An Empirical Study on Quantitative Investment Stock Selection Strategy Based on Fundamental Factors and Coefficients Combined with Entropy Method

Lu Nan

lunan0120@163.com

School of economy, Anhui University, Huainan city, Anhui province, China

Abstract—In order to investigate the impact of fundamental factors on stock returns, this article selects earnings price ratio, book-to-market value ratio, price-to-market ratio, market-to-sales ratio, market capitalization, net sales margin, return on total assets, return on net assets, and gross profit margin and other 22 indicators, using the CSI 300 constituents in January 2011 - December 2020 quarterly data for empirical analysis. The results show that 13 indicators such as earnings-price ratio, book-to-market value ratio, price-to-market value, and market capitalization have a significant impact on stock returns. Based on these indicators, this article uses coefficients combined with the entropy method to assign weights to each indicator and uses historical data to rank the companies in each quarter. Then make equal investments and adjust positions in the top 15 companies of every quarter, and finally get 16.99% annualized income.

Keywords-stock selection strategy; stock market; quantitative investment; fundamentals; multi-factor

1. INTRODUCTION

Chinese and foreign investment institutions have proved through practice that quantitative investment strategies can obtain better excess returns. Compared with the traditional fundamental analysis, the quantitative analysis focuses on specific companies, with strong depth, weak breadth and relatively concentrated position control. Ding Peng (2012) defines quantitative investment as a process of realizing investment strategy based on statistical analysis tools and computer technology [1]. Fundamental quantification is the combination of fundamental analysis and quantitative investment. Many scholars have conducted in-depth research on fundamental factors. Fama & French (1992-1993) established a three-factor model, pointing out that β, market size, and book-to-market ratio can explain stock price fluctuations, and some variables can separately explain stock price returns [2][3]. Basu (1983) studied the empirical relationship among the company size, the returns of listed companies on the New

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2327127

York Stock Exchange, and the return on common shares, found that the common shares of companies with high E/P obtained higher risk-adjusted returns than those of companies with low E/P, and even if the company size difference was controlled, this effect was indigenous [4]. Albadvi & Norouzi analyzed the fundamental factors of the German market and pointed out that there were different effective factors among different industries and the degree of influence was different [5].

Domestic scholars have also conducted a lot of research on the factors affecting stock returns. Chen, Zhang and Chen (2001) conducted a cross-sectional analysis on the determinants of expected stock returns and found that the explanatory power of scale and B/P is very significant, but β, book financial leverage and price-earnings ratio cannot pass the significant test in any model [6]. Dai Qingwen (2012), based on the financial situation of listed companies, through regression analysis and factor analysis, points out that profitability is the core of financial analysis of listed companies, and points out that earnings per share and return on net assets (ROE) are the most significant indicators of stock price returns [7]. Based on the Fama-French model, Zhang Lei & Ma Yuxin (2020) studied the market data of Chinese A-shares from 2001 to 2019, found that the scale factor and value factor are still important in A-shares. At the same time, it is pointed out that the profit-to-price ratio of A-shares has a better pricing effect and explanatory power [8].

Therefore, studying the factors that affect stock returns has profound theoretical and practical value. This article will select factors from a fundamental perspective, quantify the impact of effective factors on stock returns, and conduct back-testing with historical data to explore quantitative strategies adapting to the domestic market.

2. VARIABLE SELECTION AND DATA SOURCES

2.1 Variable selection

According to the literature review, this paper takes the stock price return rate as the dependent variable (RETURN) and selects the following factors that may have a significant impact on the return rate from the valuation, profitability, operational capacity, solvency, growth ability and technical aspects as variables, as shown in Table I.

Table I. Candidate Factors

Variable		Interpretation
	RETURN	(current final closing price − last final closing price) / last final closing price
Valuation factor	EP	Earnings per share/price per share.
	BM	Stockholders' equity/company's market value.
	PCF	Price per share/operating cash flow per share
	PS	Market price per share/sales per share
	CMV	Number of Shares in Circulation × Stock Price per Share. This paper takes a logarithm.
profitabilit	NPR	The enterprise realizes net profit/sales revenue

y	ROA	Total compensation received over a given period/average total assets
	ROE	Net profit belonging to shareholders of parent company/ average net asset
	NCF	Subtotal of cash inflows from operating activities - Subtotal of cash outflows from operating activities.
	BPS	Net profit / total equity
	GPM	Sales gross profit / sales income
operation ability	TRA	(Main business cost + final inventory cost-initial inventory cost) / average accounts payable
	ART	Net main business income / average receivable
	ITR	Sales (operating) costs / average inventory
	FAT	Total operating income / average fixed assets
	CAT	Operating income / average current assets
debt paying ability	DAR	Total liabilities / total assets
	ETA	Total equity / liabilities of shareholders
Growth capacity	NPGR	(Current net profit - prior period net profit) / prior period net profit
	TA	Economic resources owned or controlled by enterprises that can be measured in currency. This paper takes a logarithm.
technology index	MACD	2× (DIF-DEA)
	WVAD	(closing price-opening price) / (highest price-lowest price) × trading volume.

2.2 Data sources

Considering that there are few annual data, and many indicators of monthly data are incomplete and unable to reflect changes, this paper selects quarterly data as the research object and selects the companies listed in CSI 300 constituent stocks before 2011 as samples. Considering the specific economic meaning of some variables, this paper makes the following screening of the samples: (1) Because the calculation and measurement of financial indicators of financial companies are special, financial companies are excluded (2) Sample companies with incomplete indicators are excluded; (3) Technical correction of individual error data. A total of 156 companies were selected as samples. All data are from the iFind database.

3. MODEL CONSTRUCTION AND FACTOR SCREENING

3.1 Model design

This paper adopts the multi-factor analysis method, takes the quarterly return rate RETURN as the explained variable, and takes the earnings /prices ratio, book market value ratio, market current rate, market sales rate, circulation market value, net sales interest rate, total asset return rate, net cash flow generated by business activities, basic earnings per share, gross sales interest rate, accounts payable turnover days, accounts receivable turnover days, inventory turnover days, fixed asset turnover rate, current asset turnover rate, asset-liability ratio, equity

ratio, a year-on-year growth rate of net profit, total assets, smooth similarities and differences average of MACD index, WVAD William variation dispersion as the explanatory variables to establish the panel data model :

$$y_t = \alpha + \beta x_{it} + \varepsilon_t \quad (1)$$

Where t = 1,2,3... 22. y_t represents the explained variable. x_i represent explanatory variables, α is as the intercept, β is the estimated coefficient of independent variables, ε_t is the model error term

3.2 Regression analysis

In this paper, the ADF unit root test is adopted. All variable sequences are stationary sequences and can be directly analyzed by regression.

This paper selects panel data of CSI 300 listed companies before 2011, first using the F test to test the mixed fixed effect model and individual fixed effect model, F results show that P-value is 0.0000, should be applied to the individual effect regression model. Hausman test shows that when the individual is fixed and time is random, the P-value is 0.4601, which does not reject the original hypothesis. Therefore, the panel model with random time and fixed individual is selected. The results of regression analysis are as Table II, which shows that F statistics are significant, meaning the model is effective, and DW coefficient is 2.028994, meaning the model does not exist autocorrelation. The regression coefficients of independent variables, EP, BM, PCF, CMV, ROA, ROE, GPM, ART, ITR, FAT, TA, MACD and WVAD were significant at 0.05 level, while PS, NPR, NCF, BPS, TRA, CAT, DAR, ETA and NPGR were not significant, and had no significant effect on TURN.

Table II. Panel Regression Results

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.4560	0.2233	2.0423	0.0412
EP	-0.5838	0.1200	-4.8644	0.0000
BM	-0.2588	0.0325	-7.9632	0.0000
PCF	0.0000	0.0000	2.0603	0.0394
PS	0.0000	0.0000	-1.8980	0.0577
CMV	0.0803	0.0093	8.6090	0.0000
NPR	0.0000	0.0001	-0.2499	0.8026
ROA	-0.0067	0.0013	-5.0527	0.0000
ROE	0.0020	0.0008	2.5201	0.0118
NCF	0.0000	0.0000	0.6962	0.4863
BPS	0.0003	0.0008	0.4120	0.6804
GPM	-0.0012	0.0005	-2.6510	0.0080
TRA	-0.0007	0.0006	-1.1660	0.2436
ART	0.0000	0.0000	2.0093	0.0445
ITR	-0.0002	0.0001	-2.5136	0.0120
FAT	0.0009	0.0004	2.3199	0.0204

CAT	0.0003	0.0052	0.0499	0.9602
DAR	-0.0002	0.0005	-0.4857	0.6272
ETA	0.0000	0.0000	0.1506	0.8803
NPGR	0.0000	0.0000	0.8488	0.3960
MACD	-0.0070	0.0036	-1.9645	0.0495
TA	-0.0888	0.0102	-8.7123	0.0000
WVAD	0.0016	0.0001	13.6808	0.0000
R-squared	0.116516	Mean dependent var		0.065844
0Adjusted R-squared	0.090720	S.D. dependent var		0.290541
S.E. of regression	0.277049	Sum squared resid		465.2954
F-statistic	4.516803	Durbin-Watson stat		2.028994
Prob	0			

4. BACK TESTING

In order to test the factors selected by regression analysis, this paper designs the following back testing:

The first step is to set scoring rules and assign scoring weights to each indicator.

The second step is to score the company's situation in each quarter according to the weight.

The third step, select the top 15 score companies of the first quarter of 2011 and invest equally in the second quarter of 2011, then calculate the quarterly yield.

The fourth step is to screen the investment targets for the second quarter of 2011 in accordance with the third step, that is, select the top 15 score companies in this quarter, if there are the same companies, keep the positions, if there are new companies, adjust the positions. The rule of position adjustment is to make the equal investment after deducting the amount of holding positions. For example, if there are 10 same companies in the two quarters, then sell out the stock of the remaining 5 companies, and buy the new companies in equal amounts.

The fifth step, repeat step four until the last quarter of the inspection period, that is, the fourth quarter of 2020.

The sixth step, compare the portfolio yield and the CSI 300index yield.

4.1 Setting of scoring rules

In this paper, when considering the setting of scoring rules, it is considered that the regression coefficient should be considered to reflect the degree and direction of the impact. At the same time, this paper believes that the importance of each indicator needs to be considered. Therefore, the weight of each indicator is determined by the entropy method, and the following Table III is obtained. The weight summary table calculated by the entropy method shows that the weights of BM, RPV, ITR and FAT are relatively high.

Table III. Summary Of Weight Results Calculated By Entropy Method

item	Information entropy (e)	Information utility value (d)	Weight coefficient of entropy method (w)
MMS_EP	0.9996	0.0004	0.14%
MMS_BM	0.972	0.028	10.96%
MMS_PCF	0.9995	0.0005	0.21%
MMS_CMV	0.9945	0.0055	2.14%
MMS_ROA	0.9966	0.0034	1.34%
MMS_ROE	0.9998	0.0002	0.09%
MMS_GPM	0.9968	0.0032	1.24%
MMS_ART	0.9032	0.0968	37.89%
MMS_ITR	0.9373	0.0627	24.55%
MMS_FAT	0.9542	0.0458	17.92%
MMS_MACD	0.9999	0.0001	0.06%
MMS_TA	0.9913	0.0087	3.42%
MMS_WVAD	0.9999	0.0001	0.06%

Combined with the coefficient of regression analysis, the scoring weight coefficient formula of each index is :

$$S_i = |w_i| \times |co_i| \tag{2}$$

where i = EP, BM, PCF, CMV, ROA, ROE, GPM, ART, ITR, FAT, MAVD, TA, WVAD. S_i is score weight for each indicator, co_i is the regression coefficient for each indicator, w_i is the weight obtained by the entropy method of each index, c is the regression constant is 0.4560.

4.2 Back testing analysis

The actual investment of this back testing started from the second quarter of 2011, and it has been invested for 39 quarters in total. The quarterly return rate of the portfolio and the quarterly return rate of CSI 300 index are shown in Table IV below, and the comparable figure of the total return rate is shown in Figure 1 below: :

Table IV. Quarterly Back Testing Return

Quarter	**Portfolio quarterly return**	**Quarterly yield of CSI 300 index**
2011.04-2011.06	4.96%	-5.56%
2011.07-2011.09	-9.36%	-15.20%
2011.10-2011.12	-12.22%	-9.13%
2012.01-2012.03	11.94%	4.65%
2012.04-2012.06	7.28%	0.27%
2012.07-2012.09	-1.97%	-6.85%
2012.10-2012.12	-3.38%	10.02%

2013.01-2013.03	11.51%	-1.10%
2013.04-2013.06	7.62%	-11.80%
2013.07-2013.09	18.43%	9.47%
2013.10-2013.12	-8.27%	-3.28%
2014.01-2014.03	-9.68%	-7.88%
2014.04-2014.06	1.94%	0.88%
2014.07-2014.09	17.44%	13.20%
2014.10-2014.12	5.30%	44.17%
2015.01-2015.03	23.61%	14.64%
2015.04-2015.06	23.07%	10.41%
2015.07-2015.09	-34.61%	-28.39%
2015.10-2015.12	14.10%	16.49%
2016.01-2016.03	-17.82%	-13.75%
2016.04-2016.06	4.72%	-1.99%
2016.07-2016.09	-4.05%	3.15%
2016.10-2016.12	-5.28%	1.75%
2017.01-2017.03	-0.11%	4.41%
2017.04-2017.06	14.30%	6.10%
2017.07-2017.09	6.00%	4.63%
2017.10-2017.12	6.55%	5.07%
2018.01-2018.03	5.11%	-3.28%
2018.04-2018.06	-1.10%	-9.94%
2018.07-2018.09	-9.10%	-2.05%
2018.10-2018.12	-4.32%	-12.45%
2019.01-2019.03	38.99%	28.62%
2019.04-2019.06	-1.50%	-1.21%
2019.07-2019.09	13.30%	-0.29%
2019.10-2019.12	11.31%	7.39%
2020.01-2020.03	3.19%	-10.02%
2020.04-2020.06	43.72%	12.96%
2020.07-2020.09	11.27%	10.17%
2020.10-2020.12	11.07%	13.60%
Temporal yield	16.99%	5.05%

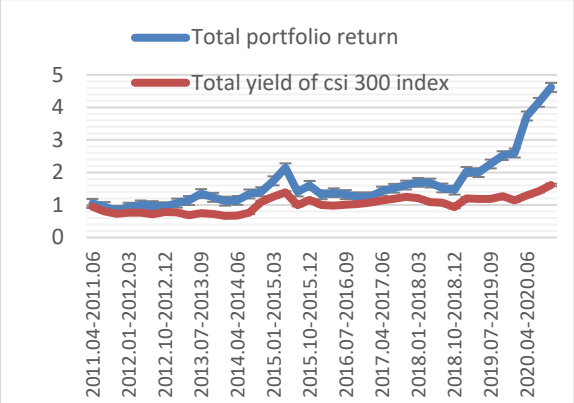

Figure 1. Comparison Chart Of Total Return Of Back Testing

From the comparison chart, it can be seen that the portfolio is not monthly returns can beat the CSI 300, but the total return rate of the investment portfolio is winning the CSI 300 index. The portfolio income is slightly higher than the CSI 300index in the early period and the overall trend is similar to the CSI 300. In the later period, it gradually shows the superiority of the portfolio, and the investment income gap is gradually widening. Combined with the last line of Table IV, the annualized income of the portfolio is 16.99 %, while the annualized income of the CSI 300 is 5.05 %.

In addition, through quarterly screening data, this paper finds that in 39 investment quarters, a few companies, such as KWEICHOW MOUTAI and WULIANGYE, continue to hold positions almost every quarter after being selected from the first round of screening. Finally, it is also the largest contribution of such companies, which proves that the stock selection strategy adopted in this paper can screen stocks with long-term investment value.

5. CONCLUSIONS

In this paper, through the regression analysis of CSI 300 constituent stock companies, it is found that among the 22 indicators selected, the price ratio, book market value ratio, market present rate, circulation market value, total asset return rate, net asset return rate, sales gross interest rate, accounts receivable turnover rate, inventory turnover rate, fixed asset turnover rate, total assets, MACD index smooth similarities and differences average, William variation dispersion have an impact on stock investment returns. In addition, this paper tests the investment credibility of stock selection based on these indicators by historical back-testing, ranking branch companies by combining the weight of coefficient and entropy method, and investing in high branch companies. The results show that the annualized income can be obtained by 16.99 % based on these indicators, and the total investment income is also higher than that of CSI 300.

Although the effect of investment returns obtained by the current model test is still good, there are still areas for improvement in the design of this paper. For example, the screening object of this paper is based on CSI 300, which is still relatively small compared with the number of

thousands of stocks in the market. In addition, the cost of warehouse adjustment is not considered in this paper. These are the directions of further research in the future.

REFERENCE

[1] Ding Peng.Quantitative investment : strategy and technology [M].Electronic Industry Press, 2012.

[2] Eugene F Fama, Kenneth R French. The Crosssection of Expected Stock Returns [J]. The Journal of Finance, 1992, (2): 427-466.

[3] Eugene F Fama, Kenneth R French. Common risk Factors in the Returns on Stocks and Bonds [J]. Journal of Financial Economics, 1993, (93): 3-56.

[4] Sanjoy, Basu. The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence[J]. Journal of Financial Economics, 1983.

[5] Albadvi A, Norouzi A . Using downside CAPM theory to improve customer lifetime value prediction in non-contractual setting[J]. Management Science Letters, 2013, 3(12):3003-3012.

[6] Chen Xinyuan, Zhang Tiancai, Chen Donghua.The cross-sectional multivariate analysis of expected stock returns : empirical evidence from China' s securities market [J].Financial research, 2001,000 (006) : 22-35.

[7] Dai Qingwen. Analysis of the influencing factors of stock price returns of listed companies in China [J]. Accounting friends, 2012 (10) : 114-116.

[8] Zhang Lei, Ma Yuxin. Empirical study based on three-factor model of profit-to-price ratio - empirical evidence from A share [J]. Tax and economy, 2020 (04) : 40-45.