

Research on the components of CSI300: Perspective from quantitative finance

Yunpeng Wang*

School of economics Lanzhou university

*Corresponding author Email: ypwang18@lzu.edu.cn

Abstract. This paper takes Shanghai and Shenzhen 300 component stocks as the research object. Specifically, this paper selects stock market data, risk data and corporate financial data from March 2020 to March 2022 for quantitative investment research, in order to obtain effective influencing factors to explain excess returns and obtain timeliness and positive investment returns. The empirical process is summarized as follows. Firstly, the factor data obtained from CSMAR are analyzed by time series single factor combination analysis to obtain three effective influencing factors: turnover rate, fluctuation and capital debt ratio. Then, Fama-MacBeth regression is carried out to verify the independent influencing factors and whether there is interaction. The regression results show that the intercept and coefficient are relatively obvious, these three factors have a good explanation effect on the stock excess return.

Keywords: single factor analysis; Fama-Macbeth regression analysis; quantitative investment analysis; turnover rate; change ratio; capital debt ratio.

1. Introduction

High speculation and volatility have always been the prominent characteristics of Chinese stock market. [1] China's stock market has been developing for several years. Major institutional investors or individual investors are trying to find effective investment profit methods. Based on the factor analysis method of quantitative investment, various information reflected by individual tickets are analyzed to obtain an effective investment model to overcome the market. According to the logic of stock value, the higher the value of the parent company is, the higher the stock price will be. The value of the parent company is related to various factors, such as asset-liability ratio, sales cash ratio, asset return rate and asset turnover rate. The stock price fluctuation of the company is related to the risk factors of the company, such as β value, correlation coefficient of risk factors, non-systematic risk and systematic risk. It may also be related to the trading information of the stock market, such as price-earnings ratio, net market rate, market sales rate, turnover rate, circulation market value and so on. The data used in previous studies are mostly 5-7 years ago. The objective of this study is to build a multi-factor investment stock selection model by univariate analysis and regression analysis based on the data of the company 's stock price impact factors. The data are the quarterly stock data of Shanghai and Shenzhen 300 index stocks from June 2020 to March 2022. The objective is to build a more reasonable and timely multi-factor stock selection model to provide investors with an updated stock selection analysis model, so as to provide help for the establishment of the correct portfolio in the current stock market and adapt to the needs of the Chinese stock market.

Globally, quantitative investment has been developed for more than 50 years. In terms of investment strategy, it almost covers the whole process of investors ' investment. It has a very wide range of research in quantitative stock selection, quantitative arbitrage, asset allocation and risk control. In China, many scholars have constructed many different types of quantitative stock selection models by drawing on foreign advanced quantitative models. Wang and Liu [2] used Shanghai and Shenzhen 180 index stocks are taken as the research objects. Based on the regression method, the factor model is quantitatively selected, the portfolio model with relatively stable positive returns is obtained, and the corresponding quantitative timing strategy is set to control investment risk. Lin [3] used python to analyze the financial data of sample stocks and constructed a quantitative portfolio to achieve more than four times the investment income within a decade, which was suitable for stock selection in the industry. Wang [4] proposed an eight-factor model index system by analyzing the

quantitative stock selection models at China and abroad and using the six-factor model and realized a more accurate prediction of stock price rise and fall in 2013 by using random forests. Taking Shanghai and Shenzhen 300 as the stock pool, Lv [5] constructed a quantitative stock selection model that continuously defeated the market, and introduced the vector algorithm to analyze the long-term dominant stocks.

The objective of this study is to find more effective influencing factors in China's stock market in the past two years and select the constituent stocks of China's CSI 300 index for research. Firstly, the research ideas and research methods of this paper are stated. After the data are processed, the single factor combination analysis is carried out in the next part. Then, the Fama-Macbeth regression is carried out to eliminate the influence between factors and obtain effective factors. The last part is the summary and reflection.

2. Research method and data

2.1 Research route

The research idea of this paper is that the volatility information of stocks depends on various market information, risk factor and financial information reflected by stocks themselves. From the perspective of market information, the information reflected by the stock market almost reflects the investors' investment attitude towards the stock. Risk information such as factor (β), market correlation coefficient and stock liquidity also reflect the rising trend of stock prices in the future and the investment of investors. The financial information released by enterprises to the society also has potential messages. The information reflected by investors or institutions will also affect the trend of stock prices. The market information, risk information and financial information of enterprises are listed below.

Table 1. Listing of Stock Impact Factors

Market information	Risk information	Financial information
price-earnings ratio (PE);	Risk factor β ;	Book-to-market ratio;
price-to-sales ratio (PS) [6];	Correlation; coefficient (COR);	Capital-liability ratio
price-to-cash flow ratio (PCF);	Non-systematic risk;	
price-to-book ratio (PB) [7];	Coefficient of determination(R^2)	
turnover rate;		
circulation market value;		
change ratio; stock liquidity		

This paper uses the quarterly data of the Shanghai and Shenzhen 300 index stocks from April 2020 to March 2022, which comes from the CSMAR database. The stocks that are listed less than 6 months are eliminated and the stocks that display extreme values of each factor are also eliminated. The financial data come from the enterprise financial statement database of CSMAR research series. The book market value ratio to asset-liability ratio is calculated by the enterprise financial data such as the number of issued shares, the number of circulating and uncirculated shares, the total debt, the total assets and the total equity.

$$BM = OE / (TS + nTS) * Cp \quad (1)$$

$$DCR = \frac{TL}{TA} \quad (2)$$

In the above equations, BM represents book market value ratio, DCR represents corporate capital liabilities ratio, OE represents owner equity. TS represents the number of tradable shares, nTS

represents the number of untradable shares, C_p represents the closing price. T_a represents total assets and T_l represents total liabilities.

2.2 Data processing and research methods

The analysis method used in this paper is the single variable grouping test and Fama-Macbeth regression. The stock is divided into nine periods according to the quarter, and the information contained in the stock is contained in it. Because some factors do not have quarterly data, for quarterly data such as PE, PS, PCF, PB, turnover rate and liquidity indicators that cannot be obtained, this paper selects the monthly data for the first three months of the time point for time series weighted average. Firstly, descriptive statistics and pre-processing are carried out on the selected factors. According to the correlation test, it is found that the market value of enterprise circulation has a high correlation with the book value ratio, COR and R^2 , risk factor β and non-systematic risk has a high correlation. Therefore, the non-systematic risk and the market value of enterprise circulation are eliminated.

The test method is the single-factor combination test.[8] Firstly, the stocks without data are eliminated, and then the stocks are divided into k groups according to the sample size. The number of samples selected in this paper is 328. After the stock information processing, the stocks are sorted according to the size of the studied factors. The yield data corresponding to the top 30 factors are selected to form the first group of data, and then the yield data corresponding to the last 30 factors are selected to form the second group of data. The average values of $Y_{n,t}$ and $Y_{1,T}$ of the two groups are calculated to obtain Y_{Diff} . According to whether the obtained time series values are significantly different from 0, one can determine whether there is explanatory power.

The formula is as follows.

$$\bar{Y}_{i,t} = \frac{\sum_{i \in k,t} W_{i,t} Y_{i,t}}{\sum_{i \in k,t} W_{i,t}} \quad k \in \{1, 2, \dots, np - 1\} \quad (3)$$

$$\overline{Y_{Diff}} = \overline{Y_{np,t}} - \overline{Y_{1,t}} \quad (4)$$

Calculation of mean value of time series:

$$\overline{y_k} = \frac{\sum_1^T \overline{Y_{k,t}}}{T} \quad (5)$$

$$\overline{Y_{diff}} = \frac{\sum_1^T \overline{Y_{Diff,t}}}{T} \quad (6)$$

The Fama-Macbeth regression is carried out on the stocks that have explanatory power for the excess return of stocks and are obviously indigenous. [9] The factor data of the first quarter and the return date of the next quarter are regressed, and the intercept term μ_i and the correlation coefficient σ_i are obtained. The time series μ_i and σ_i are obtained and the single sample T-test is carried out. Then all the effective values and excess returns are regressed and μ and σ values are recorded. Then determine the independent explanatory power of effective factors according to the regression results.

3. Empirical results analysis

3.1 Single factor combination analysis

In the analysis, PS, PE, PCF and PB are first tested by single factor grouping, and the return rate is the sum of three-month excess return rates. PE and PB have small missing values in some samples. PCF has a large number of missing values, and the number of samples is only 200, so the sample size of each group is adjusted to 20. Group1 represents the group average with the smallest ranking result,

and Group10 represents the group average with the largest ranking result. The excess return is multiplied by 100 to facilitate the result analysis.

$$\text{Return} = \text{monthly ratio} * 100 \quad (7)$$

Return is the data used for data analysis, and monthly return is the original data. The analysis results are as follows.

Table 2. PS, PE, PCF and PB single factor combination analysis results

Title	Group1	Group10	G10-G1 (diff)
PE	0.1957	1.0869	0.8911
T test	0.908	3.5189	0.0079
P value	(0.3904)	(0.0079)	(0.0616)
PS	0.3229	0.8302	0.5073
T test	1.5502	2.8537	1.9381
P value	(0.1597)	(0.0214)	(0.0886)
PCF	0.0079	0.9454	0.9375
T test	0.0265	3.2861	1.8567
P value	(0.9795)	(0.0111)	(0.1004)
PB	0.0220	0.9986	0.9765
T test	0.0935	2.4602	1.7921
P value	(0.9278)	(0.0393)	(0.1109)

The analysis results of these four factors are not very ideal, and the t value of the first group of PE, PCF and PB is very small, with almost no difference from 0. The T value obtained by subtracting the maximum group from the minimum group is not large enough. Therefore, these four groups of factors have no ability to explain the yield rate.

Next, the turnover rate, change ratio, liquidity index and risk factor β are verified. Since both the change ratio and the risk factor β are from the data of CSMAR, the data are not processed with multiple deviations. The analysis results are as follows.

Table 3. Results of turnover rate, change ratio, liquidity and β single factor combination analysis

Title	Group1	Group10	G10-G1 (diff)
Turnover	0.0982	1.8728	1.7745
T test	0.7623	3.3183	3.1744
P value	(0.4678)	(0.0106)	(0.0131)
change ratio	-2.641	5.2599	7.9012
T test	-7.9593	10.2502	14.9185
P value	(0)	(0)	(0)
liquidity	0.4034	0.2979	-0.1055
T test	1.6926	1.0752	-0.2567
P value	(0.129)	(0.3136)	(0.8039)
β	0.0779	1.8869	0.9075
T test	0.6293	1.6738	1.3608
P value	(0.5467)	(0.1327)	(0.2107)

The results of the above four factors show that the T value of the maximum group minus the minimum time series of liquidity and β is not obvious, and the diff value of liquidity is obviously negative. Therefore, these two factors cannot explain the stock return rate. In the turnover rate and

the change ratio index, it can be seen that the T value of the maximum group minus the least group time series of these two factors is very obvious, indicating that these two factors have explanatory power for the return rate.

Finally, the correlation coefficient, R^2 , Book-Market value ratio and Debt to capital ratio analysis, the results are as follows.

Table 4. COR, R^2 , BM and DCR single factor combination analysis

Title	Group1	Group10	G10-G1 (diff)
COR	0.4342	0.4011	-0.0331
T test	1.919	1.704	-0.1405
P value	(0.0913)	(0.1268)	(0.8918)
R^2	0.4342	0.3981	-0.0361
T test	1.9189	1.6985	-0.1534
P value	(0.0912)	(0.1278)	(0.8819)
BM	0.8689	0.1903	-0.6786
T test	2.2535	0.8333	-1.4271
P value	(0.0543)	(0.4289)	(0.1914)
DCR	0.5396	0.0972	-0.4423
T test	3.1066	0.5278	-2.5626
P value	(0.0145)	(0.612)	(0.0335)

The above four factor analysis results show that the diff results of the correlation coefficient, R^2 and the book-to-market ratio factor are not significantly different from 0, and they have weak explanatory power for the return rate. The T value of the Debt to capital ratio is large, and they have strong explanatory power for the return rate. So far in the initial factors to be tested there are three factors to explain the rate of return: turnover, change ratio and Debt to capital ratio.

3.2 Fama-Macbeth regression

After obtaining the initial explanatory factors, in order to further test whether the predictive ability is affected by each other and verify the predictive ability of specific factors on stock excess returns, Fama-Macbeth regression is used to control other cross-sectional information to affect returns. Returns in the next quarter were used as explanatory variables, including turnover, change ratio and DCR. Table 5 reports the summary of regression results of each period. The average values of coefficients are μ_t , σ_1 , σ_2 and σ_3 .

$$R_{i,t+1} = \mu_{0,t} + \sigma_{1,t}Turnover_{i,t} + \varepsilon_{i,t} \quad (8)$$

$$R_{i,t+1} = \mu_{0,t} + \sigma_{2,t}Changeratio_{i,t} + \varepsilon_{i,t} \quad (9)$$

$$R_{i,t+1} = \mu_{0,t} + \sigma_{3,t}Dcr_{i,t} + \varepsilon_{i,t} \quad (10)$$

$$R_{i,t+1} = \mu_{0,t} + \sigma_{1,t}Turnover_{i,t} + \sigma_{2,t}Changeratio_{i,t} + \sigma_{3,t}Dcr_{i,t} + \varepsilon_{i,t} \quad (11)$$

Table 5. summary of Fama-Macbeth regression results of turnover, change ratio and DCR

Coefficient		(8)	(9)	(10)	(11)
μ_t	mean	0.3185	1.2101	-0.2489	1.3319
	SD	0.1115	0.0013	0.0262	0.2819
	T value	0.2697	0.9617	-0.270	0.7095
	P value	(0.0251)	(0.0368)	(0.0193)	(0.0301)
σ_1	mean	0.4680			0.3720
	SD	0.4403			0.2667
	T value	1.4365			1.4418
	P value	(0.0010)			(0.0378)
σ_2	mean		-0.7758		-0.7448
	SD		1.7734		1.5519
	T value		-1.6479		-1.6910
	P value		(0.014)		(0.0135)
σ_3	mean			0.2594	-0.0019
	SD			0.4413	0.0158
	T value			1.6620	-0.1385
	P value			(0.0080)	(0.0090)
R^2		0.0125	0.0212	0.0045	0.04436
adjR ²		0.0093	0.0181	0.0013	0.02593
n		314.5	314.5	314.5	314.5

There is a cross-sectional relationship between the meaning independent variables Turnover, Change Ratio, and Dcr represented by σ_i and the excess return rate of the dependent variable. When there is more than one independent variable in the regression test, the statistical significance indicates that after controlling the effect of other independent variables, it still has a relationship with the dependent variable. Usually, the T statistic of the sequence value needs to be obvious at the 5 % statistical confidence level, that is, the p statistic is less than 0.05.

In the regression results of the three factors, the average intercept (μ_0) was 1.33, the T statistic was 0.709, the average σ_1 of turnover rate was 0.372, the standard deviation was 0.266, T statistic was 1.44, and the results were very obvious. The average value of σ_2 of the change ratio is - 0.744, the standard deviation is 1.55, the T statistic is - 1.69. The average value of σ_3 of the capital-liability ratio is - 0.00194, the standard deviation is 0.0158, the T statistic is - 0.1385. The R square of the interface regression and the adjustment of R direction are 0.044 and 0.0259, and the number of observations is 314.5.

The regression results are all significant, which can reasonably explain the cross-sectional relationship between turnover rate, Change ratio, capital-liability ratio and the return rate in the next quarter. The significant indigeneity of the intercept term also proves that the prediction ability of the other two variables on excess return is still significant when any of the three variables is controlled.

4. Conclusion

This study finds that the influencing factors of three effective image stocks' excess returns show relatively strong explanatory power from mid-2020 to March 2022. In the Fama-Macbeth regression, the intercept term and the coefficient term are significantly indigenous, showing a strong correlation. It can be seen that the Change ratio, Turnover rate and Asset-liability ratio of enterprises can all be related to the stock return of enterprises, and the correlation is strong in the quarterly range. In the period after 2020, except for ST shares, GEM and financial stocks, these factors are effective factors in China' s A-share market. It has obvious reference significance for investors in future investment operations.

The lack of this study may be reflected in three aspects: First, in terms of data, the financial data obtained is less, mostly related to the size and market value of the enterprise, not the earnings per share, net value per share, Acid-test Ratio and other details of financial factors. And in the selection of time period only picked nearly two years of data, time span is less, in the regression analysis of T test may be there is no obvious situation. The second point is that the analysis method of bivariate combination analysis is not selected in the data analysis process, and the Fama-Macbeth regression alone cannot fully explain the impact of the interaction between the two variables on the yield rate. Finally, it is found that the effect of the change ratio in the single factor analysis is positive, while the coefficient in the second regression is negative, which may be due to the large span of the analysis interval according to the quarter, which is susceptible to the reversal effect. With the concept of quantitative investment gradually rising around the world, many financial industry practitioners pay more and more attention to quantitative investment. Research on quantitative investment strategy and investment income has gradually become a hot issue of financial practitioners.[10] Many influence factors related to research is also worth learning, in the future work will continue to improve my research.

References

- [1] Yin Libo, Wei Ya, Does stock market volatility encourage market speculation? [j] Management Review, 2021,33(12): 15-29.DOI: 10.14120/j.cnki.cn11-5057/f.2021.12.013. -54
- [2] Wang Chunli, Liu Guang, Wang Qi, Multi - factor Quantitative and Timing Strategy [J]. Journal of Dongbei University of Finance and Economics. 2018(05):81-87.DOI:10.19653/j.cnki.dbcjdxxb.2018.05.011.
- [3] Lin Yudong, Analysis of financial index selection and investment strategy based on Python - - Taking China ' s stock market from 2009 to 2019 as an example [J]. Shang xun ISSN: 1671-6728, 2020(0):1-2.
- [4] Wang Shuyan, Cao Zhengfeng, Chen Mingzhi. Application of random forest in quantitative stock selection [J]. Operation Research and Management science ,2016,25(03):163-168+177.
- [5] Lv Kaichen, Yan Hongfei, Chen Chong, Quantitative Investment Strategy Based on Shanghai and Shenzhen 300 Shares [J]. Journal of Guangxi Normal University (Natural Science Edition), 2019,37(01):1-12. DOI: 10.16088/j.issn.1001-6600.2019.01.001.
- [6] DU Huifen, Ping Shitao, Empirical analysis on the use of series of market sales rate indicators in the A-share market [J]. Journal of Central University of Finance and Economics. 2005(10): 35-38.
- [7] Ji Mingfa, Peng Haodong, Comparison of China' s A-share market value investment strategies — an empirical study based on P / E, P / B, P / S and M / G stock selection indicators [J]. Commercial Times. 2013(13):52-54.
- [8] Hu Yi, Gu Ming, Buffett ' s Alpha: An Empirical Study from Chinese Stock Market [J]. Journal of Management World, 201/,34(08):41-54+191. DOI: 10.19744/j.cnki.11-1235/f.2018.08.005.
- [9] Liang Xiaoying, Research on quantitative stock selection method based on multi-factor model [J]. China Market,2021(25): 31-32.DOI: 10.13939/j.cnki.zgsc.2021. 25.031
- [10] Sun Jiao, Multi - factor Quantitative Investment Strategy and Empirical Test [D]. Nanjing University 2016