A Neural Networks Tool to Enhance the Understanding of Fraudulent Financial Reporting

Wan-Ying Lin

Rua-Huan Tsaih

Shin-Ying Huang

For NTU's talk at April/2/2010

ABSTRACT

- **This research explores financial reporting fraud via an unsupervised Neural Network tool named Growing Hierarchical Self-Organizing Map (GHSOM) to enhance the understanding of fraudulent financial reporting (FFR).**
- **Differences from the prior research**
  - ✓ **focuses on predicting the likelihood of financial fraud, financial distress or bankruptcy**
  - ✓ **less emphasis has been placed on exploring financial reporting fraud itself, and FFR techniques and knowledge**
- **An application is conducted and its results show that the proposed two-stage approach is helpful in enhancing the understanding of FFR**

Fraudulent Financial Reporting

- **Fraudulent Financial Reporting (FFR) involves the intentional misstatement or omission of material information from an organization's financial reports.**
- **FFR can lead not only to significant risks for stockholders and creditors, but also financial crises for the capital market. FFR, although with the lowest frequency, casts a severe financial impact.**
- **According to the ACFE (2008), financial misstatements are the most costly form of occupational fraud, with median losses of $2 million per scheme.**

Fraudulent Financial Reporting Research on the nature of FFR

- Most prior FFR-related research focused on the nature or the prediction of FFR.
- The nature-related FFR research often uses the case study approach and provides a descriptive analysis of the characteristics of FFR and techniques commonly used.
- For example, the Committee of Sponsoring Organizations (COSO) and the Association of Certified Fraud Examiners (ACFE) regularly publish their own analysis on fraudulent financial reporting of U.S. companies.
- Based on the FFR samples, COSO examines and summarizes certain key company and management characteristics. ACFE analyzes the nature of occupational fraud schemes and provides suggestions to create adequate internal control mechanisms.

Table 1: Research methodology and findings in nature-related FFR studies.

Research	Methodology	Findings
Beasley et al. (1999)	• Case study • Descriptive statistics	• Nature of companies involved – Companies committing financial statement fraud were relatively small. – Companies committing the fraud were inclined to experience net losses or close to break-even positions in periods before the fraud. • Nature of the control environment – Top senior executives were frequently involved. – Most audit committees only met about once a year or the company had no audit committee. • Nature of the frauds – Cumulative amounts of fraud were relatively large in light of the relatively small sizes of the companies involved. – Most frauds were not isolated to a single fiscal period. – Typical financial statement fraud techniques involved the overstatement of revenues and assets. • Consequences for the company and individuals involved – Severe consequences awaited companies committing fraud. – Consequences associated with financial statement fraud were severe for individuals allegedly involved.

ACFE (2008)	• Case study • Descriptive statistics	• Occupational fraud schemes tend to be extremely costly. The median loss was $175,000. More than one-quarter of the frauds involved losses of at least $1 million. • Occupational fraud schemes frequently continue for years, two years in typical, before they are detected. • There are 11 distinct categories of occupational fraud. Financial statement fraud was the most costly category with a median loss of $2 million for the cases examined. • The industries most commonly victimized by fraud in our study were banking and financial services (15% of cases), government (12%) and healthcare (8%). • Fraud perpetrators often display behavioral traits that serve as indicators of possible illegal behavior. In financial statement fraud cases, which tend to be the most costly, excessive organizational pressure to perform was a particularly strong warning sign.

Fraudulent Financial Reporting Research on the prediction of FFR

- **Another type of FFR research often uses the empirical approach to archival data and identifies significant variables that help predict the occurrence of fraudulent reporting.**
- **This line of research also inputs these significant variables into the fraud prediction model. Such research emphasizes the predictability of the model used. For example, logistic regression and neural network techniques are used in this line of research.**
- **The matched-sample design is typical for traditional FFR empirical studies. That is, a set of samples with fraudulent financial statements is matched with a set of samples without any allegations of fraudulent reporting.**

Fraudulent Financial Reporting Research on the prediction of FFR

- **Table 2 summarizes the research methodology and findings of FFR empirical studies most relevant to our study.**
- **The research methodology has shown a trend with an emphasis on the classification mechanization which is used as the decision support information for future risk identification.**
- **However, engagements relating to criminal matters typically arise in the aftermath of FFR and an assessment to criminal engagements requires the accumulation of FFR knowledge.**

Table 2: Research methodology and findings in FFR empirical studies.

Author	Methodology	Variable	Sample	Findings
Dechow et al. (1996)	Logistic regression	• 21 variables – Financial ratios – Other indicators: corporate governance, motivationn etc.	Matched-pairs design 92 firms subject to enforcement actions by the SEC	• To attract external financing at low cost was found an important motivation for earnings manipulation • Firms manipulating earnings are more likely to have: - insiders dominated boards, - Chief Executive Officer simultaneously serves as Chairman of the Board.
Persons (1995)	Stepwise logistic model	• 9 financial ratios • Z-score	Matched-pairs design	The study found four significant indicators: financial leverage, capital turnover, asset composition and firm size
Fanning and Cogger (1998)	Self-organizing artificial neural network	• 62 variables • Financial ratios • Other indicators: corporate governance, capital structure etc.	Matched-pairs design: 102 fraud samples and 102 non-fraud samples	• Neural network is more effective • Financial ratios such as debt to equity, ratios of accounts receivable to sales, trend variables etc are significant indicators.

Bell and Carcello (2000)	Logistic regression	46 fraud risk factors	77 fraud samples and 305 non-fraud samples	Logistic regression model outperformed auditors for fraud samples, but were equally performed for non-fraud samples.
Kirkos et al. (2007)	• Decision tree • Back-propagation neural network • Bayesian belief network	• 27 financial ratios • Z-score	Matched-pairs design: 38 fraud samples and 38 non-fraud samples	• Training dataset: neural network is the most accurate • Validation dataset: Bayesian belief network is the most accurate
Hoogs et al. (2007)	Genetic Algorithm	• 38 financial ratios • 9 qualitative indicators	51 fraud samples vs. 51 non-fraud samples	Integrated pattern had a wider coverage for suspected fraud companies while it remained lower false classification rate for non-fraud ones

Growing Hierarchical Self-Organizing Map (GHSOM)

- **GHSOM addresses the issue of fixed network architecture of Self-Organizing Map (SOM) through developing the multilayer hierarchical network structure, in which, as shown in Figure 1, each layer contains a number of SOMs.**

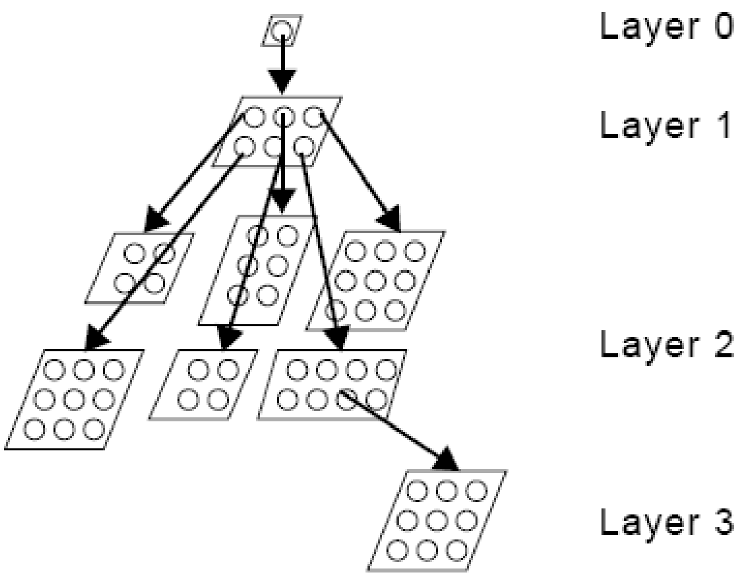


Figure 1**: GHSOM structure. (Dittenbach et al. 2000)**

Growing Hierarchical Self-Organizing Map (GHSOM)

- **The training process of GHSOM consists of the following four phases:**

1. **Initialize the layer 0: The layer 0 includes single node (mapping) whose weight vector is initialized as the expected value of all input data. Then the mean quantization error of layer 0 (MQE_0) is calculated. Hereafter, MQE of a mapping denotes the mean quantization error that sums up the deviation between the weight vector of the node and every input data mapped to the node.**
2. **Train every individual SOM: Within the training process of an individual SOM, the input data is imported one by one. The distances between the imported input data and the weight vector and all mapping are calculated. The mapping with the shortest distance is selected as the winner. Under the competitive learning principle, only the winner and its neighboring mappings are qualified to adjust their weight vectors. Repeat the competition and the training until the learning rate decreases to a certain value.**

Growing Hierarchical Self-Organizing Map (GHSOM)

3. **Grow horizontally each individual SOM:**

   Each individual SOM will grow until the mean value of the MQEs for all of the mappings on the SOM (MQE_m) is smaller than the MQE of the parent mapping (MQE_p) multiplied by τ_1. That is, the criterion for the stoppage of growth is stated in (1). If the stopping criterion is not satisfied, find the error mapping that owns the largest MQE and then, as shown in Figure 2, insert one row or one column of new nodes between the error mapping and its dissimilar neighbor.

$$MQE_m < \tau_1 \times MQE_p \qquad (1)$$

Growing Hierarchical Self-Organizing Map (GHSOM)

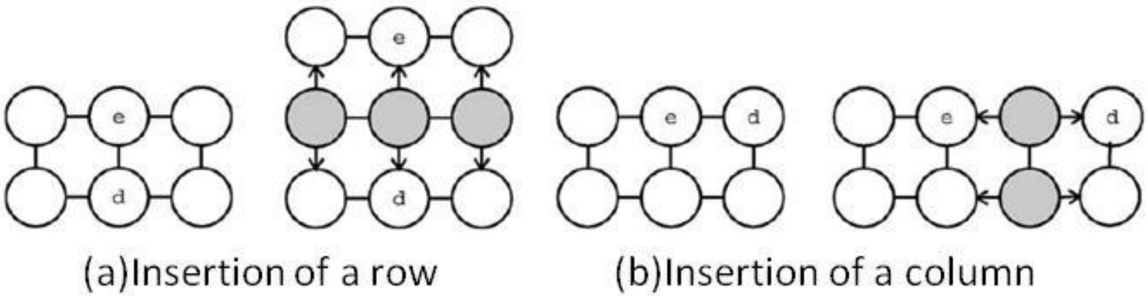


Figure 2: Horizontal growth of individual SOM. The notation e indicates the error mapping and d the dissimilar neighbor. (Dittenbach et al. 2000)

Growing Hierarchical Self-Organizing Map (GHSOM)

4. Expand or terminate the hierarchical structure: After the horizontal growth phase of individual SOM, MQE of every mapping (MQE_i) is compared with the value of MQE_0 multiplied by τ_2. The mapping with an MQE_i greater than $\tau_2 \times MQE_0$ will develop a next layer of SOM. In this way, the hierarchy grows until all of the leaf mappings satisfy the stopping criterion stated in (2). The leaf mapping means the mapping does not own a next layer of SOM.

$$MQE_i < \tau_2 \times MQE_0 \qquad (2)$$

The Proposed Approach

- A two-stage approach was developed as depicted in Figure 3

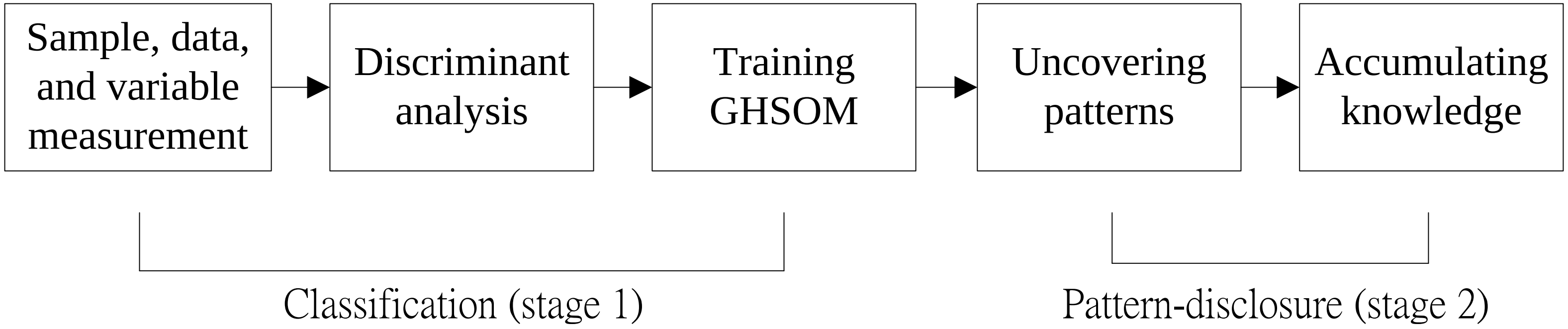


Figure 3: The two-stage approach for exploring the financial reporting fraud techniques via GHSOM.

Sample and Data

- **The following sources were used to identify the fraud sample:**
  - ✓ **indictments and sentences for major securities crimes issued by the Securities and Futures Bureau of the Financial Supervisory Commission**
  - ✓ **class action litigation cases initiated by Securities and Futures Investors Protection Center**
  - ✓ **the law and regulations retrieving system of the Judicial Yuan**

- **If a company's financial statement for a specific year is confirmed to be fraudulent by the indictments and sentences for major securities crimes issued by the Department of Justice, it is classified into our *fraud* observations, as to that company's financial statements free from fraud allegations they are classified into our *non-fraud* observations**

Sample and Data

- The matched-sample design is used to form a sample composite of 116 publicly traded companies, including 58 fraud and 58 non-fraud ones between the years of 1992 to 2006. That is, for each fraud firm, we match a non-fraud firm based on industry, total assets, and year.
- For each fraud company, we first identified the earliest year in which financial statement fraud was committed. The sample periods then cover two years before and two years after the year of the event. That is, five consecutive annual financial statements were used in our study.
- The final observations used in the study consisted of 580 firm-year observations, i.e., 580 annual financial statements were examined in the research.

Sample and Data

- For the 58 fraud firms identified, 113 annual financial statements were confirmed to have committed financial report fraud (henceforth fraud samples) and 177 annual financial statements were free of allegations of such fraud (henceforth non-fraud samples) ; on average, approximately two fraudulent financial statements (1.95 = 113/58) were included for each fraud firm.
- As to the 58 non-fraud firms, 290 non-fraud samples were included.

  →our final research samples were comprised of 113 fraud samples and 467 non-fraud samples

  →the ratio 113:467 was used as the benchmark for FFR tendency ratios, in which the FFR tendency ratio of a classification is defined as the ratio of its fraud to non-fraud samples

Sample and Data

- **It is worth noting that of the 113 fraud samples provided by the fraud firms, 78 fraudulent financial statements and 35 restated financial statements were *restated and re-announced* due to the request by government agency.**
- **The firms that provided the 35 restated statements were the ones that survived financial scandals and whose restated statements were in compliance with government regulations.**
- **The restated financial statements can be perceived as reflecting the firms' true financial positions that lead to the occurrences of the fraudulent financial reporting behavior.**
- **Such mixture of data mimics the environment of information in the real world which prevails with both true and false data.**

Variable measurement and discriminant analysis model

- **Based upon literature regarding fraudulent reporting, 25 explanatory variables are selected and incorporated into the discriminant analysis.**
- **Table 3 summarizes the definition and measurement of these variables.**
- **These are measurement proxies for attributes of *profitability*, *liquidity*, *operating ability*, *financial structure*, *cash flow ability*, *financial difficulty*, and *corporate governance* of a firm.**
- **These explanatory variables are collected from the Taiwan Economic Journal (TEJ) database.**

Table 3: Variable definition and measurement

Variable Definition	Literature	Measurement
Dependent variable:		
FRAUD	Persons (1995)	If a company's financial statements for specific years are confirmed to be fraudulent by the indictments and sentences for major securities crimes issued by the Department of Justice, the firm-year data are classified into fraud observations, and the variable *FRAUD* will be set to 1, 0 otherwise.

Independent variable

Profitability		
Gross profit margin (*GPM*)	Dechow et al. (2007)	$\frac{\text{Operating income - Operating costs}}{\text{Operating income}}$
Operating profit ratio (*OPR*)	Green (1997)	$\frac{\text{Operating income - Operating costs - Operating expenses}}{\text{Operating income}}$
Return on assets (*ROA*)	Persons (1995), Hoogs et al. (2007)	$\frac{\text{Net income} + \text{Interest expenses} \times (1 - \text{Tax rate})}{\text{Average total assets}}$
Growth rate of net sales (*GRONS*)	Stice (1991), Summers and Sweeney (1998), Dechow et al. (2007)	$(\frac{\text{Net sales}}{\text{Net sales in prior fiscal year}}) - 1$
Growth rate of net income (*GRONI)*	Summers and Sweeney (1998), Bell and Carcello (2000)	$(\frac{\text{Net sales}}{\text{Net income in prior fiscal year}}) - 1$

Liquidity		
Current ratio (CR)	Kirkos et al. (2007)	$\frac{Current\ assets}{Current\ liabilities}$
Quick ratio (QR)	Kirkos et al. (2007)	$\frac{Current\ assets - Inventories - Prepaid\ expenses}{Current\ liabilities}$
Operating ability		
Accounts receivable turnover (ART)	Green (1997)	$\frac{Net\ credit\ sales}{Average\ accounts\ receivable}$
Total asset turnover (TAT)	Persons (1995), Kirkos et al. (2007)	$\frac{\text{Net sales}}{\text{Total assets}}$
Growth rate of accounts receivable ($GROAR$)	Dechow et al. (2007)	$(\frac{\text{Accounts receivable}}{\text{Accounts receivable in prior fiscal year}}) - 1$
Growth rate of inventory ($GROI$)	Dechow et al. (2007)	$(\frac{\text{Inventory}}{\text{Inventory in prior fiscal year}}) - 1$

Growth rate of Accounts receivable to gross sales (*GRARTGS*)	Summers and Sweeney (1998)	$\frac{\text{Accounts receivable}_t}{\text{Gross sales}_t} - \frac{\text{Accounts receivable}_{t-1}}{\text{Gross sales}_{t-1}}$
Growth rate of Inventory to gross sales (GRITGS)	Summers and Sweeney (1998)	$\frac{\text{Inventory}_t}{\text{Gross sales}_t} - \frac{\text{Inventory}_{t-1}}{\text{Gross sales}_{t-1}}$
Accounts receivable to total assets (*ARTTA*)	Stice (1991), Persons (1995), Green (1997)	$\frac{\text{Accounts receivable}}{\text{Total assets}}$
Inventory to total assets (*ITTA*)	Stice (1991), Persons (1995)	$\frac{\text{Inventory}}{\text{Total assets}}$
Financial structure		
Debt ratio (*DR*)	Persons (1995), Kirkos et al. (2007)	$\frac{\text{Total liabilities}}{\text{Total assets}}$
Long-term funds to fixed assets (*LFTFA*)	Kirkos et al. (2007)	$\frac{\text{Equity + Longterm liabilities}}{\text{Fixed assets}}$

Cash flow ability		
Cash flow ratio (*CFR*)	Dechow et al. (2007)	$\frac{\text{Cash flows from operating activities}}{\text{Current liabilities}}$
Cash flow adequacy ratio (*CFAR*)	Dechow et al. (2007)	$\frac{\text{Five year sum of cash flows from operating activities}}{\text{(Five year sum of capital expenditures, inventory additions and cash dividends)}}$
Cash flow reinvestment ratio (*CFRR*)	Dechow et al. (2007)	$\frac{\text{Cash flows from operating activities - Cash dividends}}{\text{(Gross fixed assets + Long term investments + Other assets + Working capital)}}$
Financial difficulty		
Z-score	Altman (1968), Stice (1991), Summers and Sweeney (1998), Fanning and Cogger (1998)	$1.2\times(\frac{\text{Working capital}}{\text{Total assets}})+1.4\times(\frac{\text{Retained earnings}}{\text{Total assets}})+3.3\times(\frac{\text{Earnings before interest and taxes}}{\text{Total assets}})+0.6\times(\frac{\text{Market value of equity}}{\text{Book value of total debt}})+1.0\times TAT$

Corporate Governance

Stock Pledge ratio (*SPR*)#	Lee and Yeh (2004)	$\frac{\text{large shareholders' shareholdings in pledge}}{\text{large shareholders' shareholdings}}$
Sum of percentage of major shareholders' shareholdings (*SMLSR*)	Beasley et al. (1999)	Σ (Percentage of shareholdings >10%)
Deviation between CR and CFR (*DBCRCFR*)	La Porta et al. (1999), Lee and Yeh (2004)	Voting rights - Cash flow rights
Deviation between CBS and CFR (*DBCBSCFR*)	Lee and Yeh (2004), Yeh et al. (2001)	Percentage of board seats controlled - Cash flow rights

Variable measurement and discriminant analysis model

- **We first test the multi-collinearity issue between explanatory variables.**
- **The unreported results indicate that *GRITGS* should be excluded.**
- **As a result, 24 independent variables are incorporated in the Canonical Discriminant Analysis as shown in model (3).**

$$
\begin{aligned}
\mathbf{F}\mathit{RAUD} = {} & \alpha_1 \times \mathbf{GPM} + \alpha_2 \times \mathbf{OPR} + \alpha_3 \times \mathbf{ROA} + \alpha_4 \times \mathbf{GRONS} + \alpha_5 \times \mathbf{GRONI} + \alpha_6 \times \mathbf{CR} + \alpha_7 \times \mathbf{QR} + \alpha_8 \times \mathbf{ART} \\
& + \alpha_9 \times \mathbf{TAT} + \alpha_{10} \times \mathbf{GROAR} + \alpha_{11} \times \mathbf{GROI} + \alpha_{12} \times \mathbf{GRARTGS} + \alpha_{13} \times \mathbf{ARTTA} + \alpha_{14} \times \mathbf{ITTA} \\
& + \alpha_{15} \times \mathbf{DR} + \alpha_{16} \times \mathbf{LFTFA} + \alpha_{17} \times \mathbf{CFR} + \alpha_{18} \times \mathbf{CFAR} + \alpha_{19} \times \mathbf{CFRR} + \alpha_{20} \times \mathbf{Z\text{-}Score} + \alpha_{21} \times \mathbf{SPR} \\
& + \alpha_{22} \times \mathbf{SMLSR} + \alpha_{23} \times \mathbf{DBCRCFR} + \alpha_{24} \times \mathbf{DBCBSCFR}
\end{aligned}
$$

Empirical result of discriminant analysis

- **Table 4 shows the descriptive statistics of the variables in this study, including the mean, median, 25 percentiles and 75 percentiles. Column Z means one result of non-parametric test.**
- **Except *GRONS*, *GRITGS*, *DBCRCFR*, *DBCBSCFR*, other variables do have different statistical features between the fraud and non-fraud samples.**

Table 4: Descriptive Statistics of variables

	Fraud Sample (N=113)				Non-fraud Sample (N=467)				
Variable	**Mean**	**Median**	**25 Percentiles**	**75 Percentiles**	**Mean**	**Median**	**25 Percentiles**	**75 Percentiles**	**Z**
GPM	11.85	10.65	4.99	19.41	15.51	14.47	8.12	22.77	-3.19
OPR	-5.39	0.32	-7.26	6.92	-34.49	3.81	-0.24	8.60	-3.98
ROA	-13.45	-2.76	-23.48	5.29	3.40	4.19	0.39	7.97	-6.53
GRONS	8.30	7.84	-15.47	24.99	38.73	5.23	-7.77	19.89	-0.08
GRONI	47.23	-71.97	-636.91	24.49	-41.32	14.30	-44.89	80.07	-6.74
CR	109.83	104.68	60.98	141.48	190.94	150.01	110.02	210.00	-7.00
QR	57.79	45.54	21.84	77.09	110.36	75.73	38.09	124.66	-5.16
ART	7.10	4.62	3.16	7.34	8.91	5.36	3.75	8.94	-2.51
TAT	0.61	0.48	0.31	0.74	0.75	0.64	0.41	0.93	-3.69

GROAR	39.67	-5.73	-37.06	34.73	68.97	6.03	-15.15	33.86	-2.42
GROI	13.85	-1.02	-28.82	23.66	27.03	2.18	-14.80	31.14	-1.67
GRARTGS	-0.17	-1.04	-7.95	3.30	2.13	0.22	-2.75	3.11	-2.46
GRITGS	24.91	-0.34	-5.40	3.43	23.96	0.00	-3.37	4.80	-1.11
ARTTA	12.02	10.11	4.79	18.37	13.70	10.84	5.05	20.33	-1.33
ITTA	16.72	11.36	5.96	19.49	19.94	13.57	5.82	24.67	-1.74
DR	64.02	60.23	48.10	71.40	48.17	45.03	33.67	56.75	-7.59
LFTFA	452.26	165.79	95.29	399.96	482.48	225.20	146.73	427.05	-3.48
CFR	-14.91	-6.88	-21.21	6.54	13.41	8.12	-5.96	29.70	-6.26
CFAR	-18.56	-6.54	-27.97	8.65	9.36	14.52	-17.16	54.56	-5.53
CFRR	-46.73	-2.69	-14.70	3.74	0.37	2.03	-4.17	7.56	-4.59
SPR	37.44	33.44	1.83	63.26	19.32	3.58	0.00	32.49	-5.67
SMLSR	13.97	11.98	3.72	20.38	10.83	7.89	0.09	16.96	-3.16
DBCRCFR	3.47	0.47	0.00	2.76	3.62	0.56	0.00	4.09	-0.66
DBCBSCFR	46.00	45.58	22.87	67.41	44.26	43.68	26.99	63.69	-0.59
Z-Score	31.45	79.60	-91.69	166.17	198.67	194.70	120.89	270.95	-8.68

Empirical result of discriminant analysis

- **Table 5 shows the empirical results of the discriminant analysis and shows that the Wilks' Λ value equals 0.766 and x^2 equals 151.095 (both significant at p-value < 0.01), which indicates that the discriminant model employed has adequate explanatory power.**
- **Table 5 indicates that eight variables, *ROA*, *CR*, *QR*, *DR*, *CFR*, *CFAR*, *Z-Score* and *SPR*, have statistically significant effects.**
- **As shown in Table 3, these eight variables proxy a company's attributes from the aspects of profitability (*ROA*), liquidity (*CR*, *QR*), financial structure (*DR*), cash flow ability (*CFR*, *CFAR*), financial difficulty (*Z-Score*), and corporate governance (*SPR*).**
- **These eight chosen variables were collected for our sample firms and used as the training data for the GHSOM.**

Table 5: Empirical results of discriminant analysis.

Variable	Coefficient	F-value	Significance
GPM	0.14	3.51	0.061
OPR	-0.03	0.16	0.688
ROA	0.77	105.82	0.000***
GRONS	0.06	0.63	0.427
GRONI	-0.02	0.05	0.822
CR	0.34	20.59	0.000***
QR	0.28	13.42	0.000***
ART	0.09	1.58	0.210
TAT	0.19	6.38	0.012
GROAR	0.03	0.12	0.731
GROI	0.07	0.90	0.344
GRARTGS	0.00	0.00	0.997
ARTTA	0.11	2.25	0.134

ITTA	0.12	2.37	0.125
DR	-0.42	30.46	0.000***
LFTFA	0.02	0.09	0.764
CFR	0.33	19.21	0.000***
CFAR	0.24	9.89	0.002***
CFRR	0.19	6.41	0.012
SPR	-0.47	38.85	0.000***
SMLSR	-0.19	6.18	0.013
DBCRCFR	0.02	0.04	0.835
DBCBSCFR	-0.05	0.41	0.524
Z-score	0.64	72.74	0.000***
Wilks' Λ value	0.77	p-value	0.000
χ^2	151.10	p-value	0.000

Training GHSOM

- **In order to reach the goal of obtaining the multi-layer hierarchy feature and the prevention of overly clustering fraud samples, we set up the following predefined selection criteria to pick a suitable GHSOM:**

1. **There is more than one layer of SOM in the GHSOM.**
2. **Each individual leaf mapping should contain data from at least two sample firms.**
3. **Fraud or non-fraud samples of each mapping should not be overly clustered into anyone of the child mappings.**

Training GHSOM

- Table 6 shows 13 candidate GHSOM configurations conducted under different τ_1 and τ_2 setting.
- As shown in Table 6, when the depth value is 0.01, we find that a small breadth value results in a flat structure and that the number of mappings in each layer and the total number of leaf mappings converge when the breadth value is greater than 0.7.
- Then we try to increase the depth value under the breadth values 0.5 and 0.7 and find that the test No. 12 with three layers and 41 leaf mappings fits the predefined selection criteria.

Table 6: Thirteen GHSOM configurations.

No.	Parameter		Total of Layers	Number of Mappings				Total Number of Leaf Mappings
	Breadth	Depth		Layer1	Layer2	Layer3	Layer4	
1	0.1	0.01	1	144				144
2	0.2	0.01	1	63				63
3	0.3	0.01	2	21	222			231
4	0.4	0.01	2	9	125			125
5	0.5	0.01	3	6	59	4		62
6	0.6	0.01	3	4	25	85		95
7	0.7	0.01	4	4	16	54	6	63
8	0.8	0.01	4	4	16	48	4	55
9	0.9	0.01	4	4	16	48	4	55
10	1.0	0.01	4	4	16	48	4	55
11	0.5	0.02	2	6	59			59
12*	0.7	0.02	3	4	16	32		41
13	0.7	0.03	3	4	16	10		24

*: chosen GHSOM tree

The GHSOM tree

- Figure 4 shows the sample distribution of the obtained GHSOM, in which leaf mappings are marked in taint.
- In each mapping, there is a name given according to its layer number and its node order in the same SOM as well as its parent's name.
- For example, the mapping "L1m2-L2m1" means that it is developed from the second mapping of layer 1 (the first layer) and it is the first child mapping.
- In each mapping, the numbers within the parenthesis indicate the number of fraudulent financial statements, restated financial statements, and non-fraud financial statements, respectively and in that order.

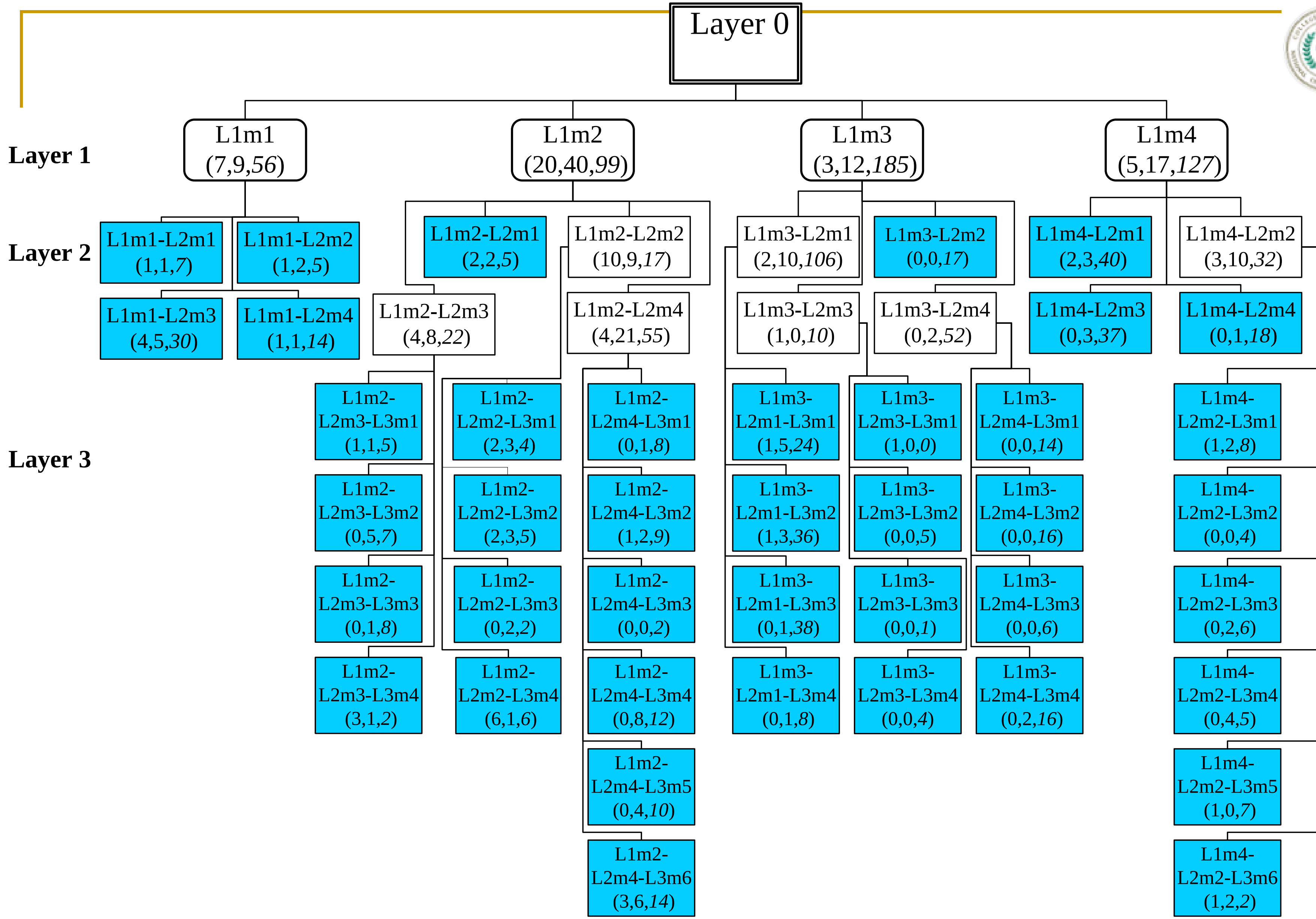


Figure 4: The sample distribution of the obtained GHSOM, in which leaf mappings are marked in taint. In each mapping, the numbers within the parenthesis indicate the number of fraudulent financial statements, restated financial statements, and non-fraud financial statements, respectively and in that order. The number of non-fraud financial statements is in italic.

The analysis of GHSOM Tree

- **Table 7 shows the FFR tendency ratio of each of the L1m1, L1m2, L1m3, and L1m4 mappings in layer 1**
- **In layer 1, the mapping L1m2 has the highest FFR tendency ratio and clusters more than half of the fraud samples; meanwhile, the mapping L1m3 has the lowest FFR tendency ratio with very few fraud samples.**

Table 7: FFR tendency ratio of each mapping in layer 1.

Layer1	The number of observation		FFR tendency ratio (%)
	Fraud	Non-fraud	
L1m1	16	56	**28.57**
L1m2	60	99	**60.61**
L1m3	15	185	**8.11**
L1m4	22	127	**17.32**

The analysis of GHSOM Tree

- **Table 8 lists the top five leaf mappings ranked by the FFR tendency ratio among 41 leaf mappings.**
- **The FFR tendency ratios of these five mappings are all greater than or equal to 100 %, and are named as high-risk mappings.**
- **For demonstration purposes, we took only the top two leaf mappings, L1m2-L2m3-L3m4 and L1m4-L2m2-L3m6, to illustrate the parts of uncovering patterns and accumulating knowledge.**

Table 8: The top five leaf mappings ranked by the FFR tendency ratio.

Leaf Mappings	The number of observation		FFR tendency ratio (%)
	Fraud	Non-fraud	
L1m2-L2m3-L3m4	4	2	200
L1m4-L2m2-L3m6	3	2	150
L1m2-L2m2-L3m1	5	4	125
L1m2-L2m2-L3m4	7	6	117
L1m2-L2m2-L3m2	5	5	100

The analysis of GHSOM Tree's leaves

- To verify whether each leaf mapping preserves its own salient features about the clustered sample, we tested the difference in financial features of L1m2-L2m3-L3m4 and L1m4-L2m2-L3m6 as shown in Figure 5 that displays the descriptive statistics of min, mean, and max.
- The nonparametric Wilcoxon signed-rank test results show that the mapping L1m4-L2m2-L3m6 has significantly higher *CR*s, *QR*s, *SPR*s and *Z-scores*, and significantly lower *DR*s, *CFR*s and *CFAR*s.
- Such significant differences provide confirmation to the statement.

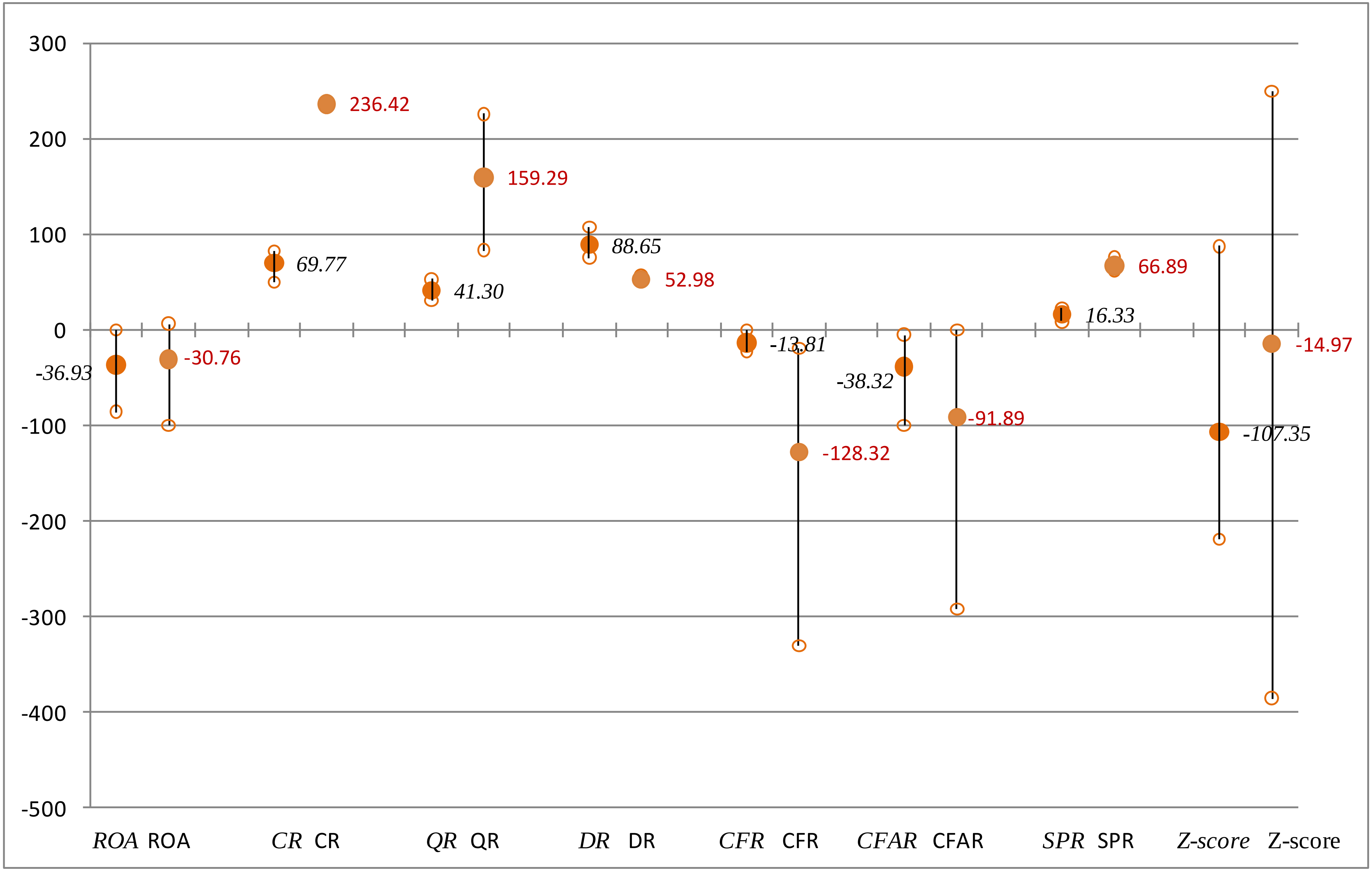


Figure 5: Descriptive statistics of L1m2-L2m3-L3m4 and L1m4-L2m2-L3m6. The corresponding numbers and variables to L1m2-L2m3-L3m4 are in italics.

Uncovering patterns of each interested leaf mapping

- For each high-risk leaf mapping of L1m2-L2m3-L3m4 and L1m4-L2m2-L3m6, we extracted the regularity of fraudulent techniques from the corresponding indictments and sentences for major securities crimes issued by the Department of Justice.
- Table 9 summarizes the fraudulent techniques commonly adopted by companies clustered in these two mappings.
- As shown in Table 9, two common fraudulent techniques found in L1m2-L2m3-L3m4 are: Recording fictitious revenues (FT1) and Misappropriation of assets (FT8).
- Specifically, some fraud samples were found using FT1 via creating fictitious transactions and defrauding export drawbacks from the Internal Revenue Service by reporting fictitious export sales.
- Some fraud samples used FT8 by processing the receipt and payment in advance.

Table 9: Common fraudulent techniques within L1m2-L2m3-L3m4 and L1m4-L2m2-L3m6.

Code_year	FT1	FT2	FT3	FT4	FT5	FT6	FT7	FT8	FT9	FT10
L-L-L4										
2328_1998 _R	○									○
3039_2004_R	○							○		
1221_2002_R	○							○		○
1601_1998_F								○		
1601_1999								○		
2005_2000	○							○		
L4-L-L6										
2407_2004_R	○			○	○		○	○		○
2017_1997_F				○				○		
8723_1998_F				○				○		
8295_1997										○

FT1: recording fictitious revenues;
FT2: recording revenues prematurely;
FT3: no description/overstated about revenues;
FT4: overstating existing assets;
FT5: recording fictitious assets or assets not owned;
FT6: capitalizing items that should be expensed;
FT7: understatement of expenses/liabilities;
FT8: misappropriation of assets;
FT9: inappropriate disclosure;
FT10: other miscellaneous techniques.

Uncovering patterns of each interested leaf mapping

- In regards to fraud samples in L1m2-L2m3-L3m6, two commonly used fraudulent techniques were: Overstating existing assets (FT4) and Misappropriation of assets (FT8).
- Specifically, some fraud samples were found to have been using the Overstating existing assets through purchasing intangible asset/long-term investment with high premiums.
- In contrast to L1m2-L2m3-L3m4, some fraud samples in L1m2-L2m3-L3m6 used FT8 through related party transactions and merger and acquisition activities to misappropriate cash.

What uncovered patterns said

- **Compared to the traditional fraudulent technique classification scheme, such a contrast demonstrates the advantage of our approach since our classification outcomes appear to be more delicate.**
- **Table 9 shows that the observed corporate behaviors (i.e., common fraudulent techniques extracted based upon the associated indictments) in different leaf mappings are distinctive even though these mappings are clustered based upon the corporate financial situations proxied by the input variables (i.e., the eight variables identified from discriminant analysis).**

Accumulating knowledge of each interesting leaf mapping

- **For each high-risk leaf mapping of L1m2-L2m3-L3m4 and L1m4-L2m2-L3m6, we investigated the causes of the observed common fraudulent techniques with the assistance of experts with domain knowledge.**
- **As shown in Table 10, the primary cause for utilizing both FT1 and FT8 fraudulent techniques in L1m2-L2m3-L3m4 may be due to the undesirable revenue situation of the firms.**
- **The primary causes for utilizing both FT4 and FT8 fraudulent techniques in L1m4-L2m2-L3m6 may be due to the bad cash flow condition of the firms and high financial pressure from management.**
- **Any pre-warning signal provided by these indicators can be used for future FFR identification.**

Table 10: Relevant indicators for L1m2-L2m3-L3m4 and L1m4-L2m2-L3m6.

Leaf mapping	Fraudulent techniques	Relevant indicators
L1m2-L2m3-L3m4	Recording fictitious revenues + Misappropriation of assets via related party transaction	Sales Growth ratio of net sales Net income Growth ratio of net income Account receivable turnover Inventory turnover Related party transaction (sale related) Net income/operating cash flow
L1m4-L2m2-L3m6	Overstating existing assets + Misappropriation of assets via manipulated cash flow	Cash flow ratio Cash flow adequacy ratio Investment cash flow Free cash flow Related party transaction (disposal of assets related) Cash flow reinvestment ratio Stock pledge ratio

The result of accumulating knowledge

- We used the sample 2328_1998 in L1m2-L2m3-L3m4 and the sample 2407_2004 in L1m2-L2m3-L3m6 as examples to show that the derived relevant indicators shown in Table 10 are reasonable.
- Figure 6 shows the pre and post restated financial indicators regarding the sample 2328_1998. It is confirmed that the sample 2328_1998 basically involved with the revenue-related manipulation which is consistent with the recording fictitious revenues (FT1), a common fraudulent technique in L1m2-L2m3-L3m4.
- Figure 7 shows financial indicators for the sample 2407_2004. It is confirmed that the sample 2407_2004 involved with the cash flow-related manipulation which is consistent with the overstating existing assets (FT4), a common fraudulent technique in L1m2-L2m3-L3m6.

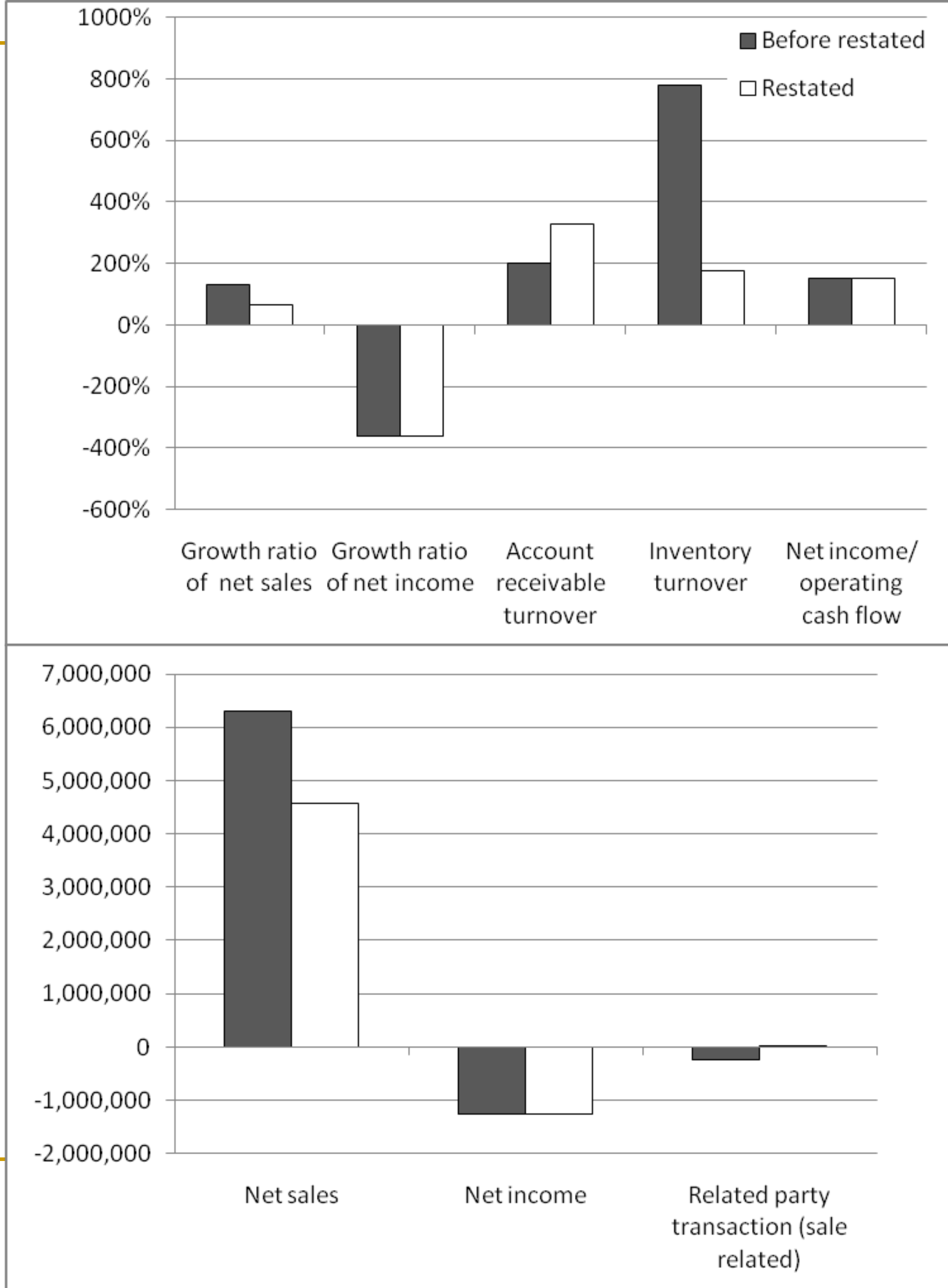
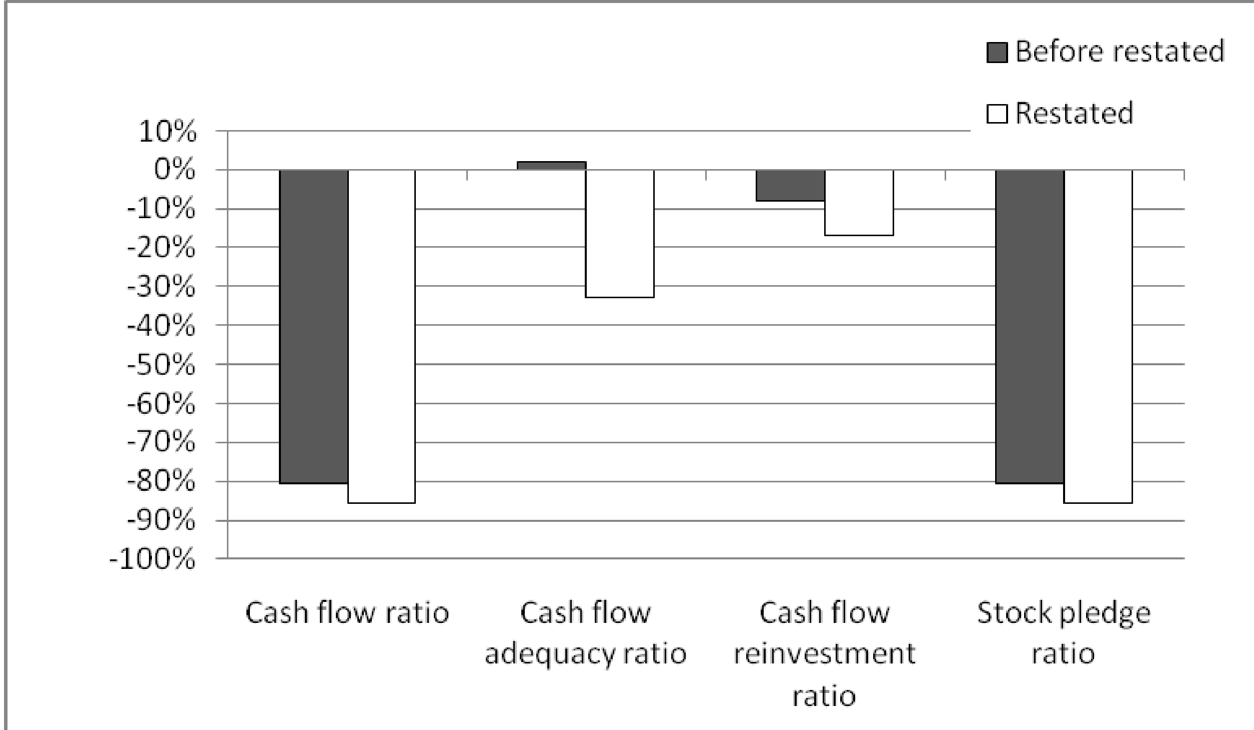


Figure 6:
Relevant indicators of sample 2328_1998 in L1m2-L2m3-L3m4



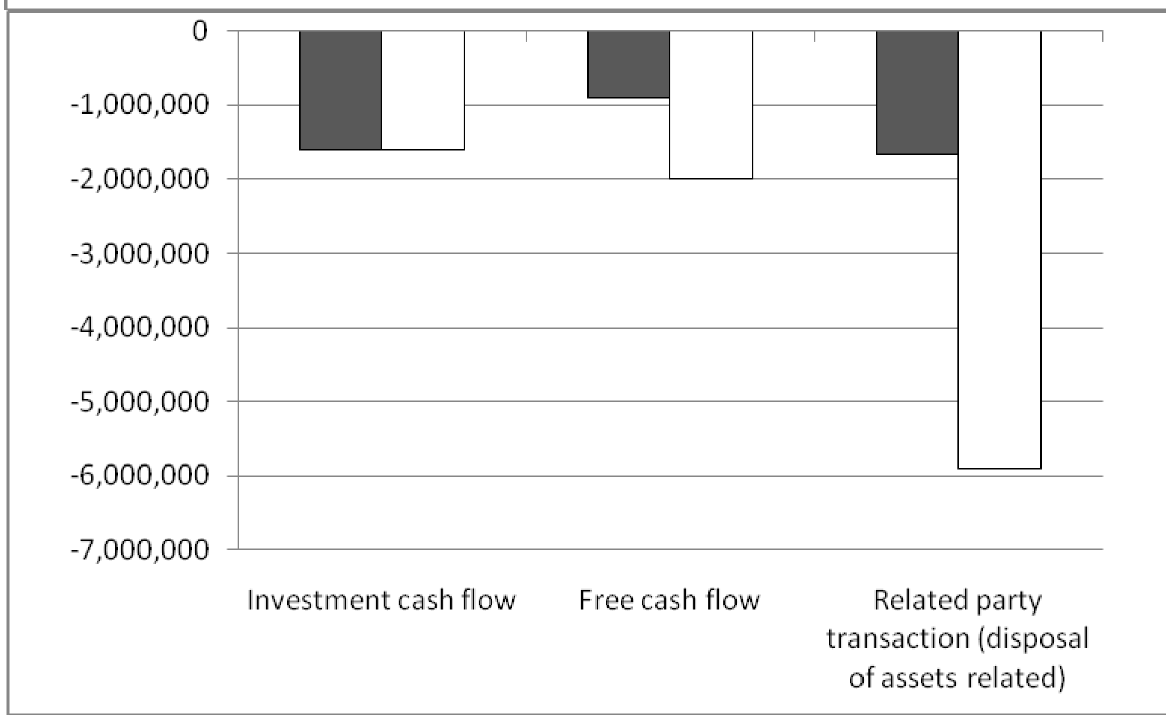


Figure 7: Relevant indicators of sample 2407_2004 in L1m4-L2m2-L3m6

Conclusion

- **This research proposes a systematic mechanism that includes (unsupervised) classification and FFR pattern-disclosure procedures.**
- **We have shown that this mechanism is helpful in obtaining knowledge that can better interpret FFR behavior.**
- **Future works are suggested as follows:**

(1) to identify better classification procedures and better FFR pattern-disclosure procedures;

(2) to investigate the generality of our approach using data from other countries

(3) to test the prediction ability of each result derived from our approach, including the fraud/ non-fraud classification, common FFR techniques, and the risk indicators.

Thanks for your Attentions.

Q & A