Probabilistic neural networks for the identification of qualified audit opinions

Chrysovalantis Gaganis

Expert Systems With Applications

Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 114–124
Expert Systems with Applications
www.elsevier.com/locate/eswa

Probabilistic neural networks for the identification of qualified audit opinions

Chrysovalantis Gaganis [a], Fotios Pasiouras [a,b], Michael Doumpos [a,*]

[a] *Financial Engineering Laboratory, Department of Production Engineering & Management, Technical University of Crete, University Campus, Chania 73 100, Greece*
[b] *Coventry Business School, Coventry University, Priory Street, CV1 5FB Coventry, UK*

Abstract

Prior studies that examine the application of neural networks in auditing investigate the efficiency of artificial neural networks (ANNs). In the present study, considering the well known disadvantages of artificial neural network, we propose the application of probabilistic neural networks (PNNs) that combine the computational power and flexibility of ANNs, while managing to retain simplicity and transparency. The sample consists of 264 financial statements that received a qualified audit opinion over the period 1997–2004 and 3069 unqualified ones, from 881 firms listed on the London Stock Exchange. The results demonstrate the high explanatory power of the PNN model in explaining qualifications in audit reports. The model is also found to outperform traditional ANN models, as well as logistic regression. Sensitivity analysis is used to assess the relative importance of the input variables and to analyze their role in the auditing process.

Keywords: Probabilistic neural networks; Auditing; Qualified audit reports

1. Introduction

In recent years neural networks (NNs) have received a lot of attention, and have been applied in various financial/accounting applications such as bankruptcy prediction (Charitou, Neophytou, & Charalambous, 2004; Luther, 1998; Pendharkar, 2005; Zhang, Hu, Patuwo, & Indro, 1999), bond trading (Huang, Chen, Hsu, Chen, & Wu, 2004), volatility forecasts (Hamid & Iqbal, 2004), portfolio management (Hung, Liang, & Liu, 1996), country risk rating (Yim & Mitchell, 2005) and auditing (Fanning & Cogger, 1998; Hansen, McDonald, & Stice, 1992). Calderon and Cheh (2002), Wong, Lai, and Lam (2000) and Wong and Selvi (1998) provide overviews of the research on NNs with applications in business and finance conducted over the last years. The vast majority of the studies has used multilayer feed-forward artificial neural networks (ANNs) trained with the back-propagation learning algorithm. However, numerous researchers document the disadvantages of this approach. For example, Calderon and Cheh (2002) mention that standard back-propagation networks are subject to problems of local minima, and can be tedious and extremely time-consuming to build. Results can also be very sensitive to specification of learning rates, momentum and other processing elements, and there is no clear guidance on the selection of those parameters. Salchenberger, Cinar, and Lash (1992) also point out the inability to explain conclusions or how they are reached, and the lack of formal theory which imposes a need for expertise on the user.

An alternative NN architecture, the probabilistic neural networks (PNNs; Specht, 1990) constitute a classification methodology that combines the computational power and flexibility of ANNs, while managing to retain simplicity and transparency. The main advantages of PNNs over ANNs include their simplified architecture which overcomes the difficulty of specifying an appropriate ANN

* Corresponding author.
E-mail address: mdoumpos@dpem.tuc.gr (M. Doumpos).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.11.003

model, as well as their easy implementation during training and testing. Calderon and Cheh (2002) point out that these advantages make PNNs a potentially attractive alternative in auditing.

So far PNNs have been applied in only a few studies in finance such as bankruptcy prediction (Etheridge & Sriram, 1997; Yang, Platt, & Platt, 1999), short-term liquidity modeling (Li, Shue, & Shiue, 2000) and stock index forecasting (Chen, Leung, & Daouk, 2003; Kim & Chun, 1998).

This study uses PNNs for the development of a model that explains qualifications in audit reports. Laitinen and Laitinen (1998) classify prior studies on qualified audit report information relevant to the present study into three categories. Studies from the first category use audit report information for the construction of bankruptcy prediction models (Keasey & Watson, 1987). Studies falling in the second category mainly deal with the construction of bankruptcy models for making audit opinions, relative to going concern (Koh, 1991). Studies from the third category develop models to explain (or predict) qualifications in audit reports (Dopuch, Holthausen, & Leftwich, 1987; Keasey, Watson, & Wynarzcyk, 1988; Laitinen & Laitinen, 1998).

The present paper falls into the third category of the studies mentioned above. The purpose of the study is to explore the potential of PNNs for the development of models that explain qualifications in audit reports. Bell and Tabor (1991) as well as Chen and Church (1992) note that auditors can use the output of such models to plan specific auditing procedures that can be applied to achieve an acceptable level of audit risk. These models can also be used as a quality control tool in the final or review stage of an engagement and for contingency analyses on how changes in specific variables could add or detract from the probability of obtaining a qualified opinion (Kleinman & Anandarajan, 1999).

The analysis is based on a large sample of firms listed in the London Stock Exchange. Both financial and non-financial variables are introduced in the models and principal components analysis is employed to build a compact set of independent variables which have a clear interpretation for the auditing domain. Special emphasis is given on the specification of the smoothing parameter of the PNN model, as well as on the interpretability of the model. The latter issue is often overlooked in the application of non-parametric techniques such as network models, to financial and auditing problems. Actually, most past studies have been mainly focused on the predictive power of the developed models. However, a good model, expect for its predictive power, should also be able to provide meaningful information to the decision maker. To address this issue we perform sensitivity analysis which provides useful insight on the model's outputs, thus designating that a complex non-linear network model is not a simple "black-box" prediction tool, but it can also be used to highlight the important factors that describe qualifications in audit reports. Furthermore, the results of a comparative analysis with ANNs and logistic regression support the superiority of the PNN modeling framework in explaining qualifications in audit reports.

The remaining of the paper is structured as follows: Section 2 briefly outlines the PNN methodology. Section 3 describes the data and the variables used in the analysis, whereas Section 4 presents the empirical results. Finally Section 5 concludes the paper and discusses some possible future research directions.

2. Probabilistic neural networks

Probabilistic neural networks possess the simplicity, speed and transparency of traditional statistical classification models along with much of the computational power and flexibility of back-propagated neural networks (Specht, 1990).

A PNN can be realized as a network of four layers (Fig. 1). The input layer includes N nodes, each corresponding to one input attribute (independent variable). The inputs of the network are fully connected with the M nodes of the pattern layer. Each node of the pattern layer corresponds to one training object. The $1 \times N$ input vector $\mathbf{x}_i$ is processed by pattern node j through an activation function that produces the output of the pattern node. The most usual form of the activation function is the exponential one:

$$o_{ij} = \exp\left(-\frac{\|\mathbf{x}_j - \mathbf{x}_i\|^2}{\sigma^2}\right)$$

where σ is a smoothing parameter. The result of this activation function ranges between 0 and 1. As the distance $\|\mathbf{x}_j - \mathbf{x}_i\|$ between the input vector $\mathbf{x}_i$ and the vector $\mathbf{x}_j$ of the pattern node j increases, the output of node j will approach zero, thus designating the small similarity between the two data vectors. On the other hand, as the distance $\|\mathbf{x}_j - \mathbf{x}_i\|$ decreases, the output of node j will approach unity, thus designating the significant similarity between the two data vectors. If $\mathbf{x}_i$ is identical to $\mathbf{x}_j$, then the output

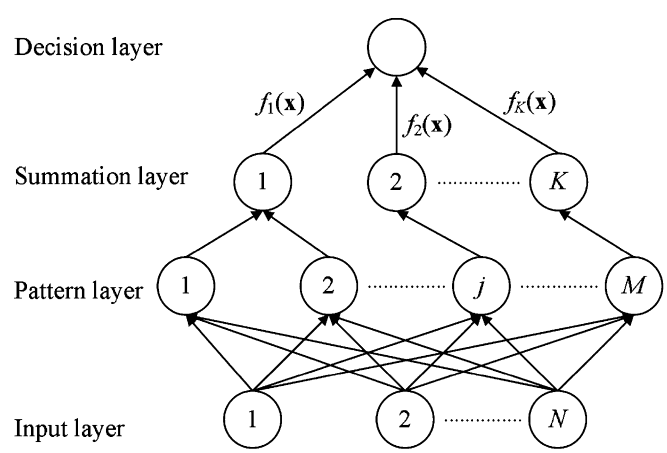

Fig. 1. Architecture of a probabilistic neural network.

of the pattern node j will be exactly one. The parameter σ controls the width of the activation function. As σ approaches zero, even small differences between $\mathbf{x}_i$ is identical to $\mathbf{x}_j$ will lead to $o_{ij} \approx 0$, whereas larger values of σ produce more smooth results.

The outputs of the pattern nodes are passed to the summation layer that consists of K competitive nodes each corresponding to one class. Each summation node k is connected to the pattern nodes that involve training objects that belong to class k. For an input vector $\mathbf{x}_i$, the summation node k simply takes the outputs of the pattern nodes to which it is connected with to produce an output $f_k(\mathbf{x}_i)$ as follows:

$$f_k(\mathbf{x}_i) = \frac{1}{M_k} \sum_{\forall \mathbf{x}_j \in y_k} o_{ij}$$

where y_k is the class label corresponding to the summation node k and M_k is the number of training objects that belong to this class. Assuming that all data vectors are normalized to unit length (i.e., $\|\mathbf{x}\| = 1$), $f_k(\mathbf{x}_i)$ can equivalently be written as:

$$f_k(\mathbf{x}_i) = \frac{1}{M_k} \sum_{\forall \mathbf{x}_j \in y_k} \exp\left(\frac{\mathbf{x}_j\mathbf{x}_i^\top - 1}{\sigma^2}\right) \tag{1}$$

and the outputs of the summation nodes can be easily transformed to posterior class membership probabilities:

$$P(y_i = k \mid \mathbf{x}_i) = \frac{f_k(\mathbf{x}_i)}{\sum_{k=1}^{K} f_k(\mathbf{x}_i)}$$

On the basis of these probabilities a classification rule is employed at the decision layer, which consists of a single node, to assign the input vector $\mathbf{x}_i$ to a particular class. The obvious approach is to assign $\mathbf{x}_i$ to the class where it is most likely to belong (i.e., the class with the maximum $P(k|\mathbf{x}_i)$). In a two class case with $y = \{0, 1\}$ it is possible to define a cut-off probability point c, such that $\mathbf{x}_i$ is assigned to class 0 if and only if $P(y_i = 0|\mathbf{x}_i) \geqslant c$. The specification of this cut-off point is based on the prior probabilities of class membership and the misclassification costs.

3. Sample and variables

3.1. Sample

The sample of the study consists of the population of firms listed at the London Stock Exchange with available data in Financial Analysis Made Easy (FAME) database, a total of 1363 firms. After excluding 146 firms with many missing values and 336 financial firms, the sample was reduced to 881 firms. Financial and non-financial information as well as the auditors' opinion were all collected from FAME, for the period 1997–2004, resulting in a total of 3333 firm-year observations.

The only audit information available in FAME is whether the auditor issued a qualified or unqualified opinion. Hence, we had no further information to distinguish whether qualifications are due to disagreements (e.g. accounting treatment or disclosure), limitations on scope (i.e. lack of audit evidence) or going-concern issues. It should be mentioned at this point that explanatory paragraphs are not considered qualified audit opinions in the UK. In particular, in forming their opinion on financial statements, auditors should take into account whether the view given by the financial statements could be affected by inherent uncertainties, which, in their opinion, are fundamental. When an inherent uncertainty exists which (i) in the auditor's opinion is fundamental, and (ii) is adequately accounted for and disclosed in the financial statements, the auditors should include an explanatory paragraph referring to the fundamental uncertainty in the section of their report setting out the basis of their opinion. When adding such an explanatory paragraph, auditors should use words, which clearly indicate that their opinion on the financial statements is not qualified in respect of its contents (SAS 600.6). In that case, an unqualified opinion indicates that the auditors consider that appropriate estimates and disclosures relating to fundamental uncertainties are made in the financial statements. It remains unqualified notwithstanding the inclusion of an explanatory paragraph describing a fundamental uncertainty. The explanatory paragraph is included as part of the basis for the auditor's opinion so as to make clear that it describes a matter which the auditors have taken into account in forming their opinion, but it does not qualify that opinion. When the auditors conclude that the estimate of the outcome of a fundamental uncertainty is materially misstated or that the disclosure relating to it inadequate, they issue a qualified opinion.

From the 3333 financial statements in the sample, 264 received a qualified audit opinion. Table 1 presents the composition of the sample by industry and auditors opinion. To ensure the proper validation of the models, the sample was split into a training sample and a holdout sample. The former (i.e. training) consists of 2215 firm-year observations, with 175 qualified cases over the period 1997–2002 and 2040 unqualified ones over the same period. The holdout sample consists of 1118 firm-year observations, with 89 qualified cases over the period 2003–2004 and 1029 unqualified ones.

3.2. Variables

The list of the explanatory variables used to predict audit qualification is presented in Table 2. The set of variables consists of absolute measures, financial ratios and non-financial information, in an effort to cover various aspects of firms' characteristics. Absolute variables measure the fees paid to the auditor (x_1), the numeration paid to the firm's directors (x_2), and the size of the firm (x_3, x_4). Financial ratios measure liquidity (x_5, x_6), solvency (x_7), employees' productivity ($x_8, x_9, x_{22}, x_{23}, x_{24}$), and efficiency ($x_{10}$–$x_{21}$). Non-financial variables involve the credit rating of the firm, whether the auditor is a big one or not, and the business sector in which the firm operates. The use of the credit rating is

Table 1
Observations by industrial sector and audit opinion

UK SIC Section	Description	Unqualified	Qualified	Total
A	Agriculture, hunting & forestry	38	9	47
C	Mining & quarrying	72	33	105
D	Manufacturing	1127	67	1194
E	Electricity, gas & water supply	13	0	13
F	Construction	98	8	106
G	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods	402	26	428
H	Hotels & restaurants	89	3	92
I	Transport, storage & communication	132	2	134
K	Real estate, renting & business activities	921	95	1016
M	Education	12	1	13
N	Health & social work	26	3	29
O	Other community, social and personal service activities	138	17	155
P	Private households employing domestic staff and undifferentiated production activities of household for own use	1	0	1
Total		3069	264	3333

Table 2
List of variables

y	Auditor's opinion
x_1	Log audit fee
x_2	Log directors' remuneration
x_3	Log number of employees
x_4	Log total assets
x_5	Current ratio
x_6	Liquidity ratio
x_7	Solvency ratio
x_8	Working capital per employee
x_9	Total assets per employee
x_{10}	Profit margin
x_{11}	Return on shareholders funds
x_{12}	Return on capital employed
x_{13}	Return on total assets
x_{14}	Debtors turnover
x_{15}	Debtor collection period
x_{16}	Creditors payment period
x_{17}	Net assets turnover
x_{18}	Fixed assets turnover
x_{19}	Salaries/turnover
x_{20}	EBIT margin
x_{21}	EBITDA margin
x_{22}	Turnover per employee
x_{23}	Average remuneration per employee
x_{24}	Profit per employee
x_{25}	Credit risk assessment
	High risk rating
	Caution rating
	At least normal rating
x_{26}	Business sector
	Services
	Trade
	Primary and production
x_{27}	Auditor's size (dummy variable for big 4)

justified on the basis of previous studies which indicate that firms with a high probability of default are more likely to receive qualified reports because their ability to continue is in greater doubt (Bell & Tabor, 1991; Krishnan & Krishnan, 1996; McKeown, Mutchler, & Hopwood, 1991; Reynolds & Francis, 2001). We use the credit ratings issued by CRIF Decision Solutions Limited, which are also available in FAME. The QuiScore provided by CRIF measures the likelihood of default (in a 0–100 scale) for the 12 months following the date of its calculation. On the basis of their QuiScore firms are classified into the following five rating groups: secure, stable, normal, unstable (or caution) and high risk group. In the present study we model the credit risk variable in three levels: high, caution, and at least normal (e.g., normal, stable or secure). To introduce this variable in the analysis three dummy 0–1 variables are used. The first dummy variable (high risk) indicates whether a firm belongs in the high risk group or not, whereas the second (caution rating) indicates whether a firm belongs in the caution risk group or not. In a similar manner, the third dummy variable (at least normal rating) indicates whether a firm belongs in one of the normal, stable and secure groups or not. DeAngelo (1981) hypothesizes that audit firms differentiate themselves on quality and that larger audit firms are expected to provide higher quality audits. Most of the audit quality literature also indicates that quality differences between Big auditors and non-Big auditors are inferred by both users of financial statements and companies selecting an auditor. Therefore, we introduce a dummy variable indicating whether the auditor is a big (Deloitte & Touch, Ernst & Young, KPMG PricewaterhouseCoopers) one or not. Finally, three dummy 0–1 variables are used to denote: (a) whether the firm operates in a primary/production sector (i.e. sectors A, C, D, E, F) or not, (b) whether the firms operates in a services sector (i.e. sectors H, I, K, M, N, O, P) or not, and (c) whether the firm operates in the trade (i.e. sector G) sector or not.

4. Empirical results

4.1. Univariate analysis

Table 3 presents some descriptive statistics (mean, standard deviation) and the results of a Kruskal–Wallis (KW)

Table 3
Descriptive univariate statistics

		Mean		KW χ^2
		Unqualified	Qualified	
Panel A: Financial variables				
x_1	LOG audit fee	1.754	1.625	13.836*
x_2	LOG directors' remuneration	2.685	2.549	18.310*
x_3	LOG number of employees	2.434	1.990	47.636*
x_4	LOG total assets	4.481	4.030	55.880*
x_5	Current ratio	1.981	1.691	39.941*
x_6	Liquidity ratio	1.573	1.537	17.600*
x_7	Solvency ratio	49.381	42.787	2.320
x_8	Working capital per employee	18091.827	4614.321	93.072*
x_9	Total assets per employee	132867.958	113681.640	12.522*
x_{10}	Profit margin	−0.118	−5.701	26.778*
x_{11}	Return on shareholders funds	−0.201	−75.304	127.901*
x_{12}	Return on capital employed	−1.297	−67.918	149.241*
x_{13}	Return on total assets	−0.364	−44.906	177.657*
x_{14}	Debtors turnover	17.069	18.252	4.63**
x_{15}	Debtor collection period	65.382	64.715	4.569**
x_{16}	Creditors payment period	40.416	69.908	6.684*
x_{17}	Net assets turnover	2.787	2.250	17.554*
x_{18}	Fixed assets turnover	6.780	5.383	20.905*
x_{19}	Salaries/turnover	29.231	36.647	12.392*
x_{20}	EBIT margin	2.424	−2.279	20.282*
x_{21}	EBITDA margin	7.526	0.135	20.609*
x_{22}	Turnover per employee	142011.474	89239.700	66.062*
x_{23}	Average remunuration per employee	27184.253	29167.815	0.244
x_{24}	Profit per employee	4877.721	−51885.190	166.377*
Panel B: Categorical variables				χ^2
x_{25}	Credit risk rating			182.81*
	High risk	156	68	
	Caution	325	34	
	At least normal	1559	73	
x_{26}	Business sector			0.97
	Services	842	75	
	Trade	262	18	
	Primary/production	936	82	
x_{27}	Auditor's size			1.96
	Big 4	1320	104	
	Non-Big 4	720	71	

* Significant at the 1% level.
** Significant at the 5% level.

test of the significance of the continuous variables in discriminating between qualified and unqualified cases in the training sample. For the categorical variables, the distribution of the data is presented, along with the result of the Pearson's χ^2 test on the differences between the two groups. What becomes quickly apparent is that firms with unqualified financial statements tend to be larger and in general in better position either in terms of profitability, efficiency, liquidity, activity or solvency. Ireland (2003) also found that companies with poor liquidity and high financial risk that report material contingent liabilities or which do not pay dividends are more likely to receive a qualified report than other companies in the UK. Furthermore, numerous previous studies indicate that firms which receive qualified opinions are less profit (Beasley, Carcello, & Hermanson, 1999; Loebbecke, Eining, & Willingham, 1989; Spathis, Doumpos, & Zopounidis, 2002; Spathis, 2002; Summers & Sweeney, 1998). As Spathis (2002) points out, the profitability orientation is tempered by the manager's own utility maximization, defined by job security. In addition, Beasley et al. (1999) found that companies committing fraud in the US were generally small ones. As Ireland (2003) points out large companies are more likely to have good accounting systems and internal controls, thus reducing disagreements and limitations on scope. Auditors independence may be affected by the proportion of audit fees that a client contributes to the total fees earned by the auditor as well as from of non-audit service fees. This economic independence creates an incentive for auditors to compromise their independence and report favorably in order to retain clients (DeAngelo, 1981). Nevertheless, Reynolds and Francis (2001) find that big auditors do not treat large clients more favorably than smaller clients; however, research into threats to auditor independence posed from non-audit service dependence has produced mixed findings (Craswell, 1999; Wines, 1994). Our univariate results indicate that firms with unqualified financial statements pay more audit fees than firms with qualified ones. However, this could also be attributed to the larger size of the firms with unqualified statements in the sample, given that the literature on audit fees determinants documents that as client size increases, the audit fee also increases, because the auditor will have to perform more work to ensure adequate compliance and substantive testing.

As far as the three non-financial variables are concerned, it is clear that only the credit risk rating provides significant information in discriminating between qualified and unqualified cases. The auditor's size as well as the business sector are not found statistically significant.

4.2. Variable selection

The large number of the available financial variables poses some problems on the development of a useful model. In particular, developing a model with too many variables is likely to lead to overfitting, and furthermore it makes the model unpractical since its use would require the collection of too many data. Therefore, the construction and selection of a small number of variables is an important issue in developing a realistic model.

In this study principal components analysis (PCA) is used as a data reduction methodology. The objective of PCA is to reduce the dimensionality (number of variables) of the dataset while retaining most of the original variability (information) in the data. This is performed through the construction of a set of principal components which act as a new reduced set of variables. Each principal component is a linear combination of the original variables and they are all orthogonal to each other. For a given data set with N variables, the first principal component has the highest explanatory power (it describes as much of variability of the data as possible), whereas the Nth principal component has the least explanatory power. Thus, the n first principal components are supposed to contain most of the information implicit in the attributes.

In this analysis PCA with varimax rotation is performed and the principal components with eigenvalues higher than one are retained. Overall, eight principal components are constructed that explain more that 75% of the overall variability in the original data set. Table 4 provides details on the principal component loadings, with coefficients higher than 0.6 (in absolute terms) marked in bold. These results indicate the interpretation of the constructed components. In particular, the first component is related to profit margins, the second involves the firms' size, the third measures profitability, the fourth involves liquidity and solvency, the fifth involves productivity, the sixth represents assets turnover, whereas the last two components are related to the credit policy of the firm with regard to its debtors and creditors.

Table 4
Principal component loadings

Financial variables	Components							
	1	2	3	4	5	6	7	8
EBIT margin	**0.948**	0.035	0.201	−0.022	0.071	−0.017	−0.003	−0.025
Profit margin	**0.921**	0.051	0.226	−0.011	0.013	0.026	0.015	−0.021
EBITDA margin	**0.903**	0.017	0.235	0.003	0.069	−0.076	−0.018	−0.006
Profit per employee	0.550	0.046	0.348	−0.029	0.239	0.060	−0.060	0.282
LOG audit fee	0.004	**0.931**	0.006	−0.056	−0.025	0.019	0.043	−0.018
LOG total assets	0.123	**0.903**	0.174	−0.046	0.157	−0.096	−0.083	−0.040
LOG directors' remuneration	0.014	**0.871**	−0.042	0.040	0.089	−0.003	0.047	0.101
LOG number of employees	0.002	**0.610**	0.241	−0.268	−0.437	0.078	−0.205	−0.055
Return on capital employed	0.257	0.059	**0.868**	0.037	0.020	−0.053	0.028	0.043
Return on shareholders funds	0.259	0.053	**0.840**	0.031	0.010	−0.048	0.045	0.042
Return on total assets	0.370	0.099	**0.773**	0.018	0.010	0.021	−0.007	0.073
Current ratio	−0.004	−0.042	0.015	**0.961**	0.044	0.030	0.068	−0.014
Liquidity ratio	−0.029	−0.048	−0.024	**0.960**	−0.013	0.006	0.095	−0.017
Solvency ratio	−0.009	−0.070	0.132	**0.614**	0.096	−0.456	−0.049	0.124
Turnover per employee	0.193	0.004	0.023	0.011	**0.758**	0.334	−0.018	0.261
Total assets per employee	0.051	0.105	0.020	0.065	**0.692**	−0.269	0.007	−0.154
Salaries/turnover	−0.295	−0.076	−0.246	0.093	−0.493	−0.353	0.266	0.331
Net assets turnover	−0.009	0.046	−0.102	−0.210	−0.128	**0.804**	−0.025	−0.016
Fixed assets turnover	−0.045	−0.099	0.066	0.127	0.173	**0.764**	0.029	0.118
Debtor collection period	−0.028	−0.089	0.039	0.121	0.035	−0.073	**0.812**	−0.267
Debtors turnover	−0.011	−0.028	0.024	−0.012	0.101	−0.006	**−0.720**	−0.078
Working capital per employee	−0.018	0.014	0.336	−0.015	0.350	0.208	0.472	0.011
Creditors payment period	−0.069	−0.044	−0.306	0.012	0.092	−0.100	0.238	**−0.688**
Average remuneration per employee	−0.030	−0.019	−0.271	0.142	0.426	−0.023	0.415	0.533
Eigenvalue	5.038	3.197	2.439	2.124	1.572	1.523	1.190	1.052
Variance explained (%)	20.99	13.32	10.16	8.85	6.55	6.35	4.96	4.38
Cumulative variance explained (%)	20.99	34.31	44.47	53.32	59.87	66.22	71.18	75.56

With the extracted principal components, the original set of financial variables is transformed into a new set of eight uncorrelated variables with the aforementioned interpretation. The component scores provide the description of the firms in the new set of variables. In addition to the eight extracted components, the three categorical variables regarding the credit rating, the auditor's size and the business sector are also used as independent variables in the analysis. Overall, the data involve 15 input variables (eight principal components and seven dummies for the categorical variables). All the input variables are normalized to zero mean and unit variance.

4.3. Model development

The development of a PNN model requires the specification of the smoothing parameter σ. The specification of this parameter in the analysis is based on the generalized cross validation (GCV) criterion, which is defined as follows (Craven & Wahba, 1979):

$$\mathrm{GCV}(\sigma) = M\frac{\mathrm{SSE}(\sigma)}{[M-\lambda(\sigma)]^2}$$

where $M = 2215$ is the number of training observation, SSE is the sum of squared error of the model and λ is the effective number of parameters of the model (degrees of freedom). The effective number of parameters can be easily specified, since a PNN model is a linear fit on the training data of the form $\hat{\mathbf{y}} = \mathbf{D}\mathbf{y}$, where $\hat{\mathbf{y}}$ is vector with the estimated probabilities of non-qualification for the training observations, $\mathbf{y}$ is the $M \times 1$ column vector with the actual classification of the training observations ($y_i = 1$ for any observation i in the unqualified group and $y_i = 0$ otherwise), and $\mathbf{D} = \mathbf{B}\mathbf{W}$ is the design matrix, with

- $\mathbf{B}$ a $M \times M$ matrix with entries
$$b_{ij} = \frac{1}{f_0(\mathbf{x}_i)+f_1(\mathbf{x}_i)}\exp\left(\frac{\mathbf{x}_i\mathbf{x}_j^\top - 1}{\sigma^2}\right)$$
where f_0 and f_1 are the outputs of the summation nodes corresponding to the qualified and the unqualified group, respectively, as defined in (1).
- $\mathbf{W}$ a diagonal matrix such that $w_{ii} = 1/M_k$ for any observation i in group k ($k = 0$ for unqualified observations with $M_0 = 175$ and $k = 1$ for unqualified observations with $M_1 = 2040$).

In this setting, the effective number of parameters of a model developed with some value for the smoothing parameter σ is $\lambda(\sigma) = \mathrm{tr}[\mathbf{D}(\sigma)]$ (Hastie, Tibshirani, & Friedman, 2001).

Fig. 2 presents the GCV criterion, as well as the mean squared error (MSE) and the number of parameters as functions of σ. The minimum value of the GCV criterion (0.13) is obtained at $\sigma = 0.25$ corresponding to a model whose effective number of parameters is approximately 388. However, it is clearly evident that a considerably simpler model with significantly smaller number of parameters can be developed with only a minor increase in GCV. Therefore, σ was set to 0.4, which corresponds to a model with $\lambda = 56.6$ and GCV = 0.15.

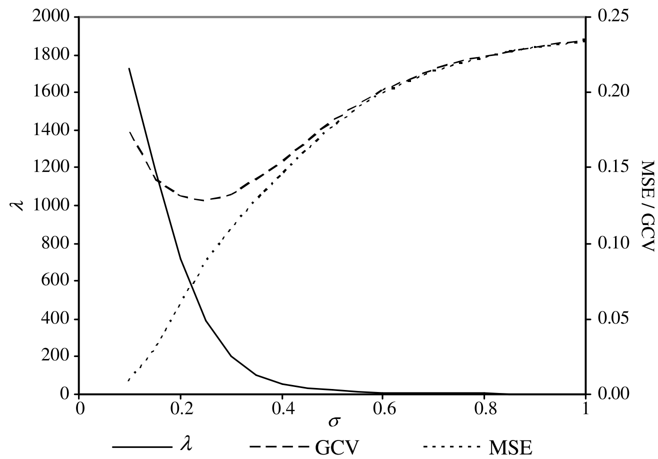

Fig. 2. Mean square error, GCV, and number of parameters as functions of the PNN smoothing parameter.

On the basis of the developed PNN model with the selected setting for the smoothing parameter, an observation is classified in the unqualified group if and only if its estimated probability of receiving a clean (unqualified) report exceeds a fixed cut-off point. This cut-off point is selected on the basis of the Kolmogorov–Smirnov statistic to maximize the difference between the cumulative distribution functions of the two groups of training observations (qualified, unqualified). Table 5 summarizes the obtained classification results, for both the training and the holdout samples. The results involve the classification accuracies for each group of firms, the average of the two accuracies, the overall accuracy, as well as the Gini index (GI) which is defined as GI = 2AUC − 1, where AUC is the area under the receiver operating characteristic (ROC) curve (Fawcett, 2003). The average of the accuracies for the two groups implicitly assumes equal misclassification costs. Despite this assumption this criterion may be more appropriate for problems with considerable imbalance in the class sizes. In this study the qualified cases account for 7.9% of the total number of training cases and 7.96% of the holdout cases. Therefore, a naïve model assigning all the observations as unqualified would have an overall accuracy of 92.1% for the training sample and 92.04% for the holdout sample. However, it is clear that such a model does not

Table 5
Classification results for the PNN model

Sample	Accuracies (in %)				Gini index
	Unqualified	Qualified	Average	Overall	
Training	85.15	84.00	84.57	85.06	81.12
Holdout	84.45	83.15	83.80	84.35	78.22

provide any useful information in discriminating between the two groups. The average accuracy for such a model would be 50% for both the training and the holdout sample, thus designating the failure of the model. On the other hand, the Gini index taken from a ROC analysis provides an assessment of the discriminating power of the model over all possible misclassification costs.

The obtained classification results indicate the satisfactory performance of the developed model, since in the holdout sample it classifies correct 84% of the cases. It is also worth noting that the performance of the model is similar for both the qualified as well as the unqualified cases and furthermore, no significant differences are observed between the training and the holdout sample. Prior studies in auditing, that deal with the application of different forms of neural network models in fraud detection, have achieved rather unbalanced Type I and Type II error rates. For example the models of Fanning, Cogger, and Srivastava (1995) had a 4% Type I error rate and a 70% Type II error rate in the testing sample, while the corresponding errors in the study of Fanning and Cogger (1998) were 41% and 34%, based on the testing sample.

Despite the satisfactory classification performance of PNNs in the context of this study, such models are often criticized as "black-boxes" that do not enable the decision maker and the analyst to make inferences on how the input variables affect the models' result. One way to address this issue is sensitivity analysis, which is performed for each independent variable j as follows. The description of all training observations on the variable j, is fixed at a specific value t_j and the model's results are recorded for all observations. For the variables corresponding to each of the constructed principal components, five different values for t_j are tested, each defined as $p_{j^*} + k(p_j^* - p_{j^*})/4$, with $k = 0, 1, 2, 3, 4$, where p_{j^*} and p_j^* represent the minimum and maximum component scores, respectively. For each categorical variable, k different values are tested, each corresponding to one level of the variable under consideration. For each training case the maximum difference in the output of the model is recorded across all t_j. This maximum difference for each training case is then normalized over the total change for all variables. The sensitivity of the model's results to a change in a specific variable is then measured as the corresponding normalized maximum difference averaged across all training observations. These sensitivities are all positive and sum up to 100%. Table 6 presents the results of this analysis.

Table 6
Sensitivities (in %) of the PNN model on the input variables

Variables	Sensitivity
Profit margins	6.45
Size	8.10
Profitability	24.68
Liquidity & solvency	8.09
Productivity	6.10
Assets turnover	4.95
Credit policy (debtors)	8.03
Credit policy (creditors)	9.10
Credit rating	17.85
Business sector	4.63
Auditor's size	2.04
Total	100

The results show that the most important input variables for the PNN model are the principal component corresponding to profitability as well as the credit rating variable. On the other hand, the assets turnover, the business sector as well as the auditor's size have small influence on the results of the model.

Fig. 3 provides some further details on the influence of the four more important variables on the model's outputs. The results demonstrate that the probability that a qualified audit report will be issued is considerably high for firms with low profitability as well as for firms rated as high risk. The probability of a qualified report is also decreasing with the size of the firm. On the other hand, a non-monotone relationship is observed between the probability of a qualified report and the creditor payment policy that they follow. In particular, the probability of a qualified report is higher for firms that exhibit "abnormal" behavior in terms of their credit payment period. In general the results of the multivariate analysis are consistent with the univariate results of the Kruskal–Wallis test. They also support the findings of previous studies which indicate that firms that receive qualified opinions or have falsified financial statements are less profit (e.g. Laitinen & Laitinen, 1998; Loebbecke et al., 1989; Spathis et al., 2002; Spathis, 2002; Summers & Sweeney, 1998), smaller (e.g. Beasley et al., 1999), and more likely to default (e.g. Bell & Tabor, 1991; Laitinen & Laitinen, 1998; Reynolds & Francis, 2001; Spathis, Doumpos, & Zopounidis, 2003; Spathis, 2002).

4.4. Comparative analysis

The results of the developed PNN model are compared to other techniques which have been widely used in finance and accounting. Feed-forward artificial neural networks (ANN) and logistic regression are used in this comparison. Different ANN architectures are considered including one and two hidden layer back-propagation networks. For the one layer networks 5, 10 and 20 neurons are tested, whereas for the two layer networks 5 and 10 neurons are used at each hidden layer. For each network configuration, 10 train and test runs are performed with different initial random settings for the weights and biases. The results are averaged over the 10 runs.

Table 7 summarizes the results of the comparative analysis for the holdout data. The PNN outperforms all the ANN models both in terms of classification accuracy as well as in terms of the Gini index. For the ANN models, no significant differences are observed among the different architectures, but two hidden layers with five neurons each, is found slightly better than the others. The LR model performs well for the group of unqualified cases, but its

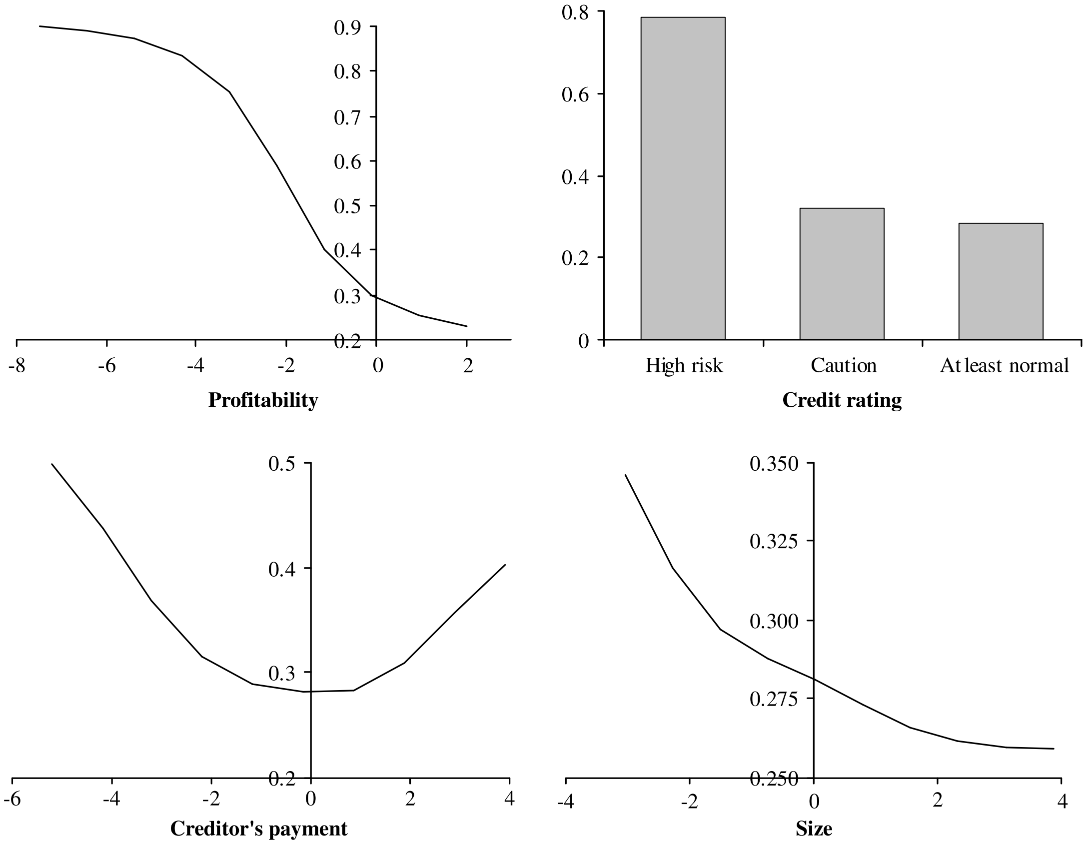

Fig. 3. The probability of qualified opinion in terms of the most significant variables.

Table 7
Comparative classification results (holdout sample)

	Accuracies (in %)				Gini index
	Unqualified	Qualified	Average	Overall	
PNN	84.45	83.15	83.80	84.35	78.22
ANN (1,5)	77.93	81.57	79.75	78.22	72.74
ANN (1,10)	76.12	83.03	79.58	76.67	72.79
ANN (1,15)	79.73	78.65	79.19	79.64	71.78
ANN (1,20)	78.27	80.00	79.14	78.41	70.26
ANN (2,5)	80.51	81.01	80.76	80.55	72.88
ANN (2,10)	76.60	82.47	79.54	77.07	72.41
LR	87.27	73.03	80.15	86.14	67.95

performance for the qualified group is considerably lower than the PNN and ANN models. This unbalanced performance of the LR model leads to an overall accuracy which is higher compared to the rest of the models, but this is simply due to its superior performance for the unqualified cases. On the other hand, in terms of the average accuracy, the LR model is inferior compared to the PNN model as well as to the ANN model with two hidden layers each with five neurons. The performance of the LR model is also considerably lower to the other models in terms of the Gini index.

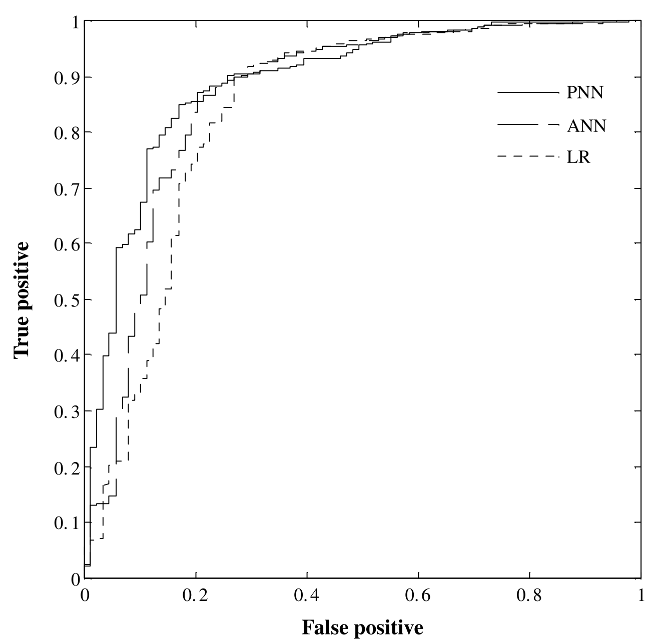

Fig. 4. ROC curves for the PNN, ANN and LR models.

Fig. 4 presents the receiver operating characteristic (ROC) curve for the three models. In this graph, false positive refers to the percentage of qualified cases classified as unqualified, whereas true positive refers to the percentage of unqualified cases that are classified correctly. As the cost of issuing a wrong qualified report (i.e., when actually an unqualified report should be issued) increases, the false positive rate also increases and vise versa. According to the results of the ROC curve, when this cost is low to medium, the PNN model dominates the ANN and LR model, whereas for higher costs all models provide similar results.

5. Conclusions

This study explored the potential of using probabilistic neural networks in developing a model for explaining qualifications in audit reports. The auditors may use such a model to plan specific auditing procedures to achieve an acceptable level of audit risk. Such a model can also be used as a quality control tool for the auditing process.

The analysis was based on a large sample of UK listed firms for the period 1997–2004. Expect for the discriminating and predictive power of the model its interpretability was also explored through sensitivity analysis, which provided useful and meaningful results on the factors that are related to qualifications in audit reports. Thus, in contrast to past studies, the PNN model was not simply used for prediction, but also as an analysis methodology. The performance of the model was found superior compared to other popular techniques which have been used in this field, namely artificial neural networks and logistic regression.

Future research could be directed towards various directions. First, of all due to data availability, it was not possible in the present study to discriminate between the reasons that resulted in the qualified opinion. Hence, future research could for example develop models to classify financial statements in more groups based on whether the qualification was due to disagreements or limitations on scope. Second, the inclusion of additional non-financial variables such as the number of directors, or the complexity of the firm could be examined. Finally, the application of additional techniques, and the combination of them into an integrated model, is another potential extension of the present research.

References

Beasley, S. M, Carcello, J. V., & Hermanson, D. R. (1999). Fraudulent financial reporting: 1987–1997: an analysis of US public companies, Research Report, COSO.

Bell, T., & Tabor, R. (1991). Empirical analysis of audit uncertainty qualifications. *Journal of Accounting Research, 29*, 350–370.

Calderon, T. G., & Cheh, J. J. (2002). A roadmap for future neural networks research in auditing and risk assessment. *International Journal of Accounting Information Systems, 3*(4), 203–236.

Charitou, A., Neophytou, E., & Charalambous, Ch. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review, 13*(3), 465–497.

Chen, K., & Church, B. (1992). Default on debt obligations and the issuance of going concern opinions. *Auditing: A Journal of Practice and Theory*(Fall), 30–49.

Chen, A.-S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research, 30*(6), 901–923.

Craswell, A. (1999). Does the provision of non-audit services impair auditor independence? *International Journal of Auditing, 3*, 29–40.

Craven, P., & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik, 31*, 377–403.

DeAngelo, L. (1981). Auditor size and audit quality. *Journal of Accounting and Economics, 3*, 183–199.

Dopuch, N., Holthausen, R. W., & Leftwich, R. W. (1987). Predicting audit qualifications with financial and market variables. *Accounting Review, LXII*(3), 431–453.

Etheridge, H., & Sriram, R. (1997). A comparison of the relative cost of financial distress models. *International Journal of Intelligent Systems in Accounting Finance and Management, 6*(3), 235–248.

Fanning, K., & Cogger, K. O. (1998). Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting Finance and Management, 7*(1), 21–41.

Fanning, K., Cogger, K. O., & Srivastava, R. (1995). Detection of management fraud: a neural network approach. *International Journal of Intelligent Systems in Accounting Finance and Management, 4*(2), 113–126.

Fawcett, T., (2003). ROC graphs: notes and practical considerations for researchers, Technical Report HPL-2003-4, HP Laboratories. Available from www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf.

Hamid, S. A., & Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research, 57*(10), 1116–1125.

Hansen, J. V., McDonald, J. B., & Stice, J. D. (1992). Artificial intelligence and generalized qualitative-response models: an empirical test on two audit decision-making domains. *Decision Sciences, 23*(3), 708–723.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: data mining, inference and prediction*. New York: Springer.

Huang, Z., Chen, H., Hsu, Ch.-J., Chen, W-H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems, 37*(4), 543–558.

Hung, S.-Y., Liang, T.-P., & Liu, V. W.-C. (1996). Integrating arbitrage pricing theory and artificial neural networks to support portfolio management. *Decision Support Systems, 18*(3–4), 301–316.

Ireland, J. (2003). An empirical investigation of determinants of audit reports in the UK. *Journal of Business Finance & Accounting, 30*(7 & 8), 975–1015.

Keasey, K., & Watson, R. (1987). Non-financial symptoms and the prediction of small company failure: a test of Argenti's hypotheses. *Journal of Business Finance and Accounting, 14*(3), 335–354.

Keasey, K., Watson, R., & Wynarzcyk, P. (1988). The small company audit qualification: a preliminary investigation. *Accounting and Business Research, 18*(72), 323–333.

Kim, S. H., & Chun, S. H. (1998). Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index. *International Journal of Forecasting, 14*(3), 323–337.

Kleinman, G., & Anandarajan, A. (1999). The usefulness of off-balance sheet variables as predictors of auditors' going concern opinions: an empirical analysis. *Managerial Auditing Journal, 14*(6), 273–285.

Koh, H. C. (1991). Model predictions and auditor assessments of going concern status. *Accounting and Business Research, 21*(84), 331–338.

Krishnan, J., & Krishnan, J. (1996). The role of economic trade-offs in the audit opinion decision: an empirical analysis. *Journal of Accounting, Auditing and Finance, 11*(4), 565–586.

Laitinen, E. K., & Laitinen, T. (1998). Qualified audit reports in Finland: evidence from large companies. *The European Accounting Review, 7*(4), 639–653.

Li, S.-T., Shue, L.-Y., & Shiue, W. (2000). The development of a decision model for liquidity analysis. *Expert Systems with Applications, 19*(4), 271–278.

Loebbecke, J., Eining, M., & Willingham, J. (1989). Auditor's experience with material irregularities: frequency, nature, and detectability. *Auditing: A Journal of Practice and Theory, 9*, 1–28.

Luther, R. K. (1998). An artificial neural network approach to predicting the outcome of Chapter 11 bankruptcy. *The Journal of Business and Economic Studies, 4*, 57–73.

McKeown, J. C., Mutchler, J. F., & Hopwood, W. (1991). Towards an explanation of auditor failure to modify the audit opinions on bankrupt companies. *Auditing: A Journal of Practice and Theory, 10*, 1–13.

Pendharkar, P. C. (2005). A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem. *Computers & Operations Research, 32*(10), 2561–2582.

Reynolds, J., & Francis, J. (2001). Does size matter? The influence of large clients on office-level auditor reporting decisions. *Journal of Accounting and Economics, 30*, 375–400.

Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). A new tool for predicting thrift failures. *Decision Sciences, 23*, 899–916.

Spathis, Ch. (2002). Detecting false financial statements using published data: evidence from Greece. *Managerial Auditing Journal, 17*(4), 179–191.

Spathis, Ch., Doumpos, M., & Zopounidis, C. (2002). Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. *The European Accounting Review, 11*(3), 509–535.

Spathis, Ch., Doumpos, M., & Zopounidis, C. (2003). Using client performance measures to identify pre-engagement factors associated with qualified audit reports in Greece. *The International Journal of Accounting, 38*, 267–284.

Specht, D. (1990). Probabilistic neural networks. *Neural Networks, 3*, 109–118.

Summers, S. L., & Sweeney, J. T. (1998). Fraudulently misstated financial statements and insider trading: an empirical analysis. *The Accounting Review, 73*(1), 131–146.

Wines, G. (1994). Auditor independence, audit qualifications and the provision of non-audit services: a note. *Accounting and Finance, 34*, 75–86.

Wong, B. K., Lai, V. S., & Lam, J. (2000). A bibliography of neural network business applications research: 1994–1998. *Computers & Operations Research, 27*(11–12), 1045–1076.

Wong, B. K., & Selvi, Y. (1998). Neural network applications in finance: a review and analysis of literature (1990–1996). *Information & Management, 34*(3), 129–139.

Yang, Z. R., Platt, M. B., & Platt, H. D. (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research, 44*, 67–74.

Yim, J., & Mitchell, H. (2005). Comparison of country risk models: hybrid neural networks, logit models, discriminant analysis and cluster techniques. *Expert Systems with Applications, 28*(1), 137–148.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. *European Journal of Operational Research, 116*(1), 16–32.