables.

Application of Data Mining

The decision tree model in data mining was available from SQL Server 2005 Analysis Services. Since each configuration could provide different results, the same configuration was used for all three groups of variables: the default configuration in the Analysis Services. Data mining would be particularly useful to get new insights into the variables used in the studies by Doyle et al. and Ogneva et al. because they hypothesized the relationships between their dependent variables and independent variables.

For the purpose of training the model, one-year lagged data were used. For example, the 2004 data were used for training and then the 2005 data were used for prediction, and the 2005 data for training and the 2006 data for prediction. In that way, the decision rules were actually predicting MW of the sample companies with unknown data of the next year. This process provided with four sets of data that could be used for both training and testing, which will be explained in the following section.

ANALYSES AND FINDINGS

It appears that in some cases, the financial variables from two studies by Doyle et al. (2007) and Ogneva et al. (2007) on Sarbanes-Oxley Act of 2002 internal control MW add value to the predictability of the financial variables which have been used in the study by Cheh et al. (2006), although this finding may become somewhat less clear in special cases. Results also provide preliminary evidence that all of the variables collectively can contribute to construction of a good predictive model. Detailed discussions on the results follow in the subsequent paragraphs.

Table 2 Training 2004 data and testing 2005 data using all 35 variables

Panel A: Actual Classification

Predicted	0 (Actual)	1 (Actual)	Total
0	21	4	25
1	149	174	323
Total	170	178	348

Panel B: Prediction Accuracy

Predicted	0 (Actual)	1 (Actual)
0	12.35% (Prediction Accuracy)	2.25% (Type II Error)
1	87.65% (Type I Error)	97.75% (Prediction Accuracy)
Total	100%	100%

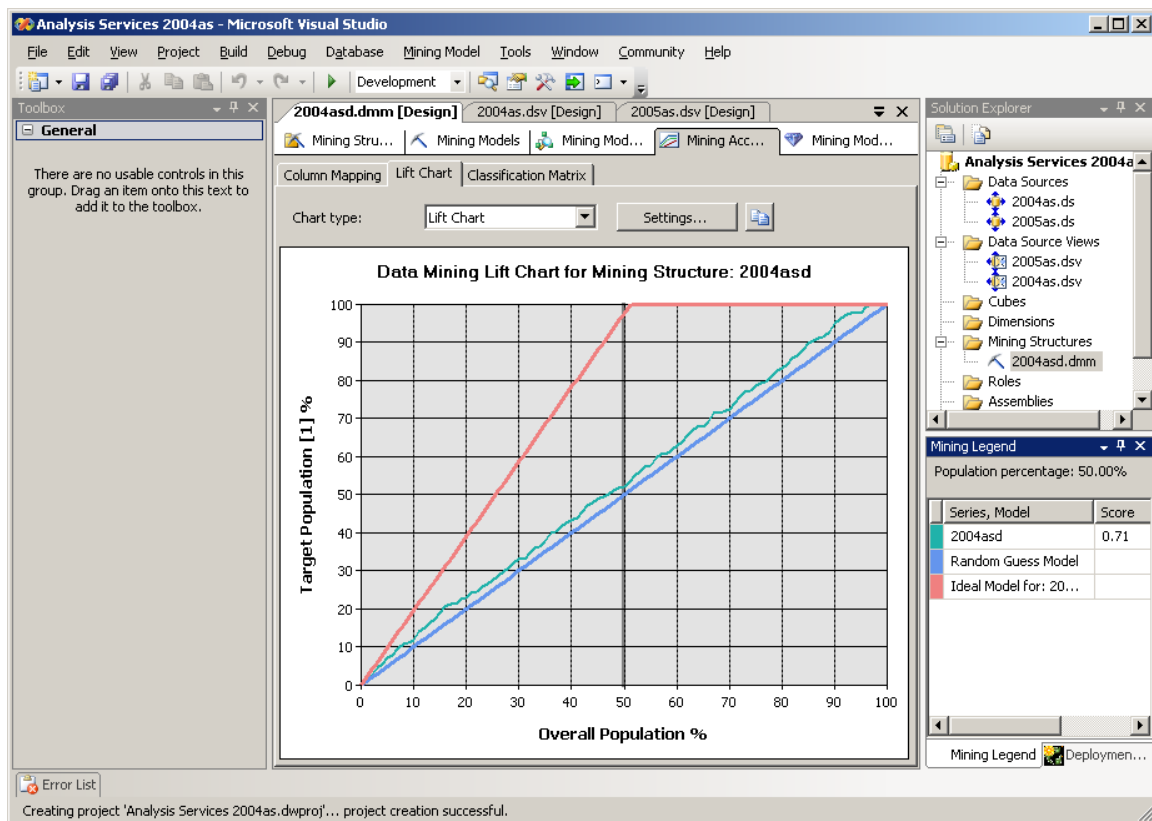


Figure 1 Graphic results for Table 2

Using All 35 Variables

As the first test, we used the 2004 fiscal year data for training to predict MW companies in the 2005 fiscal year, using all 35 variables. As is shown in Table 2, when there is no MW, the prediction accuracy is only 12.35% while the error of misclassifying MW as non-MW is 87.65%. However, when there is actually MW, the prediction accuracy is quite impressive: 97.75% with the error of misclassifying MW as non-MW being only 2.25%. It should be noted that in this data set and also subsequent date sets, we matched both sets of non-MW and MW companies with standard industry code (SIC) and also sales within 25% plus or minus.

The lift chart shown in Figure 1 displays a lift which is an indication of any improvement from random guess and shows a graphical representation of the change in the lift that a mining model causes (Tang and MacLennan 2005). For example, the top line shows that an ideal model would capture 100% of the target using a little over 50% of data. The 45-degree random line across the chart indicates that if a data miner were to randomly guess the result for each case, the person would capture 50% of target using 50% of data (Tang and MacLennan 2005). The data mining model is depicted in the line between the ideal line and random line. According to the graph,

the decision tree model is doing only a little better than a random model. This also means that there isn't sufficient information in the training of the data to learn the patterns about the target.

To predict MW companies with the 2006 fiscal year based on the 2005 data, our data mining model trained the 2005 fiscal year data for all 35 variables of the sample companies. As we did with the first test, we matched both sets of non-MW and MW companies with SIC and sales within 25% plus or minus. The final results are presented in Table 3 and Figure 2, which are similar to those of the previous test. When there is actually MW, the prediction accuracy is as high as 98.47%, and the error rate of misclassifying MW as non-MW is 1.53%. When there is no MW, the prediction accuracy is 11.02%, and the error rate of misclassifying non-MW as MW is 88.98%.

Table 3 Training 2005 data and testing 2006 data using all 35 variables

Panel A: Actual Classification

Predicted	0 (Actual)	1 (Actual)	Total
0	13	2	15
1	105	129	234
Total	118	131	249

Panel B: Prediction Accuracy

Predicted	0 (Actual)	1 (Actual)
0	11.02% (Prediction Accuracy)	1.53% (Type II Error)
1	88.98% (Type I Error)	98.47% (Prediction Accuracy)
Total	100%	100%

It is interesting to see that, even though there is a slight increase in the prediction rate of MW companies when there is MW from 97.75% to 98.47%, the prediction rate of non-MW companies when there is no MW has somewhat deteriorated from 12.35% to 11.02%. It appears that there isn't more information in the 2006 data than in the 2005 data sufficient enough to make a difference in training the data and learning patterns about the target. As expected, the lift chart in Figure 2 is similar to the previous lift chart with only a slightly better result in predicting MW companies.

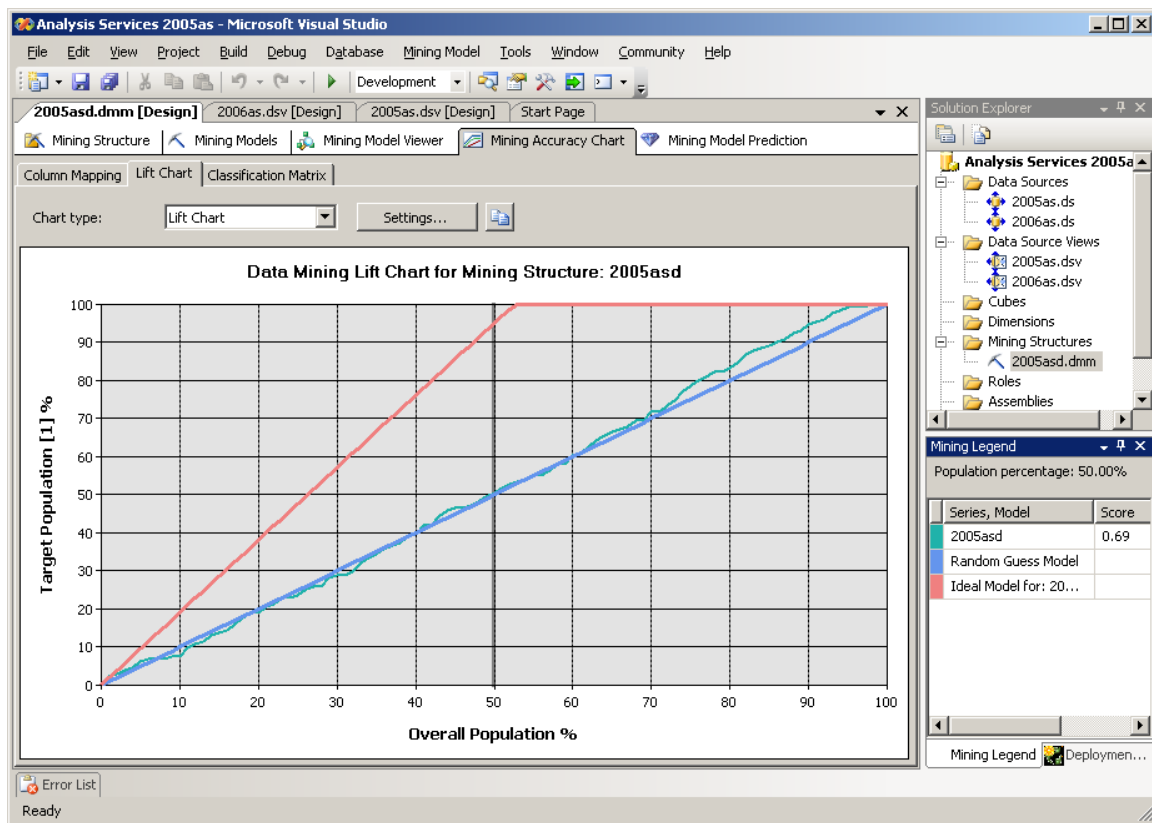


Figure 2 Graphic results for Table 3

Using the Doyle et al.'s Variables

We repeated the process of training the data and predicting MW with the nine variables used by Doyle et al. As mentioned before, Doyle et al.'s (2007) used two additional variables that we did not use in our study: number of special purpose entities (SPE) associated with a firm under study and its governance score. Doyle et al. do not find a significant relation between MW disclosures and corporate governance. Although its prediction rate of non-MW companies increased slightly, its MW prediction rate decreased to below 95 % for training in 2004 data to predict 2005 MW companies as shown in Table 4. The significant decrease in the predictive power of our data mining model was probably caused by, among others, the use of the smaller number of independent variables from 35 to nine.

Figure 3 provides two types of information about the model used in this study. The first type of information presented in Panel A is what the trees reveal. The decision rules in rectangular boxes or nodes present which rule will more likely generate a desired outcome. For example, if a company's log of market capitalization is greater than or equal to 1.775, but less than 3.833, then it is highly likely that the

company would have material weakness. The darker the node is, the more cases it contains.

Table 4 Training 2004 data and testing 2005 data using Doyle et al.'s variables

Panel A: Actual Classification

Predicted	0 (Actual)	1 (Actual)	Total
0	27	10	37
1	158	168	326
Total	185	178	363

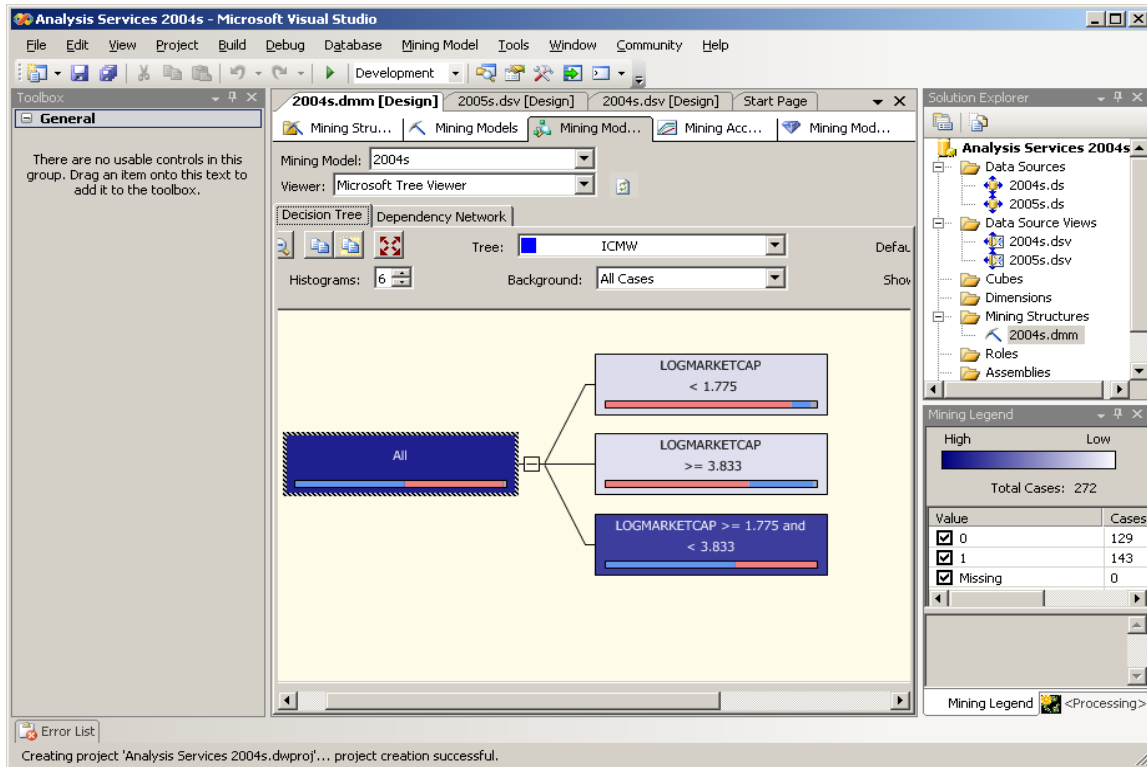
Panel B: Prediction Accuracy

Predicted	0 (Actual)	1 (Actual)
0	14.59% (Prediction Accuracy)	5.62% (Type II Error)
1	85.41% (Type I Error)	94.38% (Prediction Accuracy)
Total	100%	100%

In Panel B of Figure 3, the top line of the lift chart shows that almost 50% of the data indicates the desired target. According to Tang and MacLennan (2005), the bottom line is the random line, and the random line is always a 45-degree line across the chart. What this means is that if a business analyst were to randomly guess the results for each case, the analyst would capture 50% of the target using 50% of the data.

An interesting part is the line between these two lines of ideal line and random line. We may call this middle line a "model line" which represents the data mining model that we ran. In this particular chart, the model line was merged with the ideal line up to almost 50% of the data. Then, it dipped a little and paralleled with the ideal line of the remaining data until a little after 90% of the data. Then, eventually, it merged with the ideal line again. Thus, it indicates that this data mining model worked rather well, almost as if it were an ideal model. Hence, in terms of data efficiency point, we can conclude that the variables used by Doyle et al. (2007) are quite efficient because the decision tree model based on the nine variables is close to the ideal model.

Panel A:



Panel B:

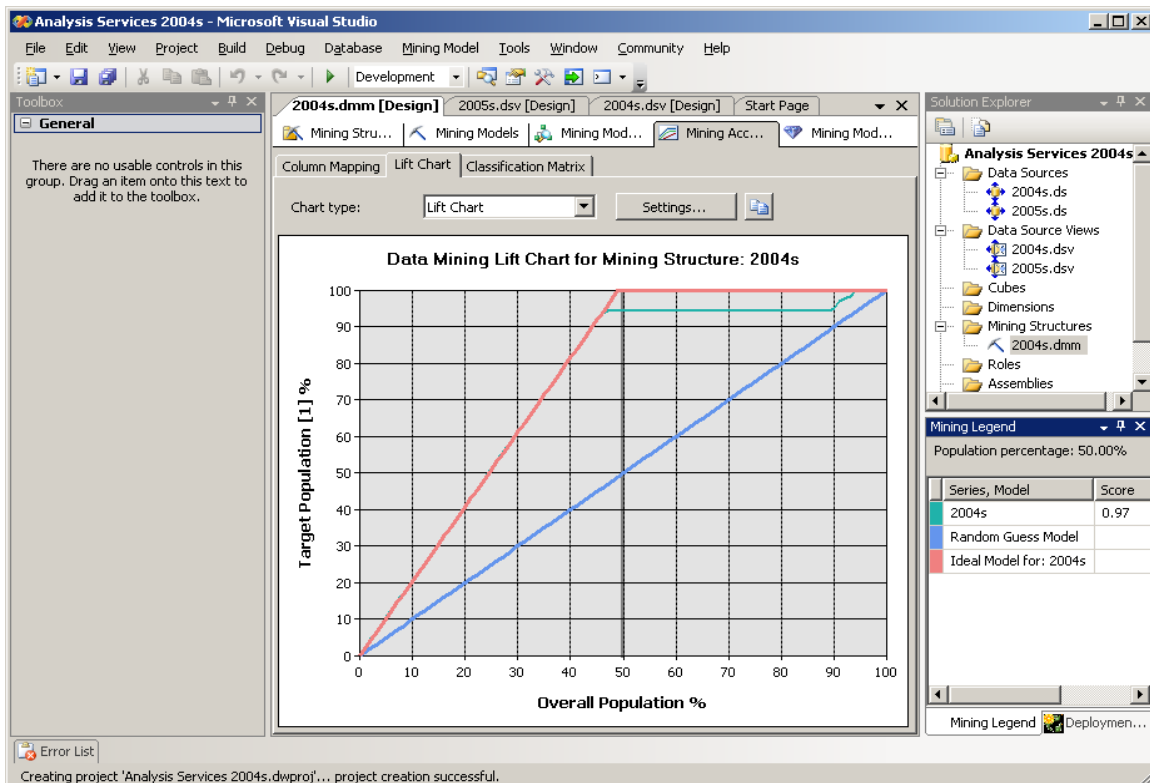


Figure 3 Graphic results for Table 4

Similar results are shown in Table 5 and Figure 4 for using 2005 data to predict 2006 MW companies. The MW prediction rate increased slightly in predicting 2006 MW companies to 96.21%, but is still below 98.47% as demonstrated using all 35 variables in Panel B of Table 3. Thus, using all 35 variables is a powerful reminder that more independent variables increase the data mining model's predictive power. The lift chart for 2005 training data in predicting 2006 MW companies using the variables by Doyle et al. shows some surprising finding that data efficiency is even for all distribution parts of population. Apparently, the decision model does better than the random model only in the end extremes.

Cost of Misclassification Errors

Using the following cost of misclassification model that was originally developed by Masters (1993), the cost of misclassification errors can be incorporated into the process of data analyses:

$$\text{Cost of misclassification errors (CME)} = (1 - q) p_1 c_1 + q p_2 c_2,$$

where q is a prior probability that the information risk associated with a company's internal control weakness was high, p_1 and p_2 are the probabilities of type 1 (false positive) and type 2 (false negative) errors, and c_1 and c_2 are the costs of type 1 and type 2 errors.

Table 5 Training 2005 data and testing 2006 data using Doyle et al.'s variables

Panel A: Actual Classification

Predicted	0 (Actual)	1 (Actual)	Total
0	17	5	22
1	113	127	240
Total	130	132	262

Panel B: Prediction Accuracy

Predicted	0 (Actual)	1 (Actual)
0	13.08% (Prediction Accuracy)	3.79% (Type II Error)
1	86.92% (Type I Error)	96.21% (Prediction Accuracy)
Total	100%	100%

The actual cost of misclassification error can be significant to audit firms. For example, by classifying a good client without MW as a bad one with MW, an audit firm makes a Type I error and, as a result, could potentially lose a client. To avoid the Type 1 error, the audit firm may assign the weight of unwarranted MW risk too much in its equation of audit engagement and planning processes. On the other hand, by mistakenly classifying a bad client with MW as a good one without MW, an audit firm makes a Type II error and, as a result, could miss MW in the client firm. In an effort to avoid the Type 2 error, the audit firm may assign the weight of MW risk too little in its equation of audit engagement and planning processes.

Accordingly, Cheh et al. (1999) applied the CME model to determine the cost of misclassification in predicting takeover targets. Later, Calderon (1999) and Calderon and Cheh (2002) also used this idea in auditing and risk assessment applications. The prior probability, q , can be estimated by using training data, as Cheh et al. (1999) did in their takeover target study. For example, using the data shown in Panel A of Table 2, we can compute that $q = 178/(170 + 178) = 0.5115$. Then, using the information in Panel B of Table 2, the cost of misclassification errors can be computed as following:

$$\text{CME} = 0.4885 * 0.8765 * c_1 + 0.5115 * 0.0225 * c_2.$$

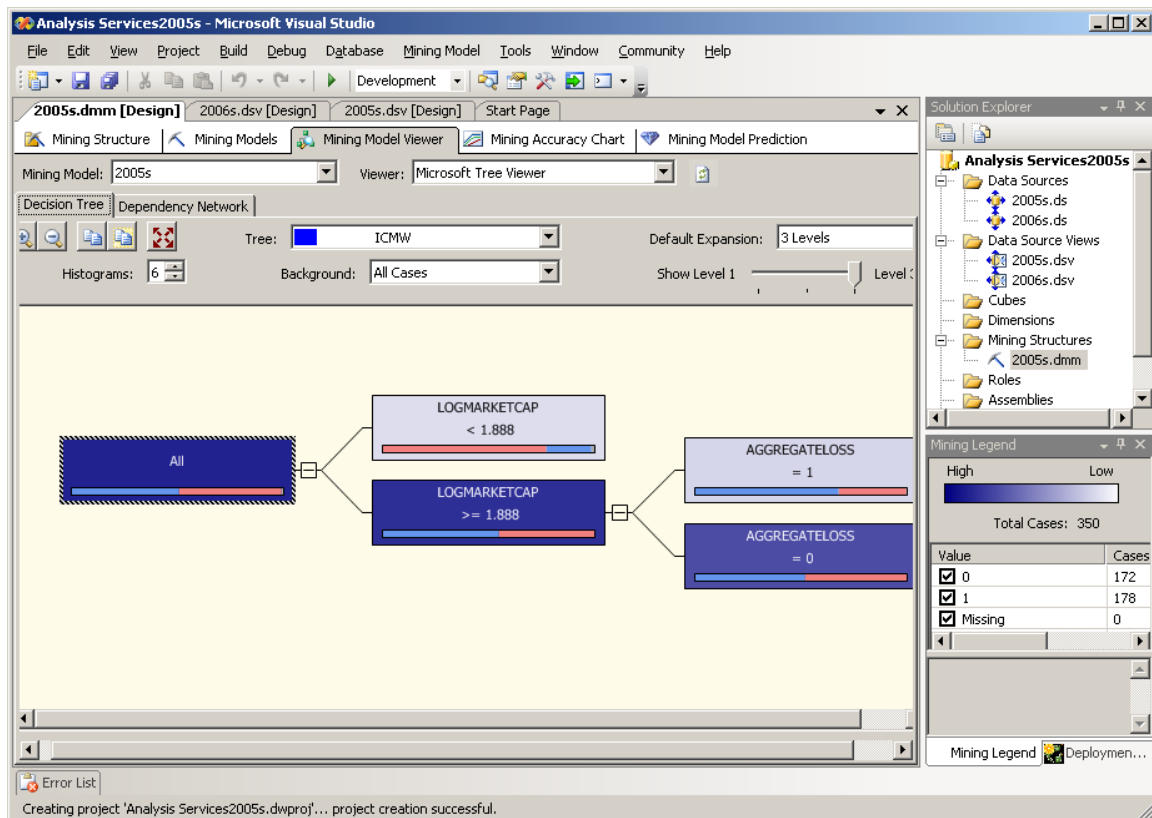
Since this predictive model can provide p_1 and p_2 , as shown above, each audit firm will be able to customize the cost model by using values on c_1 and c_2 based on the firm's professional experiences and past cost data. With this cost model based on the data mining prediction model, each audit firm can prepare an expected net benefit report before the firm engages in its potential audit client. With assistance from computer professionals, the audit firm can create another useful audit risk engagement tool, based on this type of cost analysis.

CONCLUSIONS AND FUTURE RESEARCH

The purpose of this paper was to find the financial and nonfinancial variables which would be useful in predicting Sarbanes-Oxley Act internal control material weaknesses. By examining a number of independent variables used in three published studies, we were able to construct a highly predictive data mining model with the prediction rate of MW of 98.47%. It appears that each set of variables from three groups of nine researchers are complementary to each other and strengthens the prediction accuracy. Apparently, each set of variables brings different types of strength to the prediction of material weaknesses.

The motivation of this research project was to lay out a ground work for a similar type of studies in the future that will help an audit firm build a customized audit engagement tool which can effectively predict MW companies. Hence, this work can be further extended to find whether in predicting MW companies 35-variables set dominates for a combined set of 12 variables by both Doyle et al.' (2007) and also Ogneva et al. (2007). By adding three additional variables from Ogneva et al. (2007), it is conceivable that the additional variables may tip the scale in favoring the smaller set by improving its prediction accuracy over Doyle at al.'s 9 variables-set so that neither set could dominate the other and that both sets of variables might be able to produce equal percent in predicting MW companies. If this holds, then a smaller set of variables may produce a similar quality of prediction accuracy.

Panel A:



Panel B:

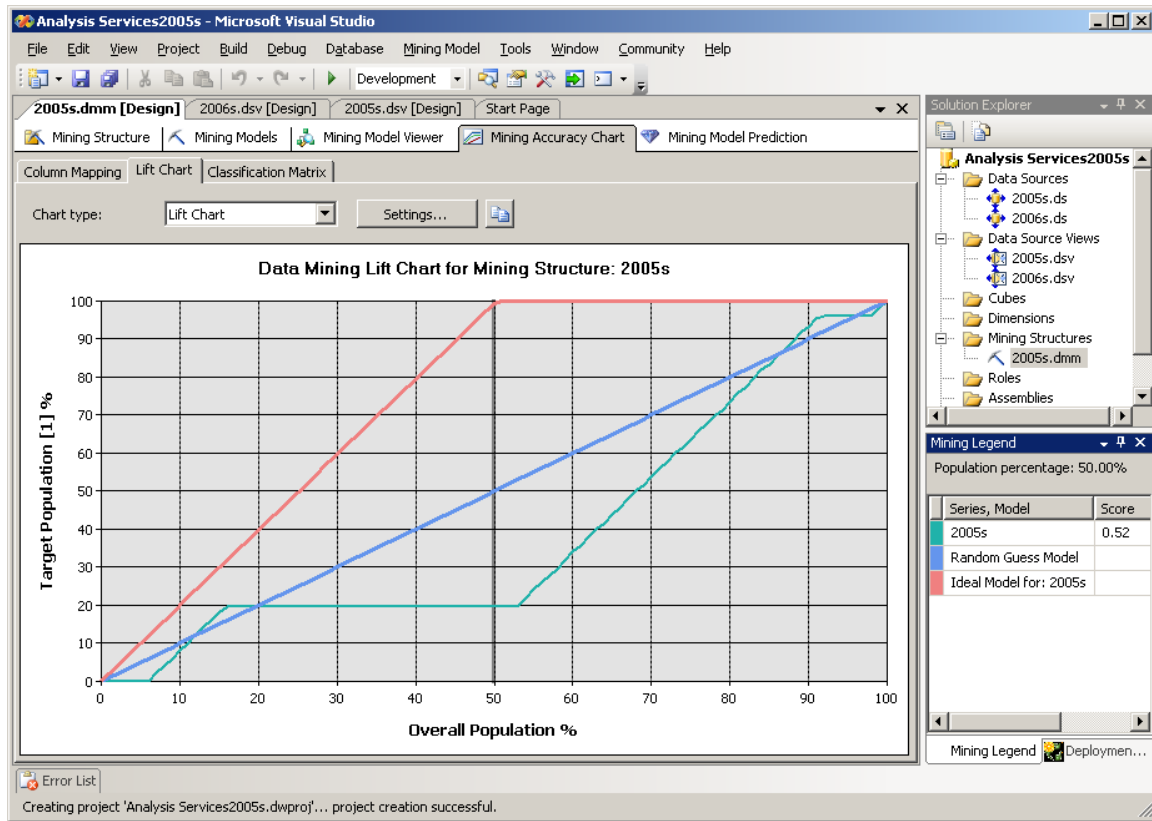


Figure 4 Graphic results for Table 5

REFERENCES

- Calderon, T. G. (1999). Neural Networks and Preliminary Information Risk Assessment in An Auditing Environment. *Accounting Inquiries Research Journal*, 8(2), 243-297.
- Calderon, T. G. and Cheh, J. J. (2002). A Roadmap for Future Neural Networks Research in Auditing and Risk Assessment. *International Journal of Accounting Information Systems*, 3(4), 203-236.
- Cheh, J. J., Kim, I., and Moon, H. (2006). *Preliminary Findings on Predicting Material Weakness Companies under Section 404 of the Sarbanes-Oxley Act – Comparison of Data Mining Models*. Paper presented at the meeting of the 15th Annual Research Workshop on Artificial Intelligence and Emerging Technologies in Accounting, Auditing and Tax, Washington, D. C..
- Cheh, J. J., Weinberg, R. S., and Yook, K. C. (1999). An Application of an Artificial Neural Network Investment System to Predict Takeover Targets. *Journal of Applied Business Research*, 15(4), 33-46.

- Doyle, J., Ge, W., and McVay, S. (2007). Determinants of Weaknesses in Internal Control over Financial Reporting. *Journal of Accounting & Economics*, 44, 193-223.
- Kim, J., Song, B. Y. and Zhang, L. (2009). Internal Control Weakness and Bank Loan Contracting: Evidence from SOX Section 404 Disclosures. Social Science Research Network Working Paper Series #: 1329114.
- Li, Chan, Rupley, K. H., and Johnstone, K. (2008). Internal Governance, External Governance, and Internal Control Material Weakness Remediation. Social Science Research Network Working Paper Series #: 1091685.
- Masters, T. (1993). *Practical Neural Network Recipes in C++*. Boston, Massachusetts: Academic Press.
- 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*, 82(5), 1255-1298.
- O'Sullivan, K. (2006). The Chase for Clarity. *CFO*, September, 65-69.
- Patterson, E. R. and Smith, J. R. (2007). The Effects of Sarbanes-Oxley on Auditing and Internal Control Strength. *The Accounting Review*, 82(2), 427-455.
- Tang, Z. and MacLennan, J. (2005). *Data Mining with SQL Server 2005*. Indianapolis, Indiana: Wiley Publishing Inc.
- Tang, A. P. and Xu, L. (2008). Institutional Ownership, Internal Control Material Weakness and Firm Performance. Social Science Research Network Working Paper Series #: 1031270.
- Witten, I. H., and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques Amsterdam*. Netherlands: Morgan Kaufmann Publishers.
- Xu, L. and Tang, A. P. (2008). Internal Control Material Weakness, Analysts' Accuracy and Bias, and Brokerage Reputation. Social Science Research Network Working Paper Series #: 105506.