Risk Assessment in Internal Auditing: A Neural Network Approach

Sridhar Ramamoorti[1], Andrew D. Bailey, Jr[2]* and Richard O. Traver[2]
[1]Arthur Andersen LLP, Chicago, IL, USA
[2]University of Illinois at Urbana-Champaign, Champaign, IL, USA

ABSTRACT **Risk assessment is a systematic process for integrating professional judgments about relevant risk factors, their relative significance and probable adverse conditions and/or events leading to identification of auditable activities (IIA, 1995, SIAS No. 9). Internal auditors utilize risk measures to allocate critical audit resources to compliance, operational, or financial activities within the organization (Colbert, 1995). In information rich environments, risk assessment involves recognizing patterns in the data, such as complex data anomalies and discrepancies, that perhaps conceal one or more error or hazard conditions (e.g. Coakley and Brown, 1996; Bedard and Biggs, 1991; Libby, 1985). This research investigates whether neural networks can help enhance auditors' risk assessments. Neural networks, an emerging artificial intelligence technology, are a powerful non-linear optimization and pattern recognition tool (Haykin, 1994; Bishop, 1995). Several successful, real-world business neural network application decision aids have already been built (Burger and Traver, 1996). Neural network modeling may prove invaluable in directing internal auditor attention to those aspects of financial, operating, and compliance data most informative of high-risk audit areas, thus enhancing audit efficiency and effectiveness. This paper defines risk in an internal auditing context, describes contemporary approaches to performing risk assessments, provides an overview of the back-propagation neural network architecture, outlines the methodology adopted for conducting this research project including a Delphi study and comparison with statistical approaches, and presents preliminary results, which indicate that internal auditors could benefit from using neural network technology for assessing risk. Copyright © 1999 John Wiley & Sons, Ltd.**

Keywords: internal audit; risk assessment; internal control; neural networks

BACKGROUND AND MOTIVATION

Risk assessment represents a critical aspect of internal audit planning. As a systematic process for the identification and analysis of relevant risks threatening the achievement of an entity's objectives, risk assessment is helpful for assessing and integrating professional judgments about probable adverse conditions and/or events (COSO, 1992). The process of risk assessment includes identification of auditable activities, identification of relevant risk factors, and determination of their relative signifi-

* Correspondence to: Andrew D. Bailey, Jr, Department of Accounting, University of Illinois at Urbana-Champaign, 1206 South Sixth Street, Champaign, IL 61820, USA

CCC 1055-615X/99/030159-22$17.50

cance (IIA, 1995, SIAS No. 9). An efficient and effective audit program is responsive to risk assessment, and is designed to ensure that proper controls are in operation that minimize or eliminate risk and exposure (Sawyer and Dittenhofer, 1996).

Risk assessment in auditing involves pattern recognition because an unexpected deviation or variation is symptomatic of risk. Significant deviations, error distributions, or anomalous data and discrepancies indicate underlying risk (Brown and Solomon, 1990, 1991; Libby, 1985; Bedard and Biggs, 1991). Unfortunately, the internal auditor works in data-rich environments where the sheer volume of data encountered renders the risk-assessment process overwhelming and unmanageable (Hackenbrack, 1992). It is in such contexts that neural network technology can assist in directing the auditor's hypothesis generation and search process to those aspects of financial, operating, and compliance data that are most informative about error or hazard conditions (Coakley and Brown, 1996; Bedard and Biggs, 1991).

Neural network technology represents an ensemble of powerful techniques that can be used for modeling, forecasting, signal processing, and pattern recognition (Chen, 1996). Over the last decade, neural networks have been used in diverse advanced applications such as defense systems, process control, oil and gas exploration, character and speech recognition, industrial inspection, and antibody detection (Coleman, 1991; Ward Systems Group, 1993). Real-world business applications currently exist for credit authorization screening, corporate loan portfolio risk evaluation, and fraud detection (American Express, Chase Manhattan, Mellon Bank, First USA Bank), mortgage risk assessment (J.P. Morgan), financial and economic forecasting (Merrill Lynch, Salomon Brothers, Citibank, World Bank), tracking stock market movements and futures trading (LBS Capital Management, Gerber Baby Foods), data mining and target marketing (Veratex Corporation, Spiegel), and bankruptcy prediction (Trippi and Turban, 1996; Burger and Traver, 1996; Deboeck, 1994). In all these and other real-world business applications, neural network technologies appear to have significantly outperformed conventional technologies both in terms of cost savings as well as improved quality and productivity (cf. Green and Choi, 1997). The power of neural networks derives from their ability to model nonlinear relationships and their robustness in dealing with noisy and incomplete data commonly found in real-life situations (Tam, 1994). Given this new technology's impressive track record and future promise, it is not surprising that applications of neural network modeling in business, including the auditing domain, have been vigorously championed (Coakley and Brown, 1996; Massaad and Winkler, 1995; Doherty, 1993; Garceau and Foltin, 1995; Coderre, 1993; Zarowin, 1995; Hansen and Messier, 1991).

Neural networks can do more than help internal auditors cope with 'information overload'. In the context of expanded definitions of risk, performing risk assessment becomes a non-trivial task that requires professional experience and expertise. Considerable research in cognitive science has shown that expertise in any domain is gained only after undergoing significant instruction, engaging in practical applications, and through several years of experience (Chi *et al.*, 1988; Shanteau, 1988). Presently, we expect that only seasoned internal auditors with broad and varied professional experience ('experts') will exhibit superior performance on diagnostic tasks such as risk assessment in audit planning. An important aspect of professional experience is the superior ability displayed in associative reasoning and (error) pattern recognition (Massaad and Winkler, 1995; Tubbs, 1992). Neural networks act on data much as experienced experts act on data, by detecting an existing, if often hidden, underlying organization. Thus, neural network technology can help knowledge transfer and provide less experienced individuals with the impounded expertise previously only attained with a significant amount of training and experience.

Neural network technology is based on the premise that, given a set of input variables and associated expert judgments, it is possible for the neural network model to extract the essential input–output relation from a reasonably large number of exemplars. We only need one or more experts' final assessments on a task, and it is left to the network to infer the combining process used by the expert and build a suitable model that mimics expert judgment fairly well. Of course, as a 'function mapping device', the neural network model is only as good as the quality of the input–output association pairs presented to it (cf. GIGO: 'garbage in, garbage out' syndrome that is well known in the computer science literature). Power's (1995) questioning of 'assumed expertise' is relevant: in the absence of objective task performance measures, there is a natural tendency to focus on and bolster the **process** of judgment. Although this altered emphasis on process may be viewed as an adequate rationalization, from a modeling perspective, and in the context of this study, the quality of expert performance constitutes an upper bound for neural network models.

While the knowledge–performance relationship in auditing continues to be investigated by cognitive scientists and judgment and decision-making researchers (for reviews, see Arnold and Sutton, 1997; Ashton and Ashton, 1995), computer or cognitive modeling, including the use of expert systems, remains a viable approach to capturing expertise (Srivastava *et al.*, 1996; Van Dijk and Williams, 1990; Vasarhelyi, 1990). Cognitive modeling using advanced technology and aimed at developing decision aids is an inherently interdisciplinary endeavor that draws from diverse disciplines such as cognitive psychology, artificial intelligence, management and computer science, and statistics (Brown and Eining, 1997). Cognitive/computer modeling utilizes 'knowledge engineering' techniques (e.g. protocol analysis) to capture the knowledge of one or more experts in order to make decisions that are comparable to experts in quality and approach (see Vasarhelyi, 1995; Bailey *et al.*, 1989; Bouwman and Bradley, 1997). If the design and implementation of decision aids is grounded in theory and employs sound methods, such systems may add to both developmental and empirical knowledge (Mock and Vertinsky, 1985; Brown and Eining, 1997). By undertaking such research, we hope to obtain insights into the actual process of risk assessment and a perhaps better understanding of the normative aspects of risk and its modeling potential for improving decision quality.

The data for this research study consisted of 'real-world' qualitative and quantitative risk factor information about the academic/administrative departments at the University of Illinois at Chicago. We conducted a Delphi study featuring experienced internal auditors who used the risk factor information to assign risk rankings for 141 departmental units. As part of our modeling, these risk factor values (inputs) along with the overall risk rank (target or output) for each department, were fed into three neural network vendor software packages to 'train' and 'test' the models. Subsequently, the risk rankings from the neural network models were compared with the Delphi rankings as well as with statistical models. Our preliminary analyses indicate that we were quite successful in building neural network models that are suitable for risk ranking applications in internal audit settings. We should note that neural network models are currently being used by the Office of University Audits at the University of Illinois to perform risk assessment and initial results have been more than satisfactory. A significant reason for carrying out this study was to support the process of continuous improvement at the University.

Among academic researchers, the ability of neural networks to make bankruptcy predictions and detect management fraud successfully has captured the lion's share of research attention (e.g. Fanning and Cogger, 1998; Green and Choi, 1997; Fanning *et al.*, 1995; Fletcher and Goss, 1993; Tam and Kiang, 1992; Odom and Sharda, 1990; BarNiv and Hershbarger, 1990; Bell *et al.*, 1990), and almost no research exists on using neural networks for risk assessment in internal auditing (for a notable exception in the context of external auditing, see Davis, 1996). Following the recommendation made by Green and Choi (1997), we include qualitative information in our broadened risk conceptualization and neural network implementation and thus extend research in this area. The findings from this study indicate that internal auditors could utilize neural networks to perform risk rankings of auditable units. We are also able to identify several fruitful lines of research that would productively utilize this newly available technology.

The balance of this paper is as follows. The next section defines risk in general and discusses current approaches to risk assessment. The third section provides an overview of neural network architectures with special emphasis on the Multi-Layer Perceptron (MLP) using backpropagation. The fourth section describes the quantitative and qualitative risk factor data and their collection. The fifth section details a Delphi study involving the participation of experienced internal auditors. The sixth section presents preliminary results. The seventh section compares neural network models with traditional statistical models such as multiple linear regression and logistic regression. The final section provides a summary discussion and conclusion.

RISK ASSESSMENT IN INTERNAL AUDITING

The Concept of Risk in Internal Auditing

Although risk arises in virtually all fields of endeavor, each field appears to develop its own unique ways of assessing and handling them. There are basically two dimensions to risk: the amount or severity of an unfavorable outcome (magnitude), and the probability of occurrence (frequency) (Moore, 1983). In this paper, we adopt Shakespeare's (1996, p. 4) definition of risk as 'the compound cost estimate of loss frequency, loss severity (including public perception of harm), and risk control measures'. Fischhoff *et al*. (1983) suggest three decision-making approaches to characterizing and measuring risk: relying on **professional judgments** of technical experts to devise solutions; searching for historical precedents to guide future decisions, i.e. **bootstrapping**; and employing theory-based **formal analysis** for modeling problems and calculating the best decision. We will be primarily concerned with the use of professional judgment as well as neural network modeling in addressing the issue of risk assessment in internal auditing.

Internal auditors are concerned with the various risks facing an organization. Organizational risks include anything from lost market share, environmental liabilities, customer dissatisfaction, low employee morale, violation of laws and regulations, to fraudulent financial reporting (Colbert, 1995). In the context of these broadly defined organizational risks, the internal auditor evaluates the controls established by management to assess their adequacy in appropriately limiting the occurrence of adverse conditions or mitigating their impact. Paragraph 7 of SIAS No. 9 (IIA, 1995) enumerates several risk factors that are relevant for internal auditors.[1] While some of these factors are capable of quantification, others, such as 'competency, adequacy, and integrity of personnel', can at best elicit only subjective, qualitative judgments.

In addition to the risk of misstatements in financial statements, Shakespeare (1996) enumerates other types of risks divided into five categories that are more relevant to the internal auditor (see Table 1). We use the risk categories presented in Table 1 to justify the inclusion of qualitative and quantitative risk factor input variables in our risk-ranking models.

Contemporary Approaches to Risk Assessment in Internal Auditing

Shakespeare (1996) defines the following six steps in a risk assessment process: (1) develop loss scenarios; (2) identify exposures and controls; (3) define risk categories; (4) assess frequency and severity of possible losses; (5) develop risk control costs; and (6) rank exposures. In step 1, with the objective of antici-

[1]For instance, risk factors include the ethical climate and pressure on management to meet objectives; competency, adequacy, and integrity of personnel; financial and economic conditions; impact of customers, suppliers, and government regulations; date and result of previous audits; degree of computerization; adequacy and effectiveness of the system of internal control; management judgments and accounting estimates; and geographical dispersion of operations (partly based on Patton *et al*., 1982).

Table 1. Definitions developed for illustrative purposes for five risk categories. Adapted with permission from Shakespeare (1996)

Risk category	Definition	Representative potential losses
Operational	Risk association with equipment breakdowns, operator errors, product quality, damage to facilities	• Cost of modifying process • Cost of corrections • Cost of plant repair, business interruption
Fraud/criminal	Risks derived from opportunities in system and processes for employees/non-employees to steal or commit other fraudulent or criminal acts	• Loss of funds • Recovery costs
Legal/professional	Risks associated with becoming the target of lawsuits because of actual or alleged actions	• Legal fees/court costs • Diversion of management time and attention
Image/marketing	Risks associated with declining public and individual student perceptions as a result of actions of the university	• Unwanted adverse attention and visibility • Incremental costs to re-establish image/reputation
Compliance	Risks associated with failure to comply with applicable laws and regulations	• Penalties/fines • Increased regulatory scrutiny

pating possible adverse effects and designing preventive controls, loss scenarios are developed by experienced and knowledgeable people familiar with the organization's financial, operating, and compliance aspects, relying on both historical precedent, 'bootstrapping' (Fischhoff *et al*., 1983), as well as professional judgment (expert knowledge, imagination, and experience). In step 2, these previously developed loss scenarios prove very helpful in determining exposures to risk and identifying loss control measures. Risk category definition is undertaken in step 3. In addition to labeling different classes of risk, this includes identifying the responsible personnel within the organization who may be used as a resource in designing risk controls and estimating risk control costs. In steps 4 and 5, when assessing frequency and severity of possible losses, dollar value and other ranges used should reflect impact levels appropriate to the organization; this consideration extends to the range of risk control cost estimates as well. Finally, in step 6, using available information on the severity, probability, and cost of installing preventive controls, a risk ranking may be attempted. The ranking and evaluation of identified risks enables management to determine their significance and optimize expenditures designed to control risks.

McNamee (1996) defines risk assessment as a three-step process consisting of risk identification (what the risks are), risk measurement (how big are the risks), and risk prioritization (which risks are the most important). He suggests enumerating risk factors (see IIA, 1995, SIAS 9) for 'macro risk assessment' and using weighted or sorted matrices for 'micro risk assessments'. Macro risk assessment looks at the risks facing the enterprise as a whole (major goals, products, processes, issues, etc.), while 'micro risk assessment' concerns the internal auditor's audit program and testing strategy, that is, which areas to audit and the extent and detail needed to be applied.

The SAS No. 47 (AICPA, 1983) audit risk model (AR = IR*CR*DR) is used in practice by external auditors. However, its narrow focus on financial statement misstatement risks makes it inappropriate for use by internal auditors whose concern extends broadly to organizational risks. Conceptually, however, it is possible to relate the fundamental ideas of inherent, control, and detection risks to cover financial, operational, compliance and other aspects of an organization.

In the context of information systems auditing, Gallegos *et al*. (1987) point out that auditors have traditionally relied upon audit judgment and intuition, dollar risk estimation using a risk formula, identifying and weighting risk attributes, and/or the use of computer software packages (e.g. IST/RAMP, PANRISK, ESTIMACS) to compute dollar risk. In their evaluation of these methods, Gallegos *et al*. (1987) first note that no current approach guarantees the correct prediction of audit risk, and second, that 'ease of use' is an important characteristic that determines whether an approach is used by practicing auditors. In particular, they warn that if auditors do not possess 'a convenient, structured method' they will be tempted to revert to more informal but readily available procedures such as intuition and professional judgment.

Going by the professional literature in auditing (e.g. Colbert, 1995; McNamee, 1996; Shakespeare, 1996), it appears that internal auditors deal with a broader set of risks than do external auditors, and consequently, must adopt a risk assessment methodology that is responsive to their specific goals. Accordingly, the next section outlines our approach in developing a risk assessment model that is used for the study.

Risk Assessment Model

As technology and globalization become the key drivers of business competitiveness, information about outside influences and other interconnected and interdependent entities yields a host of qualitative information (e.g. political/regulatory climate, innovation in the marketplace, EDI partnerships, etc.). In the context of a knowledge-based economy, the significance of assets not valued easily or carried on the books (e.g. reputation, information, human skills and experience, etc.) is also increasing. While consideration of these qualitative risk factors is critical, collecting information on them is time and labor intensive. Quantitative risk factor values are more readily available and also contain relevant and important risk information. For effective risk assessment in internal auditing, both qualitative and quantitative risk factors must figure prominently in the evaluation of risks faced (cf. Green and Choi, 1997; Ramamoorti and Hoey, 1998). SIAS No. 9 (IIA, 1995) mentions several such factors, e.g. dollars at risk, liquidity of assets, management competence and integrity, internal control effectiveness, time since last audit, etc. Accordingly, our risk assessment approach involves a combination of the quantitative as well as qualitative factors that underlie risk. One of the issues to be dealt with subsequently is the manner in which the quantitative risk factor-based rankings and the qualitative risk factor-based rankings are to be combined to achieve the best results. Figure 1 is a conceptual depiction of the risk assessment model adopted for the study.

CONDENSED REVIEW OF NEURAL NETWORKS

A neural network is a statistical information-processing mechanism composed of numerous, distributed processing units or nodes that perform simultaneous computations and communicate using adaptable interconnections called 'weights' (Davis, 1996; Lippmann, 1987). It resembles the brain in two respects: (1) knowledge is acquired by the network through a learning process, and (2) interneuron connection strengths known as synaptic weights are used to store knowledge (Haykin, 1994, p. 2). By mimicking the processing characteristics of the brain, neural networks are able to achieve knowledge representations based on the fast retrieval of large amounts of information, and the ability to recognize patterns based on experience (Medsker *et al*., 1996). Further, their adaptive nature, allowing them to 'learn by example', makes them very useful in application domains where the problem to be solved is ill-structured or ill-understood but where training data is available or can be made available (Hassoun, 1995). Neural networks operate on numeric representations, use non-linear differentiable functions, and have the capacity to generalize and learn from noisy or incomplete data (Swingler, 1996). Taken in combination, these features make neural networks not only a very distinct computing paradigm but also very attractive for practical applications in a variety of fields.

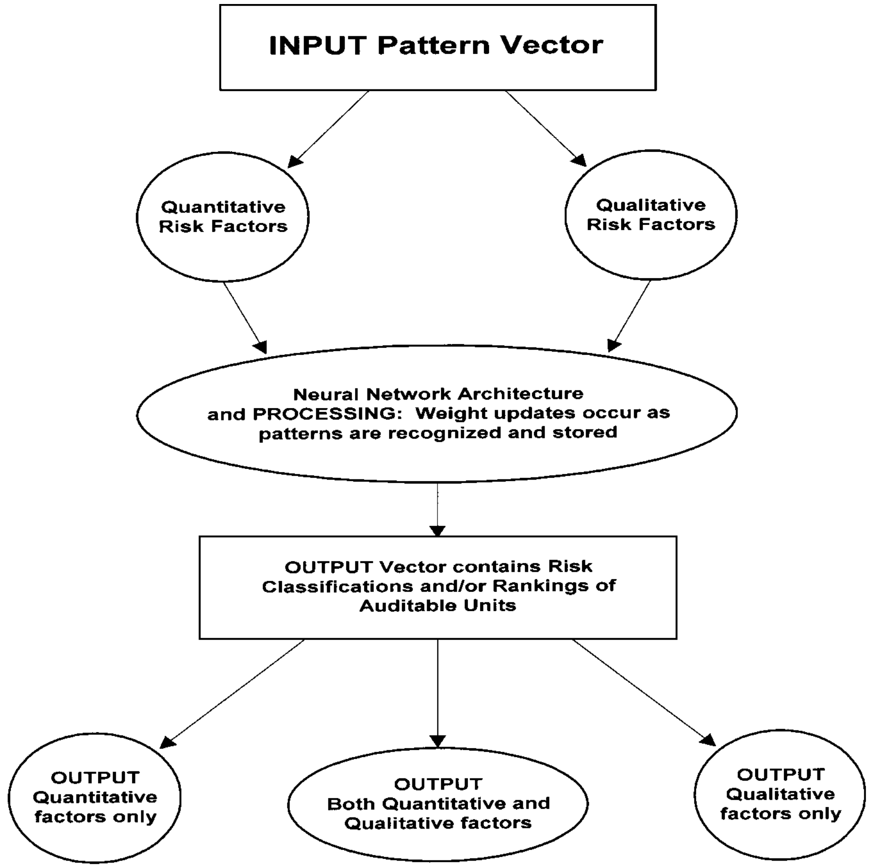

Figure 1 Conceptual overview of risk assessment model used in this study (risk factor input, neural network processing, and risk ranking output).

Neural Network Architectures

Neural network architecture typically refers to the design of neural networks, the number of nodes in different layers and the pattern of connectivity among them. Some well-known neural network architectures are multi-layer perceptrons, radial basis function networks, recurrent networks, and self-organizing systems such as the Kohonen self-organizing map, and Grossberg's adaptive resonance theory (Haykin, 1994). We limit the ensuing discussion to a description of a multi-layer perceptron that uses backpropagation. (For more detailed accounts of the biological foundations of neural computing architectures and the relevance of neural network technology to accounting and auditing, see Massaad and Winkler, 1995; Davis, 1996; and Green and Choi, 1997).

Multi-layer Perceptron

The following description of a multi-layer perceptron is based on Swingler (1996). The mathematical model of an MLP (see box below for a formal description) consists of a set of sensory units that constitute the input layer ($X_1 \ldots X_1$), one or more hidden layers of computation nodes ($h_1 \ldots h_m$), and an output layer of computation nodes ($Y_1 \ldots Y_n$). The units in the hidden layer link the inputs to the outputs. The hidden units play an important role: they extract the most useful features from the input vector and use them to predict values on the output vector. Each unit has an associated activation flowing into it from the units in the previous layer; these activation values are multiplied by the strength of the associated weight, which may be positive or negative. Because input flows from numerous inputs into each one in the next layer, these products must be summed and passed through a transfer function that 'squashes' the summed inputs into a [0, 1] range. Finally, there may also exist a 'bias' term typically set to one, that connects to all units except the input layer and whose function is to draw the inputs to the hidden and output units into the correct range for the squashing function to work smoothly. Figure 2 shows a multi-layer perceptron.

Backpropagation Algorithm

A backpropagation network is a multi-layered, feedforward neural network. Each unit in a layer is connected in the forward direction to every unit in the next layer (see Figure 2). Within this network structure, the error back-propagation algorithm (BP) is based on the error-correction learning rule popularized by Rumelhart and McClelland (1986). Basically, the error backpropagation process consists of two passes through the different layers of the network: a forward pass and a backward pass. In the forward pass, an activity pattern (input feature vector) is applied to the sensory nodes of the network, and its effect propagates through the network. During the forward pass the synaptic weights of the network are all fixed. During the backward pass, on the other hand, the synaptic weights are all adjusted in accordance with the error-correction rule: the actual response of the network is subtracted from a desired response to produce an error signal which is then propagated backward through the network, against the direction of synaptic connections. The synaptic weights are adjusted so as to make the actual response of the network move closer to the desired response. Error backpropagation is thus an efficient way of calculating the derivative of a function at the output of an MLP with respect to internal variables; its two main uses are in network training and doing sensitivity analysis (Guiver, 1997).

Structure of a multilayer perceptron: a formal description
(Partly based on lecture notes by Dr John Guiver (1997) at an advanced neural computing seminar)

A neuron is an information-processing unit (sometimes called a 'processing element' or PE) that is fundamental to the operation of a neural network. A neuron contains a set of synapses or connecting links, each of which is characterized by a weight of its own. Weights are connections of varying strength which carry activation information between network units. A signal x_j at the input of synapse j connected to neuron k is multiplied by the synaptic weight w_{kj}. There is also an activation function, a mathematical function which takes the weighted activation values coming into a unit, sums them, and translates the result to a position along a given scale, i.e. 'squashes' the summed value to within a given range (e.g. typically using a sigmoid or a hyperbolic tangent function). Neural computing derives its power in a deceptively simple way: the burden imposed by the overall processing task is shared among the massive interconnections among the PEs which have adaptive parameters called 'weights'. Neural computation is thus based on the fundamental concepts of distributed, adaptive, and nonlinear computing (NeuroDimension, Inc., 1995).

Using mathematical notation, let i, j, and k represent indices for the input, hidden, and output layers, respectively. Further, let s denote the summation value of a processing element (PE), y the output value of a PE, and f the transfer function at a PE. Then, at the jth PE in the output layer, we have:

$$S_j = \sum_{k=1}^{K} w_{jk} y_k$$

where $y_j = f(s_j)$ and x_i is the value of a PE in the input layer.

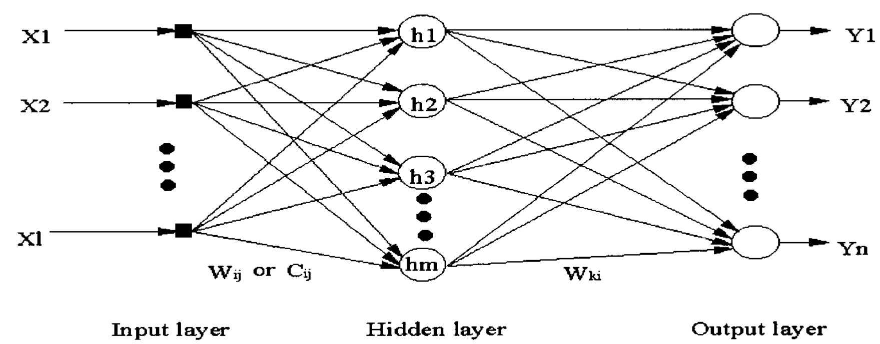

Figure 2 Structure of a multi-layer perceptron

For this study, we used the backpropagation algorithm featured in three neural network vendor software packages: NeuroShell by the Ward Systems Group in Frederick, MD, Predict by NeuralWare, Inc., a Division of Aspen Technologies, based in Pittsburgh, PA, and NeuroSolutions by NeuroDimension, Inc., based in Gainesville, FL (Ward Systems Group, 1993; NeuralWare, 1995; NeuroDimension, 1995).

RISK FACTOR DATA

The data for the study are from the Chicago campus of the University of Illinois. In total, there were 45 input variables, 26 qualitative risk factors and 19 quantitative risk factors. Table 2 presents the list of qualitative and quantitative risk factors used in this study.

Most of the quantitative data were already available in a Spreadsheet format (Microsoft Excel) compatible with vendor software requirements. These data were downloaded from the University Financial and Administrative System (UFAS). A considerable amount of our data are qualitative in nature: these data were obtained in assessments elicited from experienced Office of University Audits staff in connection with the implementation of an earlier IIA product auditMASTERPLAN (Boritz, 1986). The quantitative risk factor values were primarily dollar amounts (e.g. academic salary, tuition and fees, etc.). The qualitative risk factor values were elicited from audit staff using a pre-defined scale from 0 to 9 (0 = not applicable; 1 = lowest risk rating; 9 = highest risk rating).

DELPHI STUDY

Eliciting Expert Risk Rankings

To implement the backpropagation network (i.e. to train the network) we must have 'correct answers' as target output. We used the risk rankings assigned to various academic/administrative departments by experienced auditors. These risk rankings are based on a separate consideration of qualitative risk and quantitative risk factors that apply to specific academic/administrative departments.

Three experienced auditors from the University of Illinois at Chicago participated in a Delphi study designed to elicit risk rankings for 141 academic/administrative departments. The three Delphi participants, denoted by the initials, M, J, and S, are intimately familiar with the University of Illinois—Chicago environment and we believed they were able to assess the impact of the qualitative and quantitative risk factors and arrive at a risk ranking on a scale of 1 to 5 (with 1 = lowest risk ranking possible and 5 = highest risk ranking possible) for each department. The three authors presented the Delphi study format to the auditor-participants prior to the risk ranking exercise but no guidance was provided on how to rank the departments.

Table 2. Quantitative and qualitative risk factors used as input variables in model

Input variables in qualitative database (26 variables)		
1. Cash Risk	10. Unsolicited Interest	19. Computer Equipment
2. Other Assets Risk	11. Pronouncements	20. Sensitive Information
3. Number of Employees	12. Cash Amount	21. Purchase Relationships
4. Budget Size	13. Regulatory Impact	22. Audit Interval
5. Dollar Throughput	14. Special or Specific Risk	23. Audit Recommendations
6. Transaction Volume	15. Revolving Fund Accounts	24. External Audit
7. Last Audit	16. Grants and Contracts	25. Policies & Procedures
8. University Focus	17. Inventory	26. Management Quality
9. Solicited Interest	18. Business Administration	

Input variables in quantitative database (19 variables)		
1. Academic Salary	8. Sales and Service Revenue	16. Refunds and Discounts
2. Nonacademic Salary	9. Total Inventory Dollars	17. Cash Receipts
3. Wages	10. Cash/Over/Under	18. Cash Disbursements
4. Total Non-Payroll Expense	11. Lease Payments	19. Movable Equipment
5. Total Grants and Contract Dollars	12. State Budgeted Funds	
	13. Local Budgeted Funds	
6. Tuition and Fees	14. Total Gift Dollars	
7. Miscellaneous and Interest Income	15. Student Financial Aid	

Each Delphi study participant was given a laptop for entering data. The screen presented a Microsoft Excel spreadsheet that listed all the departments and their ratings on all the quantitative risk factors. (Qualitative data were provided in another session.) Each participant then picked academic/administrative departments that could be given a risk ranking of 1 or 5 corresponding to the minimum and maximum risk assignments. They then proceeded to assign rankings of 2, 3, or 4 to each academic/administrative department, as applicable. After this first round, the auditors switched laptops so that each auditor had access to another's ranking (but not his or her own ranking). With this additional information about peer ranking, they again entered their risk rankings for all 141 departments. The third round provided each auditor with the rankings of the other two participants.

Due to time constraints, risk ranking information based on qualitative risk factor information was separately collected from the Delphi participants by one of the research team members during the following week. The gap of one week between the elicitation of quantitative- and qualitative factor-based risk rankings had at least two effects. First, the Delphi participants did not suffer from the adverse effects of fatigue on such a cognition-intensive task. Second, the risk rankings based on the qualitative risk factors were insulated from any 'carryover effects' (positive or negative transfer).

One way to use the Delphi study risk ranking data is to take a 'grand average' of all the final rankings. Alternatively, different raters may be weighted differentially. We conducted our study using an unweighted (raw scores, i.e. M = J = S = 0.333) and two weighted schemes (single expert, i.e. M = 1, J = S = 0, and unequal weighting, i.e. M = 0.6, J = S = 0.2, respectively). From a modeling perspective, we found the (0.6,0.2,0.2) weighting scheme to be the most satisfactory. Because one of the Delphi study participants was significantly more familiar with the task performed, we employed an unequal weighting scheme, with the weights reflecting hierarchical rank and number of years of experience. For the weighting scheme (0.6, 0.2, 0.2), this produces the following formula, where R^* represents the output vector of risk rankings used for backpropagation training purposes:

$$R_M\,(0.6) + R_S\,(0.2) + R_J\,(0.2) = R^*$$

Table 3 Pairwise inter-judge correlations for Delphi study participants

Quantitative risk factors ($n = 141$)		Qualitative risk factors ($n = 141$)	
Pair	Inter-judge correlation	Pair	Inter-judge correlation
M & S	0.759	M & S	0.779
S & J	0.587	S & J	0.443
J & M	0.607	J & M	0.473

No theoretically correct weighting formula exists. Professional judgment and model sensitivity analyses are the only practical means of choosing a functional weighting scheme.

Statistical Analysis of Delphi Results

The risk rankings on a scale of 1 to 5 for each participant (denoted by M, J and S) and across all 141 departments was first correlated. Table 3 reports that inter-judge correlations pertaining to risk ranking assessments made on quantitative risk factors was on average, higher than comparable correlations pertaining to qualitative factors.

Additional descriptive statistics about the participants' risk rankings are presented in Table 4. Each Delphi participant assigned, on average, significantly ($p < 0.001$, paired t-test) higher risk rankings for the quantitative risk factors.[2] This finding is confirmed by the distribution of modal values for each Delphi participant. For Delphi participant S, the quantitative mode of 5 against the qualitative mode of 2 is particularly striking. Table 4 provides some comfort and support for using a weighting scheme rather than using raw risk rankings; in forecasting, expert opinions gathered to make predictions are typically weighted to reflect their competence (e.g. Myung *et al.*, 1996).

Table 4 Risk ranking sums, means, and modes for Delphi study participants

Quantitative risk factors ($n = 141$)				Qualitative risk factors ($n = 141$)			
	Σ ranking	Mean	Mode		Σ ranking	Mean	Mode
M	381	2.7	2	M	267	1.9	1
S	502	3.6	5	S	333	2.4	2
J	429	3.0	3	J	304	2.2	1

One reason for the significantly higher mean rankings for quantitative risk factors could be their salience: as noted before, most of the quantitative risk factors appeared in their original values (e.g. \$1,200,000) as compared to the qualitative risk factors which took on scaled values (e.g. 0 to 9). It could be argued that raw dollar amounts provide an indication of 'size' and larger size may frequently be regarded as 'material' and therefore important for risk assessment purposes. Qualitative risk factors, being independent of size, do not furnish such an information cue and may therefore, tend to be discounted. Another reason could be that internal auditors are more comfortable dealing with hard, quantitative data that are typically not subject to the same vagaries in interpretation as soft, qualitative data. At this time, we can only surmise that quantitative risk factors were somehow more salient and therefore elicited higher risk rankings although it is possible that the variables selected for the qualitative risk factors were not as important in conveying 'riskiness' as the quantitative risk factors were.

The pairwise correlation between the qualitative factor-based and quantitative factor-based risk rankings for the Delphi study was only $r = 0.51$, indicating no strong linear relationship. This correlation value suggests that the qualitative and quantitative risk factors may be tapping into different dimensions of risk and, therefore, need to be considered in conjunction with each other in making an overall risk assessment. This observation raises definitional issues about input risk factors and

[2]The paired two-sample t-test for means is appropriate wherever a natural pairing of observations in samples occurs, e.g. the Delphi participants first performed risk rankings using quantitative risk factors and subsequently, using qualitative risk factors. While the test does not assume that both populations from which the data sets are drawn have equal variances (homoscedasticity assumption), it does require that the two samples have the same number of observations. The t-tests were performed using $\alpha = 0.05$ and $df = 140$; the means were found to be significantly different in all cases.

output or organizational risk assessment that is discussed later.

PRELIMINARY RESULTS

Variable Selection

Like stepwise regression techniques in statistics, variable selection allows the neural network model to retain variables having a high degree of diagnosticity while discarding those that do not have predictive power. Among the three neural network software packages used, it turned out that only two of them, NeuroShell and NeuralWare (Predict), had specific variable selection algorithms that automatically did 'pre-processing' including the necessary variable transformations and the elimination of input variables with low predictive power. For back-propagation models from these two packages, the eventual number of variables selected to construct the models were in the 7 to 18 range out of a total possible number of 45 qualitative and quantitative risk factors presented as input variables (see Table 2).[3]

As discussed later, we made use of logistic regression and multiple linear regression techniques. The stepwise procedure was used in both cases and was able to pick a small but efficient set of predictor variables. Following Breiman *et al*. (1984), we also used Classification and Regression Trees (CARTS) to partition the data set into thresholds that take the shape of branches on a tree and where successive refinements of the qualitative and quantitative risk factors chosen can be readily seen. This procedure effectively performs variable selection through 'pruning of regression trees', i.e. the most important variables which have predictive power are chosen to partition the data set. We report the most significant variables from the first three layers in the tree branching that partitions the risk factor data according to the CARTS procedure, the stepwise regression techniques, and the neural network models, in Table 5.

Backpropagation Model Results

Using the output values obtained from the Delphi risk rankings we developed backpropagation neural network models. We used a training set size covering 70% of the data and a 'hold-out' data or test size set of 30%. Backpropagation network development still requires extensive experimentation, parameter selection, and human judgment (Hammerstrom, 1993). We also made several decisions regarding parameter choices and values based on software package default settings, advice from software vendor technical personnel, reading books and manuals, and finally, based on our limited experience with the software. Choices of this kind are frequently required to be made; we chose based on the best available information. Future studies may consider other options. This whole area of parameter selection and fine-tuning is a matter of interest in the neural network research community.

The results obtained from training backpropagation networks using the Delphi study output values for all three vendor software packages, i.e. NeuroShell, Neural Ware and NeuroSolutions are now presented. Appearing in Tables 6 though 11 below are the risk ranking predictions for the top 25 academic/administrative departments corresponding to (1) Delphi Study Assessments, (2) NeuroShell (NShell), (3) NeuralWare (Predict), and (4) NeuroSolutions (NSol), respectively. We are primarily interested in the top 25 riskiest departments denoted 'A' through 'Y', because, as maintained by experienced practitioners, efficiency precludes consideration of a larger number of departments for audit coverage. From a practical perspective, we were less concerned with the risk ranking accuracy of lower risk departments and elected to focus on the top 25 ranked departments.

From Tables 6 and 7, it appears that, for

[3] With as many as 45 input variables, there was some concern whether the smaller sample of 141 departments was adequate to yield an acceptable number of degrees of freedom and reliable parameter estimates. However, after the variable selection procedure eliminated 25 or more input variables, our concern that the sample departments may be 'much too small relative to the numbers of variables under study' was diminished (Hays, 1994, p. 723).

Table 5. Key risk factors identified through variable selection

Model	Qualitative risk factors	Quantitative risk factors
Neural networks	Cash Risk, Revolving Fund Accounts, Regulatory Impact, Inventory, Computer Equipment, Purchase Relations	Total Non-payroll, Grants & Contracts, Sales and Service Revenue
Classification and regression trees (CARTS)	Special Risks, Computer Equipment, Transaction Volume, Sensitive Information, Inventory, and Other Assets	Total Non-payroll, Equipment, Grants & Contracts, Local Funds, Sales and Service Revenue, and Non-academic Salaries
Logistic regression	Solicited and Unsolicitied Interest, Grants & Contracts, Regulatory Impact	log (1 + Sales Service Revenue)
Multiple regression	Policies and Procedures	log (1 + Sales and Service Revenue), log (1 + Grants & Contracts), log (1 + Local Funds), log (1 + Non-academic Salaries)

Table 6 Comparison of backpropagation quantitative results with Delphi risk

Dept. key	Delphi	Dept. key	NShell	Dept. key	Predict	Dept. key	NSol
A	5	D	5.00	H	4.93	H	4.57
B	5	A	4.99	D	4.81	O	4.33
C	5	C	4.96	C	4.76	R	4.31
D	5	H	4.87	R	4.73	B	4.26
E	4.6	B	4.84	K	4.72	K	4.20
F	4.4	M	4.69	O	4.71	P	4.09
G	4.4	V	4.65	N	4.71	BW	3.98
H	4.4	K	4.48	B	4.55	M	3.94
I	4.4	R	4.37	M	4.52	D	3.89
J	4.4	L	4.33	P	4.39	N	3.69
K	4.2	S	4.29	A	4.39	AF	3.63
L	4.2	N	4.28	S	4.38	L	3.55
M	4.2	O	4.27	BD	4.37	AS	3.48
N	4.2	J	4.11	I	4.21	V	3.28
O	4.2	Q	4.09	T	4.13	I	3.04
P	4.2	I	4.05	AA	4.06	AQ	3.02
Q	4.2	P	4.05	L	4.04	AR	2.97
R	4.2	T	3.91	J	4.02	AB	2.97
S	4.2	AB	3.82	V	4.00	A	2.94
T	4.2	AR	3.74	AH	4.00	AK	2.87
U	4	AG	3.73	AX	3.96	F	2.86
V	4	AD	3.69	F	3.95	AP	2.85
W	4	BW	3.67	Q	3.94	BK	2.84
X	4	AX	3.63	AZ	3.87	T	2.83
Y	4	AH	3.62	BA	3.82	DH	2.81

Table 7 Evaluation criteria for backpropagation quantitative results

	NShell—quant	Predict—quant	NSol—quant
R-squared	0.6162	0.5961	0.3040
% Delphi overlap	72%	76%	60%

the quantitative risk factors, all three vendor software packages do a reasonably good job of capturing the top 25 risky departments as determined in the Delphi study.

Tables 8 and 9 show that, for the qualitative risk factors, all three vendor software packages do a modest job of capturing the top 25 risky departments as determined in the Delphi study.

From Tables 10 and 11, it appears that, overall, all three vendor software do quite well capturing the top 25 risky departments as determined in the Delphi study.[4] This finding indicates that neural networks can function effectively in aiding internal auditors in the task of risk assessment. Further, because the neural network models so developed embed the pattern recognition expertise of experienced practitioners, they constitute a valuable component of efforts at knowledge acquisition from experts and knowledge transfer to novices.

[4]Note that the target values used for this backpropagation network were computed as the average of the separately assessed qualitative and quantitative risk factor based risk rankings elicited from the Delphi participants and weighted (0.6, 0.2,0.2) as explained earlier.

Table 8 Comparison of backpropagation qualitative results with Delphi

Dept. key	Delphi	Dept. key	NShell	Dept. key	Predict	Dept. key	NSol
A	5.00	A	4.99	A	4.78	A	4.72
B	4.80	J	4.80	E	4.61	Q	4.53
C	4.80	U	4.56	B	4.50	E	4.47
D	4.40	Y	4.36	J	4.46	G	4.42
E	4.40	R	4.26	H	3.96	H	4.31
F	4.40	K	4.22	C	3.93	BS	4.31
G	4.40	C	4.21	M	3.88	J	4.25
H	4.40	AV	4.17	D	3.84	B	4.22
I	4.20	I	4.12	G	3.77	F	4.20
J	4.20	AN	4.09	R	3.72	I	4.16
K	4.00	H	4.08	BA	3.47	AV	4.15
L	4.00	O	4.06	P	3.35	CA	4.14
M	3.80	G	3.96	V	3.34	U	4.06
N	3.80	L	3.88	AA	3.33	C	4.05
O	3.80	AO	3.86	AI	3.23	L	3.95
P	3.60	N	3.85	AG	3.16	CM	3.90
Q	3.60	AY	3.84	T	3.15	BY	3.87
R	3.60	AG	3.80	L	3.14	AG	3.83
S	3.60	E	3.78	AS	3.08	BO	3.83
T	3.40	F	3.77	I	3.08	BQ	3.81
U	3.40	Q	3.77	X	2.98	D	3.79
V	3.40	CS	3.77	U	2.91	AZ	3.67
W	3.20	AM	3.75	W	2.79	AS	3.65
X	3.20	CM	3.73	AJ	2.73	AW	3.62
Y	3.00	BS	3.73	AM	2.72	O	3.62

Table 9 Evaluation criteria for backpropagation qualitative results with Delphi

	NShell—qual	Predict—qual	NSol—qual
R-squared	0.6488	0.8573	0.4497
% Delphi overlap	64%	72%	56%

Table 10 Comparison of backpropagation qualitative/quantitative results with Delphi

Dept. key	Delphi	Dept. key	NShell	Dept. key	Predict	Dept. key	NSol
A	5.0	B	5.00	D	4.86	A	4.78
B	4.6	A	5.00	E	4.70	K	4.70
C	4.5	AF	4.97	B	4.68	G	4.52
D	4.4	F	4.89	Q	4.40	M	4.51
E	4.3	Q	4.87	L	4.32	D	4.45
F	4.3	D	4.65	F	4.31	E	4.38
G	4.3	S	4.47	G	4.20	C	4.35
H	4.2	E	4.40	AF	4.07	F	4.25
I	4.1	V	4.29	N	4.06	B	4.25
J	4.1	BG	4.22	O	4.05	Q	4.21
K	4.1	M	4.16	H	3.99	AY	4.17
L	4.0	Z	4.15	J	3.89	H	4.12
M	4.0	U	4.11	C	3.84	U	4.11
N	4.0	O	4.11	M	3.84	BB	4.10
O	3.9	N	4.10	I	3.83	BY	4.04
P	3.9	H	4.09	T	3.71	J	4.02
Q	3.8	AJ	4.04	P	3.65	N	3.97
R	3.8	L	4.02	AM	3.64	R	3.89
S	3.6	AW	3.88	U	3.58	Z	3.82
T	3.5	G	3.84	A	3.58	P	3.82
U	3.4	R	3.83	V	3.54	BH	3.80
V	3.4	J	3.82	AW	3.48	AD	3.74
W	3.4	AB	3.81	AB	3.41	AW	3.72
X	3.3	C	3.80	Z	3.41	AT	3.69
Y	3.3	I	3.80	K	3.35	W	3.65

Table 11 Evaluation criteria for backpropagation qualitative/quantitative results

	NShell—QualQuant	Predict—QualQuant	NSol—QualQuant
R-squared	0.7742	0.8834	0.5504
% Delphi overlap	80%	84%	72%

COMPARISON WITH STATISTICAL MODELS

Multiple Linear Regression and Logistic Regression

In general, neural networks are able to relax classical assumptions regarding independence of input variables as well as underlying parametric distribution requirements (Rumelhart and McClelland, 1986). In addition, they are capable of performing linear and nonlinear modeling, and handling complex hierarchical or other intermediate data relationships (Davis, 1996). As such, in line with prior research that makes comparisons of neural network models with statistical models (e.g. Yoon *et al.*, 1993; Balakrishnan *et al.*, 1994; Odom and Sharda, 1990; Bell *et al.*, 1990), we compared our neural network results with two statistical models: multiple linear regression and logistic regression. Both statistical procedures were performed with stepwise variable selection that systematically eliminates variables that have lower diagnosticity while retaining those with higher diagnosticity.

Logistic regression is of particular interest to us because the standard transfer function used in backpropagation networks, the sigmoidal function, is equivalent to the logistic function.[5] Because logistic regression works with a dichotomous response variable, we defined the following dichotomy: risk rankings ≥4 ('high risk') and risk rankings ≤3 ('low risk'). Our justification for this split is that none of the top 25 riskiest departments in the Delphi rankings had a risk rank of 3 or lower. The logistic regression chose four qualitative risk factor variables and only one quantitative risk factor variable. The proportion of the deviance explained (a measure roughly equivalent to R-squared) was a modestly high 0.6045.

Multiple linear regression, involving the stepwise procedure, resulted in the choice of four quantitative variables and only one qualitative variable. Logarithmic transformation of the quantitative risk factors led to making an effective use of them as regression variables. The multiple R-squared for the linear regression was again high, at 0.7925.

Both these statistical procedures compare favorably with the neural network results in terms of the proportion of the variance in the risk rankings explained by the quantitative and qualitative risk factors. Moreover, the statistical results from logistic regression and stepwise multiple linear regression are interesting, if only because they emphasize different dimensions of risk depending on the nature of the response variable. The logistic regression uses a dichotomous response variable: 'high risk' or 'low risk', while the multiple linear regression uses a continuum of risk rankings ranging from 1 to 5. The sensitivity of the (variable selection) results to the choice of response variable scaling is an interesting finding and needs to be taken into account in future research.

The fact that well-understood statistical techniques such as logistic regression and multiple linear regression do a reasonably good job of modeling our data set should not be surprising. Indeed, the statistical foundations of neural network models are increasingly being recognized in the technical literature (Swingler, 1996; Bishop, 1995; Sarle, 1994; Smith, 1993). However, for this study, we need more fine-tuned measures to compare the performance of statistical and neural network models. We chose Kendall's Tau, a non-parametric measure of association that makes no distributional assumptions, for this purpose.

Additional Analyses

As mentioned earlier, we are focused on the rank ordering of the top 25 riskiest departments. Thus, while we may not care about whether the 38th ranked department in the Delphi study is ranked 138th in one of the test models, we are very much interested in knowing whether the 5th ranked department in the Delphi study has been relegated to the 26th or lower by one of the test models. All the previous analyses have focused on the unordered risk rankings. In order to compute the ordered ranking matches of the top 25 riskiest depart-

---

[5]In logistic regression, the predictor variables can be a mix of continuous, discrete and dichotomous variables (Tabachnick and Fidell, 1996). The dichotomous outcome or response variable, Y', in logistic regression, is the probability of having one outcome or another based on a nonlinear function of the best linear combination of predictors. The response function, varying from 0 to 1 as $-\theta$ varies from $-\infty$ to $+\infty$, is given by:

$$Y'_i = e^{\theta}/(1 + e^{\theta})$$

where Y'_i is the estimated probability that the ith case ($i = 1,2,\ldots,n$) is in one of the categories and θ is the standard linear regression equation:

$$\theta = A + B_1X_1 + B_2X_2 + \ldots + B_kX_k$$

with constant A and regression coefficients, B_j, and predictors, X_j, for k predictors ($j = 1,2,\ldots,k$).
This linear regression equation, in effect, produces the logit or log of the odds:

$$\ln (Y'/1 - Y') = A + \Sigma\, B_j\, X_{ij}$$

In other words, we estimate coefficients using maximum likelihood to determine the best linear combination of predictors, for a linear regression equation that represents the natural logarithm of the ratio of the probability of being in one category divided by the probability of being in the other category.

ments with the Delphi rankings, we chose the Kendall's Tau measure of rank correlation.[6]

To perform the Kendall's Tau computation, the Delphi sample of 141 departments was divided into the top 25 riskiest departments (ranked 1 through 25) and the remaining departments were lumped together and given a common rank of 26 (there were thus 116 such departments). Because of this operationalization and the inevitable 'ties' that resulted, the range over which Kendall's Tau took values was (–0.102, 1.000) rather than (–1.000, 1.000).[7]

[6]The Kendall's Tau measure is based on the following computation:

(# of concordant pairs – # of discordant pairs)/total # of pairs

The denominator, 'total # of pairs' is, of course, adjusted for any 'ties'. Observation pairs, (X_i, Y_i) and (X_j, Y_j) are said to be concordant if the difference between X_i and X_j is in the same direction as the difference between Y_i and Y_j. Similarly, observation pairs, (X_i, Y_i) and (X_j, Y_j) are said to be discordant if the difference between X_i and X_j is not in the same direction as the difference between Y_i and Y_j. When either the X's or Y's are equal, the observation pairs are neither concordant nor discordant (Daniel, 1978).

[7]Suppose we denote $n = 141$ and $r = 25$ to represent the total number of departments to be risk-ranked and the top 25 riskiest departments respectively. For the purposes of focusing on the top 25 departments, we merely lumped together all non-top 25 riskiest departments into a common rank of 26. This procedure caused the Kendall's Tau lower and upper bounds to deviate from the theoretical (–1.000, 1.000) range.

In the case of a perfect match of the top 25 between the Delphi rankings and a particular model, we would have the following calculation to compute the upper bound (the numerator is the difference between concordant and discordant pairs, while the denominator represents the total number of pairs less 'ties'):

$$\left\{\binom{r}{2} + r(n-r)\right\} / \left\{\binom{n}{2} - \binom{n-r}{2}\right\} = (2rn - r - r^2)/(2rn - r - r^2) = 1$$

The equivalent numerical computation produces 3200/3200 = 1, the upper bound for Kendall's Tau measure.

Where none of the top 25 riskiest departments constitutes a match, we need to similarly compute a lower bound (again, the numerator is the difference between concordant and discordant pairs, while the denominator represents the total number of pairs less 'ties'):

$$\left\{\binom{r}{2} - r^2\right\} / \left\{\binom{n}{2} - \binom{n-r}{2}\right\} = (-r - r^2)/(2rn - r - r^2)$$
$$= -(1 + r)/(2n - r - 1)$$

The equivalent numerical computation produces –325/3200 = –0.1015, the lower bound for our operationalization of Kendall's Tau measure.

Because the lower bound is only –0.102, we expect a positive bias in the values of the Kendall's Tau measure for our operationalization of the risk rankings (see Table 12).

Table 12 presents a summary of the Kendall's Tau measure of association for the qualitative, quantitative and combined risk factor inputs across different models. For simplicity of presentation, only the diagonal elements of the matrix are presented, except for the logistic and multiple linear regression based rankings. As explained in footnote 6, there is a positive bias for Kendall's Tau values because the lower bound of –1.000 was moved up to –0.102 as a consequence of our particular operationalization of risk ranking assignments. However, it should be noted that our operationalization of the Kendall's Tau measure with a focus on the top 25 riskiest departments does not impair the validity of the comparisons with the statistical models. In fact, it slightly biases our hypotheses against finding the neural network models to be better predictors of Delphi risk rank

Table 12 Kendall's Tau rank correlation measures across models

Model used	Inputs used	Delphi rankings		
		Quant	Qual	Quant & Qual
Predict	Quantitative	0.713		
	Qualitative		0.670	
	Quant & Qual			0.750
NShell	Quantitative	0.645		
	Qualitative		0.544	
	Quant & Qual			0.676
NSol	Quantitative	0.484		
	Qualitative		0.486	
	Quant & Qual			0.631
Linear regression		0.607	0.360	0.475
Logistic regression		0.651	0.502	0.612

orderings. Moreover, subsequent analyses incorporating all 141 departmental units (non-truncated, full sample) produced qualitatively identical results.

From Table 12, based on the Kendall's Tau measures, NeuralWare's Predict appears to make the best predictions in terms of ordered ranking matches with the Delphi study risk rankings. Although the two statistical techniques, multiple linear regression and logistic regression, show relatively high and positive Kendall's Tau values for the quantitative model, their performance is significantly worse for the qualitative model. The better performance of the logistic regression model is partially explained by noting that the sigmoidal transfer function used in neural networks is equivalent to the logistic function used in logistic regression (Bell *et al.*, 1990; Swingler, 1996).[8] Nevertheless, Table 12 illustrates the slightly superior performance of neural network backpropagation models when compared with traditional statistical models using the Kendall's Tau measure. Because we were unable to perform significance tests on the Kendall's Tau measure, we proceeded to compute standardized scores (z-scores) using the Kendall's Tau values from Table 12, along with the modified lower and upper bounds (–0.102, 1.000). This produced Table 13.

Table 13 confirms our earlier remarks that NeuralWare's Predict backpropagation model is the best performer; in one case the Kendall's Tau z-score is over 1 standard deviation above the mean. Table 13 also shows that, in some cases, the performance of logistic and multiple linear regression is below the mean, across all models. Again, subsequent analyses involving the full sample produced qualitatively identical results.

These statistical analyses increase our confidence in the ability of neural networks to furnish reasonably good models for risk ranking applications in internal auditing. Overall, we remain quite enthusiastic about the performance exhibited by these neural network models and hope that future research in this area will be able to further optimize their performance.

Table 13 Standardized Kendall's Tau rank correlation measures across models

Model used	Inputs used	Delphi rankings		
		Quant	Qual	Quant & Qual
Predict	Quantitative	0.819		
	Qualitative		0.582	
	Quant & Qual			1.02
NShell	Quantitative	0.445		
	Qualitative		–0.111	
	Quant & Qual			0.615
NSol	Quantitative	–0.442		
	Qualitative		–0.431	
	Quant & Qual			0.368
Linear regression		0.235	–1.124	–0.491
Logistic regression		0.478	–0.324	0.263

DISCUSSION AND CONCLUSION

Information technology in general, and artificial intelligence tools in particular, are perceived as being critical in the future role of accountants as information providers and risk management consultants (Elliott, 1992). Integrating these into audit practice will facilitate the development of more effective training programs and decision aids that feature advanced technologies (see Brown and Eining, 1997). Neural networks, a technology inspired by human brain architec-

[8]The sigmoid function, also known as the logistic function, is given by:

$$f(x) = \frac{1}{1 + e^{-x}}$$

The function takes values from [0,1]. Its derivative conveniently turns out to be: $f'(x) = f(x)(1-f(x))$, that is, it can be written as an output of the function itself.

The hyperbolic tangent (tanh) function is given by:

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

This function takes values from [–1,1]. Its derivative, again, conveniently turns out to be: $f'(x) = (1 + f(x))(1-f(x))$, which can be written as an output of the function itself.

ture, represents one of the latest advances in artificial intelligence techniques. They have been tremendously successful in a variety of real-world applications and contain the promise to become part of the standard toolkit of the internal auditor of the twenty-first century. This paper reports the results of a study aimed at evaluating the potential of neural networks for risk assessment in internal auditing.

We began with the premise that risk assessment is complex and difficult especially because the internal auditor is faced with large amounts of both qualitative and quantitative data. A key aspect of risk assessment is the detection of patterns and trends that are indicative of noticeable discrepancies, significant anomalies, and exceptional or error conditions. Consequently, the internal auditor's ability to use emerging technologies with pattern recognition capabilities, such as neural networks, has the potential to enhance audit quality and performance. In this connection, a common observation reiterated by McNamee and McNamee (1993), McNamee (1996), and Shakespeare (1996) is that 'risk assessment needs the participation and input of management outside of internal auditing in order to succeed'. If neural network technology is not embraced by management who have the primary responsibility for risk management, or if it proves not to be 'user-friendly' or feasible from a cost–benefit standpoint, it is unlikely to succeed in practice. In addition, the anticipated benefits from using neural networks depends on satisfactory answers to the following questions: (1) do we obtain faster, cheaper, and possibly more accurate output from a neural network application than afforded by traditional methods? (efficiency); (2) are the risk rankings produced consistent upon repetition? (consistency/reliability); (3) how well do neural network models of risk assessment work in practice? (validity); (4) are the variables entering into a neural network model and the subsequent combining process used easily explained? (defensibility/transparency); and, finally, (5) how well are the model results 'received and accepted' by knowledgeable practitioners? (profession-wide acceptance). We believe neural network technology does carry significant promise in adequately addressing each of these concerns.

As noted before, the three Delphi study auditor-participants included the Director of the UIC Office of University Audits, with 21 years' experience; and two Senior Auditors, UIC Office of University Audits, with 19 years and 7 years of experience respectively. The use of these experienced auditors shields us from Graham's (1993) justified criticism that much of past expertise research has inappropriately labeled auditors with only two or three years of experience as 'experts'. With reference to external auditing, to highlight the role of seniority in accumulating professional experience, Van Dijk and Williams (1990, p. 67) observe: 'Experience gains in importance with the level in the audit hierarchy, to complement audit technical findings with rules of experience about causal relationships within the client's organization and outside. This gives the audit partner and manager the ability to understand the implications of audit findings better than field staff could.' These remarks also generally apply to the internal auditing environment (cf. Colbert, 1989).

The pervasive influence of qualitative factors in assessing risk indicates that the best approach would be a combination of 'scenario-building tools' and neural network technology. Scenario analysis, in a brainstorming session of knowledgeable in-house experts, can help identify all manner of organizational risks. Once scenarios have been elicited, neural networks can be utilized to absorb all these scenarios (input vector of risk characteristics) and their associated risks (target risk ranking) in a pattern-mapping exercise. Careful selection of high audit risk areas using such a strategy yields at least two major benefits to auditors: recipients of such audits are likely to value them; and, a focused approach is likely to shorten the overall audit duration (McNamee, 1996). If neural network based decision aids are developed and implemented successfully, internal auditors will possess a sophisticated tool that can enable them to make sound recommendations to management for strategic purposes such as process control and business process re-engineering (Stoner and Werner, 1995).

This research study is not without its limitations. We recognize that the most important limitation arises from the fact that no external cross-validation using a completely new data set was done. Obtaining good results on a fresh data set would be an unambiguous measure of network performance and generalization. However, to do that would have required us to possess an equally large data set with accompanying Delphi type evaluations that we did not possess. Nevertheless, we did perform suitable internal cross-validations by splitting the training and test sets on a (70%, 30%) basis that resulted in reasonably good network performance. A more carefully defined linkage between input risk factors and output or organizational risk may also be warranted in future studies. Our analysis of the pairwise correlation between the qualitative-factor-based and quantitative-factor-based risk rankings suggests that these two distinct types of risk factors may be tapping into different dimensions of output or organizational risk. If this is indeed the case, the neural network model performance can be further improved by specifically distinguishing qualitative and quantitative risk factors, linking them with aspects of organizational risk that they refer to, and strengthening the model specification.

The neural network models that we have developed are in the context of a public state university. Therefore, a legitimate concern is whether the model so developed applies to smaller community colleges or even to private schools. Following Van Dijk and Williams (1990), a similar question can also be raised about the portability of our model outside of the academic setting, e.g. in banks or insurance companies. We must point out that the risk assessment task is quite generic to internal auditors working in business, government or industry. Accordingly, the findings from this study should not be regarded as being too industry-specific, but generalizable. However, it is true that other customized applications for risk assessment will need to be similarly explicit about design considerations, i.e. the domain-specific data input–output structures, availability of output values representing 'correct answers', choice of network architectures, learning algorithms, and parameter values, in order to permit replicability.

Among the many alternative network architectures available, we have only implemented and reported the results of the backpropagation architecture in conjunction with error-correction learning. However, there exist a plethora of other neural network architectures such as Boltzmann machines, Hopfield networks, modular networks, radial basis functions, recurrent networks, etc. Similarly, there are numerous learning algorithms as well as transfer functions that could be used. As business applications of neural networks become more common, we are bound to see experimentation with more of these 'esoteric' network architectures and learning.

This collaborative research represents our initial foray into evaluating the potential of neural network technology for internal auditing, fuller details of which can be found in a research monograph published by the Institute of Internal Auditors' Research Foundation (Ramamoorti and Traver, 1998). The research effort has already assisted with the process of continuous improvement at the Office of University Audits, University of Illinois, by providing a richer understanding of the benefits and limitations of neural network technology. More sophisticated applications such as neural network modeling of 'pattern mapping' (e.g. for fraud detection, see also Green and Choi, 1997; Fanning and Cogger, 1998) and 'data mining' of institutional databases (e.g. for strategic decision making, see Berry and Linoff, 1997) would appear to be logical extensions of this stream of research.

AUTHOR NOTE

This research was undertaken and completed during the time the first author was on the Accountancy Faculty of the University of Illinois at Urbana-Champaign. Consequently, the personal views of the first author expressed in this paper should not, in any way, be construed as having the endorsement of Arthur Andersen LLP.

Acknowledgements

We thank Chris Wong, Ed Mascorro, Gloria Ring, and especially Mike Godsell, for their research and administrative assistance. We appreciate comments from Ramji Balakrishnan, John Guiver, Ingoo Han, David McNamee, In Jae Myung, Larry Rittenberg, and from participants at Auditing Workshops at the University of Connecticut, University of Maryland, and Wichita State University. We gratefully acknowledge the financial support received from the Institute of Internal Auditors' (IIA) Research Foundation and the University of Illinois at Urbana-Champaign.

References

AICPA, *Statement on Auditing Standards (SAS) No. 47: Materiality and Audit Risk*, American Institute of Certified Public Accountants, New York, 1989.

Arnold, V. and Sutton, S.G. (eds), *Behavioral Accounting Research: Foundations and Frontiers*, *American Accounting Association*, Sarasota, FL, 1997.

Ashton, R.H. and Ashton, A.H. (eds), *Judgment and Decision-Making Research in Accounting and Auditing*, Cambridge University Press, Cambridge, 1995.

Bailey, A.D., Jr, Hackenbrack, K., De, P., and Dillard, J., 'Foundations of artificial intelligence and expert systems in accounting and auditing', *Journal of Information Systems*, Spring 1987, 20–40.

Balakrishnan, P.V., Cooper, M.C., Jacob, V.S. and Lewis, P.A. 'A study of the classification capabilities of neural networks using unsupervised learning: a comparison with *K*-means clustering', *Psychometrika*, **59**, No. 4, 1994, 509–525.

BarNiv, R. and Hershbarger, R. 'Classifying financial distress in the life insurance industry', *The Journal of Risk and Insurance*, Spring 1990, 110–135.

Bedard, J. and Biggs, S.F. 'Pattern recognition, hypothesis generation, and auditor performance in an analytic task', *The Accounting Review*, July 1991, 622–643.

Bell, T.B., Ribar, G.S. and Verchio, J. Neural nets versus logistic regression: a comparison of each model's ability to predict commercial bank failure's, in Srivastava, R.P. (ed.), *Auditing Symposium X—Deloitte & Touche/University of Kansas Symposium on Auditing Problems*, pp. 29–53, University of Kansas, Lawrence, KS, 1990.

Berry, M.J.A. and Linoff, G., *Data Mining Techniques for Marketing, Sales, and Customer Support*, Wiley Computer Publishing, New York, 1997.

Bishop, C.M. *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1995.

Boritz, J.E., *auditMASTERPLAN: Audit Planning Software*, Institute of Internal Auditors, Altamonte Springs, FL, 1986.

Bouwman, M.J. and Bradley, W.E., 'Judgment and decision making, Part II: Expertise, consensus and accuracy', in Arnold, V. and Sutton, S.G. (eds), *Behavioral Accounting Research: Foundations and Frontiers*, American Accounting Association, Sarasota, FL, 1997.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, Wadsworth, Belmont, CA, 1984.

Brown, C.E. and Solomon, I., 'Auditor configural information processing in control risk assessment', *Auditing: A Journal of Practice and Theory*, Fall 1990, 17–38.

Brown, C.E. and Solomon, I., 'Configural information processing in auditing: the role of domain-specific knowledge', *The Accounting Review*, January 1991, 100–119.

Brown, D.L. and Eining, M.M. 'Judgment and decision making, Part IV: Information technology and decision aids', in Arnold, V. and Sutton, S.G. (eds), *Behavioral Accounting Research: Foundations and Frontiers*, American Accounting Association, Sarasota, FL, 1997.

Burger, C. and Traver, R.O., 'Applying neural networks in information system auditing', *EDPACS: The EDP Audit, Control, and Security Newsletter*, **XXIV**, No. 6, December 1996, 1–10.

Chen, C.H. (ed.), *Fuzzy Logic and Neural Network Handbook*, McGraw-Hill Series on Computer Engineering, McGraw-Hill, New York, 1996.

Chi, M.T.H., Glaser, R. and Farr, M.J. (eds), *The Nature of Expertise*, Erlbaum, Hillsdale, NJ, 1988.

Coakley, J.R. and Brown, C.E., 'Artificial neural networks applied to ratio analysis in the analytical review process', In Trippi, R.R. and Turban, E. (eds), *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*, pp. 123–162. Irwin, Chicago, IL, 1996.

Coderre, D.G., 'Computer assisted audit tools & techniques', *Internal Auditor*, February, 1993, 24–26.

Colbert, J.L. 'The effect of experience on auditors' judgements', *Journal of Accounting Literature*, **8**, 1989, 137–149.

Colbert, J.L., 'Risk', *Internal Auditor*, October 1995, 36–40.

Coleman, K.G., 'Neural networks: a high impact technology for bankruptcy prediction', *Proceedings of the Multidisciplinary Conference on Decision Making, Cognitive Science and Accounting* sponsored by The Peat Marwick Foundation, Pittsburgh, PA, 1991.

COSO (Committee of Sponsoring Organizations of the Treadway Commission), *Internal Control—Integrated Framework*. COSO, Jersey City, NJ, 1992.

Daniel, W.W., *Applied Nonparametric Statistics*, Houghton Mifflin, Boston, MA, 1978.

Davis, J.T., 'Experience and auditors' selection of

relevant information for preliminary control risk assessments', *Auditing: A Journal of Practice & Theory*, **15**, No. 1, 1996, 16–37.
Deboeck, G.J. (ed.), *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, Wiley, New York, 1994.
Doherty, S.A., 'Emerging technologies', *Internal Auditor*, December 1993, 20–25.
Elliott, R.K., 'The third wave breaks on the shores of accounting', *Accounting Horizons*, June 1992, 61–85.
Fanning, K.M. and Cogger, K.O., 'Neural network detection of management fraud using published financial data, *International Journal of Intelligent Systems in Accounting, Finance & Management*, **7**, 1998, 21–41.
Fanning, K.M., Cogger, K.O. and Srivastava, R.P., 'Detection of management fraud: a neural network approach', *International Journal of Intelligent Systems in Accounting, Finance & Management*, **4**, 1995, 113–126.
Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S.L. and Keeney, R.L., *Acceptable Risk*, Cambridge University Press, Cambridge, 1983.
Fletcher, D. and Goss, E., 'Forecasting with neural networks: an application using bankruptcy data', *Information & Management*, **24**, 1993, 159–167.
Gallegos, F., Richardson, D.R. and Borthick, A.F., *Audit and Control of Information Systems*, South-Western Publishing, Cincinnati, OH, 1987.
Garceau, L.R. and Foltin, C., 'Neural networks: new technology for accountants and the impact on IS audit and control professionals', *IS Audit & Control Journal*, **II**, 1995, 52–57.
Graham, L.E., Discussion of expertise in auditing', *Auditing: A Journal of Practice and Theory, Supplement*, **12**, 1993, 46–50.
Green, B.P. and Choi, J. 'Assessing the risk of management fraud through neural network technology', *Auditing: A Journal of Practice and Theory*, Spring, **16**, No. 1, 1997, 14–28.
Guiver, J.P. 'Lecture Notes for Seminar on Advanced Neural Computing, February 10–13, 1997, *NeuralWare, Inc.*, Pittsburgh, PA, 1997.
Hackenbrack, K., 'Implications of seemingly irrelevant evidence in audit judgment', *Journal of Accounting Research*, **30**, No. 1, 1992, 126–136.
Hammerstrom, D., 'Working with neural networks', *IEEE Spectrum*, July 1993, 46–53.
Hansen, J.V. and Messier, W.F., 'Artificial neural networks: foundations and applications to a decision problem', *Expert Systems with Applications: An International Journal*, **3**, No. 1, 1991, 135–141.
Hassoun, M.H., *Fundamentals of Artificial Neural Networks*, The MIT Press, Cambridge, MA, 1995.
Haykin, S., *Neural Networks: A Comprehensive Foundation*, Macmillan Publishing, New York, 1994.
Hays, W.L., *Statistics*, 5th edn. Harcourt Brace College Publishers, Fort Worth, TX, 1994.
IIA, *Standards for the Professional Practice of Internal Auditing*. SIAS No. 9: Risk Assessment, Institute of Internal Auditors, Altamonte Springs, FL, 1995.
Kendall (1962).
Libby, R., 'Availability and the generation of hypotheses in analytic review', *Journal of Accounting Research*, **23**, 1985, 648–667.
Lippmann, R.P., 'An introduction to computing with neural nets, *IEEE ASSP Magazine*, April 1987, 4–21.
Massaad, T.G. and Winkler, L.R., 'A basic introduction to neural networks for accountants', in Vasarhelyi, M.A. (ed.), *Artificial Intelligence in Accounting and Auditing: The Use of Expert Systems, Vol.2*, pp. 131–148, Markus Wiener Publishing, New York, 1995.
McNamee, D., *Assessing Risk. Internal Audit Tool Kit*, Institute of Internal Auditors, Altamonte Springs, FL, 1996.
McNamee, D.W. and McNamee, F.A., *Risk Assessment in the 90's*. Management Control Concepts, Walnut Creek, CA, 1993.
Medsker, L., Turban, E. and Trippi, R.R., 'Neural network fundamentals for financial analysts', in Trippi R.R. and Turban, E. (eds), *Neural Networks in Finance and Investing*, pp. 3–24, Irwin, Chicago, IL, 1996.
Mock, T.J. and Vertinsky, I., *Risk Assessment in Accounting and Auditing*. Research Monograph No. 10, The Canadian General Accountants' Research Foundation, Vancouver, BC, 1985.
Moore, P.G., *The Business of Risk*, Cambridge University Press, New York, 1983.
Myung, I.J., Ramamoorti, S. and Bailey, A.D., Jr., Maximum entropy aggregation of expert predictions', *Management Science*, **42**, No. 10, 1996, 1420–1436.
NeuralWare, *NeuralWorks Predict: Complete Solution for Neural Data Modeling*, NeuralWare, Inc., Pittsburgh, PA, 1995.
NeuralWare, *Neural Computing: A Technology Handbook for Professional II/PLUS and NeuralWorks Explorer*, NeuralWare, Inc, Pittsburgh, PA, 1996.
NeuroDimension, Inc., *NeuroSolutions Manual*, Volumes I & II, Gainesville, FL, 1995.
Odom, M. and Sharda, R., 'A neural network model for bankruptcy prediction', *Proceedings of the International Joint Conference on Neural Networks*. Washington, DC, INNS January, Vol. II, pp. 163–168, 1990.
Patton, J.M., Evans, J.H. and Lewis, B.L., *A Framework for Evaluating Internal Audit Risk*, Institute of Internal Auditors, Altamonte Springs, FL, 1982.
Power, M., 'Auditing, expertise, and the sociology of technique', *Critical Perspectives on Accounting*, **6**, 1995, 317–339.
Ramamoorti, S. and Hoey, A.L. 'Qualitative materiality in governmental audit planning', *The Government Accountants' Journal*, Spring, **47**, No. 1, 1998, 44–49.
Ramamoorti, S. and Traver, R.O., *Using Neural Networks for Risk Assessment in Internal Auditing: A*

Feasibility Study, Institute of Internal Auditors' Research Foundation, Altamonte Springs, FL, 1998.

Rumelhart, D.E. and McClelland J.L. (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, Chapter 8, MIT Press, Cambridge, MA, 1986.

Sarle, W.S., Neural networks and statistical models, *Proceedings of the Nineteenth Annual SAS, Inc. Users Group International Conference*, April, 1994, pp. 1–13.

Sawyer, L. B. and Dittenhofer, M. A., *Sawyer's Internal Auditing: The Practice of Modern Internal Auditing*, 4th edn, Institute of Internal Auditors, Altamonte Springs, FL, 1996.

Shakespeare, K., 'University risk control begins with risk assessment', *Ledger*, **40**, April 1996, 4–7.

Shanteau, J., 'Psychological characteristics and strategies of expert decision makers', *Acta Psychologica*, **68**, 1988, 203–215.

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

Srivastava, R.P., Dutta, S.K. and Johns, R.W., 'An expert system approach to audit planning and evaluation in the belief-function framework', *Intelligent Systems in Accounting, Finance and Management*, **5**, 1996, 165–183.

Stoner, J.A.F. and Werner, F.M., *Internal Audit and Innovation*. Financial Executives Research Foundation, Morristown, NJ, 1995.

Swingler, K., *Applying Neural Networks: A Practical Guide*, Academic Press, London, 1996.

Tabachnick, B.G. and Fidell, L.S. *Using Multivariate Statistics*, 3rd edn, Harper Collins, New York, 1996.

Tam, K. and Kiang, M., 'Managerial applications of neural networks: the case of bank failure predictions', *Management Science*, **38**, No. 7, July 1992, 926–947.

Tam, Y.T., 'Neural networks for decision support'. *Decision Support Systems*, **11**, 1994, 389–392.

Trippi, R.R. and Turban, E. (eds), *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*, revised edition, Irwin, Chicago, IL, 1996.

Tubbs, R.M., 'The effect of experience on the auditor's organization and amount of knowledge', *The Accounting Review*, **67**, No. 4, 1992, 783–801.

Van Dijk, J.C. and Williams, P., *Expert Systems in Auditing*, Stockton Press, New York, 1990.

Vasarhelyi, M.A., *Artificial Intelligence in Accounting and Auditing: Using Expert Systems, Vol. 1*, Marcus Werner Publishers, New York, 1990.

Vasarhelyi, M.A., *Artificial Intelligence, Vol.III.* Marcus Werner Publishers, New York, 1995.

Ward Systems Group, Inc., *NeuroShell 2*, Executive Park West, Frederick, MD, 1993.

Yoon, Y., Swales, G. and Margavio, T.M., 'A comparison of discriminant analysis versus artificial neural networks', *Journal of the Operations Research Society*, **44**, No. 1, 1993, 51–60.

Zarowin, S., 'Thinking computers', *Journal of Accountancy*, **180**, No. 5, November 1995, 55–56.