Auditing Journal Entries Using Self-Organizing Map

Argyris Argyrou
Hanken School of Economics, Helsinki, Finland., argyris.argyrou@hanken.fi

Auditing Journal Entries Using Self-Organizing Map

Argyris Argyrou
HANKEN School of Economics, Helsinki, Finland
argyris.argyrou@hanken.fi

ABSTRACT

A considerable body of regulatory pronouncements attests to the significance of auditing journal entries for ensuring that financial statements are free of material misstatements; however, existing empirical studies have paid insufficient attention to the audit of journal entries. To explore this issue further, this paper proposes a model based on self-organizing map as well as validates this model by performing experiments on a dataset containing journal entries. Empirical results suggest that the proposed model can detect "suspicious" and legitimate transactions with a high degree of accuracy. Further investigations reveal that the performance of the model is robust to varying prior probabilities of "suspicious" journal entries occurring in the population. The findings indicate that the model can assist auditors in detecting "suspicious" journal entries.

Keywords

Auditing, journal entries, self-organizing map

INTRODUCTION

Auditing journal entries has gained prominence as a result of statute and auditing standards that have been promulgated in the wake of financial malfeasance (e.g. Enron, WorldCom) in order to protect investors. In essence, Sarbanes and Oxley Act (U.S. Congress, 2002) Sections 302 and 404 require the management of companies to establish and maintain internal controls as well as to assess the effectiveness of these controls at the end of a fiscal year; and, Auditing Standard No.5 (Public Company Accounting Oversight Board (PCAOB), 2007) requires external auditors to give an opinion on the effectiveness of a company's internal controls. Further, auditors are required to test the appropriateness of journal entries recorded in the general ledger and other adjustments made in preparing financial statements in order to obtain reasonable assurance about whether the financial statements are free of material misstatements, whether caused by errors or fraud (PCAOB, 2002). In addition, the Standing Advisory Group (PCAOB, 2004) considers "... significant or unusual journal entries ... " to be an area that poses a high risk of material misstatement due to fraud. Auditors may be exposed to litigation (Palmrose, 1991) and subject to regulatory enforcement releases (Bonner et al., 1998), if they fail to detect material misstatements to financial statements. However, absent a single study by Debreceny and Gray (2010), existing literature has paid little attention to the audit of journal entries; indeed, literature acknowledges the paucity of empirical studies in this area (Hogan et al., 2008).

This paper proposes and then validates a model in order to assist auditors in detecting "suspicious" journal entries. To this end, the paper adopts an one-class classification approach (Chandola et al., 2009; Juszczak et al., 2008), because, in an audit engagement, a dataset does not contain journal entries that are pre-labelled as "suspicious" nor as legitimate. As a result, detecting "suspicious" journal entries can be achieved only with reference to information describing legitimate journal entries. The implicit assumption being the vast majority of journal entries are legitimate and generated by a probability distribution that is different form that generating "suspicious" journal entries. Given this assumption, the paper employs self-organizing map (SOM; Kohonen, 1997), an unsupervised-learning algorithm, to derive a reference model that can describe the behaviour of legitimate journal entries. The paper compares novel journal entries against the reference model to determine whether they are "suspicious" or legitimate; it does so by calculating the quantization error, a distance metric, between individual journal entries and the SOM-based reference model. A journal entry having a quantization error that exceeds a specified threshold is deemed to be "suspicious".

To validate the proposed SOM-based model, the paper conducts a series of experiments on a dataset that has been made available by an international shipping company, and contains the complete set of their journal entries for fiscal year 2006. The paper incorporates certain characteristics of an audit engagement in order to carry out as realistic experiments as possible. First, it simulates four categories of potential errors that may affect a journal entry; the number of categories is restricted by the number of variables describing a journal entry. Second, it considers three prior probabilities, or prevalence, of "suspicious" journal entries occurring in the population, because the number of "suspicious" journal entries is likely to be orders of magnitude smaller than that of legitimate. Third, it examines four cost-ratios of Type I and Type II errors, as in an audit engagement the cost of a Type II error, identifying a "suspicious" journal entry as legitimate, tends to be much higher than that of a Type I error, identifying a legitimate journal entry as "suspicious". Fourth, the paper initialises a range of thresholds and selects that which minimises the total misclassification cost as the optimum threshold.

In the following section, the paper reviews related studies; subsequently, it describes the data, introduces SOM and the reference model as well as elaborates on the simulation of "suspicious" journal entries and the experiments. Further, the paper presents and discusses the results, and then provides conclusions and suggests directions for additional research.

BACKGROUND AND RELATED STUDIES

The Statement On Auditing Standards 99 (SAS 99): Consideration of Fraud in a Financial Statement Audit (AICPA, 2002), which is the interim standard AU Section 316 of PCAOB, requires an auditor to obtain reasonable assurance that financial statements are free of material misstatements, whether caused by error or fraud. Further, AU Section 316 0.58 requires an auditor, among other things, to test the appropriateness of journal entries recorded in the general ledger as well as other adjustments made in preparing financial statements. The reason is management can misstate financial statements by recording fictitious and inappropriate journal entries, especially towards the end of a fiscal year (AU Section 316 .08, and .58). Financial statement fraud entails considerable legal and economic costs for companies, directors, and shareholders (COSO, 2010) as well as may expose auditors to litigation (Palmrose, 1991).

A research synthesis of the literature pertinent to financial statement fraud acknowledges the lack of empirical studies that address the relationship between unusual or top-level journal entries and financial statement fraud (Hogan et al., 2008). In order to test journal entries for detecting fraud, Debreceny and Gray (2010) have applied digit analysis or Benford's Law to a set of journal entries that were obtained from 27 organizations across different industry types and sizes. In brief, the paper compares the observed distribution of the first digit of US$ amounts against that expected by Benford's Law; and if the difference is statistically significant under a chi-square test, then the US$ amount is deemed to be suspicious. The results have suggested that for all organizations the observed distribution of the first digit of US$ amounts were significantly different form that expected by Benford's Law. However, the results may be an artefact of the chi-square test, as a large number of observations can induce statistically significant results (Grabski, 2010). A further explanation is that either fraudulent journal entries were the norm in the sample, or Benford's Law is not applicable to journal entries (Grabski, 2010).

The self-organizing map has been applied to detect "anomalous" behaviour in a number of diverse domains where information describing this type of behaviour does not exist in advance, or is rare, or expensive to collect; as a result, detecting "anomalous" behaviour can be based only on information describing "legitimate" behaviour. For example, SOM has been used to detect fraudulent insurance claims (Brockett et al., 1998), to develop risk-adjustment models for estimating medical expenditures (Hsu et al., 2008), and detect fraudulent credit-card transactions (Juszczak et al., 2008).

Further, Ypma and Duin (1997) has detected faults in rotating mechanical machinery and leaks in pipelines by using a variant of the SOM-based metric proposed by Kaski and Lagus (1996). SOM has been used as an intrusion detection tool for detecting " anomalous" network traffic (e.g. TCP/IP packets) that may signal a possible Denial of Service attack (Labib and Vemuri, 2002), detecting instances of buffer-overflow attacks based on a model of legitimate network traffic (Rhodes et al., 2000), and identifying "anomalous" behaviour that may exist in network services (e.g. web, e-mail, telnet) (Ramadas et al., 2003).

RESEARCH DESIGN AND METHODOLOGY

Data description

The raw data have been provided by an international shipping company in the form of a text-file containing the complete set of their journal entries, $n = 6,404$, for fiscal year 2006. The text-file consists of 25,422 lines and eight columns, representing accounting transactions and variables, respectively. The number of accounting transactions is greater than that of journal entries, because a journal entry can be made up of any number of transactions as long as the "Debit" and "Credit" sides are equal. The eight variables are as follows: (i) "Account Number" (alphanumerical), (ii) "Account Name" (text), (iii) "Posting Date" (date), (iv) "US$ amount" (numerical), (v) "Debit-Credit Indicator" (binary), (vi) "Description" (text), (vii) "Account Class" (hierarchical), and (viii) "Code" (numerical); the types of the variables are shown in parentheses.

The paper uses the individual accounting transactions as the unit of analysis and the following three variables: (i) "US$ amount", (ii) "Debit-Credit Indicator", and (iii) "Code". The paper aggregates the transactions at the "Account Class" level ($n = 27$), because there are not enough transactions at the "Account Number" level ($n = 360$) for the paper to perform statistical analyses, nor to draw meaningful conclusions. Descriptive statistics are shown in Table 1.

Self-organizing map

SOM performs two operations: first, it performs vector-quantization by representing input vectors with a much smaller, albeit representative, set of codevectors; and second, it carries out a non-linear mapping or projection from a high-dimensional input space to a regular two-dimensional grid of neurons, while preserving the original topology as faithfully as possible. For the purpose of this study, SOM is employed to perform only vector-quantization.

In the context of this study, the input dataset to SOM is denoted by $X = (\vec{x}_{ij})_{nd}$, where $n = 25,422$ represents the number of transactions, and $d = 3$ denotes the dimensionality of the dataset. Given this input dataset, SOM constructs a set of codevectors, $M = (\vec{m}_{kj})_{Kd}$, where $d = 3$ as above, and $k = 1,2,3,\ldots,810$ denotes the number of codevectors, which is approximately equal to 5 x $\sqrt{25,422}$ (Vesanto et al., 2000, p.30); the size of the SOM grid is set to 30 rows and 27 columns. A neuron k is described by a tuple $(\vec{m}_k, \vec{p}_k)$, where $\vec{m}_k \in \mathbb{R}^3$ is a codevector and $\vec{p}_k \in \mathbb{R}^2$ is a location vector on the SOM grid.

SOM is formed in three iterative processes. First, in the competition process, each input vector, $\vec{x}_i \in \mathbb{R}^3$, is compared with all codevectors, $\vec{m}_k \in \mathbb{R}^3$, and the best match in terms of the smallest Euclidean distance, $\| \vec{x}_i - \vec{m}_k \|$, is mapped onto neuron k, termed the best-matching unit (i.e. BMU), and denoted by the subscript c: $\| \vec{x}_i - \vec{m}_c \| = \min_k \{\| \vec{x}_i - \vec{m}_k \|\}$ (Kohonen, 1997, p.86).

Second, in the co-operation process, the BMU locates the centre of a neighbourhood kernel, $h_{ck}(t)$, which is usually a Gaussian function defined as: $h_{ck}(t) = exp\left[-\frac{\|\vec{p}_c - \vec{p}_k\|^2}{2\sigma^2(t)}\right]$, where $\vec{p}_c$, $\vec{p}_k \in \mathbb{R}^2$ are the location vectors of BMU and neuron k respectively, t denotes discrete time, and $\sigma(t)$ defines the width of the kernel (Kohonen, 1997, p.87).

Third, in the adaptive process, the sequence-training SOM updates recursively codevector $\vec{m}_k$ as follows: $\vec{m}_k(t+1) = \vec{m}_k(t) + a(t)h_{ck}(t)\left[\vec{x}_i(t) - \vec{m}_k(t)\right]$, where $0 < a(t) \leqslant 1$ is a learning rate at discrete time t, and $a(t)$ is a non-increasing function of time. Batch-training SOM updates the codevectors only at the end of each epoch, which is a complete presentation of input data, rather than recursively, as follows (Vesanto et al., 2000, p.9): $\vec{m}_k(t+1) = \frac{\sum_{i=1}^{n} h_{ck}(t)\vec{x}_i}{\sum_{i=1}^{n} h_{ck}(t)}$.

Finally, SOM converges to a stable state when the codevectors do not get updated any further; the convergence criterion is $E\{h_{ck}(\vec{x}_i - \lim_{t\to\infty} \vec{m}_k(t))\} = 0$, where $E\{.\}$ denotes the expectation function (Kohonen, 1997, p.113).

					Interquartile range	
Account Class	Code	N	Mean	MAD(*)	0.75	0.25
Property, Plants and Equipment: Costs	1000	25	13,202	913	982	-526
Property, Plants and Equipment: Depr.	1100	61	-3,302	386	-98	-1,139
Investments in Subsidiaries	1200	3	11,750	9,800	21,413	2,550
Cash in Bank	2110	3,744	1,800	4,913	13,983	-247
Cash at Hand	2140	259	13	84	39	-113
Sales Taxes Receivable	2170	812	69	8	42	3
Trade Debtors	2200	3,989	-1,725	4,037	1,725	-8,812
Other Trade Debtors	2204	1,048	538	1,549	1,566	-1,495
Accounts Receivable	2207	309	0	36	7	-60
Short-term Loans Receivable	2300	14	1,582,003	110,434	2,500,000	-5,000
Insurance Receivable	2400	4,963	-86	1,869	1,843	-2,000
Other Debtors and Receivables	2600	177	1,046	979	762	-1,302
Accounts Payable	3455	3,242	-75	296	112	-546
Trade Creditors	3480	1,107	-4,542	1,743	-135	-6,165
Other Income Received	6000	19	-12,671	186	-139	-430
Fees Received	6100	36	-13,333	15,000	-1,500	-28,500
Insurance Commission Received	6260	1,996	-805	332	-89	-1,125
Bank Interest Received	6300	344	-1,307	712	-215	-2,367
Other Interest Received	6305	8	-145,875	10,238	-130,668	-153,653
Exchange Difference: Gain	6440	678	-788	18	1	-42
Administration Expenses	8500	1,236	65	57	114	-6
Office Expenses	8501	466	222	176	332	-54
Salaries and Wages	8550	214	3,034	4,372	7,392	-2,516
Fees and Commissions	8600	121	817	107	592	17
Professional Expenses	8610	28	3,440	583	1,757	-52
Finance Expenses	8700	465	158	14	36	10
Depreciation Expenses	8800	58	941	388	1,139	165
		25,422				

(*) Median absolute difference. Amounts are expressed in US$.

Table 1: Descriptive statistics

Reference model

The set of stable-state codevectors, $M = (\vec{m}_{kj})_{Kd}$, can function as the reference model of input data, because the probability density function of the codevectors approximates that of input data (Kohonen, 1999, 1997, p.48). The reference model thus derived enjoys a number of properties that can enhance its application. First, it does not make any assumptions concerning the probability distribution of the transactions, as the codevectors are essentially non-parametric regressors of the transactions. Second, it is robust to changes to input data, because all the pertinent operations (e.g. calculating distances) are performed in the input space; and thus any changes in the input data would propagate corresponding changes to the estimation and update of codevectors. Third, it is derived directly from the data, and hence does not entail encoding domain-expert knowledge, as the case would have been had the paper adopted a model-driven approach. Fourth, the paper can detect transactions deviating from the reference model without having to form expectations about these transactions in advance. Fifth, the computational complexity of the model scales well with large datasets, because the number of codevectors is often chosen to be approximately equal to the square root of the number of input vectors, $K \approx \sqrt{n}$.

The paper assesses the degree of affinity between the reference model and a transaction by calculating the quantization error: $R_i = \| \vec{x}_i - \vec{m}_c \|^2$; it represents the Euclidean distance between the i^{th} transaction, $\vec{x}_i$, and the codevector corresponding to its BMU, $\vec{m}_c$. This scoring mechanism is feasible and valid, because quantization error is monotonically related to the degree of "suspiciousness" (Bolton and Hand, 2002). The larger a transaction's quantization error is, the farther away from the reference model this transaction is going to be, and hence the more likely it could be "suspicious". An equivalent interpretation is transactions having quantization errors that exceed a specified threshold, and hence are considered to be "suspicious", are generated by a probability distribution that is different from that generating legitimate transactions.

Simulating "suspicious" transactions

To represent an audit engagement as faithfully as possible, the paper addresses the following four issues: (i) categories of potential errors that may affect a transaction, (ii) prior probability, or prevalence, of "suspicious" transactions occurring in the population, (iii) asymmetrical misclassification costs of Type I and Type II errors, and (iv) optimum threshold that can distinguish "suspicious" from legitimate transactions.

Four categories of potential errors

In the present case, a transaction is described by three variables and can be considered as "suspicious" if either of these variables contain an error given the rest are error-free, or all of the variables contain errors. This combination yields four categories of potential errors that may affect a transaction. The paper assumes that the four categories occur with an equal probability.

In order to simulate errors in "US$ amount", the paper adds a noise to this variable; this noise is equal to the average of median-absolute-difference of an "Account Class" that is selected randomly excluding the "Account Class" of the transactions whose "US$ amount" are to be modified. Second, to simulate errors in "Debit - Credit Indicator", the paper reverses the binary indicator, $\{1,0\}$, thereby changing "Debit", denoted by 1, to "Credit", denoted by 0, and vice versa. This operation is equivalent to multiplying the "US$ amount" by (-1), thus converting positive amounts, Debit balances, to negative amounts, Credit balances, and conversely. Third, the paper replaces the value of "Account Class" by a different value selected randomly; and the combination of the foregoing three categories forms the fourth category of potential errors.

Prior probabilities

The paper investigates the following three prior probabilities of "suspicious" transactions : $p1 = 5\%$, $p2 = 3\%$, and $p3 = 1\%$. For each of the three probabilities, the paper sets up a dataset that contains both legitimate transactions, selected randomly from the input dataset, and simulated "suspicious" transactions that are seeded according to the foregoing probabilities; the paper elaborates on this procedure in the section describing the experiments.

Asymmetrical misclassification costs

The model produces a binary output, $T \in \{1,0\}$, where 1 and 0 denote "suspicious" and legitimate transactions, respectively; and similarly, $D \in \{1,0\}$ represents the actual classes of transactions. The model can make two types of errors: (i) Type I, or false positive, when it identifies a legitimate transaction as "suspicious", and (ii) Type II, or false negative, when it fails to identify a "suspicious" transaction as such. Table 2 describes the two types of errors, the two correct outcomes, and their respective costs; for example, $C_{s/l}$ and $C_{l/s}$ represent the costs of false negative and false positive, respectively. The paper assumes that no costs are incurred in identifying "suspicious" and legitimate transactions correctly (i.e. $C_{s/s} = C_{l/l} = 0$). Although actual costs may be difficult to estimate, in an audit engagement the cost of false negative, $C_{s/l}$, tends to be much higher than that of false positive, $C_{l/s}$. Consequently, the paper investigates the following four cost-ratios, $r = C_{l/s}/C_{s/l}$: 1:1, 1:10, 1:20, and 1:30.

	Model	
Actual	suspicious (1)	legitimate (0)
suspicious (1)	true positive $C_{s/s}$	false negative $C_{s/l}$
legitimate (0)	false positive $C_{l/s}$	true negative $C_{l/l}$

Table 2: Classification and cost matrix

Optimum threshold

In order to estimate the optimum threshold, the paper minimises the total misclassification cost that is calculated as follows:

$$C = C_{l/s}\, P(T = 1|D = 0)\, P(D = 0) + C_{s/l}\, P(T = 0|D = 1)\, P(D = 1). \tag{1}$$

Where $P(T = 1|D = 0)$, the false positive rate, denotes the probability that the model identifies a transaction as "suspicious", given that it is legitimate; $P(T = 0|D = 1)$, the false negative rate, represents the probability that the model identifies a transaction as legitimate, given that it is "suspicious"; $P(D = 1)$ represents the prior probability of "suspicious" transactions occurring in the population; and, by definition, $P(D = 0) = 1 - P(D = 1)$ represents the prior probability of legitimate transactions.

The model's binary output, $T \in \{1,0\}$, can be expressed as: $T_i = \begin{cases} 1, R_i \geq u \\ 0, \mathit{otherwise} \end{cases}$. Where u is a threshold, $R_i = \| \vec{x}_i - \vec{m}_c \|^2$ represents the " suspiciousness" score the model estimates for each transaction, and $i = 1, 2, \ldots, 25,422$ denotes the transactions. A range of thresholds is initialised, and that which minimises the total misclassification cost, Equation 1, is declared to be the optimum threshold.

Experiments

The paper conducts the experiments in five steps depicted as an input-process-output diagram in Figure 1. First, it uses bootstrap, $B = 100$, to select one hundred random samples with replacement from the empirical distribution of the input dataset, $X = (\vec{x}_{ij})_{nd}$. Second, for each bootstrap dataset (i.e. $X^1, \ldots, X^{100}$), the paper simulates and then seeds "suspicious" transactions according to three prior probabilities: $p1 = 5\%$, $p2 = 3\%$, and $p3 = 1\%$. For example, Y^1_{p1}, Y^1_{p2}, and Y^1_{p3} denote the three datasets that correspond to the first bootstrap dataset (i.e. X^1) and contain simulated "suspicious" transactions in the order of $5\%, 3\%$, and 1%, respectively. Third, for each bootstrap dataset, the paper uses SOM-toolbox for Matlab (Vesanto et al., 2000) to train a SOM in batch mode with hexagonal grid of neurons and Gaussian neighbourhood. Each SOM produces a set of codevectors that constitutes the reference model describing the legitimate behaviour of the corresponding bootstrap dataset. For example, $M^2 = (\vec{m}_{kj})_{Kd}$ forms the reference model of dataset $X^2 = (\vec{x}_{ij})_{nd}$, where $K = 810$ denoting the number of neurons and that of codevectors, $n = 25,422$ reflecting the number of transactions, and $d = 3$ representing the dimensionality of input dataset. It is worth repeating that the paper trains SOMs on the bootstrap datasets that contain only legitimate transactions, and consequently the sets of codevectors do not include any information concerning "suspicious" transactions. In the fourth step, the paper applies the sets of codevectors, derived in the preceding training step, on the datasets containing both legitimate and simulated "suspicious" transactions in order to calculate the quantisation error, $R^b_i = \| \vec{y}_i{}^b - \vec{m}^b_c \|^2$; this error represents the "suspiciousness" score of each transaction. Fifth, the paper initialises a range of candidate thresholds, and that which minimises the total misclassification cost, Equation 1, is chosen to be the optimum threshold.

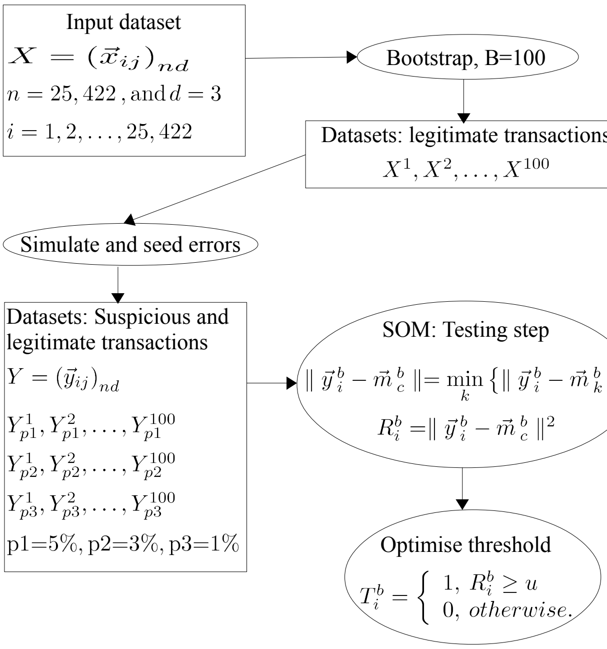
Input dataset
$X = (\vec{x}_{ij})_{nd}$
$n = 25,422$, and $d = 3$
$i = 1, 2, \ldots, 25,422$
Bootstrap, B=100
Datasets: legitimate transactions
$X^1, X^2, \ldots, X^{100}$
Simulate and seed errors
SOM:Training step
Datasets: Suspicious and legitimate transactions
$Y = (\vec{y}_{ij})_{nd}$
$Y^1_{p1}, Y^2_{p1}, \ldots, Y^{100}_{p1}$
$Y^1_{p2}, Y^2_{p2}, \ldots, Y^{100}_{p2}$
$Y^1_{p3}, Y^2_{p3}, \ldots, Y^{100}_{p3}$
p1=5%, p2=3%, p3=1%
SOM: Testing step
$\| \vec{y}^{\,b}_i - \vec{m}^{\,b}_c \| = \min_k \{\| \vec{y}^{\,b}_i - \vec{m}^{\,b}_k \|\}$
$R^b_i = \| \vec{y}^{\,b}_i - \vec{m}^{\,b}_c \|^2$
Reference Models
$M^b = (\vec{m}_{kj})_{Kd}$
K $= 810$, $d = 3$
b $= 1, 2, \ldots, 100$
$k = 1, 2, \ldots, 810$
Optimise threshold
$T^b_i = \begin{cases} 1, & R^b_i \geq u \\ 0, & otherwise. \end{cases}$

Figure 1: An input-process-output representation of the experiments

RESULTS PRESENTATION AND DISCUSSION

To recapitulate, the paper proposes and validates via experiments a SOM-based model in order to detect "suspicious" transactions. To conduct the experiments, the paper considers twelve scenarios by combining three prior probabilities of "suspicious" transactions and four cost-ratios of Type I and Type II errors. For each scenario, the paper addresses four categories of potential errors that may affect a transaction and performs one hundred experiments by using bootstrap.

	Cost	True Negative Rate						True Positive Rate					
Prior	Ratios	Mean			Stdev			Mean			Stdev		
			95% C.I			95% C.I			95% C.I			95% C.I	
5%	1:1	0.957	0.954	0.960	0.015	0.013	0.019	0.820	0.788	0.847	0.151	0.131	0.176
	1:10	0.910	0.861	0.938	0.185	0.102	0.281	0.883	0.854	0.907	0.130	0.105	0.158
	1:20	0.898	0.840	0.927	0.211	0.139	0.311	0.889	0.859	0.911	0.130	0.104	0.156
	1:30	0.897	0.844	0.930	0.211	0.136	0.299	0.891	0.859	0.912	0.130	0.108	0.163
3%	1:1	0.953	0.940	0.959	0.041	0.019	0.079	0.881	0.843	0.908	0.168	0.136	0.207
	1:10	0.914	0.876	0.938	0.149	0.087	0.248	0.936	0.907	0.955	0.116	0.086	0.159
	1:20	0.893	0.836	0.920	0.191	0.124	0.278	0.945	0.918	0.963	0.111	0.073	0.153
	1:30	0.885	0.831	0.916	0.206	0.136	0.290	0.946	0.919	0.964	0.111	0.075	0.159
1%	1:1	0.962	0.959	0.965	0.016	0.013	0.020	0.812	0.781	0.840	0.149	0.136	0.172
	1:10	0.917	0.871	0.943	0.181	0.099	0.271	0.860	0.830	0.885	0.148	0.128	0.177
	1:20	0.905	0.849	0.937	0.204	0.137	0.298	0.862	0.834	0.889	0.147	0.127	0.178
	1:30	0.897	0.840	0.931	0.216	0.137	0.298	0.862	0.829	0.890	0.148	0.127	0.177

Table 3: Statistical analyses of results

The results are depicted in Table 3; each line corresponds to a scenario and represents the outcome of one hundred experiments. For example, the second line describes the performance of the model when the prior probability of “suspicious” transactions is 5% and the cost of a Type II error is 10 times that of a Type I error. In this scenario, the model can detect on average 91% ($C.I_{0.95} = 0.861 - 0.938$) of the legitimate transactions as being legitimate as well as 88.3% ($C.I_{0.95} = 0.854 - 0.907$) of the “suspicious” transactions as being “suspicious”; the respective 95% confidence intervals are shown in parentheses.

A closer examination of the results reveals that within each of the three prior probabilities, as the cost of Type II error increases relative to that of Type I error, the mean true positive rate increases, whereas the mean true negative rate decreases. This pattern is expected, because a more costly Type II error would shift the threshold, optimised according to Equation 1, in favour of the model identifying more transactions as “suspicious”. Consequently, the true positive rate would be higher, albeit only at the expense of a higher false positive rate.

To analyse further the performance of the proposed model, the paper presents in Figure 2 the averaged Receiver Operating Characteristic Curve (i.e. ROC curve) that corresponds to the four scenarios using a 5% prior probability of “suspicious” transactions. In brief, the ROC curve plots the model’s true positive rate (y-axis), or sensitivity, against the false positive rate (x-axis), or 1 - specificity. The ascending diagonal stands for the random or non-informative classifier, and any classifier that appears below this diagonal performs worse than chance. Further, points (0,0) and (1,1) represent classifiers that identify all transactions as “suspicious” and as legitimate, respectively; whereas, point (0,1) marks the perfect classifier that can identify all “suspicious” and legitimate transactions correctly. The area under the ROC curve, $AUROC = 0.975$, denotes the probability with which the model can identify correctly random pairs of legitimate and “suspicious” transactions, and thus it summarises the discriminatory qualities of the model (Hanley and McNeil, 1982).

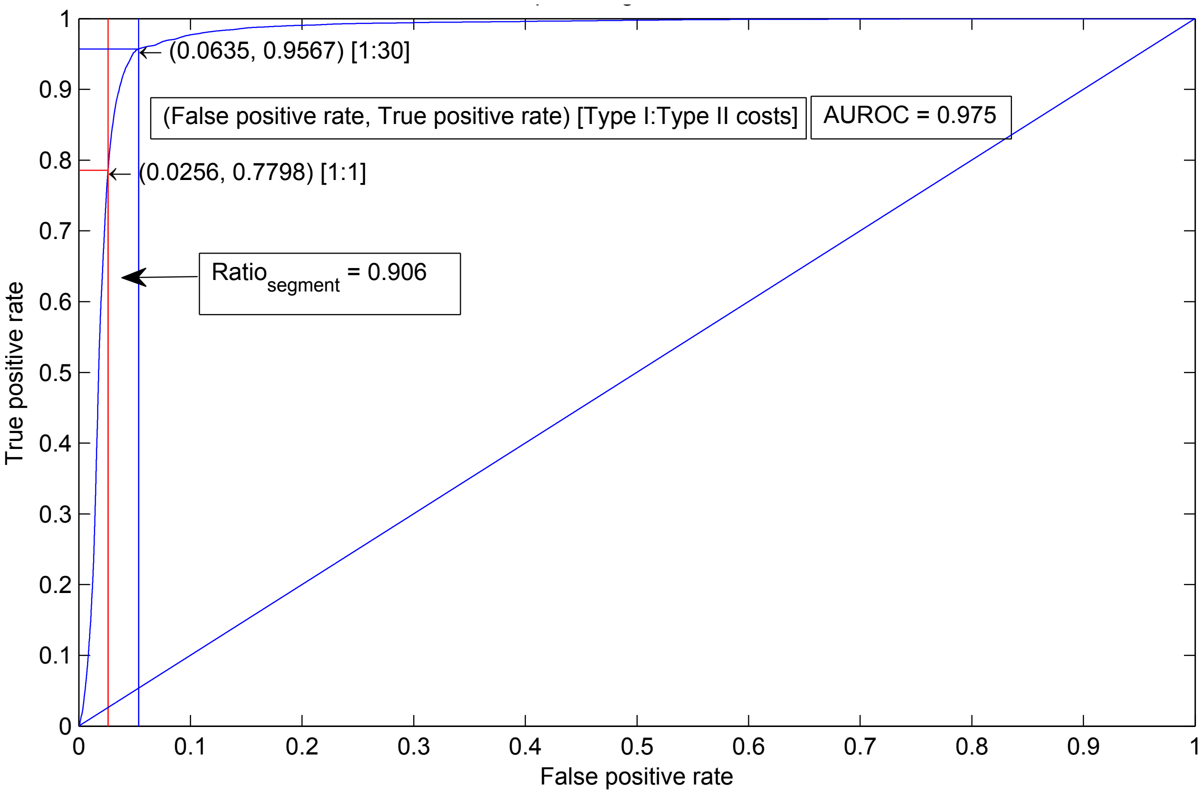

Figure 2: Averaged ROC curve for the four scenarios using a 5% probability of “suspicious” transactions

The presence of asymmetrical misclassification costs induces different operating points on the ROC curve that identify different trade-offs between true positive and false positive rates. For example, the operating points (0.0635, 0.9567) and (0.0256, 0.7798) minimise the total misclassification cost, Equation 1, when the cost-ratios of Type I and Type II errors are 1:30 and 1:1, respectively. The uncertainty about cost-ratios can be incorporated in the ROC curve by considering the range the two operating points delineate. The area corresponding to this range can function as a measure of the model's performance that is more robust than the AUROC, because only this segment of the ROC curve is useful for decision making. In a process analogous to estimating the AUROC, the paper defines $Ratio_{segment}$ as the ratio of the area under the segment of ROC curve that is of interest to the area of the corresponding rectangle of unit length. The model exhibits a $Ratio_{segment} = 0.906$, whereas, by definition, the perfect classifier would have $Ratio_{segment} = 1$.

CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

This paper proposes a SOM-based model for detecting "suspicious" journal entries. While statute and auditing standards require auditors to consider the complete set of journal entries in planning and performing audits, existing literature, absent a single study (Debreceny and Gray, 2010), has paid insufficient attention to the audit of journal entries.

Empirical analyses suggest that the proposed model enjoys a high true positive rate in detecting "suspicious" journal entries. Further investigations reveal that the performance of the model is robust to varying prior probabilities of "suspicious" journal entries occurring in a population as well as to asymmetrical misclassification costs of Type I and Type II errors. These findings allow the paper to infer that the model can be generalised to datasets beyond the present case. As a result, the model can have a practical application in the domain of accounting in that it can aid auditors to detect "suspicious" journal entries.

Ongoing research aims at developing a probabilistic model that can assign a probability, rather than a score, to a journal entry. This probability would indicate the degree to which a journal entry is considered to be "suspicious", and be updated by using Bayesian analysis, given different estimates of audit risks. An additional line of research could be examining whether the model can detect "suspicious" journal entries auditors fail to do so. A further focal point of research could be coding and implementing the model in a database environment.

REFERENCES

1. AICPA (2002) Statement On Auditing Standards 99 (SAS 99): Consideration of Fraud in a Financial Statement Audit, *American Institute of Certified Public Accountant*.
2. Bolton, R. J. and Hand, D. J. (2002) Statistical fraud detection: A review, *Statistical Science*, 17, 3, 235–249.
3. Bonner, S., Palmrose, Z. and Young, S. (1998) Fraud type and auditor litigation: An analysis of SEC accounting and auditing enforcement releases, *Accounting Review*, 73, 4, 503–532.
4. Brockett, P. L., Xia, X. and Derrig, R. A. (1998) Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud, *The Journal of Risk and Insurance*, 65, 2, 245–274.
5. Chandola, V., Banerjee, A. and Kumar, V. (2009) Anomaly detection: A survey, *ACM Computing Surveys*, 41, 3, 1–58.
6. Debreceny, R. S. and Gray, G. L. (2010) Data mining journal entries for fraud detection: An exploratory study, *International Journal of Accounting Information Systems*, 11, 3, 157–181.
7. Grabski, S. (2010) Discussion of "Data mining journal entries for fraud detection: An exploratory study", *International Journal of Accounting Information Systems*, 11, 3, 182–185.

8. Hanley, J. A. and McNeil, B. J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143, 1, 29–36.

9. Hogan, C. E., Rezaee, Z., Riley, R. A. and Velury, U. K. (2008) Financial Statement Fraud: Insights from the Academic Literature, *Auditing: A journal of Practice and Theory*, 27, 2, 231–252.

10. Hsu, S., Lin, C. and Yang, Y. (2008) Integrating Neural Networks for Risk-Adjustment Models, *Journal of Risk and Insurance*, 75, 3, 617–642.

11. Juszczak, P., Adams, N. M., Hand, D. J., Whitrow, C. and Weston, D. J. (2008) Off-the-peg and bespoke classifiers for fraud detection, *Computational Statistics and Data Analysis*, 52, 9, 4521–4532.

12. Kaski, S. and Lagus, K. (1996) Comparing Self-Organizing Maps, in *Proceedings of the 1996 International Conference on Artificial Neural Networks*, vol. 1112 of *Lecture Notes in Computer Science*, Springer-Verlag, Bochum, Germany, pp. 809–814.

13. Kohonen, T. (1997) *Self-Organizing Maps*, Springer Series in Information Sciences,Volume 30, 2nd edn., Springer-Verlag, Heidelberg, Germany.

14. Kohonen, T. (1999) Comparison of SOM point densities based on different criteria, *Neural Computation*, 11, 8, 2081–2095.

15. Labib, K. and Vemuri, R. (2002) NSOM: A Real-Time Network-Based Intrusion Detection System Using Self-Organizing Maps, *Network Security*.

16. Palmrose, Z. (1991) An Analysis of Auditor Litigation Disclosures., *Auditing: A Journal of Practice and Theory*, 10, Supplement, 54–71.

17. Public Company Accounting Oversight Board (PCAOB) (2007) Auditing Standard No. 5: An Audit of Internal Control Over Financial Reporting That Is Integrated with An Audit of Financial Statements.

18. Public Company Accounting Oversight Board (PCAOB) (2004) Standing Advisory Group Meeting: Financial Fraud, Available at: http://pcaobus.org/News/Events/Documents/09082004_SAGMeeting/Fraud.pdf.

19. Public Company Accounting Oversight Board (PCAOB) (2002) AU Section 316: Consideration of Fraud in a Financial Statement Audit.

20. Ramadas, M., Ostermann, S. and Tjaden, B. (2003) Detecting anomalous network traffic with self-organizing maps, in *Recent Advances in Intrusion Detection*, pp. 36–54.

21. Rhodes, B. C., Mahaffey, J. A. and Cannady, J. D. (2000) Multiple self-organizing maps for intrusion detection, in *Proceedings of the 23rd National Information Systems Security Conference*, pp. 16–19.

22. The Committee of Sponsoring Organizations of the Treadway Commission (COSO)(2010) Fraudulent Financial Reporting 1998-2007: An Analysis of U.S. Public Companies,
Available at: http://www.coso.org/documents/COSOFRAUDSTUDY2010_001.pdf

23. U.S. Congress (2002) Sarbanes-Oxley Act of 2002, H.R.3763.

24. Vesanto, J., Himberg, J., Alhoniemi, E. and Parhankangas, J. (2000) SOM Toolbox for Matlab 5, *Tech. Rep. A57*, SOM Toolbox Team, Helsinki University of Technology, Espoo, Finland. Available at: http://www.cis.hut.fi/somtoolbox/.

25. Ypma, A. and Duin, R. P. W. (1997) Novelty detection using self-organizing maps, in *Proceedings of International Conference on Neural Information Processing. ICONIP '97*, vol. 2, Springer-Verlag, Dunedin, New Zealand, pp. 1322–1325.