Multi-factor Stock Selection Strategy Incorporating Liquidity Risk

Die Wang

School of Finance and Statistics, Hunan University, China

714173401@qq.com

Abstract. As an emerging market dominated by retail investors, China's stock market is greatly affected by external shocks such as market noise and policy changes, and the traditional fundamental factors have limited ability to predict the future returns of stocks. Stock liquidity is also affected by noise factors and macro factors, which can be regarded as one of their external performance. Based on the multi-factor stock selection model, this paper uses the 2010-2022 CSI 300 stock data to construct a fundamental fusion factor and another fusion factor considering both fundamental and liquidity risk factors, then compares the stock selection effect of the two fusion factors through a variety of test indicators. The empirical results show that the IC value of the fusion factor considering both fundamental and liquidity risk factors is 0.0267, which is greatly improved compared with the fundamental fusion factor with IC value of only 0.0195, and its excess yield of 64.728% is 13.134% higher than the fundamental fusion factor, while the benchmark win rate is increased by nearly 3% and the maximum retracement is reduced by nearly 12%. The research shows that the effectiveness and stability of the fusion factor after adding the liquidity index to predict the future return of the stock and obtain the excess profits are better than the fundamental fusion factor. Therefore, considering the measurement and inclusion of stock liquidity level can improve the yield of stock selection strategy in the Chinese stock market to a certain extent.

Keywords: Multi-factor stock selection model; liquidity risk; fusion factor; stock selection strategy.

1. Foreword

Liquidity is the premise for stocks and other financial assets to obtain income, and also the basis for realizing the security of financial assets. However, at present, China's A-share market is still dominated by retail investors and there are a large number of noisy investors. The irrational investment behaviors of these noisy investors are easy to cause the block of stock liquidity and form a certain liquidity risk. But the traditional concept of value stock selection does not consider the stock liquidity risk caused by the noise factor, so this paper will add liquidity factors to the traditional factor stock selection model and compare the stock selection effect, to verify and supplement the application of the traditional stock selection factor and multi-factor model in the Chinese stock market.

In Multi-Factor research field, a large number of scholars at home and abroad have explored and studied many effective factors of optimizing stock selection strategies from various angles: Fama [1] based on the asset pricing model put forward the Three-Factor model, and this model is considered by many scholars as the benchmark of the research factor model in the subsequent studies. Based on the Three-Factor model, Fama and French [2] again found that the stock excess return can also be affected by the profitability risk and investment level risk, so the Five-Factor model was constructed. Domestic scholars have also made some explorations in this field based on the Chinese stock market, based on the empirical study of all A-shares in Shanghai and Shenzhen from January 1995 to December 2015, Shengmin Zhao et al.[3] found that the explanatory ability of the model is significantly improved after the investor preference factor BMG was introduced to the Fama-French Five-Factor model; Yuxuan Liu et al.[4] found that the Multi-Factor stock selection model introducing the financial cycle index can better capture the changes of the financial market than the stock selection model introducing the economic cycle index, thus to obtain a higher excess yield.

Although domestic and foreign scholars have tried to introduce different effective factors to improve the excess return rate of stock portfolios, domestically there is relatively less research on comparing the effect of stock selection strategy before and after the introduction of stock liquidity

risk factors. However, the domestic scholars like Yunhui Sun and Weiqi Liu have all studied and proved that due to the existing institutional defects of China's stock market and the poor quality of corporate information disclosure, the liquidity level of individual stocks is uneven, and the liquidity risk of some stocks is large [5,6]. Therefore, it is of great practical significance to measure the liquidity risk and introduce it into the multi-factor stock selection model.

In conclusion, this paper's main contributions are as follows: First, use the CSI 300 data of China's A-share market to calculate the fundamental indicators and build the quality factor of stock selection. Although there are similar studies in the existing literature, this paper features the use of long-term data from 2010 to 2022 to ensure the consistency of the accuracy of factor selection and the effect of the strategy; Second, this paper reviews a large body of literature on stock liquidity risk, calculated various liquidity indicators such as Amihud index[7] and tested their effectiveness, which can make the measure of stock liquidity risk more comprehensive and perfect; Third, this paper compares the traditional fundamental fusion factor and the comprehensive factor after adding the liquidity index, intuitively verify and demonstrate the improvement of liquidity factors on the income and the strategy.

2. Research Design

By constructing the stock selection factors of fundamentals and liquidity risk, this paper uses the multi-factor stock selection model to explore the stock selection effect after including the liquidity factors. The overall study design is shown in Fig. 1, and the specific single-factor tests and factor fusion steps are shown below.

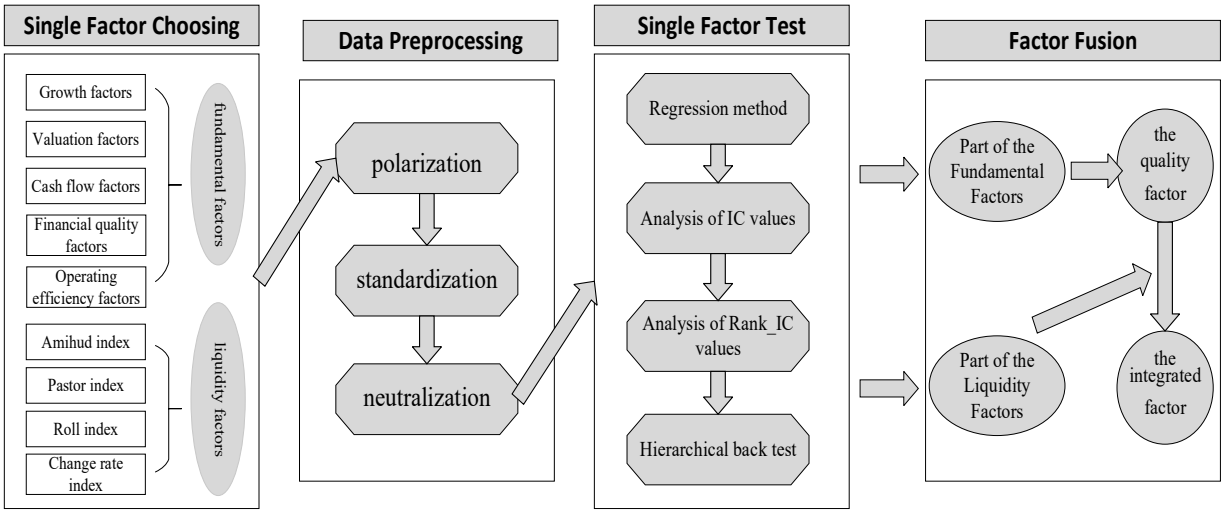


Fig. 1 Study flow chart

2.1 Single factor test

In this paper, all factors are tested separately by regression method, IC value, rank_IC value analysis, and hierarchical back-test. In the course of the above, a series of factor evaluation indicators are constructed, in order to comprehensively measure the stock selection ability of individual factors and screen out the effective factors for subsequent synthesis.

The regression method is the most commonly used method to test the effectiveness of factors by constructing a linear regression equation between the factor exposure vector in stage T and the stock return vector in stage T + 1 to obtain relevant indicators for screening stocks. The specific expression of the phase-T regression model is as follows:

$$ret^{T+1} = a^T X^T + \varepsilon^T$$

Among them, ret^{T+1} is the yield vector of all individual stocks in the T+1 period, X^T is the exposure vector of all individual stocks in the measured single factor in the T period (the factor data

that has been pretreated excluding the impact of the industry and market value), and a^T is the return rate of the measured factor to be fitted, ε^T is the residual vector of the model.

The IC value of the factor is the correlation coefficient between the exposure vector of the factor in stage T and the vector of the stock return in stage T + 1, as $IC^T = corr(r^{T+1}, X^T)$. According to the calculation method of the correlation coefficient, it can be divided into IC value calculated by Pearson correlation coefficient and Rank _ IC value calculated by Spearman rank correlation coefficient. In this paper, the IC value sequence of the factor is calculated in two ways for subsequent analysis. According to the sequence of IC value and Rank _ IC value of the factor, we construct the ratio of sequence mean, sequence standard deviation, standard deviation of sequence mean and the ratio of sequence greater than 0 to measure the factor's significance, stability, validity and the stability degree of its action direction respectively.

When doing the hierarchical back-test, we divided the stocks into 5 groups according to the factor exposure size in period T, and then counted the average return of each group in period T+1 as its revenue sequence. After that, we got the H-L portfolio revenue sequence using each period's return of the Top group minus the corresponding value of Bottom group, and the benchmark revenue sequence based on the daily closing price of the CSI 300 index. Using these revenue sequences, we calculated some evaluating indicators of 5 stock groups and the H-L portfolio respectively, like the annual return rate, annualized volatility, maximum retracement, relative benchmark win rate and so on, as supplementary judgments of the effectiveness of factor stock selection.

2.2 Factor fusion

Considering the similar structural dimensions of similar factors, there may be a strong correlation between some individual factors. For example, in the growth factors, if total operating revenue is relatively high, the operating profit growth rate is likely to be correspondingly high. Therefore, we first performed a correlation analysis on all factors that passed the single-factor test described above. For multiple factors with high correlation, only the factor with the best test results was retained, so as to prevent excessive inclusion of certain factors from causing information overlap and weight imbalance in the fusion process.

To highlight the changes in stock selection ability after the addition of liquidity factors, this paper used equal-weight composite method to get two synthetic factors: one is only fundamental factors' fusion and the other was joined the liquidity factors' fusion, and then compared their stock selection ability mainly through the above single factor test methods.

3. Factor Choosing and Data Processing

3.1 Factor Choosing

The factor is the component element of the multi-factor model and the variable that can predict the stock return rate. The financial condition of the company is the most obvious factor affecting the performance of the stock price[8,9], so the fundamental factors constructed based on the financial condition are also the most commonly used category factor. This paper adds liquidity factors based on the traditional fundamental factors and enriches the application of the multi-factor model.

3.1.1 Fundamental factor selection

To minimize the correlation between different categories of fundamental factors, through reading relevant literature and research reports [9,10], we divided the 78 fundamental factors into five categories: 13 growth factors, 15 valuation factors, 13 cash flow factors, 16 financial quality factors and 21 operating efficiency factors, and with specific introduction as following:

The company's growth is mainly reflected in the growth of revenue and profit brought about by its operating activities, so this paper selects the company's operating income, operating profit, net profit and other indicators' growth rates in the current quarter and in the same period as the growth factors.

The valuation factors mainly consider the value of the stock.In traditional enterprise value evaluation, PB, PE and some other indicators reflect investors' estimate of stock value. Considering adding the adjustment factor related to the accounting period in the denominator and the dynamic factor of the rolling window, this paper constructs three different forms of PE, PS and PCF respectively. At the same time, PB, market value tangible asset ratio and other indicators are also selected as the valuation factors.

Cash is the blood of the company, and cash flow is a necessary prerequisite for the company's profit and business development, which also provides certain information on the future earnings changes of the stock. Therefore, this paper constructs 7 cash content indicators, such as cash content of operating income and cash satisfaction investment ratio, and 6 cash flow indicators, such as net cash flow per share and net cash flow generated by operating activities per share, to be tested and studied.

Financial quality factors mainly reflects its financial stability and solvency through the financial structure. This paper mainly uses current ratio and quick ratio to reflect the short-term solvency; property right ratio and owner equity ratio to measure the long-term solvency. In addition, the asset-liability ratio and other indicators are also added as a supplement to the solvency.

The operating efficiency factors mainly reflect the operating income of the enterprise, which is also an important reference for investors to choose stocks. This paper selects 11 indicators such as return on capital and return on equity to measure the efficiency of sales operation and investment, and also selects 10 turnover rates like operating asset turnover to evaluate the efficiency of enterprise asset management.

3.1.2 Liquidity factor selection

The liquidity of stocks is how easy it is to trade bulk stocks at the minimum cost, the minimum price impact and the fastest speed [11,12], namely whether the stocks can be sold and converted into cash in time. The higher level of stock’s liquidity is, the liquidity risk faced is smaller. Foreign research on measuring stock liquidity and its potential risks originated earlier, so the content is also relatively mature: Roll R. [13] used the covariance of the first order difference of the price sequence to measure the stock liquidity with the daily trading data; Amihud, Y.[7,14] constructed the still widely used illiliquidity indicator: the greater the illiquidity this indicator represents, the weaker the liquidity of stocks; Pastor, L., Stambaugh, R.[15] measured the stock liquidity by the influence of trading volume and trading direction on stock yield, and the stock liquidity risk by the covariance of stock return and market liquidity information. Therefore, referring to the relevant foreign literature, we selected the Amihud index, Roll indexes and Pastor_Stambaugh indexes as liquidity factors for calculation and study.

Table 1. Basic information table of partial factors

Factor abbreviation	Account Form
ROE_growthA	(Return on equity in this quarter-return on equity in the previous quarter) / return on equity in the previous quarter
ROE_growthB	(Return on equity in a single quarter- -return on equity in the same period of last year) / return on equity in the same period of last year
EPS_growthA	(Basic earnings per share amount of this year- -amount of basic earnings per share] / (amount of basic earnings per share)
EPS_growthB	(Amount of basic earnings per share this year- -amount of basic earnings per share last year) / (amount of basic earnings per share last year)
opprofit_growthA	(Quarterly amount of operating profit- -quarterly amount of operating profit) / (quarterly amount of operating profit)
opprofit_growthB	(Amount of operating profit in the current period- -amount of operating profit in the same period of last year) / (amount of operating profit in the same period of last year)

Moreover, the turnover rate reflects the activity of stock trading and the purchase intention of investors, and is also an important indicator to measure stock liquidity. Domestic scholars often construct various turnover rate indexes as optional factors when studying the multi-factor stock selection model [16,17,18]. This paper also calculates the daily turnover rate data of individual stocks, the average daily turnover rate within the most recent (n=10,30,60,120) days and its deviation degree of the average turnover rate within one year as the turnover rate indicators.

In conclusion, 98 stock selection factors were selected and constructed in this paper. Due to the limited space, Table 1 only gives some factors' abbreviation and calculation methods. For a detailed understanding, you can contact the author.

3.2 Data pre-processing

This paper selects the related factor data and stock earnings of the CSI 300 index in all trading days from January 2010 to December 2022 from financial databases such as CSMAR and iFinD as all samples for subsequent empirical research.

Due to the loss of the original stock data, we processed the raw data according to the extent of missing data: delete the stock trading records with missing yield or excessive missing factor data (all data of a major factor are missing); for the partly missing stock records due to the statistical time limit, use the backward filling method to fill in. After the factor construction and data manipulation given above, we finally obtained 665,143 sample records of 98 quantified factors.

Table 2 randomly gives descriptive statistics of some factors' raw data without processing. It can be seen from the table that there are some differences in the data range among different factors, and even different orders of magnitude between some factors. Therefore, we need to do some data pre-processing operations.

Table 2. Factor description table (data not processed)

Factor category	Factor abbreviation	mean	std	min	max
Valuation factor (part)	PE_TTM	56.4012	897.4603	0.3018	92212.4558
	PCF_TTM	64.6986	563.8443	0.2081	37557.5473
	PB	4.0025	4.4728	0.0427	66.2384
	book_valueA	0.6701	0.3095	0.0292	1.6517
Operating efficiency factor (part)	opmargin_ratioTTM	0.3166	0.2518	-9.2910	5.1013
	opprofit_ratioTTM	2407.1346	194110.9036	-4878.2860	25164254.5519
	inventory_turn_TTM	12840.7507	1370789.6823	-5.1722	151053254.4151
	acpay_turn_TTM	33.8544	940.4803	-5.9348	105273.8872

We used the Median Absolute Deviation method (MAD) to identify and process the extreme values in the sample data, with specific steps given as follows: firstly, find out the median of the factor value as X_m and the absolute deviation of the factor values from the median as $|X_i - X_m|$; secondly, calculate the median of the absolute deviation values as MAD and get the reasonable range of the factor value by determining the parameter n (n=3 is selected in this paper); finally, obtain the processed factor value as X_i^* according to the formula giving as follows:

$$X_i^* = \begin{cases} X_m + n \times MAD, & \text{if } X_i > X_m + n \times MAD \\ X_m - n \times MAD, & \text{if } X_i < X_m - n \times MAD \\ X_i & \text{else} \end{cases}$$

In order to eliminate dimensional effects between data of different factors, we also adopted the sorting method to unify the dimension of the factor data to make comparative analysis among different factors. Most factors will be influenced by the total stock market value and industry preference, so we did neutral treatment as follows:

$$X_i = \beta_M \times \ln(Mkt_i) + \sum_{j=1}^{n} \beta_j \times Ind_{j,i} + \varepsilon_i$$

Among them, X_i is the factor exposure value of the i th stock, Mkt_i is the market value of this stock and $Ind_{j,i}$ is its industry exposure vector (if the stock i belongs to the j th industry, then $Ind_{j,i} = 1$; otherwise $Ind_{j,i} = 0$). β_M and β_j are the coefficient of the market value factor and the j th industry factor respectively; ε_i is the residual item of the model, that is, the factor data after excluding the influence of the industry and the market value.

The descriptive statistics of the processed factors in Table 2 are shown in Table 3. As can be seen from Table 3, the maximum value of the processed factor data is basically in the same order of magnitude, and the data range is controlled within the same interval, which indicates that the dimensional gap between factors has been well eliminated by the data pre-processing operation.

Table 3. Factor description table (after data processing)

Factor category	Factor abbreviation	mean	std	min	max
Valuation factor (part)	PE_TTM	-3.45E-06	0.770704775	-5.073712366	6.854915173
	PCF_TTM	-7.99E-07	0.783276963	-5.032881646	5.695802058
	PB	-2.90E-06	0.746159161	-4.997432637	6.701557484
Operating efficiency factor (part)	book_valueA	2.44E-06	0.687702803	-6.10359827	4.4098741
	opmargin_ratioTTM	3.80E-07	0.762452236	-10.70742717	8.574800156
	opprofit_ratioTTM	4.54E-06	0.702625918	-4.041234542	6.64518839
	inventory_turn_TTM	1.01E-05	0.790803662	-12.47827773	9.997454234
	acpay_turn_TTM	-7.76E-06	0.934706037	-18.73177079	15.01424415

To further investigate the data morphology, we randomly plotted some factors' probability density curves in Fig. 2. Through observation, the factors after data pre-processing given above are all 0-centered normal distribution, which can avoid possible interference due to the differences in the data and make the results of the single-factor test more dependent on the nature of the factors themselves, thus increasing the credibility of the test results.

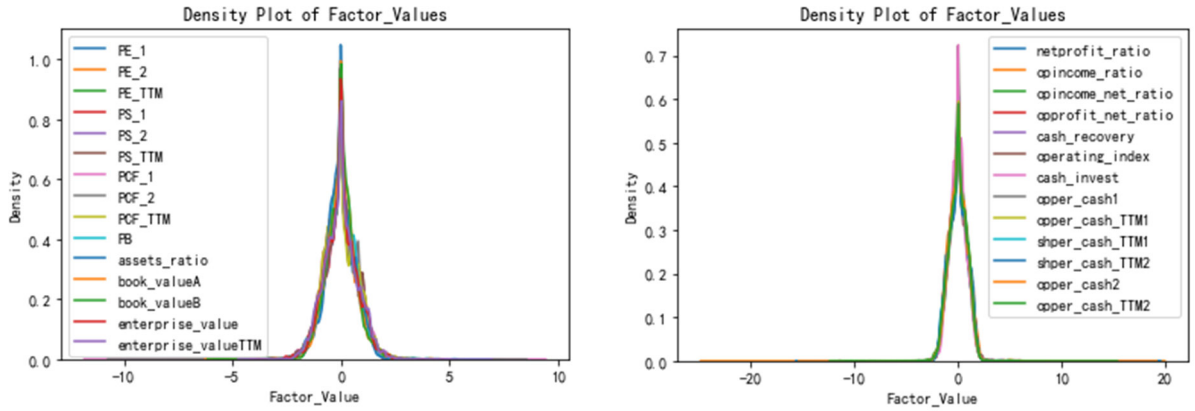


Fig. 2 Probability density plot of partial factor data

4. Empirical Analysis of the Multi-factorial Model

4.1 Single-factor test results

4.1.1 Single factor screening

In this paper, through comparing the test results of the regression method, IC value test and stratified back-test method, some relatively effective stock selection factors are preliminarily selected. Due to the large number of factors and limited space, here we only takes 15 valuation factors as an example to show the factor screening process.

Table 4 presents the specific values of the four evaluation indicators obtained by the valuation factors' regression method. It can be seen that the performance of the factors of PE_1, PE_TTM,

PS_2 and PB are the best: their absolute means of the t-value sequence are above 1.2, indicating that the significance of these factors affecting the stock yield is higher; their proportion of t value sequence greater than 2 and their means of t value sequence are both higher than the median level of the factor in this group, that is, the significance of these factors is relatively stable.

Table 4. Test results of valuation factors' regression method

	Absolute mean of t-value series	The proportion of t value sequence greater than 2	The mean of the t-value sequence	The sequence mean of the factor yield
PE_1	1.200134	0.179544	0.225625	0.000528
PE_2	1.146809	0.166878	0.173480	0.000434
PE_TTM	1.211307	0.183344	0.186898	0.000462
PS_1	1.175105	0.181444	0.260044	0.000857
PS_2	1.126861	0.165294	0.240887	0.000816
PS_TTM	1.154544	0.169728	0.242717	0.000824
PCF_1	0.970193	0.103230	0.185933	0.000665
PCF_2	0.935854	0.094680	0.134096	0.000606
PCF_TTM	0.975421	0.107980	0.164987	0.000614
PB	1.204794	0.187144	0.311147	0.001214
assets_ratio	1.157741	0.168461	0.258708	0.001244
book_valueA	1.081807	0.140912	-0.242816	-0.001295
book_valueB	1.185854	0.181761	-0.290927	-0.001295
enterprise_value	0.989814	0.111463	0.090204	0.000071
enterprise_value TTM	1.042296	0.129196	0.116682	0.000259

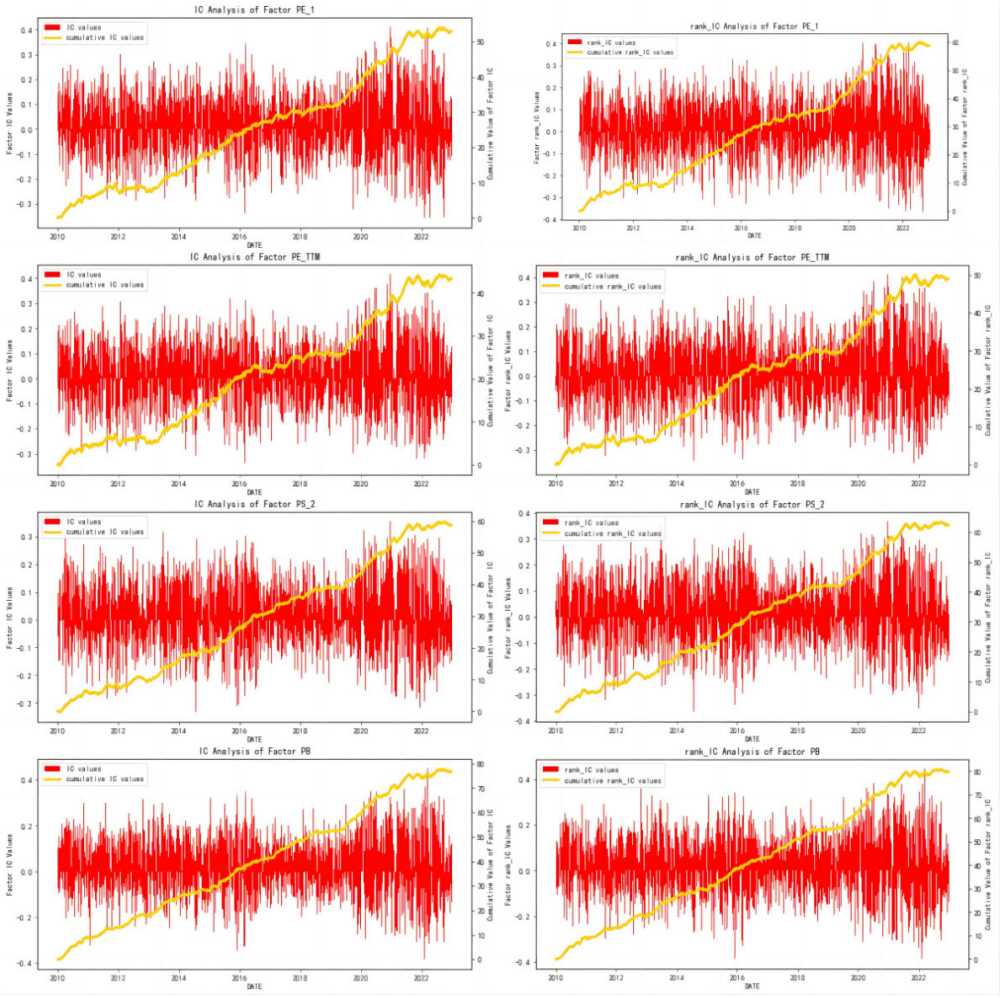


Fig. 3 The IC charts, rank_IC charts of PE_1, PE_TTM, PS_2 and PB

Fig. 3 shows the bar chart of IC, rank_IC values and cumulative IC, rank_IC values of each period of the four valuation factors selected by the regression method. Only these plots show that the IC value and rank_IC value test of these four factors all perform relatively well. Table 5 and Table 6 give some specific results of the valuation factors' IC test, from which we found that PE_TTM's IC mean, IR ratio, IR ratio, IC>0 ratio and rank_IC>0 ratio are all lower than the overall median, located at a poor level in this group. That is, the robustness and validity of the prediction yield of this factor are relatively poor. Besides, the IC standard error and rank_IC standard error of PS_2 showed relatively poor performance, namely the significance level of this factor's action is less stable, so we excluded these two factors. In conclusion, after the analysis of factor IC value, only PE_1 and PB are retained as effective factors.

Table 5. Part results of valuation factors' IC test

	IC mean	IC standard error	IR ratio	The portion of IC> 0 (%)	The portion of IC> 0.02 (%)
PE_1	0.016758	0.114916	0.145828	56.428119	49.366688
PE_TTM	0.013709	0.115684	0.118500	54.433186	48.195060
PS_2	0.018577	0.107531	0.172762	56.744775	49.778341
PB	0.024310	0.114321	0.212647	59.341355	52.501583
median	0.013709	0.109622	0.132221	55.984801	47.751742
average	0.010307	0.106001	0.096821	53.977201	46.599113

Table 6. Results of the valuation factor rank_IC test

	rank_IC mean	rank_IC standard error	The rank _ IR ratio	Proportion of the rank_IC> 0	Proportion of rank_IC> 0.02
PE_1	0.018716	0.122769	0.152446	56.174794	50.031666
PE_TTM	0.015536	0.122816	0.126498	54.718176	49.240025
PS_2	0.019709	0.113519	0.173620	57.473084	50.538315
PB	0.025289	0.121038	0.208933	59.468018	52.944902
median	0.015536	0.115607	0.139633	55.794807	48.543382
average	0.011284	0.111832	0.100036	54.188305	47.209204

According to the stratified back-test method introduced above, we obtained the net value curve for each group of PE_1 and PB respectively in Fig. 4, and their long and short return curves in Fig. 5. From Fig. 4, the net value curve of PE_1's each group is seriously crossed and there is no obvious monotonous trend; but for the factor PB, although the net value curve of each group overlapped before 2016, it's obviously layered after 2018, and the net value of Group4 with the highest factor value is significantly higher than that of other groups, thus PB is more effective than PE in stock selection. It can also be found from Fig. 5 that since 2010-2022, the factor PB's long and short net income is mainly positive proportion and its upward trend is relatively stable, so we excluded PE_1 and only retained PB at last. Table 7 presents all the results of the stratified back-test of the factor PB, revealing that all evaluation indicators have a certain monotone, which is another evidence of the prediction validity of this factor.

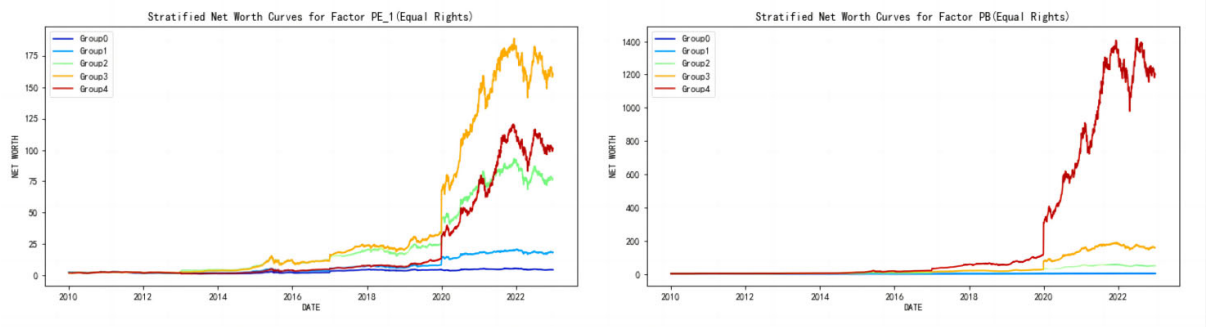


Fig. 4 Net value graph of PE_1's and PB's groups

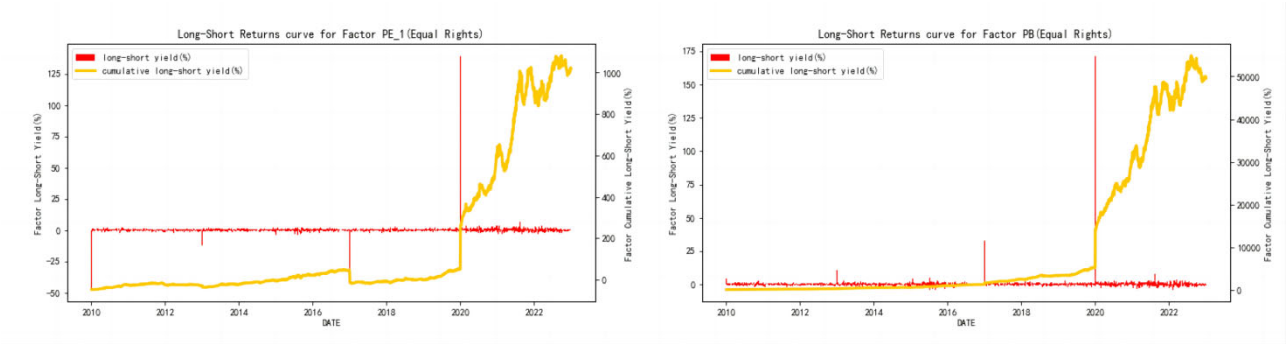


Fig. 5 Long-short return chart of PE_1 and PB

Table 7. Back-test results of the factor PB

Evaluating indicator	Group0	Group1	Group2	Group3	Group4	H-L
The alized rate of return	1.0587	3.6383	27.6072	40.0110	66.4594	63.5787
The alized volatility	25.2650	27.3374	38.0482	57.5233	55.2691	51.3716
Sharp ratio	0.0419	0.1331	0.7256	0.6956	1.2025	1.2376
Maximum retraction	-58.4053	-63.1135	-42.4912	-42.0455	-37.9731	-15.3944
Excess annualized rate of return	1.5486	4.1407	28.2257	40.6897	67.2663	64.3716
Excess annualized volatility	10.9638	14.7897	30.3937	52.7015	50.5591	55.5584
information ratio	0.1412	0.2800	0.9287	0.7721	1.3304	1.1586
Relative benchmark win rate	51.0611	52.8666	56.7627	57.4279	58.2832	53.9119
The maximum retracting	-33.9091	-45.3729	-10.3905	-12.9143	-14.9951	-49.6112

Through a single-factor testing process similar to the above steps, screening the stock selection factors of other major categories, finally we got the following nine effective factors: PB in growth factors; opprofit_growthB, opincome_all_growth and opincome_growthC in growth factors; ROE_TTM and allROE_parent_TTM in operating efficiency factors; opcash_all_debt in financial quality factor and amid and Roll_M in liquidity factors(Given the relatively poor test results for all cash flow factors, no valid factor is selected in this paper). You can contact the author for the specific test results of these factors.

4.1.2 Single factor analysis

Here is some theoretical analysis of the effective factors selected above: the growth of stock value can be summarized from two aspects and six specific perspectives as shown in Fig. 6.

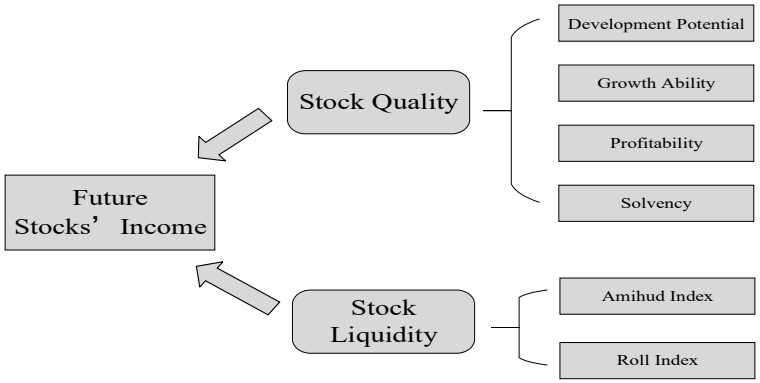


Fig. 6 Impact mechanism chart of stock returns

This paper measures the quality of stocks from the perspectives of development potential, growth ability, profitability and solvency, respectively corresponding to the valuation factor, growth factor, operating efficiency factor and financial quality factor; the better the stock quality, the higher the future value, the greater the return. The following specifically introduces the influence effect of various factors on the quality of the stock.

The valuation factor PB is the ratio of the market price per share and the net assets per share, focusing on the expectation of future profitability and reflecting the development potential of the stock. It is generally believed that the quality of the enterprise assets is better when the market price is higher than the book value, that is, when PB is high, the stock has the development potential; otherwise, the quality is poor and the development prospect is poor. The three growth factors reflect the increase and decrease of the profit compared with last year, focusing on the growth ability of the business. Specifically, if the index value of this category is greater than 0, the operating profit has increased; the higher the index value, the faster the growth rate is increased, and the higher the stock quality. ROE_TTM and allROE_parent_TTM measure the profitability by using the rate of return of the company's assets. The higher the index value, the higher the return brought by the enterprise investment and the stronger profitability of the company, which can be used as the basis for predicting the future growth of the stock value and the increase of income. The financial quality factor opcash_all_debt is the ratio of the net cash flow generated by the business activities to the total liabilities of the enterprise. In addition to borrowing new debts to repay the old debts, what can be used by enterprises to repay debts generally should be the cash inflow of business activities. Then this index directly reflects the ability of the enterprise to bear the debt and the maximum interest payment ability, and can also be regarded as one of the representatives of the stock quality.

The above stock quality is mainly measured through the fundamental factors constructed by the company's financial information. However, the empirical research of Yan Yang, RICHARDSON S and other scholars shows that the company's fundamental information of the company can only partly explain the volatility of the stock yield[19,20]. In order to optimize the stock selection strategy, this paper adds liquidity factors to the traditional fundamental research. The factor test results mentioned above show that Amid and Roll_M are the most effective for predicting stock returns among all liquidity factors, so these two indicators are used to measure stock liquidity. The greater the liquidity of the stock, the higher the recognition and the less potential liquidity risk; and the stock liquidity can increase corporate value, whose robustness has been verified by through a variety of inspection methods[21]. Meanwhile the improvement of enterprise value can also promote the growth of stock value. Therefore, stock liquidity will have a positive impact on the forecast of future stock returns, that is, the higher the level of stock liquidity, the relatively higher its future returns.

4.2 Factor fusion results

4.2.1 Correlation analysis

To eliminate some redundant factors and simplify the model, before the factor synthesis, this paper first drew the heat map of the effective factors' correlation test as Fig. 7.

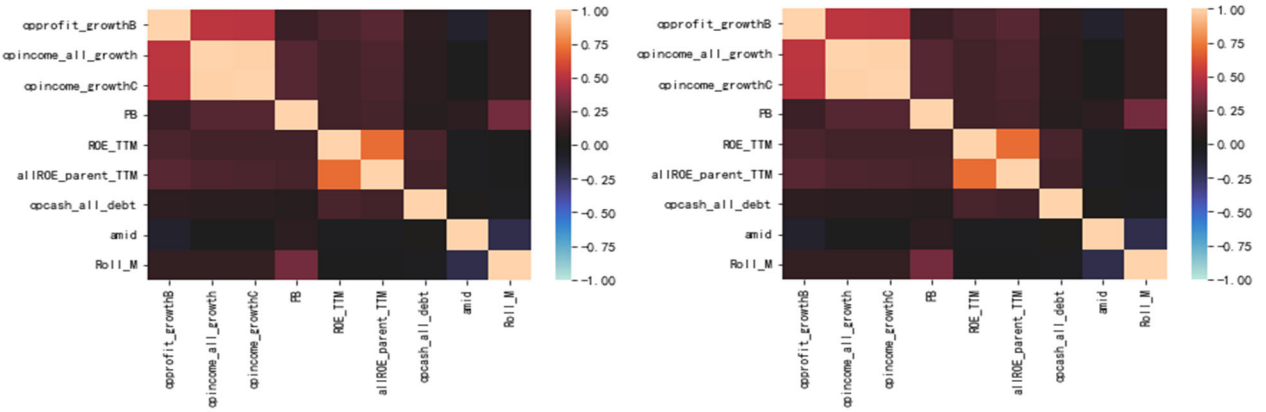


Fig. 7 Factor correlation test's heat map

As can be seen from the graph, there is a strong correlation between the three growth factors, and also a strong correlation between the two operating efficiency factors. Therefore, we excluded the opprofit_growthB, opincome_all_growth and allROE_parent_TTM through a comprehensive

comparison of the factors with strong correlation. Table 8 gives the correlation coefficient between all lastly selected factors. It can be seen that the absolute values of the correlation coefficient between these factors are all less than 0.35, and between fundamental factors or liquidity factor are all less than 0.1, which means that the information overlap between these selected factors is less and they can depict stock information from different angles.

Table 8. Table of factor correlation coefficient

Selected factors	opincome_growthC	PB	ROE_TTM	opcash_all_debt	Amid	Roll_M
opincome_growthC	1.000000	0.216447	0.279239	0.157186	-0.001183	0.113732
PB	0.216447	1.000000	0.330767	0.032836	0.094502	0.278404
ROE_TTM	0.279239	0.330767	1.000000	0.281916	-0.064497	0.084748
opcash_all_debt	0.157186	0.032836	0.281916	1.000000	-0.004049	-0.023772
Amihud metric	-0.001183	0.094502	-0.064497	-0.004049	1.000000	-0.140740
Roll metric	0.113732	0.278404	0.084748	-0.023772	-0.140740	1.000000

4.2.2 Fusion factor test

After the above screening, this paper combines the four fundamental factors of stock quality to build the stock selection quality factor; and add two liquidity factors on this base to obtain the comprehensive factor. The results of the regression method and IC values and rank_IC values tests for these two fusion factors are shown in Table 9 and Fig. 8.

Table 9. Results of two fusion factor regression methods, IC values and rank_IC values

test method	test specification	quality factor	comprehensive factor
Regression method	Absolute mean of t-value series	1.045196	1.029307
	The proportion of t value sequence greater than 2	0.132996	0.131096
	The mean of the t-value sequence	0.245472	0.343509
	The sequence mean of the factor yield	0.001616	0.002013
IC value analysis	IC mean	0.019508	0.026704
	IC standard error	0.100665	0.097114
	IR ratio	0.193793	0.274977
	The portion of IC> 0	57.378087	61.146295
	The portion of IC> 0.02	48.860038	53.261558
Rank_IC value analysis	rank_IC mean	0.018758	0.025228
	rank_IC standard deviation	0.097683	0.094648
	The rank _ IR ratio	0.192025	0.266543
	Proportion of the rank_IC> 0	58.264725	60.956301
	Proportion of rank_IC> 0.02	50.000000	52.786574

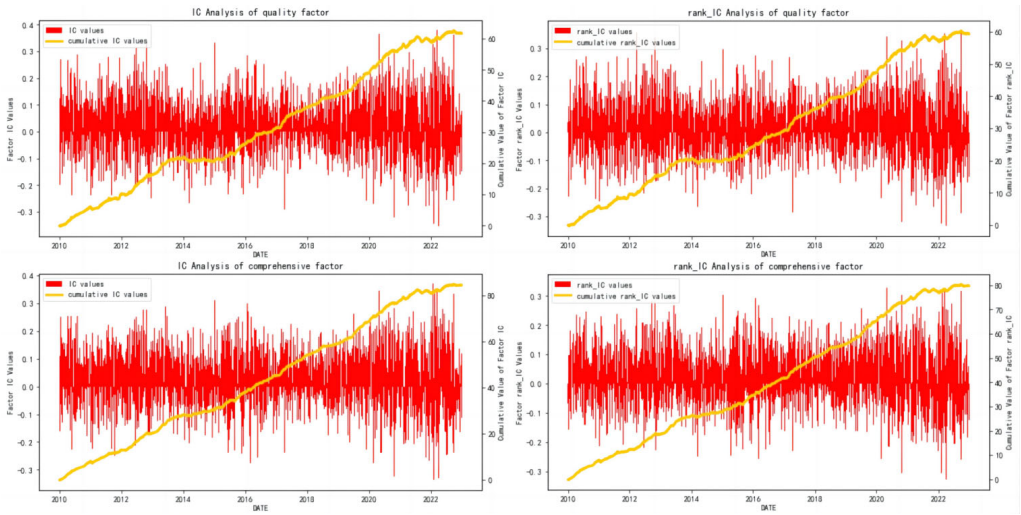


Fig. 8 IC plots of the two fusion factors and rank_IC plots

In the regression test, the sequence mean of the factor yield of the comprehensive factor is slightly higher than that of the quality factor, which indicates that the prediction coefficient of stock return rate increases and the effect is enhanced after the fusion of liquidity factor; and sequence mean of t-values of the comprehensive factor is significantly higher than that of the quality factor, but its absolute mean of t-value series and the proportion of t value sequence greater than 2 is lower, indicating that the overall significance is improved but the stability is slightly decreased. Although the cumulative IC and rank_IC values of the two factors both show an upward trend according to the IC and rank_IC plots, combined with the specific data in Table 9, we could found that the IC and rank_IC of the comprehensive factor are larger, the standard deviation is smaller, and its properties of the evaluation indicators are better than the quality factor.

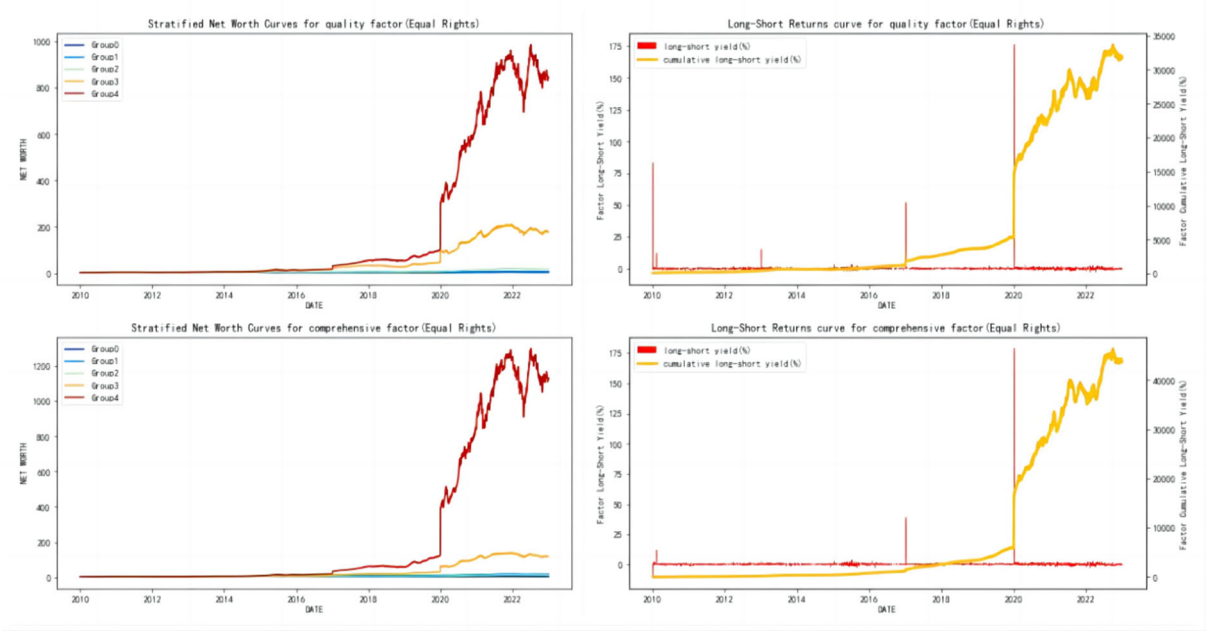


Fig. 9 The two fusion factors' stratified net worth curves and long-short returns curves

Table 10. Stratified back-test results for two fusion factors

Fusion factor name	group	The alized rate of return	Annualized volatility	Sharp ratio	Maximum retraction	Excess annualized rate of return	Excess annualized volatility	informati on ratio	Relative benchmark win rate	Excessma ximum retracting
factor of merit	Group0	6.363	30.737	0.207	-65.981	6.878	19.659	0.350	53.754	-50.196
	Group1	8.808	24.026	0.367	-44.447	9.335	8.798	1.061	54.736	-17.378
	Group2	21.129	29.510	0.716	-43.193	21.716	19.460	1.116	57.460	-16.453
	Group3	40.157	46.533	0.863	-37.413	40.836	40.430	1.010	57.681	-9.536
	Group4	59.368	67.534	0.879	-40.466	60.140	63.576	0.946	58.885	-18.884
	H-L	50.862	53.295	0.954	-16.643	51.594	57.512	0.897	51.631	-62.590
multi-stress	Group0	4.052	30.430	0.133	-57.884	4.556	19.620	0.232	50.871	-37.673
	Group1	14.043	28.240	0.497	-47.963	14.596	17.010	0.858	54.355	-9.153
	Group2	19.301	25.231	0.765	-45.570	19.880	11.522	1.725	57.175	-26.878
	Group3	34.939	36.777	0.950	-41.201	35.593	28.810	1.235	57.681	-9.865
	Group4	67.533	70.764	0.954	-39.169	68.345	66.953	1.021	59.899	-15.575
	H-L	63.934	52.833	1.210	-12.550	64.728	56.882	1.138	54.292	-50.203

Fig. 9 and Table 10 give the results of the two fusion factors' stratified back-test. Fig. 9 shows that the net value of each group of the both two factors has certain monotony, and their cumulative long-short returns increase in the last 12 years as a whole. And the specifically comparative analysis between these two factors can be made according to the values of each evaluation index in Table 10. Since the actual decision usually only pays attention to the return of the group with the largest factor

value and its degree of beating the worst group, the subsequent analysis mainly compares the evaluation index values corresponding to the largest group (Group4) and the long-short combination (H-L) of the two fusion factors. The data in Table 10 show, the largest group's and the long-short combination's excess annualized return and other indicators are all significantly higher than the corresponding indicators of the quality factor, which can be also seen from the vertical coordinate size of the two stratified net worth curves in Fig. 9. That is, the stock selection strategy including the liquidity factors has higher returns. Although the annualized volatility and excess annualized volatility corresponding to the comprehensive factor are slightly higher, its maximum retraction and the excess maximum retraction are lower than the quality factor, showing that the overall earnings volatility of the stock selection strategy using comprehensive factors is slightly larger while the extreme fluctuations are smaller. Therefore, in general, stock selection using the comprehensive factor has less strategic risk, higher return and greater win rate, and better effect than the quality factor composed of only fundamental factors.

5. Conclusion

Based on the multi-factor stock selection model, this paper explores the effect of the stock selection strategy after incorporating the liquidity risk based on the traditional fundamental factors. In this study, 78 fundamental factors and 20 liquidity factors were constructed from the data of CSI 300 stocks from 2010 to 2022. Also, we used the liquidity factor value of stocks to indirectly measure the potential liquidity risk of stocks, that is, the larger the factor value, the smaller the liquidity risk corresponds. The specific research process is as follows: Firstly, we have done a series of pre-processing work for the factor data, including missing value treatment, extreme value treatment, standardized treatment and neutral treatment. Secondly, we used regression method, IC and rank_IC value test and hierarchical back-test method for the single-factor validity test. Based on the results of these test methods and inter-factor correlation analysis, we screened out 4 fundamental factors and 2 liquidity factors. Finally, we used the above fundamental factors and all effective factors to construct the stock quality factors and comprehensive factors respectively, and then tested and compared the stock selection effect of the two different fusion factors by the above three methods. The conclusion is obtained as follows:

In general, the effectiveness and stability of the fusion factor after adding the liquidity factor are better than the fundamental fusion factor in predicting the future returns of stocks and obtaining excess profits. That is to say, the liquidity risk of the stock itself will affect its future earnings, so adding the liquidity risk factor constructed directly or indirectly in the actual stock selection can effectively increase the earnings of the strategy. In this paper, the effective factors used to measure the stock liquidity risk are Amihud and Roll, which measure the stock liquidity from the perspectives of volume-to-price ratio and buy-sell spread. The larger the index value, the better the stock liquidity, and the smaller the liquidity risk. Since existing studies and strategies rarely consider and apply these two indicators, Amihud and Roll indicators have not been overused, so still have a good stock selection effect. In the future, the effectiveness of these two indicators may decrease, and the stock selection ability of other liquidity factors may increase. However, the empirical analysis of this study opens up a new idea for the application of multi-factor stock selection: The liquidity of stocks has a certain role in predicting future returns, which means that measuring and incorporating the liquidity level of stocks can improve the strategic returns to a certain extent.

References

[1] Fama, E. F. & K. R. French. The Cross-Section of Expected Stock Returns[J]. The Journal of Finance. 1992, 47(2), p.427-465.

[2] Fama, E. F. & K. R. French. A Five-factor Asset Pricing Model[J]. Journalof Financial Economics. 2015, 116(1), p.1-22.

[3] Shengmin Zhao, Xiaotian Liu. A multi-factor model of investor preference —— an analysis based on the perspective of prospect theory [J]. China's Economic Problems,2019(02),p.106-121.

[4] Yuxuan Liu, Weize Jin, Liang Yuan. Multi-factor quantitative stock selection model optimization and empirical research —— Analysis of introducing financial cycle indicators [J]. Price Theory and Practice,2022(04), p.141-145.

[5] Yunhui Sun, Chuanzhe Liu. Liquidity risk analysis of China's stock index futures market [J]. The Price Monthly Journal, 2007(09), p.51-53.

[6] Weiqi Liu, Danfeng Li. Analysis of the impact of information disclosure quality on stock liquidity risk [J]. Economic Issue,2014(11),p.25-32.

[7] Amihud, Y., (2002). Illiquidity and stock returns: cross-section and time series effects. Journal of Financial Markets 5,p.31–56.

[8] Ran Zhang, Rongfei Wang. Fundamental quantitative investment: the use of financial analysis and quantitative strategies to obtain excess returns. Peking University Press.2017.7.p.3-20

[9] Xiaohui Hou, Bo Wang. Quantitative investment based on Fundamental Analysis: Research Review and Outlook [J]. Journal of Northeast Normal University (Philosophy and Social Sciences Edition),2021(01),p.124-131+141

[10] Liang Zhou. Study on multifactor stock selection strategy based on quantile regression [J]. Journal of Southwest University (Natural Science Edition),2019,41(01),p.89-96.

[11] Mingtao Wang, Yaming Zhuang. Research on the liquidity risk measurement model of the stock market [J]. Chinese Management Science,2011,19(02), p.1-9.

[12] Goyenko, R., Y., Holden, C., Trzcinka, C., Do liquidity measures measure liquidity? Journal of Financial Economics, (2009). 92, p.153-181.

[13] Roll R., A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. Journal of Finance, (1984). 39, p.1127–1139.

[14] Amihud, Y., Mendelson, H., Asset pricing and the bid-ask spread. Journal of Financial Economics, (1986). 17, p.223–249.

[15] Pastor, L., Stambaugh, R., Liquidity Risk and Expected Stock Returns. Journal of Political Economy, (2003). v111, p.642-685.

[16] Rui Zhao. Research on the learning strategy of the picking robot based on the multifactor quantitative stock selection model [J]. Agricultural Mechanization Research,2023,46(04),p.188-192.

[17] Wenxuan Wang, Lu Li. SCDF based on feature change [J / OL]. Computer Engineering and Application, [2023-08-02], p.1-12

[18] Xiaoyan Wang, Ying Zhou, Tingting Tang, etc. Multifactorial quantitative stock selection study based on Knockoff-Logistic [J]. Statistics and Information Forum,2023,38(04), p.19-32

[19] Yan Yang. Empirical study on the impact of accounting information on stock prices [J]. Modern Economic Information, 2009(16), p.206-207

[20] RICHARDSON S, SLOAN R, YOU H. What makes stock prices move? Fundamentals vs. Investor recognition[J]. Financial Analysts Journal,2012,68(2), p.30-50

[21] Can Huang, Qingshan Jiang. Equity liquidity in China: retesting based on the impact mechanism [J]. Management Science,2023,36(02), p.144-158.