Frontiers in Economics and Management
Volume 2 Issue 6, 2021
ISSN: 2692-7608
DOI: 10.6981/FEM.202106_2(6).0038

Study on Liquor Sector Quantitative Stock Selection based on Principal Component Analysis and Cluster Analysis

Minghui Zhang, Lu Cao, Qianyu Xu

Anhui University of Finance and Economics, Bengbu 233000, China.

Abstract

In view of the investment value of the liquor sector, due to the large number of listed companies involved in the liquor sector and the complicated financial situation, it is very important to analyze and study the performance of the liquor sector companies. In this paper, 8 financial indicators were selected based on the financial statement data to construct the investment value evaluation index system. 20 liquor listed companies were selected as the research objects, and the comprehensive evaluation was carried out by using principal component analysis. Secondly, 20 listed liquor companies were classified by systematic cluster analysis. Finally, it provides feasible investment suggestions for investors to help them narrow their investment range and maximize their returns.

Keywords

Liquor Sector; Principal Component Analysis; Cluster Analysis; Quantitative Stock Selection.

1. Introduction

Chinese liquor culture has a long history. With the rapid development of China's economy and the substantial improvement of people's consumption power, some medium and large liquor enterprises have seized the opportunity to continuously innovate the industrial structure and successfully carried out transformation and adjustment, effectively driving the steady development of China's liquor industry and gradually getting rid of the influence of factors such as the eight-point regulation. Even in the current epidemic situation, the international economic downturn under the big economic environment, investors from all sides are still optimistic about the liquor sector, investment enthusiasm is not reduced but increased. At present, the heat of liquor plate is not reduced, and the listed companies involved in liquor plate make investors dazzling and difficult to choose. Based on this background and from the perspective of investors, this paper makes full use of the financial statement data of listed companies to conduct quantitative analysis on the stocks of 20 popular listed companies in the liquor sector. Through empirical analysis, this paper studies the investment value of each stock and reduces the investment scope, so as to provide some references for investors.

2. Establishment of quantitative investment evaluation index system

2.1. Construction of evaluation index system

Whether a listed company has investment value, can start from the consideration of the company's financial information. First of all, we need to build an index system for the investment value analysis of listed companies in the liquor sector to measure and evaluate the analysis from multiple perspectives.

For this reason, 20 listed enterprises in liquor industry were selected as the research samples. "Choose two indicators that represent the solvency of an enterprise: quick ratio (X1) and cash flow ratio (X2). Select two major indicators representing the operating capacity of the

ISSN: 2692-7608 DOI: 10.6981/FEM.202106_2(6).0038

enterprise: inventory turnover (X3) and total assets turnover (X4). Select two major indicators representing the profitability of the enterprise: return on equity (ROE) (X5) and net profit rate on sales (X6). Two indicators representing the growth ability of enterprises are selected: basic earnings per share (X7), total operating income (X8) and other eight indicators to construct the evaluation index system of the investment value of listed liquor industry enterprises [1].

2.2. Data source and processing

The data in this paper came from the financial indexes of 20 listed liquor companies in 2020 selected by the iFinD data terminal of flush. Due to the dimensional and numerical differences among selected data, "SPSS25.0 was used to standardize the original data in order to eliminate possible errors. The standardized treatment was to standardize the 8 variables using the Z-score method [2]." The formula is ZX=(X-μ)/S, where μ is the mean value of variable data and S is the standard deviation. Data in this paper do not need to be cotrended.

3. Comprehensive evaluation of liquor stock investment value based on principal component analysis

3.1. Principal component analysis

"Principal component analysis (PCA), also known as factor analysis, is a common dimensionality reduction method in multivariate statistical analysis. The goal is to turn multiple indicators into aggregate indicators with little loss of information [3]." Usually called main component, the comprehensive indicators of the transformation to generate each principal component is a linear combination of the original variables, the vast majority of information to be able to reflect the original variables, and unrelated between each main component, the winner of the ingredients with some of the more superior performance than the original variables, to achieve the effect of dimension reduction, more scientific and effective data and information, make things simple [4].

3.2. Suitability test

KMO test and Bartlett sphericity test can be used to determine whether the selected financial indicators are suitable for factor analysis to extract common factors. Studies have shown that when KMO<0.5, factor analysis is not suitable, and when KMO>0.9, factor analysis has the best effect. A large Bartlett sphericity test means that factor analysis can be performed. The suitability test was carried out with the help of SPSS25.0 software, and the output results were shown in Table 1."As can be seen from Table 1, the KMO value is 0.632, greater than 0.5, and the significance level of Bartlett's spherical test is 0, less than 0.05. The null hypothesis is rejected, and the correlation matrix is not the identity matrix. The above two statistical tests show that the original data are suitable for factor analysis [5]."

Table 1. KMO and Bartlett tests

KMO sampling fitness measure		.632
Bartlett's test for sphericity	chi-square	111.968
	Degrees of freedom	28
	significant	.000

3.3. Determine the number of principal components

The financial data of 20 liquor listed companies were analyzed by SPSS25.0 statistical software.on the basis of data standardization, the eigenvalue of the correlation coefficient matrix and the variance contribution rate of each factor were obtained, according to the calculation results, the cumulative variance contribution rate of the three common factors was

ISSN: 2692-7608 DOI: 10.6981/FEM.202106_2(6).0038

81.93%. This shows that the extracted 3 common factors can explain 81.93% of the 8 original variables. The original variable information is fully retained, and the extracted common factors have good representativeness, as shown in Table 2.

Table 2. Contributions of characteristic roots and major factors

composition	Initial eigenvalue			The sum of squares of the extraction load		
	total	Percentage of variance	Cumulative %	total	Percentage of variance	Cumulative %
1	4.362	54.526	54.526	4.362	54.526	54.526
2	1.421	17.766	72.292	1.421	17.766	72.292
3	0.771	9.641	81.933	0.771	9.641	81.933
4	0.627	7.842	89.775			
5	0.424	5.302	95.077			
6	0.302	3.781	98.858			
7	0.060	0.745	99.603			
8	0.032	0.397	100.000			

3.4. Name the principal components

As can be seen from Table 3, "ROE and net profit rate on sales have a large load on the first principal component, that is, they have a high correlation coefficient with the first principal component, so they can be defined as profitability factor (F1). The absolute value of the load on the second principal component of inventory turnover is larger, that is, the correlation degree is higher, which can be defined as the operating capacity factor (F2). The quick ratio has a large load on the third principal component, which can be defined as the solvency factor (F3) [6]."

Table 3. Factor load matrix

indicators	composition		
	1	2	3
Quick ratio (ZX1)	0.649	-0.289	0.687
Cash flow ratio (ZX2)	0.775	-0.161	0.004
Inventory turnover (ZX3)	0.254	0.886	0.307
Total asset turnover (ZX4)	0.804	0.386	-0.381
ROE (ZX5)	0.898	-0.283	-0.206
Net profit rate on sales (ZX6)	0.860	-0.391	-0.066
Basic earnings per share (ZX7)	0.682	0.344	-0.072
Gross operating income (ZX8)	0.788	0.168	0.090

3.5. Calculation and ranking of principal component scores

Although the factor load matrix Table 3 can be obtained directly, the load matrix is not the eigenvector of the principal component, that is to say, it is not the coefficient of the principal component. The principal component coefficient can be calculated by dividing the respective factor load vector by the square root of the respective eigenvalue. The results are shown in Table 4

Table 4. Principal component score coefficient matrix

indicators	composition		
	1	2	3
Quick ratio (ZX1)	0.311	-0.242	0.782
Cash flow ratio (ZX2)	0.371	-0.135	0.005
Inventory turnover (ZX3)	0.122	0.743	0.350
Total asset turnover (ZX4)	0.385	0.324	-0.434
ROE (ZX5)	0.430	-0.237	-0.235
Net profit rate on sales (ZX6)	0.412	-0.328	-0.075
Basic earnings per share (ZX7)	0.327	0.289	-0.082
Gross operating income (ZX8)	0.377	0.141	0.102

ISSN: 2692-7608 DOI: 10.6981/FEM.202106_2(6).0038

The scoring function of each principal component can be obtained from Table 4

$$F1=0.311*ZX1+0.371*ZX2+0.122*ZX3+0.385*ZX4$$
$$+0.430*ZX5+0.412*ZX6+0.327*ZX7+0.377*ZX8$$
$$F2=-0.242*ZX1-0.135*ZX2+0.743*ZX3+0.324*ZX4-0.237*ZX5$$
$$-0.328*ZX6+0.289*ZX7+0.141*ZX8$$
$$F3=0.782*ZX1+0.005*ZX2+0.350*ZX3-0.434*ZX4-0.235*ZX5$$
$$-0.075*ZX6-0.082*ZX7+0.102*ZX8$$

Taking the variance contribution rate of the three principal components as the weight, the weighted average of the three principal components was used to construct the comprehensive evaluation model of the investment value of the listed companies in the liquor sector:

$$F=0.545*F1+0.178*F2+0.096*F3$$

According to the above three principal component score functions and comprehensive value evaluation model, the ranking of each principal component score and comprehensive value score can be obtained by substituting standardized financial data. The results are shown in Table 5.

Table 5. Principal component scores and comprehensive ranking of 20 listed companies in the liquor sector

The listed company	F1	F2	F3	F	Comprehensive ranking
600519.SH	3.9325	-1.8534	0.6988	1.88	1
000568.SZ	2.9418	-0.1904	-0.0811	1.56	2
000858.SZ	2.2655	-0.0169	1.9600	1.42	3
600809.SH	1.7537	0.4116	-1.4396	0.89	4
000860.SZ	0.7490	2.7968	-0.3122	0.88	5
002304.SZ	1.3457	-0.3457	-0.5550	0.62	6
000596.SZ	1.0776	0.3159	-0.3491	0.61	7
603369.SH	0.8970	-0.1989	0.1548	0.47	8
600197.SH	-.0380	0.8247	1.4234	0.26	9
600559.SH	0.1634	1.2528	-0.8006	0.24	10
000799.SZ	0.6221	-1.0082	0.3153	0.19	11
600702.ST	0	-0.4255	-0.8825	-0.09	12
603919.SH	-0.2994	0.6421	-0.7104	-0.12	13
603589.SH	0.0742	-0.7686	-0.3675	-0.13	14
600238.SH	-2.0934	2.5753	0.6334	-0.62	15
603198.SH	-1.2923	-0.8135	-0.5126	-0.90	16
600199.SH	-2.3060	0.0302	1.3496	-1.12	17
600779.SH	-2.2572	-1.1105	-0.9654	-1.52	18
002646.SZ	-3.8940	-0.8490	0.2072	-2.25	19
600766.*ST	-3.7629	-1.2688	0.2335	-2.25	20

3.6. Principal component analysis

According to Table 5, when the principal component score results are positive or negative, those with a score greater than 0 indicate good financial performance of the company, while those with a score less than 0 indicate poor financial performance, and the larger the score, the better. According to the scores of each principal component, as far as profitability factor (F1) is concerned, 600519.SH(3.9325), 000568.SZ(2.9418), 000858.SZ(2.2655) scored higher, It shows that these three companies have stronger profitability than other listed liquor companies. In terms of the operating capacity factor (F2) score, 000860.SZ has a score of 2.7968, which is best among listed companies, indicating that 000860.SZ is better than other listed companies in the liquor sector in terms of operating capacity. In terms of solvency (F3), the

ISSN: 2692-7608 DOI: 10.6981/FEM.202106_2(6).0038

scores of 000858.SZ and 600199.SH were 1.9600 and 1.3496, respectively, which were relatively outstanding, showing strong solvency. In the ranking of comprehensive score (F), the comprehensive competitiveness of 600519.SH was 1.88, the most competitive among listed liquor companies, followed by 000568.SZ and 000858.SZ. A closer look at the statement can be found that 600519.SH continued the double-digit growth trend of its main financial data. From January to September in 2020, even though the epidemic situation is under high pressure, 600519.SH still achieved operating income of 67.2 billion yuan, up 10% year on year.Net profit reached 33.8 billion yuan, up 11% year on year, fulfilling the company's annual revenue growth target of 10%. 600519.SH, 000568.SZ and 000858.SZ, as the leading stocks in the liquor industry, have good development prospects and deserve investors' attention. However, the profitability of 0002646.SZ and 600766.*ST, which are at the bottom and have much difference from the previous scores, is not satisfactory, and other financial indicators also appear abnormal, so it is not recommended for investors to start.

4. Classification of liquor sector stock investment value based on cluster analysis

4.1. Cluster analysis method

Cluster analysis refers to the analysis process of grouping a set of physical or abstract objects into multiple classes composed of similar objects. Clustering method, also known as group analysis, is a multivariate statistical analysis method for quantitative analysis of multiple samples. Clustering methods mainly include: system clustering method, K-means clustering method, hierarchical clustering method, SOM clustering method, FCM clustering method [7].

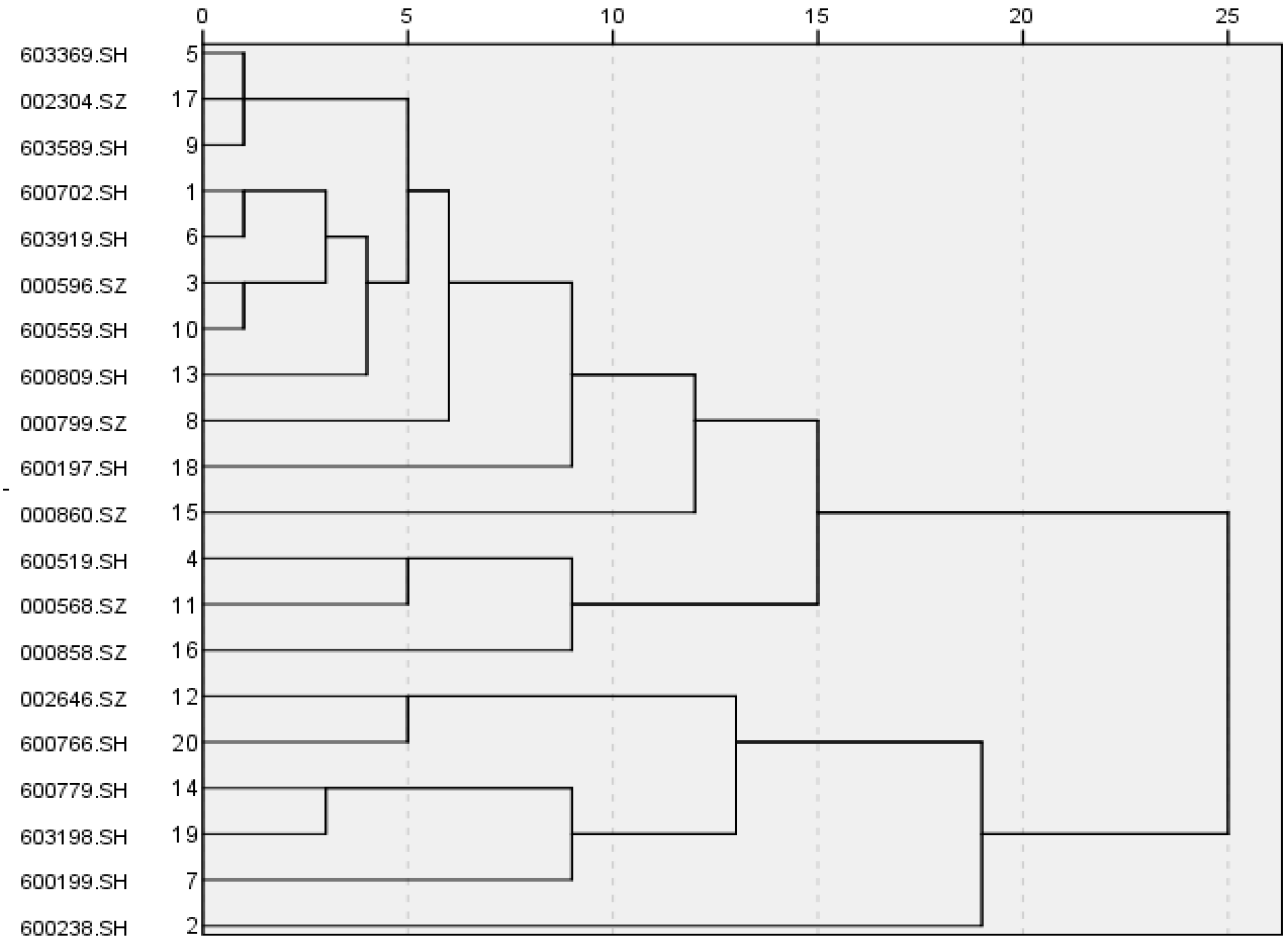

Figure 1. System cluster pedigree diagram

ISSN: 2692-7608 DOI: 10.6981/FEM.202106_2(6).0038

4.2. Empirical analysis process

In this paper, the systematic clustering method is used. Firstly, the data of each major factor is standardized; secondly, the Euclidean distance is used to measure the similarity between different major factors; finally, the class average method is used to calculate the distance between different major factors.Samples close to each other are clustered into clusters, and those far apart are clustered into clusters. The process continues until each sample can always be clustered into a suitable category. Finally, all samples are clustered into one category. SPSS25.0 software was used to complete the above process, and the cluster pedigree diagram was output, as shown in Figure 1.

The pedigree diagram shows the merged classes and their coefficient values in each step, which can reflect the whole process of clustering and convert the distance between the classes into a value between 1 and 25. Generally, the best classification scheme can be considered when the ruler is vertically placed on the map for translation and stops at the interval around the maximum distance between the combined vertical lines. Each horizontal line that intersects the ruler is then a class, and the items to the left of the horizontal line are members of the class.By observing the systematic cluster pedigree diagram in Figure 1, the 20 listed companies are divided into three categories, as shown in Table 6

Table 6. 20 liquor listed companies classification

category	number	Stock code
1	14	603369.SH, 002304.SZ, 603589.SH, 600702.ST, 603919.SH, 000596.SZ, 600559.SH, 600809.SH, 00079.SZ, 600197.SH, 000860.SZ, 600519.SH, 000568.SZ, 000858.SZ
2	5	002646.SZ, 600766.*ST, 600779.SH, 603198.SH, 600199.SH
3	1	600238.SH

4.3. Result analysis

As a result of cluster analysis, the stocks of the largest class must be the "minor majority", while the stocks of the smallest class are likely to be the "important minority". This "minority" may be either the "junk stock" with poor performance or the "dark horse" in the future [8], It must belong to one of the two factors, The analysis of these two types of stocks can be distinguished by observation: those with good financial indicators are likely to be the dark horse, while those with bad financial indicators are likely to be "rubbish". Under the clustering analysis results of this paper, 600238.SH, as a single stock classified into a class, is worthy of investors' sufficient understanding. From the previous principal component analysis, it can be seen that 600238.SH lacks profitability and ranks low. Besides, through the in-depth analysis of data, In 2019, for example, the net profit of Hainan Coconut Island (600238.SH) belongs to the shareholders of the listed company is negative, the decline is about 760%, the strength is unsightable, the financial operating situation is not too optimistic, investors should be cautious investment.

5. Conclusions and limitations

The comprehensive evaluation of the business performance of listed companies is a complicated systematic evaluation process involving many index factors [9]. In this paper, through selecting financial indicators of listed companies to establish a comprehensive evaluation index system, the principal component analysis and cluster analysis model is applied to the liquor plate stock research to quantify the performance evaluation of listed companies, to analysis the current hot liquor stocks based on the analysis of investment value and make the results more intuitive concise, to a certain extent, and to provide the reference for the decision of investors,Narrating the investment range of investors and reducing investment

ISSN: 2692-7608 DOI: 10.6981/FEM.202106_2(6).0038

risks can also provide a method of comparison for investors who understand the market information of the securities industry.

Quantitative, of course, these two methods can mining potential investment value of listed companies, just according to the comprehensive evaluation on the company's historical financial data analysis, so the body in a rapidly changing capital market, investors are still reasonable to assess their risk to bear ability, grasps the market tendency, reduce the blind investment losses.

References

[1] Yujun Sun, Xiuqin Huang: Construction of Financing Constraint Model for Chinese SMEs——Based on Binary Logistic Regression Method. Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition), Vol. 21 (2019) No.1, p .76-86.

[2] Meng Wu: Transformation and development of the bordering area of Shanxi, Shaanxi and Mongolia under the constraints of resources and environment (MS.D., ShanXi University, China 2015)

[3] YuanYuan Guo, LiWei Cheng and DongHui Zhang: The evolution and enlightenment of country risk assessment methods for commercial banks. Financial Regulation Research, 2012(04): 33-46.

[4] Zhang Qi. Research on the motivation and performance of the overall listing [D]. Donghua University, 2017.

[5] Wang Xiaozhe, Chai Liangqi, Wu Jie. Evaluation of sustainability of listed agricultural companies in my country——Based on principal component analysis, factor analysis and cluster analysis [J]. Journal of Yangtze University (Social Science Edition), 2016, 39(11): 50-55.

[6] Yuan Xiaowen. An Empirical Study on the Impact of Listed Company Investor Relations Management on Company Value [D]. Central South University, 2009.

[7] Ma Tiannan. Research on icing disaster prediction and risk management of power grid based on big data[D]. North China Electric Power University (Beijing), 2017.

[8] Chen Chen. Using cluster analysis to select listed companies with investment value [J]. Zhejiang Finance, 2004(11): 32-33.

[9] Xie Si. Analysis of the operating performance of financial listed companies[J]. Guangxi Financial Research, 2007(10): 46-50.