Empirical Research on Stock Selection based on Principal Component Analysis

-- A Case Study of Baijiu Industry Stocks

Ruman Liu*, Xinran Zhang, Xue Jiang, Xuanzhi Zheng

Anhui University of Finance and Economics, Bengbu, 233030, China

Abstract

In order to help the majority of small and medium-sized investors obtain higher returns and avoid losses in the venture capital of the stock market, and also to provide a universally applicable accurate stock selection method based on industry analysis, this paper selects six financial indicators after comprehensive consideration, including total net asset interest rate, return on net assets, earnings per share, net assets per share, undistributed profits per share, and predicted net profit growth rate, The principal component analysis method is used to evaluate the investment value of 20 stocks in the Baijiu sector that have highlighted their value in the recent market bottoming process, and SPSS is used to process the original data. Finally, it is concluded that the top three stocks of Baijiu that have been better recently are Guizhou Moutai (600519.sh), shanxifen (600809.sh) and shede (600702.sh).

Keywords

Principal Component Analysis; Baijiu Industry; Stock Selection.

1. Introduction

Baijiu has a long history in China, and the history of liquor making has lasted for thousands of years. The customs handed down by the Chinese people for thousands of years are festivals, family and friends get together, weddings and funerals, celebrations and birthday celebrations. As long as it is at the dinner table, wine is essential. The word "dinner table" has gradually evolved into "wine table", and the wine table culture has also become a special culture in China. Baijiu has been an indispensable part of Chinese life, with great rigid demand, close to condiments, and plays a very key role in the development of the national economy. Therefore, the research on it can be said to have important practical significance.

In the stock venture capital, Baijiu shares have always been one of the focuses of investors. Judging from the recent stock market, the Baijiu sector is relatively depressed, but the decline of the market does not mean the disappearance of investment opportunities and value preservation opportunities. On the contrary, I believe that the Baijiu industry still has considerable investment value from both fundamental and technical analysis.

First of all, from the fundamental point of view, the just past national day Mid Autumn Festival has driven the rise of Baijiu stocks, and it is expected that the heat will drop after the peak. Moreover, with the recent sharp drop in the national temperature, there is a trend that Baijiu will soon enter a cold winter. As a top beverage in winter, it is expected that the sales volume of Baijiu will rise. Moreover, the Spring Festival sales also make the market of liquor plate promising in the short term at least; Secondly, from the perspective of technical analysis, I think the value correction of Baijiu plate in the decline since 3478 points has been basically in place; Finally, due to the dietary preference of Chinese people since ancient times and the

characteristics of Baijiu industry, Baijiu stocks have always been a non cyclical stable industry with a defensive role.

2. Literature Review

In dingqi's research on stock multi factor quantitative investment strategy based on principal component regression analysis, it is proposed that the value factors and technical factors constructed by principal component analysis and multi factor stock selection model can accurately reflect the advantages and disadvantages of stocks under various market conditions [1].

In the research on comprehensive quantitative stock selection strategy based on improved neural network and principal component analysis, liuyanhong proved that adding principal component analysis as one of the stock selection methods can effectively improve investment income [2].

In the application of principal component analysis in stock selection of agricultural stock selection, Yu Wei used principal component analysis to select three better agricultural stocks, namely Beidahuang, Denghai seed industry and guannong, and analyzed the rationality and effectiveness of the stock selection method according to the actual market trend and the principal component analysis of fundamentals [3].

Based on the research conclusions drawn by the above scholars and experts, it can be effectively concluded that the principal component analysis method has great credibility in stock selection.

3. Theoretical Basis--Principal Component Analysis Model

3.1. Basic Thought

The idea of principal component analysis is to try to integrate the original numerous indicators with certain relevance (such as P indicators). It is recombined into a group of new comprehensive indicators that have no correlation with each other to replace the original indicators. Usually, the mathematical treatment is to make a linear combination of the original P indicators. If the first linear combination selected is recorded as E. Naturally, I hope that the more information e expresses, the better. The most classical method is represented by VAR (E). The larger the%r (f), the more information it contains. Therefore, f selected in all linear combinations should have the largest variance, so f is called the first principal component. If the first principal component is not enough to represent the information of the original P indicators. Then consider selecting only the second principal component. Of course, we require cov (F, e) =0, and so on. It is conceivable that these principal components are not only uncorrelated but also the variance decreases accordingly. Although we may miss some information, this method enables us to grasp the main contradiction and further extract some new information from the original data. Therefore, in some practical problems, the benefits outweigh the losses.

3.2. Mathematical Model

Set n samples, observe P indicators of each sample: X_1, X_2,..., X_P, and get the original data,

$$X=\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{n2} & \dots & x_{np} \end{bmatrix}=(X_1,X_2,\dots,X_P)$$

Of which.

$$X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ nj \end{bmatrix}, \ j=1,2,\dots,p$$

Use the P vectors X_1, X_2,..., X_P of the data matrix X to make a linear combination as follows.

$$\begin{cases} F_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p \\ \qquad \dots \dots \\ F_p = a_{1p}X_1 + a_{2p}X_2 + \cdots + a_{pp}X_p \end{cases}$$

Abbreviated as.

$$Fi = a1iX1 + a2iX2 + \dots + apiXp \ , \ i=1,\dots,p$$

And the coefficient is determined by the following principles,

First, F_i is not related to F_j,

Second, F1 has the largest variance among all linear combinations of X_1, X_2,..., X_P. F_2 is the largest variance in all linear combinations of X_1, X_2,..., X_P in which F_1 is not linearly related, and FP is the largest variance in all linear combinations of X_1, X_2,..., X_P in which F_1 F_2,..., F_{P-1} are not linearly related.

3.3. Computational Procedure

There are n samples. Observe the P indicators of each sample, X_1, X_2,..., X_P, and write the original number into a matrix

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

Standardize raw data, Establish the correlation coefficient matrix of variables, R= (rij) $_{p*p}$ let R=X'X

Finding the characteristic root of R ,$\lambda_1 >= \lambda_2 >= \lambda_3 >= \dots \lambda_p > 0$ and the corresponding unit eigenvalue vector,

$$a1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix}, \ a2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix}, \dots, \ ap = \begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{bmatrix}$$

Write the principal components,

$$Fi = a_{1i}X_1 + a_{2i}X_2 + \dots + a_{pi}X_p \ , \ i=1,\dots,p$$

Calculate comprehensive score

A comprehensive evaluation function is constructed: f=f (X_1, X_2,..., X_P), and the samples are sorted by the value of evaluation function f=f (X_{1j}, X_{2j},..., X_{pj}) (j=1,2,... N). According to the relative importance of each index (X_1, X_2,..., X_P), determine the weight W_1, W_2,... W_P, and the comprehensive evaluation function can be, f=f (X_1, X_2,..., X_P) =$W_1X_1 + W_2X_2 + \dots W_pX_p$.

With the principal component as the weight, let the information content of the first m principal components $Z_1, Z_2, ... Z_m$ be $\lambda_1, \lambda_2, ... \lambda_m$. And the cumulative contribution rate is very large (more than 85%).

$$So, Wk = \frac{\lambda_k}{\sum_{j=1}^{m} \lambda_j}, k=1,2,...m, W_k > 0, and \sum_{j=1}^{m} W_j = 1$$

The weight of Z_K can be taken as W_k. Therefore, the comprehensive evaluation function can be constructed as

$$f = f(Z_1, Z_2, ... Z_m) = \sum_{k=1}^{m} W_k Z_k$$

By substituting the score Z_K (j=1,2,..., n, k=1,2,..., m) of the principal component into the above formula, the comprehensive score of each sample is $f_j = \sum W_k Z_{kj}$, j=1,2,..., n.

Rank $f_1, f_2,..., f_n$ from small to large to get the ranking of the samples.

4. Sample Selection and Index Selection

4.1. Sample Selection

We selected 20 stocks in the Baijiu sector in the industry options of flush software: Guizhou Maotai (600519.sh), Shanxi Fenjiu (600809.sh) and Shuijingfang (600779.sh); Kouzijiao (603589.sh); Yingjia tribute wine (603198.sh); Highland barley wine (002646.sz); Shunxin agriculture (000860.sz); Ilite (600197.sh); Jinhui liquor (603919.sh); Laobaigan liquor (600559.sh); Golden seed wine (600199.sh); Huangtai liquor (000995.sz) shede liquor (600702.sh); Yanghe Co., Ltd. (002304.sz); Wuliangye (000858.sz); Luzhou Laojiao (000568.sz); Jiugui Liquor (000799.sz); Gujinggong liquor (000596.sz); Gujinggong B (200596.sz); Jishiyuan (603369.sh) as a sample.

4.2. Indicator Selection

To determine whether the selected listed companies in the Baijiu industry are worth investing, we must first determine which indicators of these listed companies to analyze. How to select correct and effective indicators? We believe that whether an enterprise is worth investing should be considered from the following three aspects, namely, the company's growth, profitability and equity expansion ability reflect the investment value of listed companies, and these three upper indicators are respectively embodied by the three lower indicators, combined with the statistical analysis theory of securities investment, We finally selected the following six indicators in the latest period of 20 samples: total net asset interest rate, predicted net profit growth rate, return on net assets, earnings per share, net assets per share and undistributed profit per share. In the following analysis, these indicators are set as $ZX_1, ZX_2, ZX_3, ZX_4, ZX_5$ and ZX_6 respectively.

Net interest rate of total assets: refers to the percentage of the company's net profit to the average total assets. This indicator reflects the profit level obtained by the company from the use of all assets, that is, the average profit that the company can obtain for each asset of 1 yuan. The higher the index, the higher the input-output level of the company, the more effective the asset operation and the higher the cost control level. Reflect the level of enterprise management.

Forecast net profit growth rate: the net profit growth rate refers to the growth rate of the net profit of the current period compared with the net profit of the previous period. The larger the index value, the stronger the profitability of the enterprise. The predicted value is selected here.

Return on net assets: roe, also known as return on shareholders' equity / return on net worth / return on equity / return on equity / return on equity / return on net assets, is the percentage

of net profit and average shareholders' equity. It is the percentage obtained by dividing the company's after tax profit by net assets. This indicator reflects the level of return on shareholders' equity and measures the efficiency of the company in using its own capital. The higher the index value, the higher the return of investment. This indicator reflects the ability of self owned capital to obtain net income, and is an important financial indicator to measure the efficiency of the use of shareholders' funds.

Earnings per share: earnings per share is earnings per share (EPS), also known as after tax profit per share and earnings per share, which refers to the ratio of after tax profit to total share capital. It refers to the net profit of the enterprise or the net loss of the enterprise to be borne by ordinary shareholders for each share they hold. EPS is usually used to reflect the operating results of enterprises, measure the profit level and investment risk of common shares, and is one of the important financial indicators for investors and other information users to evaluate the profitability of enterprises, predict the growth potential of enterprises, and then make relevant economic decisions. Among many tools for basic analysis of stock investment, EPS is also one of the most common reference indicators, like P / E ratio, P / B ratio, discounted cash flow and other indicators.

Net assets per share: net assets per share refers to the ratio of shareholders' equity to the total number of shares. The calculation formula is: net assets per share = total shareholders' equity / total number of shares in capital stock. This indicator reflects the present value of assets per share. The higher the net assets per share, the more the value of assets per share owned by shareholders; The less the net assets per share, the less the value of assets per share owned by shareholders. Generally, the higher the net assets per share, the better. Among the various indicators of basic analysis, net assets per share is one of the most important reference indicators to judge the intrinsic value of enterprises.

Undistributed profit per share: net assets per share refers to the ratio of shareholders' equity to total shares. The calculation formula is: net assets per share = total shareholders' equity / total number of shares in capital stock. This indicator reflects the present value of assets per share. The higher the net assets per share, the more the value of assets per share owned by shareholders; The less the net assets per share, the less the value of assets per share owned by shareholders. Generally, the higher the net assets per share, the better. Among the various indicators of basic analysis, net assets per share is one of the most important reference indicators to judge the intrinsic value of enterprises.

5. Empirical Analysis

5.1. Principal Component Analysis Process

Import the data found in the flush into SPSS software and use factor analysis to obtain the relevant data of 6 initial eigenvalues that are numerically equal to the correlation number matrix as follows:

Table 1. Initial eigenvalue

COMPONENT	TOTAL	PERCENT VARIANCE	CUMULATIVE (%)
1	3.485	58.080	58.080
2	1.591	26.521	84.600
3	.881	14.681	99.282
4	.030	.499	99.780
5	.009	.153	99.933
6	.004	.067	100.000

The principal component analysis method is used to extract the sum of squares of loads, and the relevant data of the two principal components are as follows:

Table 2. Extract sum of squares of load

COMPONENT	TOTAL	PERCENT VARIANCE	CUMULATIVE (%)
1	3.485	58.080	58.080
2	1.591	26.521	84.600

It can also be seen that the fourth λ Value is an obvious turning point, but according to the principle that the cumulative contribution rate can exceed 80%, combined with the previous cumulative contribution rate of eigenvalues, it is appropriate to extract two principal components.

"Component matrix" is a factor load matrix, and each load represents the correlation coefficient of the principal component corresponding to the original variable. Therefore, the component matrix of the two principal components extracted by SPSS analysis is as follows:

Table 3. Component matrix

PROJECT	1	2
ZSCORE: NET INTEREST RATE OF TOTAL ASSETS ROA [REPORTING PERIOD] LATEST PERIOD (MRQ) [UNIT]%	0.667	0.715
ZSCORE: FORECAST NET PROFIT GROWTH RATE [YEAR] 2021 [AS OF] LATEST [UNIT]%	-0.277	0.384
ZSCORE: RETURN ON NET ASSETS ROE [REPORTING PERIOD] LATEST PERIOD (MRQ) [UNIT]%	0.546	0.822
ZSCORE: EPS - BASIC [REPORTING PERIOD] LATEST PERIOD (MRQ) [UNIT]YUAN	0.957	-0.239
ZSCORE: NET ASSETS PER SHARE BPS [REPORTING PERIOD] LATEST PERIOD (MRQ) [UNIT]YUAN	0.930	-0.331
ZSCORE: UNDISTRIBUTED PROFIT PER SHARE [REPORTING PERIOD] LATEST PERIOD (MRQ) [UNIT]YUAN	0.941	-0.301

After calculation, the eigenvectors of the two eigenroots are as follows:

Table 4. The eigenvectors of the two eigenroots

T1	T2
0.292	0.652
0.498	-0.262
0.504	-0.239
0.357	0.567
0.513	-0.189
-0.148	0.304

So the principal component expression is:

$$Y_1 = 0.292ZX_1 + 0.49Z8X_2 + 0.504ZX_3 + 0.357ZX_4 + 0.513ZX_5 - 0.148ZX_6$$

$$Y_2=0.652ZX_1-0.262ZX_2-0.239ZX_3+0.567ZX_4-0.189ZX_5+0.304ZX_6$$

The calculated scores of each main component Y1 and Y2 are:

Table 5. The calculated scores of each main component

SORT	Y_1	Y_2
1	6.527	3.459
2	1.267	2.598
3	0.580	1.402
4	0.817	0.638
5	0.664	0.761
6	0.511	0.880
7	0.168	0.921
8	0.291	0.050
9	0.203	-0.130
10	-0.158	0.378
11	-0.369	0.337
12	-0.433	-0.380
13	0-.559	-0.227
14	-0.829	-0.446
15	-0.913	-0.967
16	-0.955	-0.970
17	-1.027	-0.996
18	-1.312	-1.345
19	-1.999	-2.599
20	-2.473	-3.366

Results the comprehensive rating function is:

$$Y=0.581Y_1+0.265Y_2$$

(the coefficient before Y1 and Y2 is the variance contribution rate of the extracted two principal components)

5.2. Principal Component Analysis Results

Bring in the values of Y_1 and Y_2, and calculate the comprehensive score Y of the principal component and the order is:

Table 6. Calculate the comprehensive score Y

SORT	SECURITIES CODE	NAME OF SECURITIES	Y
1	600519.SH	Moutai, Guizhou	4.702
2	600809.SH	Shanxi Fenjiu	1.423
3	600702.SH	Shede liquor industry	0.708

4	002304.SZ	Yanghe Co., Ltd	0.643
5	000858.SZ	Wuliangye	0.586
6	000568.SZ	Luzhou Laojiao	0.529
7	000799.SZ	Jiugui Liquor	0.341
8	000596.SZ	Gujing Gong wine	0.182
9	200596.SZ	Gujinggong B	0.083
10	603369.SH	Present life margin	0.009
11	600779.SH	Shuijingfang	-0.125
12	603589.SH	Kouzi cellar	-0.352
13	603198.SH	Yingjia tribute wine	-0.384
14	002646.SZ	Highland barley wine	-0.599
15	000860.SZ	Shunxin agriculture	-0.786
16	600197.SH	Elite	-0.811
17	603919.SH	Golden Emblem Wine	-0.860
18	600559.SH	Laobaigan wine	-1.118
19	600199.SH	Golden seed wine	-1.848
20	000995.SZ	Huangtai liquor	-2.326

6. Research Conclusion

6.1. Stock Selection Conclusion

The latest financial data published on flush software can be used to, Investors should actively allocate and flexibly operate Guizhou Maotai (600519.sh) and Shanxi Fenjiu (600809.sh).

Investors should sell Shuijingfang (600779.sh); Kouzijiao (603589.sh); Yingjia tribute wine (603198.sh); Highland barley wine (002646.sz); Shunxin agriculture (000860.sz); Ilite (600197.sh); Jinhui liquor (603919.sh); Laobaigan liquor (600559.sh); Golden seed wine (600199.sh); The 10 stocks of Huangtai liquor (000995.sz).

Stock investors can pay attention to shede liquor (600702.sh); Yanghe Co., Ltd. (002304.sz); Wuliangye (000858.sz); Luzhou Laojiao (000568.sz); Jiugui Liquor (000799.sz); Gujinggong liquor (000596.sz); Gujinggong B (200596.sz); The 8 stocks of Jinshiyuan (603369.sh).

6.2. Conclusion Analysis

From the ranking results, we can find the effectiveness of principal component analysis for stock screening of fundamental analysis: take the stocks with a score of more than 1.4 as the recent asset allocation portfolio. They are Guizhou Maotai (600519.sh) and Shanxi Fenjiu (600809.sh). As we all know, Guizhou Moutai is China's national liquor, which is well-known throughout the country and even the world. Due to the special brewing process and brewing environment, the annual output is quite limited. However, with the improvement of people's living standards, the annual demand of Guizhou Moutai is increasing day by day, the supply exceeds the demand, and the price rises. As a result, the value of Guizhou Moutai per share is quite high, and it has been growing steadily over the years. It is expected that Guizhou Moutai ranks first in the analysis results. Judging from the recent market trend, Guizhou Moutai has declined, but we are still optimistic about the long-term investment value of Guizhou Moutai. If investors can hold shares for a long time, they believe that they will make a lot of profits. As for Shanxi Fenjiu, although the general market of Baijiu stocks has shown a downward trend recently, according to the vote, most shareholders still hold a bullish attitude towards Shanxi

Fenjiu, and can carry out appropriate short-term operation arbitrage. Recently, investors increase their holdings of the above two stocks, which can probably preserve their value and make a small profit. Moreover, if they continue to hold shares, they are expected to have greater returns in the future. The above middle points illustrate the rationality and effectiveness of the principal component analysis method in this paper.

We abandon Shuijingfang with negative scores; Kouzi cellar; Yingjia tribute wine; Highland barley wine; Shunxin agriculture; Elite; Golden Emblem Wine; Laobaigan wine; Golden seed wine; Among the stocks of the 10 Baijiu companies with poor performance and low popularity, some of them were also found to have mixed liquor. It is suggested that stock investors sell these stocks as soon as possible to effectively avoid losses. The above also fully illustrates the rationality and effectiveness of the principal component analysis method.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Q.Ding Research on stock multi factor quantitative investment strategy based on principal component analysis [j] Times finance, 2020 (17): 74-76.

[2] Y.H Liu Research on comprehensive quantitative stock selection strategy based on improved neural network and principal component analysis [d] Zhejiang industrial and Commercial University, 2019.

[3] W. Yu Application of principal component analysis in stock screening of agricultural sector [j] Investment and cooperation, 2011 (5): 167-169.

[4] Y.R Liu Brand status and development strategy of Moutai in Guizhou [j] Management observation, 2019 (14): 49-50.