Evaluation of Management Fraud Risk Based on Grey Correlation Model

—Evidence from China's ST Companies

Zexia Wang
School of Accounting, Hangzhou Dianzi University
HangZhou, China

Jing Xiang
School of Accounting, Hangzhou Dianzi University
HangZhou, China

Zhaoyun Ye
School of Accounting, Hangzhou Dianzi University
HangZhou, China
Email:390642635@qq.com

***Abstract*—This paper evaluates risk of management fraud based on the grey correlation model. In this paper, 442 samples are applied to build index system according to the fraud triangle theory; 68 ST companies with motivation of preventing delisting are chosen as training samples to calculate risk scores; Another 20 companies have been chosen to test the effect of the model recognition. Research results show that the optimal threshold to identify fraud and non-fraud company is 0.465, comprehensive recognition rate of testing samples is 85%.The grey correlation model has a good performance in evaluating fraud risk.**

Keywords -management fraud; grey correlation model; risk score; the optimal threshold

I. INTRODUCTION

Fraud cases from Enron, WorldCom, Xerox, Yinguangxia in china and soonhave caused heavy loss of investors and the capital market.Quantitative methods evaluating fraud risk have great significance for auditors and government regulators.

The existing researches on management fraud mainly include three aspects. They are analysis of fraud motivation, identification of fraud signals and methods of signal recognition. In the field of identifying fraud signals, theories such as financial factors theory, corporate governance theory and the fraud triangle theory are often used. Scholars at home and abroad try to use different models to verify the identification capability of the signals. The very significant models are single variable model, multiple linear analysis, logistic regression analysis, logistic regression analysis and neural network model.

However in terms of evaluating management fraud risk, researches are relatively small. The current methods of evaluating risk such as fuzzy comprehensive evaluation method, analytic hierarchy process, artificial neural network model and the grey correlation model are applied in evaluating financial and audit risk but seldom in fraud risk assessment. This is due to the following two primary reasons. One is that fraud is complex and variable, thus make the original information unreliable. The other is that general regression analysis request the original data has statistical characteristics, not suitable for all types of indexes. The grey correlation model can use the known information to determine the unknown information and modification of an individual index has little impacts on the system. In addition, the model has no strict requirement on sample size and data distribution, it is also suitable for small-sample and poor-information analysis. Considering all the advantages,this paper tries to apply the grey correlation model in the field of fraud risk evaluation.

II. GREY CORRELATION MODEL

Step 1: This paper assumes the 'n' index of the 'm' companies as the research object. $X_i = \{X_{i1}, X_{i2}, X_{i3}, \dots, X_{in}\}$ is the data of the ith(i=1, 2, 3⋯,m) company's nth index, that is to say X_{ij} stands for the data value of the ith company's jth index. $X_0^* = \{X_{01}, X_{02}, X_{03}, \cdots, X_{0n}\}$ is the reference sequence, and X_{0k}（k=1,2,⋯,n）stands for the kth index of the listed company as the reference. All the indexes compose a matrix containing（m+1）rows and n lines.

$$X_{(m+1)\times n} = \begin{bmatrix} X_{01} X_{02} \cdots X_{0n} \\ X_{11} X_{12} \cdots X_{1n} \\ X_{21} \quad X_{22} \cdots X_{2n} \\ \vdots \quad \vdots \quad \vdots \\ X_{i1} \quad X_{i2} \cdots X_{in} \\ \vdots \quad \vdots \quad \vdots \\ X_{m1} \quad X_{m2} \cdots X_{mn} \end{bmatrix}$$

Step2: Using dimensionless methods to make the indexes value from 0 to 1. Define the original index as X_{ij}, and the standardized index as X'_{ij}.

Step 3: Applying coefficient of variation method to calculate the weight of each index:

$$V_j = \frac{\sigma_j}{\overline{X}_j} (j = 1,2,\cdots,n) \dots\dots \dots\dots \dots\dots \dots\dots.. (1)$$

$$\omega_j = \frac{V_j}{\sum_{j=1}^{n} V_j} (j = 1,2,\cdots,n) \dots\dots\dots\dots\dots\dots\dots (2)$$

V_j stands for the variable coefficient of the jth index, σ_j stands for the standard deviation of the jth index, $\bar{X}_j$ stands for the arithmetic mean of the jth index. ω_j stands for the weight of the jth index, each weight of the "n" index compose a weight vector $\overrightarrow{W}, \overrightarrow{W} = (\omega_1, \omega_2, \cdots, \omega_j, \cdots, \omega_n)$.

Step 4: According to the grey correlation theory, define $X_0^* = \{X_{01}, X_{02}, X_{03}, \cdots, X_{0n}\}$ as the reference sequence, and $X_0^{**} = \{X'_{01}, X'_{02}, \cdots, X'_{0j}, \cdots, X'_{0n}\}$ as its dimensionless sequence. As mentioned in step 2, $X'_j = (X'_{1j}, X'_{2j}, \cdots, X'_{mj})$, $j = (1,2,\cdots,n)$ is the standardized value of the jth index. TheCorrelation coefficient formula of the ithcompany'sjth index and the reference sequence's jth index is:

$$R_{ij} = \partial(X'_{0j}, X'_{ij}) = \frac{\min_{k\in I}\min_{j\in J}|X'_{0j} - X'_{ij}| + \xi \max_{i\in I}\max_{j\in J}|X'_{0j} - X'_{ij}|}{|X'_{0j} - X'_{ij}| + \xi \max_{i\in I}\max_{j\in J}|X'_{0j} - X'_{ij}|} \ldots (3)$$

ξ is the distinguish coefficient, $\xi \in (0, 1)$,this paper sets ξ value to 0.5. $R_{m\times n}$ is the correlationcoefficientmatrix,containing m rows and n lines:

$$R_{m\times n} = \begin{bmatrix} R_{11} R_{12} \cdots R_{1n} \\ R_{21} \quad R_{22} \cdots R_{2n} \\ \vdots \quad \vdots \quad \vdots \\ R_{i1} \quad R_{i2} \cdots R_{in} \\ \vdots \quad \vdots \quad \vdots \\ R_{m1} \quad R_{m2} \cdots R_{mn} \end{bmatrix}$$

Step 5: Calculating the score of the ith company's management fraud risk, callitF_i:

$$F_i = (R_{i1}, R_{i2}, \cdots, R_{in}) \times \overrightarrow{W}^T \ldots\ldots\ldots\ldots\ldots (4)$$

III. DATA SOURCES AND VARIABLE SELECTION

Data in this paper mainly comes from public punishmentsof CSMAR, China securities regulatory commission (CSRC)website, Shanghai stock exchange website and the Shenzhen stock exchange website. This paper chooses 7 types of irregularities as the selection standard of fraud. They are fictitious profits, inflated assets, false statement, delayed disclosure, major omissions, illegal guarantee, and other.Research samples include fraud and non-fraud companies. Fraud company samplescontain 221 companies punished by CSRC from 2003 to2010. And the selection ofthe paired company (or call it non-fraud company) samples obeys to the Beasley principle:

- Fraud and paired companies belong to the same stock market;
- The two belong to the same industry;
- No fraud occurred in Paired company during fraud company' fraud period;
- At the end of the year before fraud period, the chosen paired company's assets is the most close to fraud company, and the difference is less than 30%.

According to CSRC's industry classification standard, the 221 fraud company samples' industry distribution is showed in table I. Data is progressed by SPSS17.

A. Index System

According to the fraud triangle theory, risk evaluation index system is composed by pressure, opportunities and excuses. We choose variable with reference to the risk factors listed by SAS NO.99 and research achievements of scholars both at home and abroad. The index system is tested by T test and Wilcoxon's sign rank test to test itscomprehensiveness. In general, data from the last year before fraud has better prediction effect. So this paperchooses variable of the last year before fraud as research object. The index system and itsdescriptive statistics results are showed in tableII.Table II tells that the 21 indexes are significantly different between fraud and non-fraud companies. In other words, the index system can be used as significant fraud warning signals to identify the fraud and non-fraud companies.

TABLEI. INDUSTRY DISTRIBUTION OF FRAUD COMPANY SAMPLES

Industry types	**A**	**B**	**C0**	**C1**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**D**	**E**	**F**	**G**	**H**	**J**	**K**	**L**	**M**	**total**
Company numbers	8	4	18	9	3	31	10	18	30	13	11	1	2	17	12	27	5	1	7	221

TABLEII. THE INDEX SYSTEM AND DESCRIPTIVE STATISTICAL ANALYSIS

Standard level	**Sub-standard**		**Index**	**Average of fraud samples**	**Average of paired samples**	**Average of samples**	**T test for paired samples**		**Wilcoxon's sign rank test**	
							t	***Sig.***	***Z***	***Asymp.Sig***
Pressures	***X_{11}Financial Stability***	X_{111}	Gross profit margin	0.208	0.253	-0.0447	-2.5793	0.011**	-2.123	0.034**
		X_{112}	Current asset turnover	1.1507	1.3656	-0.215	-2.0061	0.046**	-3.6128	0.000***
		X_{113}	Total assets growth rate	0.0861	0.1984	-0.1123	-2.5411	0.012**	-2.0045	0.045**
		X_{114}	Account receivable/Total assets	0.1455	0.1237	0.0218	2.3565	0.019**	-2.1936	0.028**

		X_{115}	Cash flows from operating activities / Total assets	0.0227	0.0512	-0.0284	-3.6075	0.000***	-3.4965	0.000***
	X_{12}Financing Pressures	X_{121}	Financial leverage	1.6583	1.3315	0.3268	1.2222	0.136	-0.0417	0.767
		X_{122}	Time interest earned ratio	21.3806	5.6267	15.754	1.4305	0.154	-5.1001	0.000***
	X_{13}Individual wage pressures of management	X_{131}	Proportion of management stock ownership	0.0024	0.0051	-0.0027	-1.4956	0.136	-0.3011	0.763
		X_{132}	The average of the top three executives' pay	182184	192867	-10683.	-0.481	0.631	-1.8891	0.059*
Opportunity	***X_{21}Corporate governance***	X_{211}	The first majority shareholderdirector	1.7524	2.219	-0.4667	-3.0912	0.002***	-3.4787	0.001***
		X_{212}	Outside director	0.4231	0.3866	0.0364	1.9625	0.051*	-2.2543	0.024**
		X_{213}	Number of outgoing directors	1.7463	1.0049	0.7415	3.4048	0.001***	-3.4166	0.001***
		X_{214}	Shareholding proportion of board of supervisors	0.0019	0.0003	0.0016	1.978	0.049**	-1.4615	0.144
		X_{215}	Number of outgoing executives	1.0977	0.7907	0.307	2.2302	0.027**	-2.2521	0.024**
		X_{216}	Be both chairman and CEO	0.1166	0.1794	-0.0628	-2.0349	0.043**	-1.7868	0.074*
	X_{22}Ownership structure	X_{221}	State –owned share proportion	0.2169	0.2857	-0.0687	-3.4336	0.001***	-2.3822	0.017**
		X_{2222}	H10Share concentration	0.1507	0.1778	-0.0271	-2.3143	0.022**	-2.8614	0.004***
	X_{23}Special Transaction	X_{231}	Related-party transaction / Total assets	4.7612	9.7202	-4.959	-0.4483	0.654	-2.1105	0.035**
		X_{232}	External guarantee/ net assets	29.5269	0.1181	29.4088	1.0178	0.310	-3.2864	0.001***
Excuse	***X_{31}Relationship with auditors***	X_{311}	Audit opinion type	0.2242	0.0762	0.148	4.5609	0.000***	-4.4525	0.000***
		X_{312}	Auditor switch	0.5142	0.6698	-0.1557	-3.0325	0.003***	-3.0738	0.002***

***, **,*represent significance levels at 1%, 5% and 10% accordingly

TABLEIII. INDUSTRY DISTRIBUTION OF FRAUD ST COMPANY SAMPLES

Industry types	A	B	C0	C4	C6	C7	C8	D	F	G	H	J	K	total
Company numbers	4	4	2	12	2	8	8	2	2	8	6	8	2	68

B. Selection of Training Samples and Testing Samples

This paper makes a further selection of the samples. Retained samples have strong motivation of avoiding ST. Getting rid of samples with incomplete data, this paper gets 88 samples (44 fraud companies and 44 paired companies) from 2004 to 2009. Among the 88 samples, 68 are chosen as training samples to build grey correlation model and evaluate fraud risk. The Industry types of the 68 chosen companies are showed in tableIII. The last 20 companies are used as testing samples to test the effect of model recognition.

IV. RISK SCORE AND THE OPTIMAL THRESHOLD

A. Calculating Risk score

According to the research idea showed before,this paper puts the 68 samples into the grey correlation model. Through the 5 steps described in part II, this paperobtains the weight vector $\overrightarrow{W}$and the 68 samples' risk score.

$$\overrightarrow{W}^{T} = \begin{matrix}(0.0239,0.0417,0.0163,0.0200,0.0097,0.0099,\\ 0.0261,0.1531,0.0590,0.0375,0.0185,0..0357,\\ 0.0934,0.0399,0.0689,0.0308,0.0195,0.0371,\\ 0.1896,0.0466,0.0225)^{T}\end{matrix}$$

Each sample's risk score falls within the range from 0.35 to 0.63. In order to further explore score differences between fraud samples and non-fraud samples, this paper divides the range which from 0.35 to 0.63 into 28 intervals and makes probability distribution statistics of each interval. Results are showed in tableIV.

In terms of the risk scores for fraud companies, range 0.46 -0.57 stands out in tableIV. 24 companies fall in this interval and cumulative probabilityachieves 71%. In addition,interval 0.48-0.49 is the most significant one, with 6 companiescounting for 17.647% of the total amount.Risk scores of non-fraud companies mainly fall in the range from 0.35 to 0.47, which frequency is 25 and cumulative probability is 73%. And interval 0.42-0.43, 0.43-0.44, 0.44-0.45 are the three most significant ranges, each with 5 companies counting for 14.706% of the total amount.

As risk scores of fraud and non-fraud companies have different focus ranges, this paper tries to find out theoptimal threshold to identify fraud and non-fraud companies.

B. Determing the Optimal Threshold

From a statistical point of view, when setting a point of division, each model has two kinds of mistakes including mistake I (false reject) and mistake II (false accept). In this paper, mistake I represents that fraud companies are misjudged as non-fraud companies; mistake II represents that non-fraud companies are misjudged as fraud companies. When determining the point of division, reduce one kind of the mistakes will increase the other. This paper defines "p" is the probability of the event that the company "i" is a fraud company. $\{p(i = fruad)\}$is the conditional probability of the event that company "i" is in fact a fraud company; $f_2\{p(i = non\ fruad)\}$is the conditional probability of the event that company "i" is in fact a non-fraud company. In the forecast model, $f_1\{p\}$ stands for the probability distribution of the event that fraud companies have been correctly identified; $f_2\{p\}$ is the probability distribution of the event that non-fraud companies are misjudged as fraud companies. The two probability distribution curves are as follows, see in figure I.

According to the condition of determining the division point: $f_1\{p\} = f_2\{p\}$, this paper tries to get the minimum value of the absolute value$|f_1\{p\} - f_2\{p\}|$. To enhance the effect of the model recognition,the cumulative probability distribution in table IV should be combined with the condition mentioned before. Results are showed in tableV.

Taking theminimum value of the absolute value$|f_1\{p\} - f_2\{p\}|$and comprehensive recognition rate into consideration, the optimal threshold should be 0.465.When a company's fraud risk score is higher than 0.465, the company has higher possibility of fraud and its fraud risk is also higher. On the contrary, when the score is less than 0.465, the company has lower fraud risk. And then this paper uses the 68 samples to observe the accuracy of the model when the optimal threshold is 0.465. The results show that 9 fraud companies in 34 are misjudged as non-fraud companies which means the probability of mistake I is 26.47% and the accuracy is 75.53%. 10 non-fraud companies in 34 are misjudged as fraud companies, telling the probability of mistake II is 27.94% and accuracy is 70.59%. The comprehensive recognition rate is 72.06%.Results are showed in table VI.

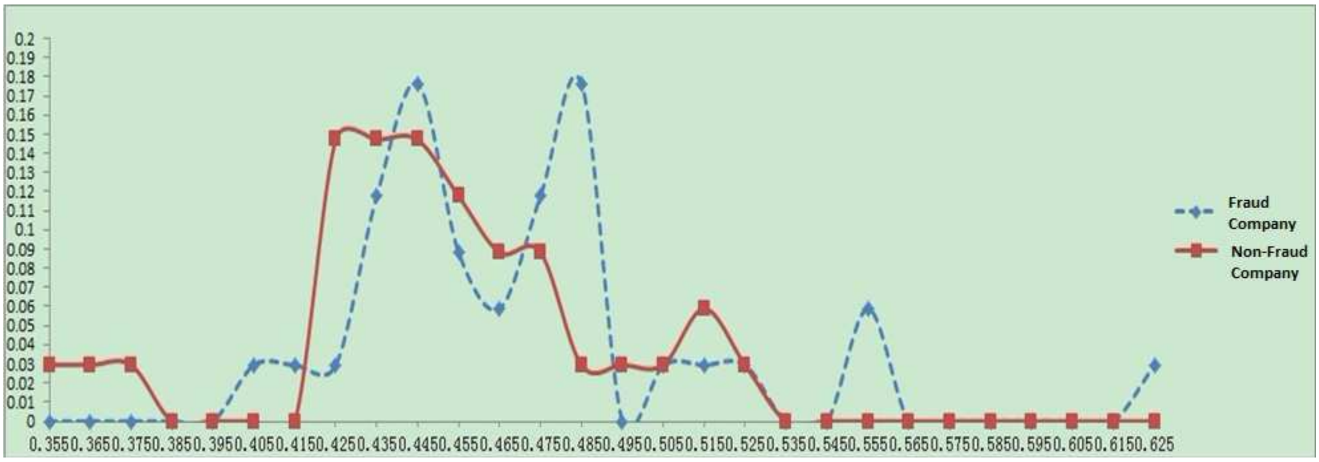

FIGURE I. PROBABILITY DISTRIBUTION CURVES OF FRAUD AND NON-FRAUD COMPANIES

TABLEIV. PROBABILITY DISTRIBUTION OF TRAINING SAMPLES' RISK SCORE

Interval	Mean	*Fraud Company*			*Non-Fraud Company*		
		frequency	probability distribution	cumulative probability distribution	frequency	probability distribution	cumulative probability distribution
0.35～0.36	0.355	0	0	0	1	0.02941	0.02941
0.36～0.37	0.365	0	0	0	1	0.02941	0.05882
0.37～0.38	0.375	0	0	0	1	0.02941	0.08824
0.38～0.39	0.385	0	0	0	0	0	0.08824
0.39～0.40	0.395	0	0	0	0	0	0.08824

0.4～0.41	0.405	1	0.02941	0.02941	0	0	0.08824
0.41～0.42	0.415	1	0.02941	0.05882	0	0	0.08824
0.42～0.43	0.425	1	0.02941	0.08824	5	0.14706	0.23529
0.43～0.44	0.435	1	0.02941	0.11765	5	0.14706	0.38235
0.44～0.45	0.445	2	0.05882	0.17647	5	0.14706	0.52941
0.45～0.46	0.455	3	0.08824	0.26471	4	0.11765	0.64706
0.46～0.47	0.465	2	0.05882	0.32353	3	0.08824	0.73529
0.47～0.48	0.475	4	0.11765	0.44118	3	0.08824	0.82353
0.48～0.49	0.485	6	0.17647	0.61765	1	0.02941	0.85294
0.49～0.50	0.495	4	0.11765	0.73529	1	0.02941	0.88235
0.50～0.51	0.505	1	0.02941	0.76471	1	0.02941	0.91176
0.51～0.52	0.515	2	0.05882	0.82353	2	0.05882	0.97059
0.52～0.53	0.525	1	0.02941	0.85294	1	0.02941	1
0.53～0.54	0.535	1	0.02941	0.88235	0	0	1
0.54～0.55	0.545	0	0	0.88235	0	0	1
0.55～0.56	0.555	2	0.05882	0.94118	0	0	1
0.56～0.57	0.565	1	0.02941	0.97059	0	0	1
0.57～0.58	0.575	0	0	0.97059	0	0	1
0.58～0.59	0.585	0	0	0.97059	0	0	1
0.59～0.60	0.595	0	0	0.97059	0	0	1
0.60～0.61	0.605	0	0	0.97059	0	0	1
0.61～0.62	0.615	0	0	0.97059	0	0	1
0.62～0.63	0.625	1	0.02941176	1	0	0	1
total		34	1	——	34	1	——

TABLEV. CONDITIONAL PROBABILITY DIFFERENCE AND MODEL RECOGNITION

Interval	**Mean**	$\lvert \boldsymbol{f_1}\{\mathbf{p}\} - \boldsymbol{f_2}\{\mathbf{p}\} \rvert$	**Fraud company recognition rate**	**Non-fraud company recognition rate**	**Comprehensive recognition rate**
0.40～0.41	0.405	0.0294	97.06%	8.82%	52.94%
0.41～0.42	0.415	0.0294	94.12%	8.82%	51.47%
0.42～0.43	0.425	0.1176	91.18%	23.53%	57.35%
0.43～0.44	0.435	0.1176	88.24%	38.24%	63.24%
0.44～0.45	0.445	0.0882	82.35%	52.94%	67.65%
0.45～0.46	0.455	0.0294	73.53%	64.71%	69.12%
0.46～0.47	0.465	0.0294	67.65%	73.53%	70.59%
0.47～0.48	0.475	0.0294	55.88%	82.35%	69.12%
0.48～0.49	0.485	0.1471	38.24%	85.29%	61.76%
0.49～0.50	0.495	0.0882	26.47%	88.24%	57.35%
0.50～0.51	0.505	0.0000	23.53%	91.18%	57.35%
0.51～0.52	0.515	0.0000	17.65%	97.06%	57.35%

TABLEVI. MODEL RECOGNITION RATE-TRANING SAMPLES

68 samples		**Accuracy**		
		Type		**Percentage**
		Fraud	***Non-fraud***	
Type	***Fraud***	25	9	73.53
	Non-fraud	10	24	70.59
Average		——	——	72.06

TABLEVII. MODEL RECOGNITION RATE-TESTING SAMPLES

20 samples		**Accuracy**		
		Type		**Percentage**
		Fraud	***Non-fraud***	
Type	***Fraud***	8	2	80
	Non-fraud	1	9	90
Average		——	——	85

V. TEST OF THE MODEL RECOGNITION EFFECT

20 companies from 2007 to 2010 are chosen as the new data in order to test the accuracy of the model. Among the 20 companies, there are 10 fraud ST companies and 10 pair companies. Results are showed in table VII. Table VIItells that 2 fraud companies in 10 are misjudged as non-fraud companies, which means the probability of mistake I is 20% and the accuracy is 80%; 1 non-fraud company in10 is misjudged as fraud company, which means the probability of mistake II is 10% and the accuracy is 90%. To sum up, the comprehensive recognition rate of the testing samples is 85%, better than the training samplesof 72.06%.

VI. CONCLUSION

This paper evaluates the fraud risk of ST companies with strong motivation of preventing delisting based on the grey correlation model. Research results show that company has higher probability of fraud when its fraud risk score is higher than 0.465(theoptimal threshold of the risk score is 0.465). And the 20 testing samples' results show that the model's comprehensive recognition rate is 85%, much higher than the training samples'. That is to say, the grey correlation model has a good performance in evaluation of the fraud risk.

ACKNOWLEDGMENT

This paper is supported by the National Natural Science Foundation of China (Grant No.71172098).

REFERENCE

[1] Beaver W H,"Financial ratios as predictorsof failure" [J]. Journal of Accounting Research, 1966(supplement), pp.71—111.

[2] Altman E I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy[J]. Journal of Finance, 1968(4), pp.589-609.

[3] Shinong Wu, Xianyi Lu.A study of models for predicting financial distress in china's listed companies[J].Econmic Research Journal, 2001(6), pp.46-56.

[4] Xiuhua Jiang, Qiang Ren, Zheng Sun, Research on models for predicting financial distress in china's listed companies[J].Forecasting, 2002(3), pp.56-61.

[5] Shouhua Zhou, Jihua Yang, Ping Wang, Predicting analysis of financial disress-F fraction model[J]. Accounting Research,1996(8), pp.8-11.

[6] Shu'e Yang, Li Huang, Fiancial crisi warning model based on BP Neural network[J].System Engineering-Theory & Practice, 2005(1), pp.12-18.

[7] Huijing Wang, Evaluation of audit risk based on dynamic fuzzy comprehensive evaluation model[J]. Accounting Research, 2011(9), pp.89-95.

[8] Ting Shen, Shuhui Wang, Application of anlytic hierarchy process in financial risk evaluation[J]. Econmic Resarch Guide, 2010(13), pp.81-82.

[9] Jiaxiang Zhu, Qingmei Tan, Xiangyuan Jing, Two phases grey fiancial alert of public corpotations based on degress[J].Soft Science, 2008(4), pp.43-48.