Note: This paper has been published in **International Journal of Intelligent Systems in Accounting, Finance, and Management,** vol. 4, 113-126, 1995.

Detection of Management Fraud: A Neural Network Approach

Kurt Fanning[1], Kenneth O. Cogger[2], Rajendra Srivastava[3]

[1] School of Business, SUNY New Paltz, New Paltz, NY 12561
[2, 3] School of Business, University of Kansas, Lawrence, KS 66045-2003
[1] Fanningk@snynewvm [2] Cogger@kuhub.cc.ukans.edu [3] Rsrivastava@pobox.ukans.edu

Abstract

The detection of management fraud is an important issue facing the auditing profession. A major contributor to this issue is the Loebbecke and Willingham (1988) conceptual model for the detection of management fraud. A cascaded Logit approach using the Loebbecke and Willingham model was developed in Bell et al. (1993). The present study offers an alternative approach using Artificial Neural Networks (ANN's). This paper develops a successful discriminator of management fraud using both the generalized adaptive neural network architectures (GANNA) and the Adaptive Logic Network (ALN) approaches to designing neural networks. The discriminant functions can distinguish between fraudulent and non-fraudulent companies with superior accuracy to the cascaded Logit results of Bell et al. (1993). Finally, the discriminant function provides a parsimonious set of questions useful for detecting management fraud.

INTRODUCTION

The issue of management fraud is increasingly important to auditors given the current litigious environment for the profession. Prior research suggests that auditors increase in their ability to detect management fraud by using expert systems (Eining et al. 1994). Other research suggests that ANN’s are ideally suited for the detection of management fraud (Ethridge and Brooks 1994). The present study suggests that Artificial Neural Networks (ANN’s) can be used as a decision aid (analytical procedure) in detecting instances of management fraud.

Management fraud is a leading cause of the litigation crisis facing the auditing profession (Palmrose 1991; Palmrose 1987; St. Pierre and Anderson 1984). The auditing profession has change the way it operates due to the large number and severity of the awards from litigations (O'Malley 1993; Elgin 1992; Fuerman 1992). For example, firms in high risk industries are finding it extremely difficult to obtain audit services even at high prices (Freedman 1993; Holland et al. 1993; Sullivan 1992). The above problem motivates the present study and we propose analytical procedures helpful in identifying the risk of management fraud. We use rapidly self-developing (evolutionary) ANN's (Cogger 1992; Armstrong et al. 1991) for developing the analytical procedures to detect management fraud. In addition, the study provides a parsimonious group of questions that has a high probability of detecting management fraud.

The paper proceeds in the following manner. The next section discusses the conceptual model for detecting the risk of management fraud developed in Loebbecke and Willingham (1988) and Loebbecke et al. (1989). Section III discusses the cascaded Logit model of Bell et al. (1993). Section IV introduces Artificial Neural Networks (ANN's), including generalized adaptive neural network architectures (GANNA) (Cogger 1992) and Adaptive Logic Networks (ALN) (Armstrong et al. 1991). Section V explains the sample and the procedures of implementing the analytical procedures developed by the ANN's. The results are presented in Section VI. Finally, Section VII reviews the conclusions with suggestions for further research.

THE LOEBBECKE AND WILLINGHAM CONCEPTUAL MODEL

A leading conceptual model for detecting management fraud was first presented in Loebbecke and Willingham (1988) and further developed in Loebbecke et al. (1989). The Loebbecke and Willingham model attempts to provide an analytical procedure for detecting the risk of management fraud. While the Statement on Auditing Standards (SAS) 47 (AICPA 1983) provides a model for audit risk, it does not directly address the audit risk associated with management fraud (Srivastava et al. 1993). With the high cost of litigation associated with management fraud, most auditors associate a high level of risk to management fraud. The

Loebbecke and Willingham model provides a viable analytical procedure that allows auditors to set the inherent level of risk associated with management fraud.

The Loebbecke and Willingham model divides the process of assessing the likelihood of the existence of management fraud into three components: conditions, motivation, and attitude. According to the model, management fraud occurs when conditions exist for fraud to occur and management has the motivation and attitude to commit fraud. This model can be expressed as:

$$P(MF) = f(C, M, A) \qquad (1)$$

Where P(MF) represents the auditor's assessment of probability of a material misstatement due to fraudulent financial reporting, and C, M, and A represent the client's conditions, management's motivation, and management's attitudes, respectively. An important facet of the Loebbecke and Willingham model is that when all three components exist simultaneously, it is extremely likely that management fraud exists. The model also suggests that when only one component is present, there is little likelihood of management fraud. Loebbecke et al. (1989) assert that:

> . . . for management fraud to occur (1) the conditions of the entity must be such that a material management fraud could be carried out; (2) the person or persons who would commit the fraud must have a reason or motivation for doing so; and (3) the person or persons who would commit the fraud must be of a character that would allow them to knowingly commit a dishonest, criminal act. If all three of these requirements exist in a given situation, it would be deemed highly likely that a material management fraud has occurred or will occur in the future. If any one of the requirements is absent, then it would be deemed highly unlikely that a material management fraud has occurred or is likely to occur.

The Loebbecke and Willingham model is consistent with prior findings regarding management fraud. Albrecht and Romney (1980) state the following:

> . . . they [managers] become involved [with fraud] because: (1) they are placed in situations where they are faced with a high degree of *situational pressure*; (2) they are faced with attractive *opportunities* to commit, conceal or not be punished for their illegal acts; or (3) they have a low level of *personal integrity or honesty*. These three forces interact to determine whether a person will commit fraud. A person with a high level of integrity and little opportunity and pressure to commit fraud will most likely behave honestly. But criminal acts become increasingly likely as individuals with lower levels of personal honesty are placed in situations with increased pressure or convenient opportunities to commit a crime.

Finally, this agrees with the official position of the AICPA. Paragraph 10 of SAS 53, "The Auditor's Responsibility to Detect and Report Errors and Irregularities" (AICPA 1988) states the following:

> The factors considered in assessing risk should be considered in combination to make an overall judgement: the presence of some factors in isolation would not necessarily indicate increased risk.

These works suggest that management fraud is a multi-dimensional issue. There is no one aspect or predictor that auditors can focus on in developing their audit plan to detect management fraud. Therefore, designing an analytical procedure that uses a combinatorial approach for detecting the audit risk associated with management fraud is necessary.

In the original model presented in Loebbecke and Willingham (1988) the authors based their model for the risk of management fraud on the suggested factors in SAS 53. To test these factors they examined the incidence of these 31 factors in 71 Securities and Exchange Commission's (SEC) Accounting and Auditing Enforcement Releases (AAER's) deemed to contain management fraud. The result of Loebbecke and Willingham (1988) was a modified list of "Red Flag" questions. Some suggested "Red Flags" from SAS 53 were changed and additional indicators were added to the list. In this process, the frequencies of the factors appearance in the AAER's classified them into primary and secondary questions. Finally, the authors tested a subset of 51 AAERs from the 71 AAERs that had sufficient information to validate the Loebbecke and Willingham model. The results showed that only 1 out of 51 (2%) cases did not have any of the three components (Conditions, Motivation, Attitude) as an indicator. More important, in 36 out of 51 cases (71%) all three components were present. These results suggested that the Loebbecke and Willingham model had potential worth exploring further.

Loebbecke et al. (1989) continued the development of the model and further refined the list of "Red Flag" questions for detecting management fraud. The authors collected a survey that questioned partners of KPMG Peat Marwick. Instead of the AAER's, Loebbecke et al. (1989) used 77 actual fraud cases mentioned by the partners in the survey to validate the Loebbecke and

Willingham assessment model. The most important result of Loebbecke et al. (1989) was the finding that in 86% of the fraud cases at least one factor from each of the three components was present. This strongly suggests that having all three components is a robust indicator for the existence of management fraud.

While the results of Loebbecke et al. (1989) suggest validation of the Loebbecke and Willingham assessment model, they provided no useful analytical procedure for planning an audit in the study. In particular, they provided no proportional weighing scheme capable of determining the relative importance of the individual factors indicating management fraud. The development of such an analytical procedure by Bell et al. (1993) is addressed in the following section.

A CASCADED LOGIT APPROACH

In Bell et al. (1993), the authors propose a working discriminant function for the Loebbecke and Willingham conceptual model. They formed their logistic regression model from the questions (Appendix 1) used in the Loebbecke and Willingham (1988) and in Loebbecke et al. (1989) studies. Since the prior works have used and discarded many questions, the numbering system for the questions is confusing. Therefore, the list in Appendix 1 presents a combination of the questions of the above two papers. We renumbered the questions for the sake of clarity. These 47 questions are the same questions used in Bell et al. (1993).

In Bell et al. (1993) the authors started their development of the logistic regression equations by testing the frequency questions occurring in the cases in Loebbecke et al. (1989). Building on this, the authors moved through several logistic regression equations using a cascaded approach to find the combination that would give the strongest discriminant function. To accomplish the above task, the authors added 305 non-fraud cases to the 77 fraud cases collected in Loebbecke et al. (1989). The 305 cases were collected to correspond to the industry composition and other factors of the 77 fraud cases. After adding the 305 non-fraud cases, the authors split their sample into an experimental group of 37 fraud and 143 non-fraud cases and a hold-out sample of 40 fraud and 162 non-fraud cases.

Information used in the logistic regression equations involved a set of questions (Appendix 1) answered by partners of KPMG Peat Marwick. The partner's responses showed whether the individual question existed as a factor for their individual case. The partners coded the answers as one when this factor was in the case and zero if the case did not involve this factor. As the next step, Bell et al. (1993) combined the questions having the strongest t-values obtained from a set of original logistic regression equations into their final logistic regression equation. Their final combination consisted of a cascaded two tier model.

The authors chose Logistic regression (Logit) because the dependent variable was binary; either the firm issued fraudulent financial statements or it issued non-fraudulent financial statements. Logit is one possible choice in dealing with binary information (Hansen et al. (1991); Elliot and Kennedy (1988); Amemiya (1981)). Logit by definition is based on a linear relationship between the variables. Such a relationship may not necessarily exist among the actual variables. An alternative approach to model development that can take advantage of non-linear relationships is Artificial Neural Networks (ANN's).

The results of Bell et al. (1993) show very high classification rates in an area where classification seems difficult. Their results include the correct classification in the estimation sample of 95% of the fraud cases, while still correctly classifying 84% of the non-fraud cases. The results for the hold-out sample were equally impressive. They correctly classified 88% for the fraud cases and 78% of the non-fraud cases. The success of the Bell et al. (1993) model has influenced a Big Six firm to incorporate it into their practice (Eining et al. 1994).

ARTIFICIAL NEURAL NETWORKS

The use of Artificial Neural Networks (ANN's) is undergoing a growth period in accounting and finance literature. Fanning and Cogger (1994), Lin and Lin (1993), Yoon et al. (1993), Liang et al. (1992), Coates and Fant (1991-1992) and Bell et al. (1990) are among the authors who have investigated ANN's ability to find the "rules" associated with the data available in finance and accounting. For complex equations with multiple variables, ANN's strive to improve on conventional problem solving techniques such as discriminant analysis and logistic regression. ANN's create their rules by learning encounters through a learning rule that adapts connection weights of a network in response to input examples and, optionally, the desired outputs of the examples.

A strength of ANN's is their use of inductive logic. The ANN's train on the provided data and can continue to grow as more data become available. ANN's are often suggested for problems in pattern recognition, and in areas where a model-based theory is not available. Smith (1993), Schalkoff (1992), and Hawley et al. (1990) among many others point out that ANN's are most effectively applied to classification, associative memory and clustering. ANN's are constructed using learning paradigms that are analogous to statistics. It is convenient to think of an ANN as a statistical model that does not impose an *a priori* model on the data; unlike discriminant analysis, no limitations are imposed on the complexity of the predictive model.

One disadvantage of standard ANN approach (backpropagation) is that the investigator must decide upon the physical architecture of the network. This is often done by trial and error, varying the number of layers, the number of processing elements in each layer, the nature of the connection patterns, etc. Some heuristics exist to guide the selection of the network architecture, see for example Smith (1993), Zahedi (1993), Klimasauskas (1991a,b,c), Caudill (1991). Simply connected networks of two or three layers of sigmoidal processing elements combined with standard backpropagation estimation of connection weights have done quite well in many settings. Still, the time, effort and lack of an easily obtainable stopping point hinder efforts to use backpropagation ANN's efficiently.

Lohmann (1992) addresses this fundamental issue of efficient design of ANN's.

> The success of neural networks in informational processing tasks depends mostly on well-designed structures of the systems. Structures have to be defined before learning algorithms are applied. Although, in practice, determination of structures takes most of the effort in adjusting a neural system to particular tasks, there are still no effective rules or systematical approaches to determine structures in neural systems.

Lohmann (1992) formalizes the measurement for creating an effective structure to ANN’s. In ANN's, the usual objective is to minimize the prediction error between the target output (T) and the system output (P). A common practice is to minimize the standard mean error (Smith 1993). In a study that has N inputs a quality function Q represents the mean square of all distances as:

where T_i is the target value and P_i is the predicted value.

As with most optimization problems, the user is trying to control several factors in the quality function. Following the logic in Lohmann (1992) the true function of the ANN designer can be expressed in the following formula.

$$Q = G(S, C, W) \tag{3}$$

Q is the quality function that measures how well the ANN is meeting the desired objective, which is usually minimizing the mean square error. S represents the input to the ANN. C is the structure of ANN chosen and W represents the weights assigned in the network. The Q function in (3) captures the various attributes the ANN designer must deal with while finding an efficient ANN. This is not an easy task. Lohmann (1992) in arguing for an evolutionary strategy for constructing the network states the following:

> It makes no significant difference whether one looks at a neural system structure in its graphical representation with a number of units, their connections, and the transition functions of units or at the mathematical equivalent where the structure is some sort of grammar that gives the instructions to calculate a signal from the set of variables and the stimulus. If these instructions do not fit the task, even the most sophisticated efforts to adjust the weights in Eq. [3] will be in vain.

The present study looks at two proposed means of producing evolutionary ANN's, a GANNA processor and an ALN. The first method is the generalized adaptive neural network architectures (GANNA) structure explained in Cogger (1992) and evaluated in Fanning and Cogger (1994). This method develops ANN's with a particular type of processing element where the architecture of the network is itself learned by trial and error. These networks grow to fit the problem at hand and do not require *a priori* specification of the number of layers, numbers of nodes, and other design choices. The particular implementation of GANNA employed in this paper uses processing elements that compute quadratic functions of their inputs. With many such elements connected, of course, there is no restriction on the form of the input-output function.

Briefly, this approach uses an evolutionary mechanism to grow networks. New layers are added provided they offer improved performance, measured in squared output error. To protect against over-specialization, performance is validated on a validation (hold-out) sample within the training set. When performance in this validation sample declines, network evolution ceases.

The merit of the GANNA approach was tested against its ability to predict bankruptcy in Fanning and Cogger (1994). The results suggest that the GANNA processor, AutoNet©, has prediction accuracy superior to standard backpropagation ANN's and is comparable to theory-based models. The strength of a GANNA processor is its speed and its self determination of the correct network structure. The efficient use of standard backpropagation ANN's requires artistic skill and largely depends on trial and error. This can be extremely time consuming and presents difficulties in finding a global minimum. By using GANNA, the user can bypass these problems and use the benefits provided by ANN's. Thus the GANNA method seems to fulfill the criteria suggested in Lohmann (1992).

Also seeming to meet the criteria for an evolutionary network suggested in Lohmann (1992) is the Adaptive Logic Network (ALN). As the tree in the ALN learns which branches are not providing any additional information, it evolves by deleting those branches. Since the rules and heuristics for ALN are not as well-known as for backpropagation, it is felt helpful to describe them briefly. In practice most of these rules and designer choices are transparent and easily operationalized making an ALN easy to use. The ALN was initially described in Armstrong and Gecsei (1979) with the current working version of ALN described in Armstrong et al. (1991). The working version, ATREE 2.7, of the ALN is used for the data processing in this study. It is currently available on Internet at menaik@cs.ualberta.ca [129.128.4.241] while a commercial version is forthcoming. While there have been only a few applications of ALN in business research (Cogger et al. 1994; Cogger 1993a) using ALN, it has been successfully used in applications such as optical character recognition, atomic particle discrimination, prostheses control design, spectra analysis and meat texture analysis.

The purely binary nature of the independent variables motivates the use of an ALN in this study. Since the framework of the ALN is based on Boolean logic, binary variables are a natural fit. However, we strongly note that ALN's are also very efficient and quick at handling continuous variables. To handle continuous variables the program divides the continuous variable into as many parts as the researcher needs and processes each segment as a binary representation (Cogger et al. 1994). The second reason for using ALN's is that they are easy to use. They avoid many of the difficulties associated with constructing networks using other forms of ANN architectures. Being an example of an easy-to-use ANN's, ALN could become an important research tool for individuals frustrated with the time and difficulty associated with traditional ANN's if they are demonstration to have success as classifiers. Finally, in contrast to the GANNA Processor, AutoNet©, which grows the tree from ground up only adding branches if they provide additional information, ALN's start from the top of the tree and prune off branches that do not provide additional information. Given these contrasting styles, it is an interesting empirical question regarding the results produced by each program. In theory they should provide similar results.

ALN's are comparable to standard backpropagation ANN’s in their formation but have significant differences. These differences allow the ALN to achieve equilibrium far quicker than standard backpropagation. This suggests that ALN's could be of great benefit to the auditing process where difficult choices have to be made in a short period. Second, because of the way it processes data, computing power in not a necessity. Therefore, auditors equipped with laptops could easily avail themselves of the benefits of ALN's.

The difference in speed results from the way the ALN processes data. ALN's process data through binary trees with simple logic functions. In particular, the ALN used in this study takes the form of simple logic functions AND, OR, RIGHT and LEFT. Since the data is handled in a tree like fashion the network truncates branches that do not provide any additional information. One example of this process is where a logical zero (one) is inputted to an AND (or) node in the logical tree. Therefore, the output of the node can be computed without using the other input. Another example is the complete truncating of branches when a top node no longer provides any information. This saves the time in processing this branch. These rules combined with other heuristics allow for massive time savings compared with standard processors. The ALN uses a training rule similar to backpropagation but the rules follow Boolean logic since the data is pre-processed into binary input. We briefly describe the ALN in Figure 1 to help explain the terminology involved with an ALN tree. Figure 1 represents a three-input parity function. This tree has 8 leaves, 4 leaf nodes, a total of 7 nodes, and 3 layers.

****

Figure 1 about here

****

The root of the tree represents the output of the network. As explained in Cogger 1993a it is important that the ALN tree be specified that is large enough for the computing task. In practice, the speed of processing in the ALN allows the doubling or tripling the value obtained from formula 4 as the starting value for the tree. By having such a large tree, all possible leaf nodes and potential leaf combination inputs are covered. Thus assuring that the output node has the possibility of being influenced by all variables. The ALN used in this study is described in Cogger et al. (1994), Cogger (1993a, b, c) and Armstrong et al. (1991). Anyone interested in additional details regarding ALN’s are directed to these sources.

One last note on ANN's and their possible advantages to accounting research comes from Etheridge and Brooks (1994). The authors strongly suggest that ANN's show promise in fraud detection for due to their ability to process a large number of data items. Specifically, the authors state the following:

> NNs also may be useful in detecting fraud by patterns that typically accompany fraud. Real and hypothetical examples of fraud may be used to train NNs. Any technique that reduces the probability of major, undetected fraud should assist boards of directors in exercising their governance responsibilities and decrease auditor exposure to litigation.

With that advice and the prior discussions on the advantages on the proposed forms of ANN network construction, the next sections discuss the sample and the results of the empirical tests.

DATA

The data used in this study is identical to the data used to develop the cascaded Logit model in Bell et al. (1993). The answers to the questions in Appendix 1 are coded in binary fashion (yes, no) that allowed the use of standard spreadsheets. The GANNA analysis used AutoNet©, which is available by contacting Dr. Ken Cogger, at the University of Kansas (Tel.: (913)-864-7554), while the ALN analysis used the ATREE 2.7 program.

Since the program used by the GANNA processor, AutoNet©, had an upper limit of 24 inputs, breaking the 47 questions into groups was necessary. A simple algorithm overcomes this input limitation. The GANNA processor, AutoNet©, automatically selects those inputs that are most useful in minimizing the mean squared error function in the individual groupings. The inputs selected in each pass are those questions that are most useful in discriminating between fraudulent and non-fraudulent financial statements. Therefore, the algorithm for the present study was to pass a set of at most 24 questions through the GANNA processor. Then we passed a second separate set through the GANNA processor. We then combined those questions that were winners in these two runs for a third run. This pattern can be repeated as often as necessary.

Using the questions in Appendix 1, the present study finds a discriminant function capable of distinguishing between fraudulent and non-fraudulent companies with superior accuracy to the Logit results. This discriminant function consists of a parsimonious list of "Red Flag" questions. This parsimony should be useful to auditors. For example, Pincus (1987), in a controlled experiment, found that a group of auditors using a check list of “red flags” did poorly in comparison to a group of auditors not using a check list. This was contradictory to standard auditing practices that suggests check lists help as an analytical procedure. Pincus (1987) suggested that the problem with the check list was in its length and type of questions. Therefore, the development of an efficient check list of "Red Flag" questions should help focus auditors to the important issues.

RESULTS

We applied two types of self-generating ANN's, GANNA processor and ALN, to the data used in Bell et al. (1993). For the present study, the ANN’s used only the individual questions in Appendix 1. No attempt to form combination patterns or cascaded nodes as used in the Logit model in Bell et al. (1993) was attempted at this stage. The strong results given by the two ANN's, as shown in Table 2, on such a preliminary data support the use of the two ANN's as analytical procedures. The two ANN’s produce superior results to the Logit model using only the questions without the benefit of any combination information. Therefore, it would be difficult to suggest that given the additional information of combination factors the two ANN's could not achieve better results.

The training sample consists of 150 companies to facilitate use of AutoNet©. The sample consists of 77 fraudulent companies and 305 non-fraudulent companies for a total of 382 companies. This was split into 150 companies in the training sample and 232 in the hold-out sample. This differs from Bell et al. (1993) that had a training sample of 180 companies. While training the ALN using the same training sample as the Logit model is possible, having the ALN use the same training sample as AutoNet© facilities comparisons of the two ANN’s. Given that the two ANN's use the same underlying principles but address the problem from opposite ends, the results are expected to be similar. This study tests this theory.

Another difficulty in comparing the ANN’s to the Logit model is a lack of knowledge regarding which companies are in the training sample of the Logit model. Therefore, a true comparison of the results of the ANN's against the Logit model escapes the present study. The results of any training and hold-out sample endeavor are strongly based on the relationship between information in the training sample and the hold-out sample. If there is a significant difference between the ANN’s training sample and the Logit training sample, this could effect the results. Therefore, any results of the present study need to be viewed with that caveat. However, given the large size of the training sample, it should be safe to assume that the results provide a reasonable comparison.

Given a training sample of 150 companies, it is necessary to decide how many fraud cases to include in the sample. Since the Bell et al. (1993) study chose 37, this is one choice. For exploratory purposes regarding the effectiveness of AutoNet©, various other values for the number of fraud companies in the training sample are used in the present study. As a form of identification of the different models in the tables, the results are labeled by their training splits. For example, the result using the training sample with 37 fraud companies is labeled as 37-113 with 113 representing the number of non-fraud companies.

Table 1 shows the results of the final cascaded Logit presented in Bell et al. (1993). As previously mentioned, these results are extremely good. Table 2 represents the comparison of the results of the three models. This represents the results of Logit and AutoNet© using their best total prediction accuracy as the criterion for choosing the cutoff value. The ALN does not use a cutoff value and its results represent those reported by the ALN model.

****

Tables 1 & 2 about here

****

Again, the results should be viewed with the following caveats. The training sample in the Logit model is larger than the ANN's. This means that the hold-out sample for the ANN's is larger. Having a larger hold-out sample, with a lower expected predicting ability in a hold-out sample, biases the comparison against the ANN's. Also, there is the difference in training data available to the Logit model and ANN's. The ANN's, as previously discussed, only train on the 47 questions without the benefit of any combinations as used in the final Cascaded Logit model. Table 2 strongly suggests that both AutoNet© and ALN can predict the hold-out sample with a higher degree of accuracy than the Logit model even when using the limited data as described above. The results make a strong argument for their use in investigating management fraud.

Another argument for these ANN’s is their rapid determination of the final model. The time invested in developing the cascaded Logit Model is unknown but it is probably much greater than either of the ANN's. Any modeling using Logit requires the examination of many different runs to investigate the influence of the individual variables. Therefore, a cascaded Logit process can safely be assumed as a lengthy process. The time involved with the two ANN's consisted mainly of the manipulation of the spreadsheet data input. The actual computer time for AutoNet© is quite small. For the present study, it was approximately 12 minutes evenly split among the three runs necessary to handle all 47 questions. The computer time for the ALN is even faster, about 3 minutes. The data preparation time for inputting, coding and using the results is roughly 35 minutes for AutoNet© and about 10 minutes for the ALN. The difference is directly related to the need for three runs for AutoNet©. Therefore, the accuracy and speed of the two ANN’s in the present study argue for their use in the detection of management fraud.

The results of the GANNA processor and the ALN in Table 2 are quite similar. Such similar accuracy suggests support for the contention that they are using the same underlying theory. Additional information regarding what variables are used by the ALN should aid in determining the similarity of the two.

The rest of the results and tables relate to AutoNet© alone. This is only the second empirical work using AutoNet©, Fanning and Cogger (1994) being the first. Therefore, there are many unanswered questions regarding certain choices available to the network designer. The discussion and results reported in the following tables should answer several of these questions. This is not to suggest that ALN's are not as potentially important for research as the GANNA processor. Merely that for the present study, we more closely examined the GANNA processor. Future work will develop the ALN in more detail.

Figure 2 shows the model of the 37-113 AutoNet© network reported in Table 2. The nineteen nodes on the left side represent the 19 questions selected from the previous independent runs on the primary and secondary questions. The eleven nodes with lines attached represents the questions (listed in table 7) that AutoNet© selected for this final model. The larger circles in Figure 2 indicate that the weights on these nodes are greater than the smaller circles.

****

Figure 2 about here

****

Tables 3 to 7 report results from various tests of certain aspects of AutoNet©. Table 3 shows the various AutoNet© models and the Logit model at various cutoff points. The results used in Table 2 are highlighted in bold. While we report only the hold-out results, the training samples give similar results. The fact that this pattern is not perfectly consistent is an inherent problem to ANN's since they train to the specific data. Since each of the AutoNet© models is randomly generated, there are different companies in each of the training samples. Therefore, part of the result is attributable to the different data in the individual training samples.

Table 3 suggests that generally, as the proportion of fraud companies included in the training sample increases, the ability of the model to detect the fraud companies in the hold-out sample increased. Exceptions such as the 40-110 model suggest the designer needs to be careful of relying on only one model. The fact that there is only 9 fraud cases in the hold-out sample tempers the results of the 68-92 model. Therefore, the results reported in the present study are from the 37-113 model. It is also the model closest to the Logit training sample, thus simplifying the comparison of the two.

****

Table 3 about here

****

One critical component of the GANNA processor is its self-validation. As discussed previously, AutoNet© uses part of the training sample to validate its choice in variables, number of layers, and strengths of connections. Table 4 reports on the robustness of the GANNA processor to variations in the composition of the self-validation sample. Table 4 clearly suggests that the results are robust to changes in the self-validation sample. For the present study, the self-validation sample is the last 26 companies in the training sample of 150 companies.

As AutoNet© develops its network, it checks its progress against the self-validation model. As a working heuristic, the designer, by roughly maintaining the same split in the self-validation sample as in the total training sample need not worry about the best split. Only the extreme variations such as the 1 fraud and 25 non-fraud cases show any degradation in prediction accuracy. The results weakly suggest that if the designer is more concerned with predicting one choice over the other, increasing the number of this choice in the self-validation sample may increase the predicting accuracy for this choice. The model 58-92 was used instead of the 37-113 model to vary the exploration so that reported results would not be model specific.

****

Table 4 about here

****

Another way to favor one choice over the other is to increase the number in the training sample in relationship to the other choice. Table 5 clearly suggests that the ability to predict fraud cases is much greater when the training sample has a higher proportion of fraud cases. The results of row 1 and 5 in Table 5 (37-20 versus the 37-113) clearly show this fact. However, this is a two-edge sword and costly when used for a sample that is highly weighted with the opposite choice as the hold-out sample as Table 5 shows. Also, limiting the total size of the training sample due to a deficit of one of the choices, such as number of fraud cases, reduces the effectiveness of an ANN to learn. Therefore, the designer should be careful in using this procedure.

****

Tables 5 and 6 about here

****

Table 6 reports several different models prediction accuracies for certain types of questions. The Loebbecke and Willingham model depends on the contention that most management fraud involves at least one indicator from each of the three aspects of fraud: attitude, motivation, and condition. As shown in Table 6, the best results are produced by models using questions from all three categories. However, the results using only questions from each category show minimal degradation in prediction accuracy. This is surprising but may suggest the robustness of ANN’s. It could also be model specific. Finally, there is a higher predication accuracy associated with the attitude questions. This agrees with the model in Bell et al. (1993) that consists mainly of attitude questions.

A check list of questions selected by AutoNet© is presented in Table 7. These are the questions in the model 37-113 with the highest predicting ability as reported in Table 2 and shown in Figure 2. Instead of 47 questions, the auditor now has a reduced set of 11 questions that can predict management fraud with a high level of accuracy. Many questions are related to the person’s attitude. This suggests that the presence of a criminal mind is an important factor for detecting management fraud.

SUMMARY AND CONCLUSION

A difficult task for auditors is the detection of management fraud. One suggested method is the cascaded Logit approach (Bell et al. 1993). The present study adds two new approaches that can produce excellent results. The results suggest that researchers should consider using either of these ANN’s for detecting management fraud. The development time of a cascaded Logit approach is costly and time consuming. Since management fraud is constantly changing, it seems that evolutionary ANN’s are more suited for their detection.

The success of the evolutionary ANN’s in the detection of management fraud argues for the strong possibility of their success in other discrimination venues. As these tools become more widely recognized and the strengths and limitations understood, they will be an strong addition tool for the researcher. We hope that the present study provides such additional information for these ANN’s.

There are several issues in the present study that need additional research. In the present study neither of the ANN's trained on any combination information available to the cascaded

Logit model. If this information is obtainable from Bell et al. (1993) or by additional manipulation of the questions into combinations, then we could make a true comparison between the models. Also, we developed only one ALN for the present study. Increasing the training sample or different data may allow the ALN to increase its predicting ability. Finally, there are several other issues in the Bell et al. (1993) study such as an estimated level of risk not used in the present study. Using this information may help further demonstrate the effectiveness of ANN's in detecting management fraud.

TABLE 1

CASCADED LOGIT MODEL PREDICTIONS
(From Table 8 in Bell et al. 1993)

(Model Applied to Hold-out Sample of
40 Fraud and 162 Non-fraud Cases)

Model Hit Rates at Alternative Cutoff Values

Cutoff	Fraud	Non-fraud
.9	37.5%	98.1%
.8	50.0%	95.7%
.7	50.0%	95.1%
.6	57.5%	93.8%
.5	62.5%	92.6%
.4	70.0%	90.7%
.3	80.0%	88.3%
.2	82.5%	85.2%
.1	87.5%	78.4%
.05	90.0%	70.4%
.025	100.0%	55.6%

Table 2

Percentage Prediction Accuracy of the Three Models[1]

	Fraud	Non-fraud	Total
Logit	70%	91%	87%
AutoNet©	75	91	89
ALN	75	92	90

[1]Based on best cutoff value for Logit and AutoNet© 37-113 model.

Table 3

Percentage Prediction Accuracy on Hold-out Sample for Different AutoNet© Models in Comparison to the Cascaded Logit Model at Different Cutoff Points

Cutoff	0.2			0.3			0.4			0.5		
Model	F[1]	NF	T	F	NF	T	F	NF	T	F	NF	T
Logit	83	85	85	80	88	87	**70**	**91**	**87**	63	93	87
10-140[2]	46	96	82	46	96	82	43	96	81	30	98	78
22-128	58	93	84	58	94	85	58	94	85	55	94	84
37-113	75	89	87	75	91	89	**75**	**91**	**89**	65	93	89
40-110	65	65	65	32	96	86	30	96	86	30	96	86
47-103	80	80	80	77	83	82	77	83	82	57	93	88
58-92	74	78	78	74	80	79	74	80	80	68	90	88
68-82	89	70	71	89	73	74	89	81	81	89	91	91

[1] F = fraud sample NF = non-fraud sample T = total

[2] The models reported are the different concentrations of fraud cases in the 150 company training samples for AutoNet©. For example, the 10-140 model includes 10 fraud cases and 140 non-fraud cases. The percentage prediction accuracy refers to the hold-out sample. Since the total sample consists of 77 fraud cases and 305 non-fraud cases, the hold-out sample is the residual after subtracting the training sample. For the 10-140 model, this is a hold-out sample of 67 fraud cases and 165 non-fraud cases. For the 68-82 model the hold-out sample is 9 fraud and 223 non-fraud. The cascaded Logit model training sample consisted of 37 fraud and 143 non-fraud companies. The holdout sample for the Logit model was 40 fraud and 162 non-fraud companies. The values for the Logit models correspond with Table 1.

Table 4
Prediction accuracy (%) for Different Validation
Samples of AutoNet© at 50 % Cutoff Rate

Model[1]	Training			Hold-out		
F-NF-F-NF	F[2]	NF	T	F	NF	T
58-92-1-25	69%	93%	84%	63%	90%	88%
58-92-4-22	88	91	91	68	79	78
58-92-7-19	86	92	90	63	80	78
58-92-10-16	84	99	95	68	90	88
58-92-13-13	84	97	92	68	89	88
58-92-16-10	79	97	90	68	90	88
58-92-19-7	80	96	90	63	90	88
58-92-22-4	76	100	91	53	91	88
58-92-25-1	81	96	90	63	88	86

[1] The model 58-92-1-25 consists of 58 fraud cases and 92 non-fraud cases. It trains on the first 124 cases and uses the last 26 as a self-validation sample. For the 58-92-1-25 case the self-validation sample is 1 fraud and 25 non-fraud cases. The others also correspond to this system so that the 58-92-25-1 model has 25 fraud cases and 1 non-fraud case in its self-validation sample. The hold-out sample for all models in this table consists of 19 fraud and 213 non-fraud cases.

[2] F = fraud sample NF = non-fraud sample T = total

Table 5
Percentage Prediction Accuracy of Different Sample Mixes of AutoNet© at 50% Cutoff

Model		Training			Hold-out		
F NF		F[2]	NF	T	F	NF	T
37-20[1]		100%	80%	93%	80%	56%	59%
37-40		92	83	87	72	74	74
37-60		89	95	93	65	84	82
37-80		70	96	88	60	91	86
37-113		62	95	87	65	93	88

[1] The models all have 37 fraud companies with differing number of non-fraud companies in the training sample. The corresponding hold-out sample has 40 fraud companies. The number of non-fraud companies in the holdout sample is 305 - number of non-fraud companies in the training sample.

[2] F = fraud sample NF = non-fraud sample T = total

Table 6
Percentage Prediction Accuracy of The Various Categories of Questions Using AutoNet© on the Training Sample of 40-110 Reporting the Hold-out Sample of 37-195 at 50% Cutoff

	Fraud	Non-fraud	Total
Primary	40%	93%	84%
Secondary	3	99	84
Motivational	43	92	84
Conditional	30	93	83
Attitude	40	94	86
Condition and Attitude	38	97	87
Motivation and Condition	30	95	84
Motivation and Attitude	30	95	85
Final	30	96	86

Table 7
Check list of Questions from the AutoNet© 37-110 Model
C = Condition M = Motivation A = Attitude

1) Q2. Does management place undue emphasis on meeting earnings projections or other quantitative targets? M A

2) Q6. Does management display significant disrespect for regulatory bodies? A

3) Q8. Do client personnel display significant resentment of authority? A

4) Q10. Do key mangers exhibit strong personality anomalies? A

5) Q13. Does the client have a weak control environment? C A

6) Q14. Are there frequent and significant difficult-to-audit transactions or balances? C

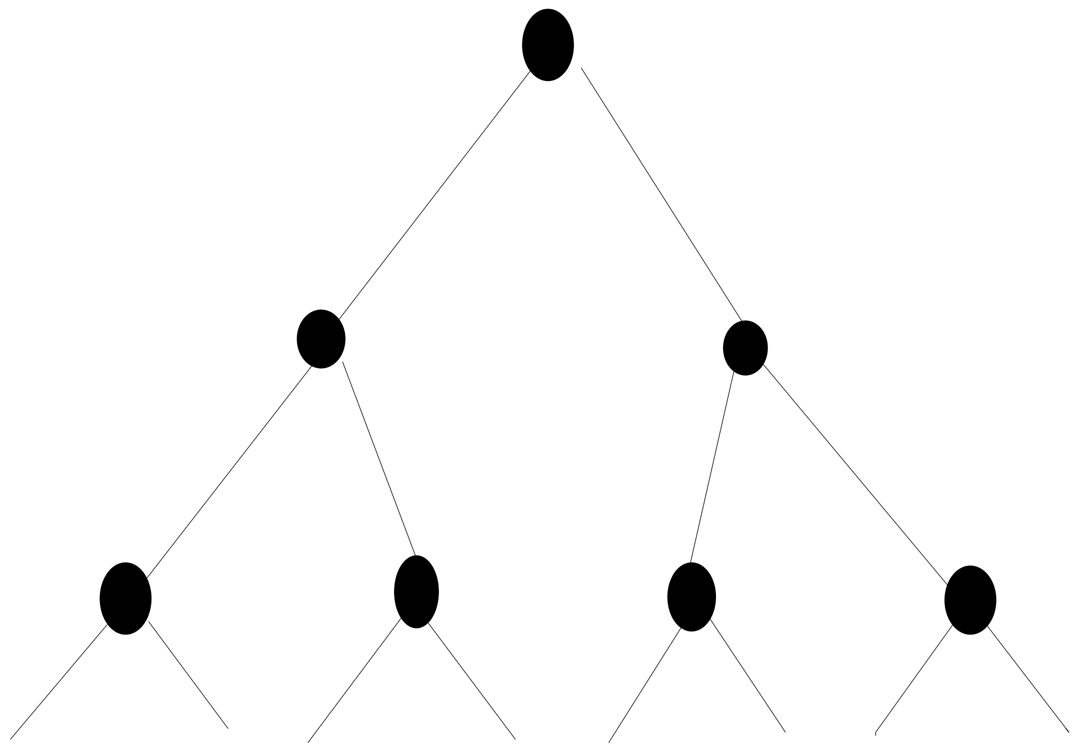
7) Q23. Is the client confronted with adverse legal circumstances? M

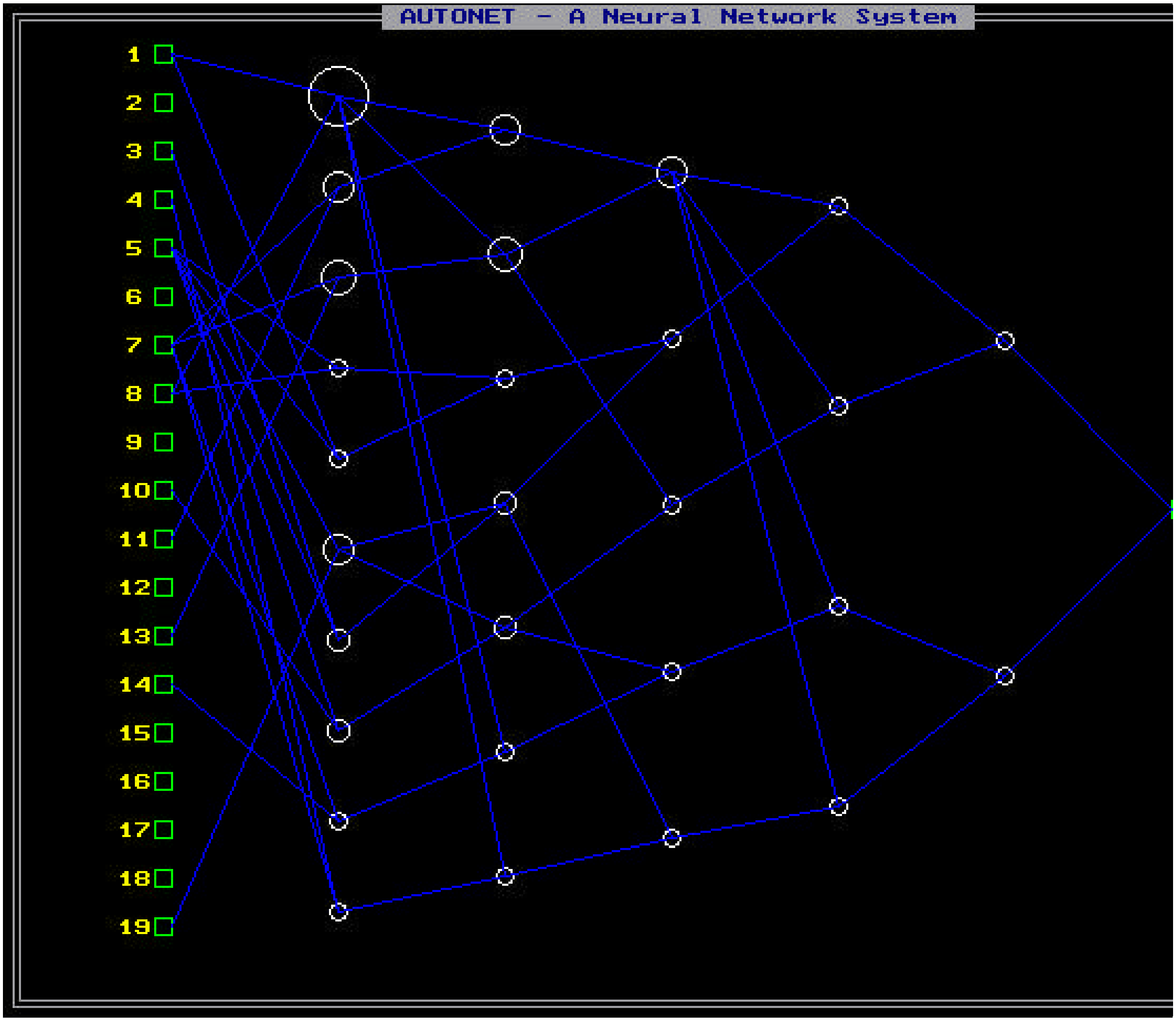
8) Q31. Is the direction of change in the client's industry declining with many business failures? M

9) Q33. Does a substantial portion of management compensation depend on meeting quantified targets? M

10) Q37. Is a significant portion of management's personal wealth in the form of holdings in the client entity? M

11) Q46. Are key managers considered highly unreasonable? A




AUTONET - A Neural Network System
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

APPENDIX 1

Management Fraud Questions
C = Condition M = Motivation A = Attitude

Q1. Are management operating and financial decisions dominated by a single individual? C

Q2. Does management place undue emphasis on meeting earnings projections or other quantitative targets? M A

Q3. Have mangers recently entered into collusion with outsiders? A

Q4. Does your experience with management indicate a degree of dishonesty? A

Q5. Does management display a propensity to take undue risks? A

Q6. Does management display significant disrespect for regulatory bodies? A

Q7. Have mangers lied to the auditors or been overly evasive in responses to audit inquires, or have they shown some other indications of dishonesty? A

Q8. Do client personnel display significant resentment of authority? A

Q9. Is management's attitude toward financial reporting unduly aggressive? A

Q10. Do key mangers exhibit strong personality anomalies? A

Q11. Is this a new client? C

Q12. Is there an attempt to cover up an illegal act? A

Q13. Does the client have a weak control environment? C A

Q14. Are there frequent and significant difficult-to-audit transactions or balances? C

Q15. Is the client a public company?

Q16. Is a significant amount of judgement involved in determining the total of an account balance or class of transactions? C

Q17. Have there been instances of material management fraud in prior years? A

Q18. Is the client's organization decentralized without adequate monitoring? C

Q19. Is the client in a period of rapid growth? M C

Q20. Does the client have solvency problems? M

Q21. Does a conflict of interest exist involving the client entity and/or its personnel? C A

Q22. Do accounting personnel exhibit inexperience or laxity in performing their duties? C

Q23. Is the client confronted with adverse legal circumstances? M

Q24. Is the client's profitability relative to its industry inadequate or inconsistent? M

Q25. Has the client entered into one or a few specific transactions that have a material effect on the financial statements? C

Q26. Has the client entered into a significant transaction or transactions with one or more related parties? C

Q27. Is management and/or key accounting personnel turnover high? C

Q28. Is management inexperienced? C

Q29. Is the client currently or was the client recently involved in a purchase, sale, or merger transaction with another company? C

Q30. Has the company recently entered into a significant number of acquisition transactions? C

Q31. Is the direction of change in the client's industry declining with many business failures? M

Q32. Is the client's industry in a state of rapid change? M

Q33. Are the client's operating results highly sensitive to economic factors (inflation, interest rates, unemployment, etc.)? M

Q33. Does a substantial portion of management compensation depend on meeting quantified targets? M

Q35. Are there adverse conditions in the client's industry or external environment? M

Q36. Is the client subject to significant contractual commitments? M

Q37. Is a significant portion of management's personal wealth in the form of holdings in the client entity? M

Q38. Does management perceive their job is threatened by poor performance? M

Q39. Does management exhibit undue concern with the need to maintain or improve the reputation/image of the entity? M

Q40. Do managers appear to engage in an inappropriate lifestyle, to have personal financial difficulties, or to live beyond their means? M

Q41. Is management's reputation in the business community poor? A

Q42. Has management engaged in frequent disputes with the auditors, particularly about aggressive application of accounting principles that increase earnings? A

Q43. Does management place undue pressure on the auditors, particularly through the fee structure or the imposition of unreasonable deadlines? A

Q44. Has the client engaged in opinion shopping? A

Q45. Do managers display a hostile attitude toward the auditors? A

Q46. Are key managers considered highly unreasonable? A

Q47. Do key managers display a significant lack of strength of character? A

REFERENCES

Albrecht, S. and M. Romney. 1980. Auditing Implications Derived from a Review of Cases and Articles Relating to Fraud. In *Proceedings of the 1980 Touche Ross/University of Kansas Symposium on Auditing Problems.* 101-119.

Amemiya, T. 1981. Qualitative Response Models: A Survey. *Journal of Economic Literature.* 19. December:1483-1536.

American Institute of Certified Public Accountants. 1983. Audit Risk and Materiality in Conducting an Audit. *Statement on Auditing Standards No. 47.* New York: Auditing Standards Board.

_________. 1988. The Auditor's Responsibility to Detect and Report Errors and Irregularities. *Statement on Auditing Standards No.* 53.New York: Auditing Standards Board.

Armstrong, W. and J Gecsei. 1979. Adaption Algorithms for Binary Tree Networks. *IEEE Transactions on Systems, Man & Cybernetics.* 9. 276-285.

________, W., A. Dwelly, J. Liang, D. Lin, S. Reynolds. 1991. Learning and Generalization in Adaptive Logic Networks. In *Artificial Neural Networks, Proceedings of the 1991 International Conference on Artificial Neural Networks.* 1173-1176.

Bell, T., S. Szykowny, and J. Willingham. 1993. *Assessing the Likelihood of Fraudulent Financial Reporting: A Cascaded Logic Approach.* Working Paper, KPMG Peat Marwick. Montvale N.J.

________, Ribar G. and J. Verchio. 1990. Neural Nets Versus Logistic Regression: A Comparison of Each Model's Ability to Predict Commercial Bank Failures. In *Proceedings of the 1990 Deloitte & Touche/University of Kansas Symposium on Auditing Problems.* 29-53.

Caudill, M. 1991. Neural Network Training Tips and Techniques. *AI Expert.* January:56-61.

Coates, P. and L. Fant 1991-2. A Neural Network Approach to Forecasting Financial Distress. *The Journal of Business Forecasting*. Winter:9-12.

Cogger, K. 1993a. Neural Forecasting of Stock Market Turning Points. *Working Paper.* University of Kansas.

_________. 1993b. Binary Trees and Artificial Neural Networks. A Nexus. *Working Paper.* University of Kansas.

_________. 1993c. Minimum Coverage Trees in Adaptive Logic Networks. *Working Paper.* University of Kansas.

_________. 1992. A Learning Paradigm for Neural Network Architecture. *Euro XII/TIMS XXXI Joint International Conference, Helsinki*. June. Working Paper No.237, School of Business, University of Kansas, Lawrence, KS, 66045.

_________, P. Koch, and D. Lander. 1994. Neural Network, ARIMA, and Linear Regression Models: Comparing the Ability to Forecast Through the Market Crash of 1987. *Working Paper.* University of Kansas.

Eining, M., D. Jones, and J. Loebbecke. 1994. An Experimental Examination of the Impact of Decision Aids on the Assessment and Evaluation of Management Fraud. In *Proceedings of the Twelfth Annual USC\Grant Thorton Audit Judgement Symposium.*

Elgin, P. 1992. Huge Liability Judgements Pressure CPAs to Raise Prices. *Corporate Cashflow.* 13. July:12-14.

Elliot, J. and D. Kennedy. 1988. Estimation and Prediction of Categorical Models in Accounting Research. *Journal of Accounting Literature.* 7. 202-242.

Etheridge, H. and R. Brooks. 1994. Neural Networks: A New Technology. *The CPA Journal.* 64. 3. March:36-39, 52-55.

Fanning, K. and K. Cogger. 1994. A Comparison Analysis of Artificial Neural Networks for Financial Distress. *International Journal of Intelligent Systems in Accounting Finance & Management*. *Forthcoming.*

Freedman, E. 1993. Enough is Enough: Why C&L Sued Phar-Mor. *Accounting Today.* May 3:54

Fuerman, R. 1992. The Accounting Profession's Litigation Crisis. *Ohio CPA Journal.* October:39-40.

Hansen, J., J. McDonald and J. Stice. 1991. Artificial Intelligence and Generalized Qualitative-Response Models: An Empirical Test on Two Audit Decision Domains. *Decision Sciences.* 23. 3. May/June:708-723.

Hawley, D., J. Johnson, and D. Raina. 1990. Artificial Neural Systems: A New Tool for Financial Decision-Making. *Financial Analysis Journal*. November-December:63-72.

Holland, K., L. Light, and M. Galen. 1993. Big Six Firms are Firing Clients. *Business Week.* March 1:76-77.

Klimasauskas, C. 1991a. Applying Neural Networks: Part I: An Overview. *PCAI.* January/February:30-33.

_________. 1991b. Applying Neural Networks: Part II: A Walk Through the Application Process. *PCAI.* March/April:27-34.

_________. 1991c. Applying Neural Networks: Part III: Training a Neural Network. *PCAI.* May/June:20-24.

Liang, T., J. Chandler, I. Han, and J. Roan. 1992. An Empirical Investigation of Some Data Effects on the Classification Accuracy of Probit, ID3, and Neural Networks. *Contemporary Accounting Research.* 9. Fall:306-328.

Lin, F. and M. Lin. 1993. Analysis of Financial Data Using Neural Nets. *AI Expert.* February:36-41.

Loebbecke, J., M. Eining. and J. Willingham. 1989. Auditor's Experience with Material Irregularities: Frequency, Nature, and Detectability. *Auditing: A Journal of Practice & Theory.* Fall:1-28.

_________, and J. Willingham. 1988. Review of SEC Accounting and Auditing Enforcement Releases. Unpublished Working Paper.

Lohmann, R. 1992. Structure Evolution in Neural Systems. In *Dynamic, Genetic, and Chaotic Programming.* John Wiley & Sons.

O'Malley, S. 1993. Legal Liability is Having a Chilling Effect on the Auditor's Role. *Accounting Horizons.* 72. June:82-87.

Palmrose, Z. 1987. Litigation and Independent Auditors: The Role of Business Failures and Management Fraud. *Auditing: A Journal of Practice and Theory.* 6. 2:90-102.

__________. 1991. Trials of Legal Disputes Involving Independent Auditors: Some Empirical Evidence. *Journal of Accounting Research.* 29. Supplement:149-185.

Pincus, K. 1989. The Efficacy of a Red Flags Questionnaire for Assessing the Possibility of Fraud. *Accounting Organizations and Society.* 14 (½):153-163.

Schalkoff, R. 1992. *Pattern Recognition: Statistical, Structural and Neural Approaches.* John Wiley & Sons.

Smith, M. 1993. *Neural Networks For Statistical Modeling.* Van Nostrand Reinhold. New York.

Srivastava, R., K. Harrison and S. Dutta. 1993. The Audit Risk Model Under the Risk of Fraud. *Working Paper 252.* University of Kansas.

St. Pierre, K. and Anderson, J. 1984. An Analysis of the Factors Associated with Lawsuits Against Public Accountants. *The Accounting Review.* 59. 2:242-263.

Sullivan, J. 1992. Litigation Risk Broadly Considered. *Proceedings of the 1992 Deloitte & Touche/University of Kansas Symposium on Auditing Problems.* 49-59.

Yoon, Y., G. Swales Jr.,and T. Margavio. 1993. *A Comparison of Discriminant Analysis Versus Artificial Neural Networks.* Journal of Operational Research Society. 44. 1:51-60.

Zahedi, F. 1993. *Intelligent Systems for Business: Expert Systems with Neural Networks.* Belmont, Ca. Wadsworth.