the principle components prediction model of accounting information distorted of listed companies

Mengyi Wang
Tianjin University of Finance and Economics

Accounting distortion has brought loss to the enterprises and country.As long as economic business exists, then the identification and management of accounting information distortion is always a eternal topic to discuss. Therefore, only by the most effective solution can we inhibit spread and development of the financial fraud and rectify the mistakes and existing errors in the economic information to a great extent,all of which enable the economic activities of enterprises reflect in a timely manner, fully and accurately. Moreover, the national economic policy, the effective implementation of laws and regulations will be strengthened to ensure the safety of the assets of the stakeholders, improve the level of internal management and financial workers quality of the enterprise and ensure the accounting information accurate and accounting work more efficient.

We use the method of principal component analysis in the statistics discipline to identify accounting distortion of listed companies. Principal component analysis (PCA) is the method to analyze multiple related financial indexes at first and then select several principal component index among them which can represent other weak correlated indicators.These a few principal component indexes are not only a linear combination of the evaluation indexes , but also still retain main information of the original indexes on the condition of no serial correlation and they can reduce a large amount of sample calculation.

We use the disclosed financial statements of listed companies from 2009 to 2013 as the research objects. And we find object sample from the auditing opinions database of CSMAR database and financial indicators of listed companies to establish sample model. The total samples include 28 listed companies of distorted accounting information and 28 listed companies of non-distorted accounting information.We divide total samples into two groups: one is a group of estimated samples, another one is a group of test samples.We regarded the audit opinions of annual reports of listed companies:"inexpressible opinions", "refused to commit herself", "negative opinions" as distorted samples, and define the audit opinion "no reservations" as a non-distorted sample.In the next step, we use T statistical testing to screen out significant financial indicators. And then we use Principal Component Analysis(PCA) and SPSS statistical software to establishing prediction model of accounting information distortion. Eventually, we test the accuracy and effectiveness of the model.

1. **Determine the financial indicators**

The selected financial indicators include four aspects such as solvency, operation ability, profit ability and development ability of the listed companies to identify accounting information distortion.

We defined Liquidity ratio as X1, quick ratio as X2, cash ratio as X3, asset-liability ratio as X4, property ratio as X5,tangible equity debt ratios as X6 to describe solvency;

Inventory turnover ratio as X7, accounts receivable turnover ratio as X8, current assets turnover as X9, fixed asset turnover as X10, total assets turnover as X11 to represent operating capacity;

Asset returns as X12, net profit margin on sales as X13, gross profit margin on sales as X14, ratio of profit to cost as X15 on behalf of profitability

Asset growth X16, sales growth X17 to stand for development capacity.

More than 17 indicators of financial ratio are selected above, which except the asset-liability ratio is inversely proportional relationship with financial risk, the rest of them exists direct proportional relationship with financial risk.

2. Discriminate the significant index

Estimate samples are selected form top 20 samples of distorted samples and from the top 20 samples of non-distorted samples, select the last 8 non-distorted samples and the last distorted samples as a prediction group, from the solvency, operation ability, profitability and development ability those four aspects to select 17 financial ratio indicator to establish principal components of prediction model of accounting information distortion, after input those samples into SPSS software and found degree of freedom(Df) and adjoint probability(Sig.) as follows:

Table 1 Df and Sig.

Index	T-value	Df	Sig.
X1	-3.739	38	0.001
X2	-3.083	38	0.004
X3	-2.739	38	0.009
X4	2.157	38	0.037

X5	0.229	38	0.82
X6	3.574	38	0.001
X7	-1.584	38	0.122
X8	-0.223	38	0.825
X9	-0.576	38	0.568
X10	-0.091	38	0.928
X11	-2.122	38	0.04
X12	-2.586	38	0.014
X13	-1.096	38	0.28
X14	-1.109	38	0.274
X15	-2.416	38	0.021
X16	-2.601	38	0.013
X17	-1.714	38	0.095

3. The T-statistic significance test: (significance level of 0.05)

According to the results of the T-value, analysis is as follows:

For X1, namely the liquidity ratio, the result of T test is 0.001, which is less than 0.05, so P-value falls in the rejection region, namely that there is a significant difference between distorted samples and non-distorted samples. Hence X1 should be retained.

Likewise, for X5, namely equity ratio, the result of T test is 0.82, which is greater than 0.05, so P-values falls in the acceptance domain, namely there is no significant difference between distorted samples and non-distorted samples. Hence, X5 should be removed.

According to this rule, X1, X2, X3, X4, X6, X11, X12, X15, X16 the nine variables should be retained. But there are still some correlations between these variables, which lead to double-counting of some information and increase the complexity of analyzing problems in the meanwhile. Therefore, the results of the prediction model will be inaccurate. In order to solve this problem,we will analyze the nine variables by principal component analysis.

4. **The determination of principal component factors**

Software(SPSS20.0)operation：

choose analysis--dimension reduction---factor analysis (select principal component analysis)

First, we test the applicability of principal component analysis in: KMO statistical value is 0.592, between 0 and 1, which is close to 1, so it accords with the requirement of principal component analysis;The associated probability of spherical hypothesis test is 0.000, which is less than the significance level of 0.05， so the spherical hypothesis is rejected. Therefore,correlations between indexes are not independent and the indexes are suitable for principal component analysis. Then we determined and interpreted principal components. At last, we obtained characteristic value and rate of contribution by principal component analysis.

Table2 Characteristic value of principal component and the rate of contribution

components	Initial Eigenvalues			Extract square and load		
	Total	Variance（%）	Accumulation（%）	Total	Variance（%）	Accumulation（%）
2	1.861	20.682	61.412	1.861	20.682	61.412
3	1.353	15.034	76.446	1.353	15.034	76.446
4	0.94	10.443	86.889	0.94	10.443	86.889
5	0.565	6.282	93.171			
6	0.336	3.735	96.906			
7	0.184	2.04	98.945			
8	0.079	0.881	99.826			
9	0.016	0.174	100			

In the table 2 the cumulative rate of contribution of the first four principal components has reached 86.889%, which satisfies the requirement of experience value 85%. Through the data in the Communality table, the extracted components have almost explained all variables and contained more than 80% information, so we can determined that the number of four principal component factors is four.

Obtained from the factor load matrix: factor Z1 is mainly composed of X1, X4, X6, X11, X12, factor Z2 is primarily made up of X2, X3, X15, main factor Z3 is principally composed of X4, X6, X16, factor Z4 chiefly consists of X4, X11, X12, X15

5. Construct prediction model to predict and select break point

Table 3 factor score matrix

components	1	2	3	4
X1	0.254	-0.146	-0.073	0.169
X2	0.245	-0.18	-0.063	0.168
X3	0.239	-0.173	-0.08	0.162
X4	-0.079	-0.171	0.612	0.075
X6	-0.15	0.15	-0.017	0.608
X11	0.065	0.324	-0.014	0.662
X12	0.137	0.395	0.189	-0.203
X15	0.137	0.373	-0.05	-0.352
X16	0.145	0.036	0.557	0.053

Consequently, we can gain prediction model by the table of contribution rates and the Component Score Coefficient Matrix

$$Y=0.4073Z_1+0.20682Z_2+0.15034Z_3+0.10443Z_4$$

$$\begin{cases} Z_1 = 0.254X_1 + 0.245X_2 + \cdots + 0.145X_{16} \\ Z_2 = -0.146X_1 + -0.180X_2 + \cdots + 0.036X_{16} \\ Z_3 = -0.073X_1 + -0.063X_2 + \cdots + 0.557X_{16} \\ Z_4 = 0.169X_1 + 0.168X_2 + \cdots + 0.053X_{16} \end{cases}$$

Through the formula above, we can obtain prediction value of every companies Y， the distorted data is in front, the non-distorted data is in the rear.

Table 4 the predicted value of estimate sample

Company Code	Predicted Value	Company Code	Predicted Value	Company Code	Predicted Value	Company Code	Predicted Value
000863	-0.1270	000818	-1.6901	000402	0.4407*	000089	0.2496
000517	0.2031	000673	-0.0158	000408	0.3041*	000036	0.7384*
000688	-0.0928	600155	-0.1655	000409	0.3709*	000048	0.6998*
000100	0.1685	600817	0.2451	000410	0.4479*	000059	0.3317*
600751	0.1014	002200	-0.7191	000151	0.6807*	000034	0.3547*
000719	-0.1332	000403	0.2670	000153	0.3075*	000039	1.3698*
600556	-0.2173	600691	0.2389	000157	0.5937*	000045	0.5101*
600185	0.2509	000035	0.2528	000158	0.3447*	000050	1.4730*
600988	0.1181	600462	0.2315	000005	0.3178*	000028	0.3567*

Using 0.28 as a segmentation, we can distinguish whether accounting information of the samples is distorted or not. As is shown in the table above, the predicted value marked as * is on behalf of the sample whose predicted value is more than 0.28, all of 20 samples of distortion are less than 0.28, so all the samples satisfy the rule. For the rest, 19 of the 20 non-distorted samples are greater than 0.28. We can figure out that the success rate of prediction model was 93.75%. Therefore, the prediction model has a high

degree of credibility, differentiation and accuracy.

6. Testing the effectiveness of prediction model

We calculated the sixteen prediction values by the established prediction model, and then the results are as follows:

Table 5 Prediction sample

Company Code	Prediction value	Distorted or not	Company Code	Prediction value	Distorted or not
000017	-0.1671	Y	000006	0.4080*	N
600555	-0.0155	Y	000026	0.3947*	N
600381	0.2498	Y	000029	0.3994*	N
002506	-0.0043	Y	000002	0.4068*	N
000033	0.3002*	Y	000010	0.5992*	N
002506	0.2727	Y	000014	0.3876*	N
600145	-0.1801	Y	000016	0.4607*	N
600247	0.1966	Y	000019	0.4470*	N

The predicted value with Star is on behalf of the one more than 0.28. We can see that there is only one distorted company whose predicted value is greater than 0.28 and the others are all right. That is to say, there are 15 correct predictions in the predictive results for the 16 prediction samples, so the success rate was 93.75%, which is very high. Therefore, we can clearly see that the established predicting model of the paper is very accurate.

7. Advantages and disadvantages of the model

Through the principal component analysis above, we can come to a conclusion: the selected original indexes in the model have a strong explanatory ability, so we can identify better whether the accounting information of a company is distorted or not; Compared with the predicted values, the prediction accuracy of the established model in this paper is as high as 93.75%, which demonstrate that it is highly accurate and strongly practical by using the model to identify the accounting information distortion in the company. Of course, there also exist some limitations by the principal component prediction model to predict when the accounting information distortion, because the number of selected samples is small. If we can choose more samples to establish a prediction model, the result will be more accurate.