Application of Multi-Factor Quantitative Stock Selection Strategy Based on Scoring Method: Evidence from the CSI 300 Component Stocks

Zikun Chen[1*, †], Min Hong[2†], Ruya Lin[3†],
[*] Corresponding author: zikun.chen@student.manchester.ac.uk

[1]Department of Mathematics, University of Manchester, Manchester, United Kingdom

[2]Department of Finance, Shanxi University of Finance and Economics Taiyuan, China

[3]Department of Finance, Zhejiang University of Finance & Economics, Hangzhou, China

† These authors contribute equally.

Abstract—With the wide application of computers and the continuous improvement of big data processing methods, the methods and theories of quantitative investment can better and faster adapt to the rapid development of the current financial market, and this method is gradually becoming one of the main means of financial investment analysis. This article selects CSI 300 constituent stocks as sample stocks, intercepts financial and market data from 2012 to 2021, establishes an effective factor database, and tests the effectiveness of each factor. Finally, a more robust factor affecting stock returns is obtained. This paper evaluates the performance of the quarterly returns and cumulative returns obtained by grouping stocks by the selected factors and obtains the validity of the established model. Through the empirical research of this paper, the main conclusions are as follows: through the analysis of multiple factors, a compound factor can obtain excess returns relative to the benchmark, there is a monotonic relationship between the returns of these combinations, and the return of the combination with the highest score The portfolio with the highest score has the highest rate of return, and the portfolio with the lowest score has the lowest rate of return. The results of the t-test between the portfolio with the highest score and the benchmark show significant differences. This article can select effective factors for investors that are undervalued in a certain period and can obtain investment returns and reflect the investment value of listed companies.

Keywords- quantitative investment; quantitative stock selection; scoring method; multi-factor model

1 INTRODUCTION

With the development of computer technology and the advent of the era of artificial intelligence, computer technology is more and more applied in the economic field because of its advantages of fast computing speed, convenient operation, and efficient processing instead of a manual algorithm. Quantitative investment is the process of practicing the investment concept and using computer technology and mathematical model to realize the investment strategy. Multi-factor stock selection is one of the most widely used and simple and intuitive models of quantitative investment. This model can judge the value of the stock market and the

FFIT 2022, October 28-30, Shenzhen, People's Republic of China

DOI 10.4108/eai.28-10-2022.2328426

stocks of listed companies by integrating the influence of various factors. In addition, multi-factor stock selection has a strong comprehensive type, which can combine valuation factors, technical data, and financial data, establish different models according to investors 'preferences, and then guide investors' portfolio selection. The multi-factor stock selection model largely promotes the development of quantitative investment in China.

The scoring method is to score the stocks according to the size of each factor and then weight the total score according to a certain weight, and finally screen the stocks according to the total score. For the evaluation of the multi-factor model, the stock combination yield measured back by the scoring method can evaluate the merits of the alternative stock selection model. The advantage of the scoring method is that it is relatively robust and not susceptible to extreme values. This paper hopes that through the effectiveness of candidate factor, by establishing an effective model selected effective factors, using the scoring method to select the candidate factor in the market performance is good, good stock price forecast, and can choose investors in a certain period is undervalued and can obtain investment income and reflect the investment value of listed companies.

Benjamin Graham and David Dodder on the classic edition of securities analysis was published in 1934, the stock portfolio management and trading strategy gained the development of Graham and Dodder by using the information in the financial statements and determining the basic principles, and expanding the traditional pricing method, in the early stage of stock quantitative model development made outstanding contributions [1]. Fama and French made the portfolio classification of value stocks and growth stocks based on book value ratio and earnings yield as indicators, and value stocks received positive excess returns relative to growth stocks [2]. Partha S. Mohanram took the top 20% stocks with the highest price-to-book ratio (PB index) as research samples and selected nine influencing factors to score individual stocks from three aspects of profitability, robustness of financial indicators and stability of growth, so as to construct a portfolio with good yield effect [3]. In China for the field of multi-factor quantitative investment research, Liu Yi respectively from growth, valuation, financial quality, and momentum four dimensions of stock instability, by testing the effectiveness and stability of 25 factors, selected the operating profit growth rate, cash flow, net assets, yield growth rate (ROE), return on total assets (ROA) rate, price to book (PB) eight indicators build stock selection model, and proved that the model has good performance in the market [4]. Tian Lihui found that China's stock market has prominent risks and is more sensitive to market risks than the United States. Compared with other factors, the scale factor has a stronger ability to explain the return. The two-factor model composed of market factor and scale factor is more suitable for Chinese stock market [5]. Alethea Rea, William Rea and Libin Yang used principal component analysis to maximize the diversification of the portfolio to obtain obvious benefits [6]. Wang Rui (2016) from many factors affecting stock volatility selected 17 simple candidate factors, through the effectiveness and redundancy test using the Z score method of empirical analysis of the portfolio, and through the cumulative return, excess annualized compound yield, sharp ratio, and the probability of positive returns on the evaluation of the model [7]. Li Bin and Feng Jiajie constructed quality factors (QMJ) from four aspects, including profit, growth, safety, and dividend distribution, taking all A-shares from January 1999 to December 2015. The empirical results show that adding QMJ to different pricing models improves the explanatory ability of each model to varying degrees [8]. Lu Kaichen builds a quantitative stock selection model that can beat the market with a continuous stock selection model over the market. The first step is to start from the fundamentals and select 50 long-term dominant stocks through the multi-factor scoring model. The corresponding listed companies are in good operating conditions and have certain investment value, but they may be affected by market shocks in the short term, and may not rise within a week. In the second step, the introduction of a support vector classification algorithm to conduct technical analysis of long-term dominant stocks, from which we selected the 10 dominant selected stocks with the biggest rising probability this week to buy [9]. Zhao Liping took all A-share enterprises from 2011 to 2017 as the research object, evaluated the value and growth of stocks according to PE and ROE growth rate indicators, and constructed value growth stock portfolio, value growth stock portfolio and growth stock portfolio respectively. The results show that the value growth portfolio has higher returns than the market during the model testing period, and it has higher return stability than the other two portfolios [10].

The rest of the paper is organized as follows: The second part is the research design, introduces the data sources and research ideas, presents the empirical results, conducts the factor effectiveness test and the redundancy elimination, builds the portfolio and studies the performance, and fourth, summarizes the main conclusions and analyzes the conclusions. This paper also discusses the existing defects and the problems needed for further study.

2 RESEARCH DESIGN

2.1 Data source

All samples of the CSI 300 index in December 2021 are taken as the sample period, with 10 years from 20122021. From 2012 to 20 1 9, the sample selection period from 2020 to the stock selection strategy, including rising, fall, and shock trends in the whole model period. The purpose is to test whether the performance of the model under different trends can achieve the expected effect. The sample of selected stocks is all normally traded A-shares that have been listed for more than one quarter, and the performance benchmark refers to the CSI 300 index. Candidate factor mainly includes financial data and market data, data from the database, wind financial terminal (in order to maintain data consistency, candidate factor sample stock prices are using quarterly data), this paper from the value factor, growth factor, quality factors, technical factors selected 13 factors as candidate factors, as shown in the following table.

Table 1 List of Candidate Factors

Valuation factor	Growth factor	quality factor	Technical factor
pe ratio	Return on equity	asset-liability ratio	turnover rate
Price-to-book ratio	The net interest rate on operating income	fixed assets ratio	Total market value
dividend rate	earnings per share	Shareholder equity turnover rate	
Operating profit growth rate		fixed asset turnover	

The variable symbol, calculation method, and related description of each candidate factor are shown in the following table.

Table 2 The selected variable and its description

Variable symbol	The variable name	Calculation method
P/B	Price to book ratio	Stock market value/net assets.Net assets are the net assets published in the latest periodic report.
P/E	Price-earnings ratio	Stock market value/net profit of the last four quarters.
Dividend Yield	The dividend yield	Total company dividend/stock market value.
EPS	Earnings per share	Net profit/latest total share capital attributable to the owner of the parent company.
ROE	Return on equity	Directly take the data disclosed in the periodic report of the company; If there is no disclosure value, ROE (deduction, weighting) =P/(E0+NP÷2+Ei×Mi÷M0-Ej×Mj÷M0) ×100%, where: P for the reporting period after deducting non-recurring profit and loss of profits, NP for net income during the reporting period, E0 net assets at the beginning of an Ei for the issuance of new shares or bonds convertible during the reporting period, etc. New net worth, Ej to repurchase or cash dividends to reduce net assets during the reporting period, M0 for during the month number, Mi next January for new net assets to the final copies of month during the reporting period, Monthly amount for Mi to reduce net assets from the following January to the end of the reporting period.
FulTurnR	Turnover rate	The frequency with which a stock changes hands during a holding period.
Market Capitalization	Total Market capitalisation	Share price x total share capital.
E/P	Earnings to price	Total net profits/stock market value ×100%.
Dbastrt	Asset-liability ratio	Total Liabilities/Total Assets x 100%.
Fixassrat	Turnover of fixed assets	Total operating income ×2/ (Total fixed assets at the beginning of the period + Total fixed assets at the end of the period).
Equrat	Equity turnover	Total operating income ×2/(Net assets at the beginning of the period + net assets at the end of the period).
Fixassrt	Fixed assets ratio	Net fixed assets/Total assets x 100%.
NprTOR	The net profit margin on operating income	Net profit/total operating revenue ×100%.

In this paper, the samples with serious data missing are removed and no special treatment is made for ST shares. According to the method of data mining theory to deal with some data, this paper supplements the method of weighted average before and after the use of some missing data. Due to the lack of stock data, it will not have a significant impact on the empirical results compared with large sample data. In particular, the factor index data of the previous period of this paper is used to represent the basis of stock price fluctuations in the current period to eliminate the influence of hysteresis and make the model more reasonable

2.2 Research idea

In this paper, the multi-factor scoring model is used as the research method, which includes four steps: validation of candidate factors, elimination of valid but redundant factors, construction of a comprehensive scoring model, and model verification.

(i) Validation of factors. Build a library of index factors, the current quarter stock pool of a single factor in ascending order, and all the stock can be divided into 5 groups, according to the group from high to low respectively for 5 ~ 1 minute, the stock of each group following the circulation of its stock market value-weighted combination for each group of earnings, returns, and score correlation coefficient; judging stock yields and index of correlation, and re-grade stocks based on this feature;

(ii) Redundancy factor check. Each quarterly factor scores the stock to get a score column, calculates the correlation coefficient of all factors scores, and calculates the mean of the quarter to get the effective factor;

(iii) Multi-factor model construction. Compound factors were obtained by summing the scores of effective factors according to a certain weight, sorting the composite factors of each quarter, and dividing the stocks into five groups according to the descending order of the comprehensive score. The quarterly stock return rate was calculated according to the weight of market value in circulation;

(iv) Model testing. The period from 2019 to 2021 is taken as the model inspection period to calculate the average return rate of each group of stocks during this period and compare it with the cumulative return rate of the CSI 300 index in the same period.

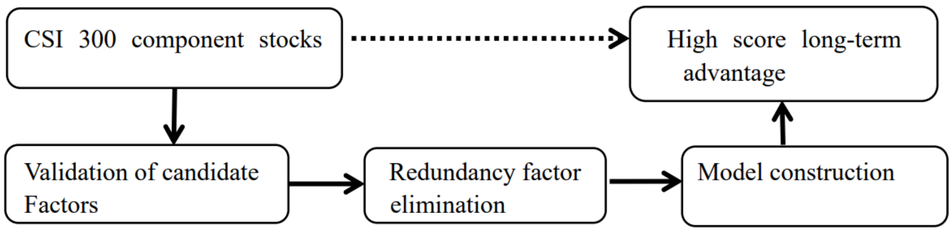

Figure 1 Multi-factor stock selection process based on the scoring method

3 EMPIRICAL RESULTS

3.1 Validation of candidate facters

Annualized compound return is calculated by annual compound interest, which standardizes the return of the best-performing stock portfolio under various factors over the entire model period. Score correlation the correlation coefficient between the compound average return (RETEXP) of each factor combination and the score sequence (ATS) was calculated. The index with the best performance is Market capitalization, with the highest annual compound return of 28.15%, while the composite average annual return of the CSI 300 index in the same period is 9.8%, and the correlation between its score and earnings is -0.993, so its change can well explain the direction of earnings change. The next best performer is EPS, with a correlation coefficient of 0. 658. The worst performance is Equrat. The correlation coefficient between score and earnings is only 0.15, so its change has little impact on stock price fluctuations and is not representative. Similarly, there are NprTOR, Dbastrt, and Fixassrt, so these indicators cannot be selected as effective factors. Accordingly, this paper takes compound return, score, and correlation coefficient of return as the basis for index selection, and the tested effective factors are E/P, P/E, P/B, Dividend yield, ROE, Fixassrat, Market Capitalisation, FulTurnR, and EPS, a total of 9 effective factors.

Table 3 Candidate factor validity test results

Factor	Annualized compound return	Excess return rate	The score is correlated with earnings
P/B	13.01	3.21	0.965
Dividend yield	19.36	9.56	-0.97
FulTurnR	15.64	5.84	0.819
P/E	16.54	6.74	-0.907
EPS	32.7	22.9	0.658
ROE	9.87	0.07	-0.633
Market Capitalisation	36.15	26.35	-0.833
Dbastrt	7.79	-2.01	-0.537
Fixassrt	4.99	-4.81	0.402
Equrat	2.43	-7.37	0.15
Fixassrat	6.84	-2.96	0.695
NprTOR	10.87	1.07	-0.349
E/P	13.66	3.86	0.846

3.2 Redundancy factor elimination

As can be seen from Table 4, P/E, P/B, and Market Capitalisation are the best to distinguish, with low correlation with other indicators, while E/P, ROE, FulTurnR, and Dividend yield are highly correlated. As E/P is derived from P/E, under the condition of selecting P/E, E/P will not be considered. After retaining the factor with the best performance in profitability, the remaining redundant factors will be removed to obtain 5 effective factors with low correlation: P/E, P/B, Market Capitalisation, EPS, and ROE. This selection is a composite model of valuation, growth, and technical style factors.

Table 4 Factor correlation coefficient matrix

	EPS	ROE	P/E	P/B	Dividend yield	FulTurnR	Market Capitalisation	E/P	Fixassrat
EPS	1								
ROE	-0.122	1							
P/E	-0.396	-0.235	1						
P/B	0.354	-0.066	0.426	1					
Dividend yield	-0.172	0.736	0.383	0.402	1				

FulTurnR	0.35	0.892	-0.127	0.137	-0.376	1			
Market Capitalisation	-0.086	0.004	-0.043	0.048	-0.862	-0.186	1		
E/P	0.216	0.854	0.094	0.273	0.864	0.894	-0.012	1	
Fixassrat	0.863	0.115	0.213	0.026	-0.037	0.097	-0.009	0.312	1

3.3 Portfolio construction and stock selection

First, score each stock that is normally traded in this period according to the scoring method, and remember that the score of factor K on stock i is ; Second, according to the formula (where n is the number of sample stocks), the comprehensive score of stock i is obtained; Third, according to the comprehensive score of stock i, all sample stocks are arranged and recombined in descending order, which is divided into five groups to reconstruct the portfolio; Fourth, according to the divided portfolio, invest the principal in the portfolio every quarter, calculate its yield rate in the holding period, and calculate the yield rate of CSI 300 index in the same period; Fifth, five effective factors with low similarity are obtained by simulating historical data. In order to test the accuracy of the model, data from the first quarter of 2020 to the fourth quarter of 2021 are used as the model test period. During the inspection period, the holding period is also calculated as the first quarter to construct the portfolio; Sixth, this paper use equal weight to weight the factor score, divide the portfolio into five categories, use the circulation market value weighting method to calculate the portfolio return, and compare with the return rate of CSI 300 index in the same period.

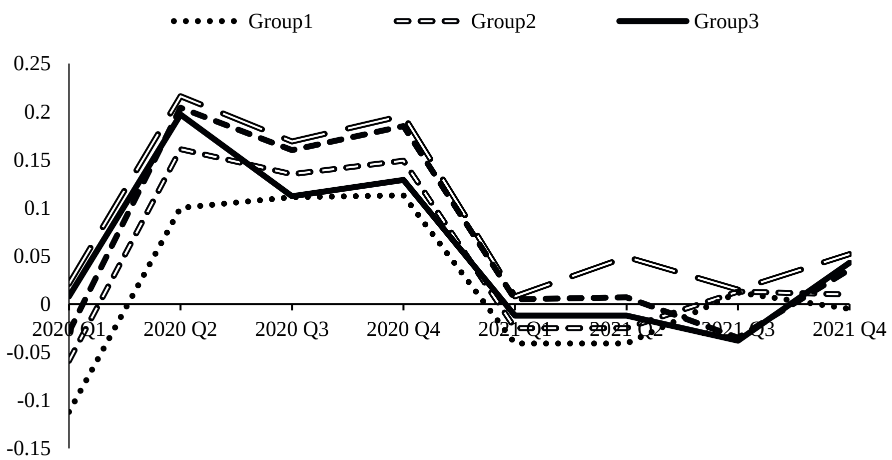

Figure 2 Trend of the average quarterly yield curve of multi-factor grouping

The lowest curve in Figure 2 is the quarterly yield curve of the lowest-scoring portfolio after grouping, and the fifth group is at the top. It can be seen from the initial model to the first quarter of 2021, that the fifth group maintained the highest yield, other groups of yield data in turn decreased. However, in the last three-quarters of the model, there is a small difference between the five portfolio returns. This may be because the equal-weight weighting method is used in this paper to calculate the compound factors, and at different stages, the effect of the factor may be different. As a result, the highest-scoring portfolio continued to outperform the others most of the time. This method can effectively group stocks, but it will still produce similar returns. In order to better illustrate the effect of the model, the fifth group with the highest score and the first group with the lowest score were compared with the benchmark CSI 300 index selected in this paper.

If the fifth group with the highest score is selected as the portfolio, the returns outperform the index, as shown in Figure 3. The statistics and summary of the above table are obtained in Table 5, which reports the quarterly rate of return of each group.

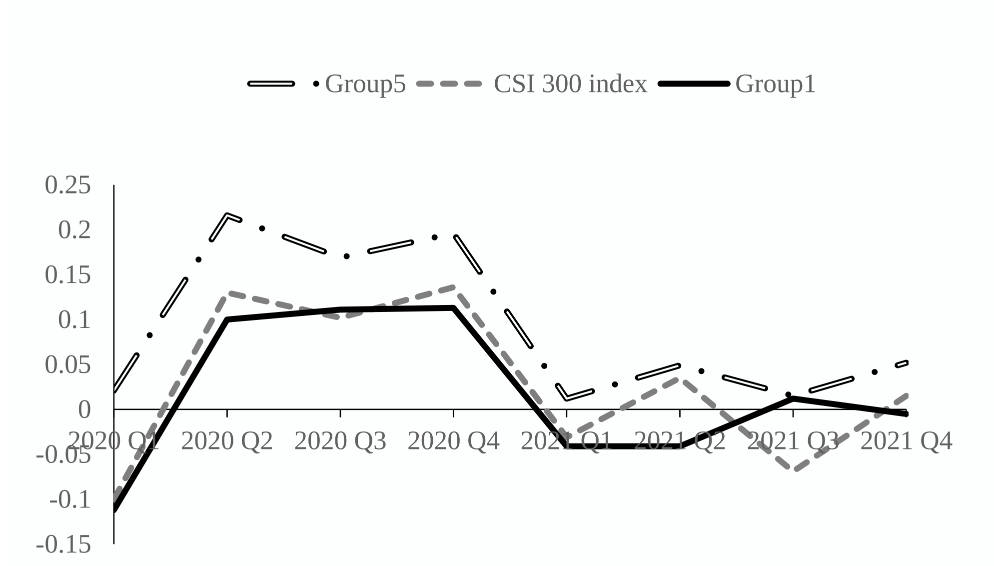

Figure 3 Comparison of high and low score portfolios with the benchmark yield curve

Table 5 Factor correlation coefficient matrix

	G1	G2	G3	G4	G5
Q1 2020	-0.112	-0.059	0.008	-0.029	0.021
Q2 2020	0.1	0.161	0.197	0.204	0.216
Q3 2020	0.111	0.135	0.112	0.16	0.169
Q4 2020	0.113	0.149	0.129	0.185	0.196
Q1 2021	-0.041	-0.025	-0.012	0.005	0.008
Q2 2021	-0.041	-0.025	-0.012	0.007	0.049
Q3 2021	0.012	0.013	-0.038	-0.035	0.015
Q4 2021	-0.005	0.01	0.043	0.036	0.052

It can be seen from Table 6 that the quarterly returns of groups 1 to 4 showed negative quarterly returns within two years, indicating that the quarterly returns were not stable enough. Only the fifth group maintained all positive returns.

Table 6 calculates the mean, standard deviation, and variance of the quarterly returns of the five groups to calculate the maximum return, maximum loss, and Sharpe ratio:

Table 6 Statistical analysis of multi-factor stock selection portfolio

	G1	G2	G3	G4	G5
Annualized rate of return	0.0694	0.1818	0.2163	0.2699	0.3677
mean	0.0171	0.0449	0.0534	0.0666	0.0908
standard deviation	0.0835	0.0888	0.0835	0.0994	0.0912
variance	0.0070	0.0079	0.0070	0.0099	0.0083
maximum return	0.1130	0.1610	0.1970	0.2040	0.2360
maximum loss	-0.1120	-0.0590	-0.0380	-0.0350	0.0120
Sharpe ratio	0.0434	0.3532	0.4773	0.5346	0.8467

Note: The portfolios divided by using indicators are constructed in the first quarter. Therefore, the data in the table are quarterly. It is assumed that the portfolio with the highest composite score constructed based on the portfolio classification method is zero differential force of quarterly return of CSI 300 index.

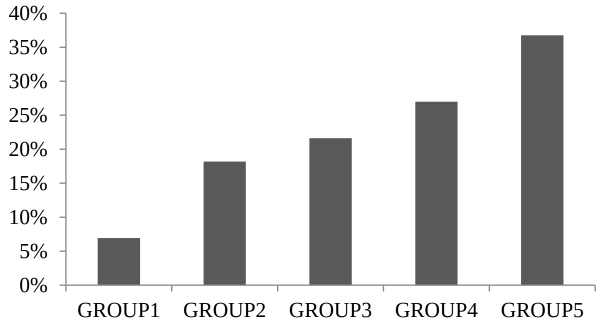

Figure 4 Annualized yield of the portfolio

Figure 4 shows the annualized return rates of five portfolios. There was an increasing trend from the first group to the fifth. The data in Figure 4 are summarized and compared with the annualized coincidence return rate of the CSI 300 index in the same period to obtain Table 7:

Table 7 Portfolio comparison with HS300 Index Annualized Compound Returns

	The annualized compound rate of return	CSI 300 Index annualized compound yield in the same period
G1	0.0694	0.1101
G2	0.1818	
G3	0.2163	
G4	0.2699	
GROUP5	0.3677	

It can be seen from the above table that the portfolio with the highest score (G5) has an annualized compound return of 36.77%. The lowest-scoring portfolio had a compound annualized return of 6.94%. The benchmark CSI 300 index has posted annualized gains of 9.8% over the same period. High-scoring combinations beat the index.

T-test results of high-score combination and CSI 300 index are as follows:

Table 8 T-test results of high-score combination and benchmarks

	G5	CSI 300 index
Mean	0.051	0.027
variance	0.632	0.104
Observations	24	24
T Statistic	4.538***	

Note: *, ** and *** represent significance levels of 10%, 5% and 1% respectively.

The results in Table 8 show that it can beat the index returns at a significance level of 1% when the fifth portfolio is the buy portfolio.

4 CONCLUSION

The CSI 300 index is A good indicator of the overall performance of the A-share market. Its constituent stocks usually have good fundamentals, which can well explain the relationship between various factors and investment returns under good operating conditions. This has strong applicability to the multi-factor stock selection model based on fundamental analysis.

It can be seen from the model results that the P/E ratio is negatively correlated with stock price fluctuations. A low P/E also indicates that the stock is undervalued. This makes it possible for investors to profit. Generally speaking, the better the business performance, the higher the net assets. When the stock price is low, the lower the relative price-to-book ratio, the higher the investment value. From the results, the inverse ratio between the price-book ratio and investment income has been well verified. The inverse ratio of roe to earnings indicates that stocks are undervalued during this period. When the market fluctuates, share prices rise, which provides the possibility for investors to profit. Total market capitalization is negatively correlated with return rate. The smaller the total market value is, the higher the return rate is, indicating that stocks with low total market value may be undervalued for a certain period in the CSI 300 index. It will go up for some time to come. As a result, investors can buy undervalued stocks to make a profit on their investments.

The model established in this paper screened out 5 effective factors, including the scoring method, using the equal weight weighting method. The construction of the portfolio selected by the comprehensive score achieved good returns, which beat the CSI 300 index of the same period and obtained relatively more excess returns. This can also obtain positive excess returns during the bear market, so the validity of the model is verified.

Multi-factor stock selection model is one of the most common and widely used fundamental analysis models in the quantitative investment stock selection model. In the process of investment practice, investors process and analyze massive historical data of listed companies through a series of mathematical formula transformations. It looks for undervalued stocks.

Through the analysis of multiple factors, a compound factor method can obtain excess returns relative to the benchmark. There is a monotonous relationship between the returns on these portfolios. The portfolio with the highest score had the highest return and the portfolio with the lowest score had the lowest return.

This paper uses quantitative means to verify that excess returns can be obtained through the analysis of financial statements. However, the design of this study can be improved. First of all, the selection of factors is not comprehensive enough, which may lead to the absence of factors with better market performance. This processing of missing data could remove some of the stocks that have performed well. Secondly, the new method is improved and verified to make it more reasonable when carrying out redundancy factor elimination. Finally, the application in the market also needs to continue the later research, such as quantitative research on the number of investment strategy selection, stock timing research.

REFERENCES

[1] B Graham and D Dodd. Security Analysis. Me Graw-Hill, 1934.

[2] EF Fama and KR French. “Value versus Growth: The international evidence.” University of Chicago,working paper, 1997.

[3] PS Mohanram. "Separating Winners from Losers among LowBook-to-Market Stocks using Financial Statement Analysis. " Review of Accounting Studies, vol.10, no.2, 2004, pp.133-170.

[4] Y Liu. “Empirical Study of the Sub-Stock Selection Model in the Chinese Market.” Fudan University, working paper, 2012.

[5] L Tian, G Wang, W Zhang. "Three-factor model Pricing: How does China differ from the US? " International Financial Research, vol.2014, no. 07, pp. 37-45.

[6] W Rea, A Rea, L Yang., "Stock Selection with Principal Component Analysis." Working Papers in Economics, 2015.

[7] R Wang. “Multi-factor quantitative stock selection research in the A-share market.” Shanxi University of Finance and Economics, working paper, 2016.

[8] B Li, and J Feng. “Research on firm quality factors in China's stock market.” Management Review, vol.31, no. 03, pp. 14-26

[9] K Lu, et al. “Research on Quantitative Investment Strategy Based on Shanghai-Shenzhen 300 Stocks.” Journal of Guangxi Normal University (Natural Science Edition), vol. 37, no. 01, 2019, pp. 1–12.

[10] L Zhao. "Research on the applicability of China's a-share market value growth strategy." Tax payment, vol.14, no.03, 2020, pp.153-155.