International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Study on Identification of Financial Fraud in Information Disclosure of Listed Companies - Based on Administrative Penalty Data of the Secondary Industry from 2001 to 2018

Yang Caifen

Xidian University, Xi 'an 710126, China

Abstract: ***This article is based on securities and futures commission due to the disclosure of financial fraud in 2001-2018 was the first time the administrative punishment of the second industry of listed companies and ultimately selected 144 listed companies as the data samples, the samples for fraud and non fraud samples 72 each, according to the theory of CRIME five factors selected 24 indicators, and the significance test, the data correlation inspection and collinearity finally get financial indicators and non-financial indicators for a total of 12, recognition rate reached 79.9%, the resulting model has good recognition effect.***

Keywords: Secondary industry; Information disclosure; Financial fraud; Logistic regression

1. Introduction

China's financial market took another big step forward on July 22, 2019 when the first batch of 25 kechuang board stocks were listed on the Shanghai stock exchange.Compared with the main board, the focus of the examination of kechuang board is no longer on the historical performance and whether the operation of the listed company is in compliance, but on whether the information disclosed by the listed company is true and complete, which shows that the country is paying more and more attention to the quality of the information disclosed by the listed company.The quality of information disclosure of listed companies is closely related to the development of the capital market and the interests of stakeholders.

In recent years, the financial fraud incidents of listed companies have only increased. Although the government regulatory authorities have long intensified their supervision, they are still unable to resist the means of fraud with numerous violators.Therefore; this paper studies the frauds of listed companies from the perspective of information disclosure, selects the representative secondary industry as the data sample, and strives to find the most suitable method for financial frauds of listed companies.

2. Relevant theories and Literature Review

Cerullo M J, Cerullo v. (1999), Krambia‐Kapardis, Maria. Yu yumiao, lv fan (2010), zhang zenglian, ya ya (2017), Yang guijun, zhou yaming (2019) and other scholars selected a series of financial indicators, non-financial indicators or both, constructed an identification model with Logistic regression, and obtained the best identification indicators to identify financial fraud of listed companies.No matter from the theoretical aspect, or from the empirical aspect, domestic and foreign scholars have deep research.

The current financial fraud theory includes four theories, namely, iceberg theory, triangle theory, GONE theory and risk factor theory. These four theories have a long history, with certain authority and practicality, and are widely spread and used for reference in the industry.In 2005, Zabihollah Rezaee put forward CRIEM five-factor theory, which was originally derived from five personality traits to analyze people's personality and needs. Zabihollah Rezaee called it CRIEM five-factor theory of crime.Guan yangwei and zhu weidong (2014) applied this method and selected 22 indicators ST company to conduct fraud behavior research, and the model recognition rate was up to 81.7%.

Zabihollah Rezaee's CRIEM five-factor theory mainly describes five aspects of financial fraud of listed companies."C" stands for forger. This shows that most financial frauds of listed companies are a series of Numbers games with the acquiescence of the management."R" represents the secret, that is, the means of financial fraud of listed companies. The common means of fraud of listed companies are nothing more than tampering with accounting elements such as assets, income and profit. "I" stands for incentive, mainly considering the motivation of fraud of listed companies. "E" stands for supervision, including internal and external aspects, such as external audit, internal audit, etc. "M" stands for the result, which is the result of the financial fraud of the listed company.

The uniqueness of this paper lies in :(1) the identification of financial fraud of listed companies from the perspective of information disclosure; (2) select the secondary industry data to subdivide the industry more targeted; (3) make an in-depth study on the internal and external environment and financial situation of listed companies based on the five-factor theory of CRIEM.

3. Selection of Samples and Variables

(I) Sample selection and data sources

This article chooses the securities and futures commission in 2001-2018 due to a violation of the "securities law" the sixty-third about "issuers, listed companies disclose information in accordance with the law, must be truthful, accurate and complete, and may not engage in any false records, misleading statements or material omissions" the provisions of the administrative punishment was first 72 listed companies, considering the missing data more ST shares and the particularity of financial companies, taking out ST shares, the listed company financial;The data of the earliest fraud year were selected for the listed companies with continuous fraud.

Data selection criteria of non-fraudulent listed companies are as follows :(1) in the same industry as fraudulent companies; (2) in the same year as the fraudulent company; (3) similar comprehensive indicators; (4) the paired samples have never been subjected to administrative punishment by the CSRC and other regulatory authorities.

After screening, a total of 144 listed companies of 72 fraudulent listed companies and 72 non-fraudulent listed companies were selected as the samples of this paper. The data were obtained from the database of China Securities Regulatory Commission and wind, and some of the data were obtained from flush flush, juchao information network, etc.

(II) Variable selection

According to the five-factor theory mentioned above, C forger, R secret, I incentive, M regulatory mechanism and E final result are selected respectively, and the variables shown in table 1 are selected in combination with the characteristics of financial fraud in information disclosure of the secondary industry.

Table 1: Variable Definitions

CRIME classification	source	code	indicators	A formula to calculate
C	Management characteristics	The X1	Board size	Board size = number of directors
		X2	The situation of the board of directors	The value of chairman concurrently general manager is 1, otherwise it is 0
		The X3	Executives at scale	Executive size = number of executives
R	External control	X4	The audit opinion	If there are no reservations, the value of the audit result is 1; Otherwise 0
		X5	If the big four	The accounting firm of this company is 4 big big 1, otherwise value is 0.
		X6	Property rights	Property right property right
I	Operation ability	X7	Inventory turnover	Inventory turnover = cost of goods sold ×2/ (beginning inventory + ending inventory)
		By 8	Accounts receivable turnover	Accounts receivable turnover = net credit sales income /[(beginning accounts receivable + ending accounts receivable) /2]
		X9	Total asset turnover	Total assets turnover = sales revenue /[(total assets at the beginning + total assets at the end) /2]
	The capital structure	X10	Asset-liability ratio	Asset-liability ratio = total liabilities/total assets
		X11	Long-term debt-to-capital ratio	Long-term gearing ratio =[non-current liabilities/(non-current liabilities + shareholders' equity)]
		X12	The rights and interests multiplier	Equity multiplier =1+ equity ratio
	profitability	X13	Return on equity	Return on equity = after-tax profit /[(beginning equity + ending equity) /2]
		X14	Net interest on total assets	Net interest rate on total assets = net profit /[(total assets at the beginning + total assets at the end) /2]
		X15	Cost utilization ratio	Cost utilization ratio = operating profit/total cost
	Debt paying ability	X16	Current ratio	Current ratio = current assets/current liabilities
		X17	Quick ratio	Quick ratio = (current assets - inventory)/current liabilities
		X18	Cash flow interest coverage	Cash flow interest coverage = net operating cash flow/interest expense
M	Internal audit	X19	Size of board of supervisors	Size of board of supervisors = number of board of supervisors
		X20	Dong ratio alone	Ratio of independent directors = number of independent directors/number of directors
E	Equity structure	X21	Proportion of state-owned shares	The proportion of state-owned shares = state-owned shares of the company/total shares of the company

		X22	Legal person share ratio	Legal person share ratio = company legal person share/company total share
		X23	Ratio of outstanding shares	Free float ratio = company free float/company total number of shares
		X24	Ownership concentration	Equity concentration = the sum of the shares of the top 10 shareholders

Identification model construction and analysis

(I) Descriptive Statistics

Table 2: Descriptive statistics of fraudulent and non-fraudulent companies

		Fraud company				Non-corrupt company			
indicators	N	The mean	The standard deviation	The minimum value	The maximum	The mean	The standard deviation	The minimum value	The maximum
The X1	72	8.500	1.768	5.000	13.000	9.569	1.992	0.000	167.280
X2	72	0.264	0.444	0.000	1.000	0.139	0.348	0.000	636.690
The X3	72	6.194	2.401	2.000	15.000	7.986	2.776	0.070	13.770
X4	72	0.472	0.503	0.000	1.000	0.903	0.298	0.040	4.660
X5	72	0.069	0.256	0.000	1.000	0.333	0.475	7.430	95.710
X6	72	0.278	0.451	0.000	1.000	0.569	0.499	2.530	64.260
X7	72	52.934	397.999	0.280	338.630	9.639	12.938	1.080	23.320
By 8	72	17.038	44.406	0.000	303.270	55.030	234.264	2.470	179.770
X9	72	0.538	0.427	0.040	2.470	0.923	0.704	3.140	95.520
X10	72	53.522	57.266	4.380	477.890	50.382	17.586	0.070	13.180
X11	72	10.858	36.704	191.880	207.430	15.732	17.341	0.070	13.180
X12	72	2.182	1.220	0.000	9.040	2.415	1.799	0.050	13.070
X13	72	5.663	12.679	32.580	31.280	12.552	13.449	119.140	169.780
X14	72	27.320	250.749	212.740	32.960	7.272	7.261	59.480	60.310
X15	72	0.857	71.208	513.710	171.930	11.429	12.515	122.940	47.090
X16	72	2.384	3.613	0.040	24.630	1.548	1.275	10.900	43.550
X17	72	2.003	3.257	0.030	21.110	1.160	1.173	88.320	94.030
X18	72	204.746	160.381	71.930	135.070	352.224	257.646	0.750	1.620
X19	72	3.625	0.941	3.000	7.000	3.972	1.150	0.100	15.080
X20	72	0.000	0.053	0.030	0.000	0.000	0.000	8.300	3.370
X21	72	0.018	0.066	0.000	0.357	0.079	0.174	0.370	3.110
X22	72	0.051	0.113	0.000	0.574	0.108	0.191	1.960	14.300
X23	72	1.000	0.000	1.000	1.000	1.000	0.000	0.000	0.790
X24	72	0.532	0.137	0.210	0.881	0.561	0.164	0.000	0.860

(II) Significance test

First, a single sample k-s test was conducted on 34 indicators to determine whether they were normally distributed. If independent sample T test was conducted, if not, an independent sample mann-whitney U test was conducted to determine whether the indicators of fraud samples and non-fraud samples were different.

Table 3: k-s test results of single sample

indicators	T value	Sig.	indicators	T value	Sig.
The X1	0.208	0.000	X13	0.131	0.000
X2	0.490	0.000	X14	0.471	0.000
The X3	0.131	0.000	X15	0.323	0.000
X4	0.437	0.000	X16	0.297	0.000
X5	0.490	0.000	X17	0.313	0.000
X6	0.380	0.000	X18	0.496	0.000
X7	0.458	0.000	X19	0.371	0.000
By 8	0.416	0.000	X20	0.213	0.000
X9	0.164	0.000	X21	0.438	0.000
X10	0.226	0.000	X22	0.339	0.000
X11	0.308	0.000	X23	0.012	0.000
X12	0.201	0.000	X24	0.048	0.200

As can be seen from table 4, except X24, the k-s test significance level of all the other indicators is significant at the level of 0.05. That is to say, we believe that all the indicators do not follow normal distribution, so the difference between the mann-whitney U test indicators of the independent samples is adopted.

Table 4: mann-whitney U test results of independent samples

indicators	N	The Mann – Whitney U	Wilcoxon W	Z	Sig.
X1	144	1885.000	4513.000	2.914	0.004
X2	144	2268.000	4896.000	1.864	0.062
X3	144	1554.000	4182.000	4.180	0.000
X4	144	1476.000	4104.000	5.554	0.000
X5	144	1908.000	4536.000	3.934	0.000
X6	144	1836.000	4464.000	3.529	0.000
X7	144	1876.000	4504.000	2.861	0.004
X8	144	1732.000	4360.000	3.436	0.001
X9	144	1453.500	4081.500	4.549	0.000
X10	144	2345.000	4973.000	0.987	0.324
X11	144	2300.000	4928.000	1.167	0.243
X12	144	2291.500	4919.500	1.201	0.230
X13	144	1838.500	4466.500	3.011	0.003
X14	144	1866.000	4494.000	2.901	0.004
X15	144	2334.000	4962.000	0.896	0.370
X16	144	2434.000	5062.000	0.631	0.528
X17	144	2290.000	4918.000	1.207	0.228
X18	144	1936.500	4564.500	2.619	0.009
X19	144	2197.500	4825.500	1.806	0.071
X20	144	2543.500	5171.500	0.202	0.840
X21	144	2178.500	4806.500	2.270	0.023
X22	144	2192.000	4820.000	1.810	0.070
X23	144	2592.000	5220.000	0.000	1.000
X24	144	2297.000	4925.000	1.179	0.239

The table 4 shows that at 0.05 significant level, the board of directors of the X1 size, X3, X4 audit executives scale opinion, four X5, X6 property right nature, X7 inventory turnover, by 8 accounts receivable turnover ratio, total asset turnover X9, X13 roe, X14 return on total assets, the multiple of interest safeguard X18 cash flow, the proportion of state-owned shares X21 these 12 indexes were significantly, so first select the 12 indicators as variables in the model.

(III) Correlation and collinearity test

Considering the collinearity of the 12 indicators, in order to ensure the efficiency of model identification, the correlation between the indicators must be removed before the establishment of the model. Therefore, spearman correlation analysis was firstly carried out on the above indicators, and the analysis results were shown in the following table:

Table 5: Spearman correlation analysis results

	X1	X3	X4	X5	X6	X7	X 8	X9	X13	X14	X18	X21
X1	1.000	286.**	359.**	0.133	310.**	0.096	0.004	0.090	0.114	0.093	0.024	258.**
X3	286.**	1.000	253.**	377.**	0.090	0.025	0.023	0.075	190.*	0.044	204.*	0.018
X4	359.**	253.**	1.000	264.**	335.**	0.126	0.103	0.132	0.109	0.130	0.095	0.102
X5	0.133	377.**	264.**	1.000	165.*	0.039	0.040	0.067	0.097	0.048	0.070	0.017
X6	310.**	0.090	335.**	165.*	1.000	0.069	0.148	0.030	0.021	0.075	0.025	277.**
X7	0.096	0.025	0.126	0.039	0.069	1.000	0.006	0.085	0.025	0.009	0.013	0.026
X 8	0.004	0.023	0.103	0.040	0.148	0.006	1.000	0.007	0.066	0.020	0.096	0.090
X9	0.090	0.075	0.132	0.067	0.030	0.085	0.007	1.000	0.137	- 862.**	0.092	0.026
X13	0.114	190.*	0.109	0.097	0.021	0.025	0.066	0.137	1.000	0.096	0.116	0.043
X14	0.093	0.044	0.130	0.048	0.075	0.009	0.020	- 862.**	0.096	1.000	0.021	0.031
X18	0.024	204.*	0.095	0.070	0.025	0.013	0.096	0.092	0.116	0.021	1.000	0.052
X21	258.**	0.018	0.102	0.017	277.**	0.026	0.090	0.026	0.043	0.031	0.052	1.000

* means the correlation is significant at the level of 0.05, while * * means the correlation is significant at the level of 0.01

It can be seen from the above table that the correlation between the above indicators is not strong according to the above spearman correlation analysis results. In order to further understand the collinearity of the indicators, multicollinearity test is carried out on the above 12 indicators.

Table 6: Multicollinearity test

The Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			How they	2 VIF is based
(Constant)	1.234	0.184		6.724	0.000		
X1	0.008	0.021	0.030	0.377	0.707	0.718	1.393
X3	0.029	0.015	0.158	1.976	0.050	0.727	1.376
X4	0.293	0.087	0.272	3.372	0.001	0.716	1.397
X5	0.198	0.096	0.159	2.065	0.041	0.783	1.277
X6	0.094	0.078	0.093	1.199	0.233	0.775	1.290
X7	0.000	0.000	0.010	0.139	0.890	0.979	1.021
X 8	0.000	0.000	0.003	0.037	0.971	0.890	1.124
X9	0.150	0.065	0.182	2.300	0.023	0.740	1.352
X13	0.005	0.003	0.122	1.600	0.112	0.799	1.251

X14	0.000	0.000	0.002	0.023	0.982	0.968	1.033
X18	0.000	0.000	0.029	0.402	0.688	0.901	1.110
X21	0.599	0.272	0.161	2.204	0.029	0.871	1.149

According to the results of the above multicollinearity test, it is found that VIF values are all less than 10, so it can be known that these 12 indicators have a low degree of collinearity and have little influence on the distortion of the model results, so they are suitable for Logistic regression.

(iv) Logistic regression results

According to the above analysis, the 12 indicators mentioned above are brought into the model. The dependent variable is whether fraud is committed. If so, it is 1, otherwise it is 0.In the regression process, the method of "input" is adopted. If the P value of the variable is greater than 0.1, it is removed from the model; if the P value of the variable is less than 0.05, it is entered into the model. Thus, the following Logistic regression results are obtained:

Table 7: Logistic regression results

Observed			Predicted		
			Whether fraud		Percentage with Correct
			0	1	
Step 1	Whether fraud	0	57	15	79.2
		1	14	58	80.6
	Overall Percentage				79.9
Log likelihood of -2			130.890[a]		
Cox and Snell R2			0.380		
Nagelkerke R2			0.506		

As can be seen from table 7, the evaluation index of the model is -2 Log likelihood, i.e. the logarithm likelihood value of -2 is 130.890, with a relatively large value, Cox & Snell R is 0.380, Nagelkerke R is 0.506, with a relatively large value, indicating that the model has a good fitting degree.22The results of the model showed that 15 of the non-fraud samples were misjudged as fraud companies, with a recognition rate of 79.2%. Among the fraud samples, 14 were misjudged as fraud companies, with an identification rate of 80.6%. The overall recognition rate of the model reached 79.9%.

Table 8: Variables in regression

		B	S.E.	Wald	df	Sig.	Exp (B)
Step 1[a]	X1	0.061	0.133	0.211	1	0.646	0.941
	X3	0.157	0.090	3.041	1	0.081	0.854
	X4	1.720	0.570	9.099	1	0.003	0.179
	X5	1.289	0.639	4.064	1	0.044	0.276
	X6	0.523	0.466	1.258	1	0.262	0.593
	X7	0.000	0.002	0.038	1	0.846	1.000
	X 8	0.000	0.002	0.010	1	0.919	1.000
	X9	0.863	0.481	3.221	1	0.073	0.422
	X13	0.027	0.024	1.301	1	0.254	0.974
	X14	0.011	0.029	0.134	1	0.715	0.989
	X18	0.000	0.000	0.393	1	0.531	1.000
	X21	4.284	2.229	3.692	1	0.055	0.014
	Constant	4.439	1.289	11.867	1	0.001	84.684

The general expression of the model is

Fraud $=\mathrm{Ln}\frac{p}{1-p}=\beta 0+\beta 1X1+\beta 2+\cdots+\beta n$, where, $P=\frac{\exp(\beta 0+\beta 1X1+\beta 2+\cdots+\beta n)}{1+\exp(\beta 0+\beta 1X1+\beta 2+\cdots+\beta n)}$

As can be seen from table 8, the final recognition model established in this paper is:

$\mathrm{Fraud}=\mathrm{Ln}\frac{p}{1-p}$=4.439-0.061*X1-0.157*X3-1.720*X4-1.289*X5-0.523*X6-0.863*X9-0.027*X13-0.011*X14-4.284*X21

4. Conclusion

From the results of identification model, it can be seen that the scale of X1 board of directors, the scale of X3 executives, X4 audit opinions, whether X5 is a big four, X6 property right nature, X9 total asset turnover, X13 return on net assets, X14 return on total assets and X21 state-owned shares are negatively correlated with financial fraud of listed companies.As the scale of the board of directors, executives size, total asset turnover, return on equity, return on total assets, the greater the proportion of state-owned shares, issued by the certified public accountants audit opinions issued by the auditing opinions without any reservations, external auditing department more is one of the big four, nature of the enterprise is state-owned enterprises, the smaller the possibility of financial fraud of listed companies.

References

[1] Cerullo M J, Cerullo v. Using neural networks to predict financial reporting fraud: Part 1[J]. Computer fraud & Security, 1999, 1999(5):14-17.
[2] Chen shaohua. Comprehensive analysis of the causes of financial fraud [J]. Economics economics,2005(01):74-76.
[3] Guan L, Kaminski K, Wetzel t. Can Investors Detect Fraud Using Financial Statements: An Exploratory Study[J]. Advances in Public Interest Accounting, 2007, 13:17-34.
[4] Edge M ,Sampaio P , Philpott O , et al. A Policy Distribution Service for Proactive Fraud Management over Financial Data Streams[C]// IEEE International Conference on Services Computing. IEEE, 2008.
[5] Wen yongjun, zhuwenjie. Empirical study on identification of fraud in financial reports of listed companies [J]. Financial communication,2009(02):36-37.
[6] Yu yumiao, lv fan. Identification of financial fraud risk -- a research perspective based on incremental information of financial indicators [J]. Economic review,2010(04):124-130.
[7] Xu H, Jiang l. Study on the Relation between the

Board Characteristics and Financial Fraud of Chinese Listed Companies[C]// International Conference on Management & Service Science. IEEE, 2011.
[8] Ye shulin. Study on financial fraud of listed companies and its countermeasures [J]. Chinese certified public accountant,2011(04):75-80.
[9] Ai hongmin, wangguichun, liuqian, Chen jie. A study on financial report fraud of listed real estate companies -- based on Logistic model empirical analysis [J]. Financial communication,2012(21):80-82.
[10] Guan yangwei, zhuweidong. Identification of financial fraud in ST company based on CRIME five factors [J]. Journal of finance and accounting,2014(16):12-16.
[11] Chengwei L, Yixiang C, Hasnain A K S, et al. Financial Fraud Detection Model: Based on Random Forest[J]. International Journal of Economics and Finance, 2015, 7(7).
[12] Kanapickien R, Grundien. The Model of Fraud Detection in Financial Statements by Means of Financial Statements [J]. Procedia-social and Behavioral Sciences, 2015, 213:321-327.
[13] Tang lingbing, luochangqing, peng pin. Influencing factors of off-balance sheet accounting fraud of listed companies based on Logistic model [J]. Systems engineering,2016,34(06):144-149.
[14] Zhang zenglian, gao ya. Construction of financial fraud identification model and empirical test [J]. Statistics and decision making,2017(09):172-175.
[15] GhafoorA ,Zainudin R , Mahdzan N S . Factors Eliciting Corporate Fraud in Emerging Markets: Case of Firms Subject to Enforcement Actions in Malaysia[J]. Journal of Business Ethics, 2018.
[16] Feng bingchun. Construction of financial fraud identification model based on data mining technology [J]. Financial communication,2019(05):93-97.
[17] Xu jinye, shi yanyi. The relationship between financial fraud and property right nature in capital market -- a test of financial data based on benford's law [J]. Financial communication,2019(02):37-41.
[18] Yang guijun, zhou ya-meng, sun ling-li, shi yu-hui. Logistic model based on Benford's law and its application in financial fraud identification [J]. Statistics and information BBS,2019(08):50-56.