

# Quantitative Stock Selection Strategy Analysis Based on Multi-factor Model Under the Epidemic Situation

Senhan Hu<sup>1,†</sup>, Yujin Kong<sup>2,\*†</sup>, and Fan Tong<sup>3,†</sup>

<sup>1</sup>Shenzhen MSU-BIT University, Computational Math and Controls Department, 518172 Shenzhen, China

<sup>2</sup>Tianjin Renai College, Finance Engineering Department, 301636 Tianjin, China

<sup>3</sup>Wenzhou Business College, Computer Science Department, 325035 Wenzhou, China

<sup>†</sup>These authors contributed equally.

**Abstract.** Contemporarily, COVID-19 undoubtedly had a profound impact on the current Chinese economy, as well as presented various challenges and opportunities for different industries in China. As a matter of fact, these changes may profoundly affect the future development trends and competitive landscape of industries. In recent years, the number of stock investors had continued to increase, surpassing 200 million in 2022. In a highly competitive market, the ability to select high-quality stocks often determines the final returns. This study selected the constituent stocks of the "CSI 500" as the "stock pool" for stock selection between 2020 and 2023. By analysing the degree of factor exposure, the correlation between factors, and IC values of factors, suitable and effective factor combinations were selected, and an evaluation model was established using multiple regression analysis to obtain a trading strategy. To ensure its feasibility and rationality, the strategy was simulated using backtesting before actual trading, ultimately aiming to achieve excess alpha returns.

## 1 Introduction

With the increasing complexity and scale of China's capital market, the number and types of investments are increasing day by day, and the investment channels and investment methods available to both individual and institutional investors are gradually increasing. Quantitative investment has started earlier in foreign countries and is relatively large in terms of market size and number of institutions, while quantitative investment strategies based on machine learning models have developed rapidly in China, and quantitative investment, as an advanced technology from overseas capital markets, has been widely sought after in the domestic market.

In a competitive market, the ability to select high-quality stocks often determines the final return. Quantitative techniques have received increasing attention and exploration, Xu verified the effectiveness of quantitative stock selection strategies based on the industry rotation multi-factor model and the multivariate regression model based on fixed effects [1]. Zou selected the constituent stocks of the SSE 300 Index as the stock pool for multi-factor model analysis, selected effective factors, and constructed a multi-factor model suitable for general investors, obtaining a conclusion that financial quality is related to the stock returns of the company [2]. Wu analyzed and summarized the development context of the multi-factor model, and made comments and expectations on it while analyzing its application [3]. Geng used the Logistics model to

select stocks on the ChiNext board and obtained the stocks that need to be combined [4]. Lin found that factor timing technology can significantly improve the effectiveness of multi-factor stock selection models, which not only increases returns but also reduces risk [5]. Liu introduced the China Financial Conditions Index (FCI) synthesized by the GARCH model into the multi-factor stock selection model and constructed a multi-factor stock selection model based on industry rotation and financial cycles [6].

It is clear from the above that there is a lack of research on the CSI 500 constituents in scholars' studies, and this paper argues that in the special case of an epidemic, returns are also affected considerably. Therefore, only by studying the trend of stocks under the epidemic can we help to explore quantitative investment in the future and thus improve the return. Therefore, this paper will explore the quantitative investment of multi-factor models under the epidemic from three aspects: adding factors, improving models, and optimizing backtesting:

- Adding factors. This paper will expand the analysis method of the multi-factor stock selection strategy and add new factors related to the impact of fundamentals on stock returns in the company's operations to broaden the selection range of effective factors.
- Improving the model. By analyzing the degree of factor exposure, the correlation between factors, and the IC value of factors, the paper selects suitable and effective factor

\* Corresponding author: [1811000232@mail.sit.edu.cn](mailto:1811000232@mail.sit.edu.cn)

combinations and uses multiple regression analysis to establish an evaluation model to ultimately obtain a trading strategy. Meanwhile, based on the stock data from 2020 to 2023, the paper obtains the most effective factor combinations in the past three years through updating time, thus establishing a model.

- Backtesting optimization. After obtaining effective factors, this paper will also backtest the research results on the Ricequant quantitative platform, repeatedly testing and improving them, to ultimately obtain excess returns.

## 2 Data & Method

First, this article obtains more universal factors from a benchmark perspective. The "CSI 300" reflects the trend of "large-cap stocks," the "SSE 50" reflects the trend of "mega-cap stocks," and the "CSI 500" reflects the overall trend of small and mid-cap stocks. To ensure that the results of the experiment are significant, this paper uses the "CSI 500," which consists of 800 stocks with higher market capitalization from the Shanghai and Shenzhen stock markets, as the "stock pool."

As the number of basic factors is too large, combining factors will greatly increase the complexity

and calculation difficulty of derivative factors. Therefore, in this data study, a multi-factor model and common factors are used to replace the characteristics of individual stocks, thereby reducing the calculation difficulty in this data study. In the selection of factors, growth, market sentiment, and valuation-related factors (which are highly active and closely watched in the market) are chosen, including *ps\_ratio*, *pcf\_ratio*, *pb\_ratio*, *pe\_ratio*, *turnover\_ratio*, and *profit growth rate*, etc. [7]. In data analysis, factor exposure analysis is used to calculate factor exposure, to display the weight of the factor in the multi-factor combination model.

As shown in Table 1, this paper analyzes 11 factors, including turnover rate, earnings per share, return on equity, asset turnover, capitalization, market capitalization, price-to-sales ratio, price-to-cash flow ratio, price-to-book ratio, price-to-earnings ratio, and year-on-year growth rate of net profit.

The data sources for this paper mainly come from Joinquant and China Securities Index Co., Ltd., covering the period from 2020 to 2023, using the "CSI 500" as the stock pool. In the stock selection process, stocks with a unit price below 20 and low activity (turnover rate less than 4%) were excluded. To avoid the impact of meaningless data on the study, the data were cleaned and appropriately standardized using Z-score normalization in the data processing.

**Table 1.** Textual selection factors.

| Market factors | Profit factors     | Operating factors    | Debt-paying factor | Valuation factors    | Growth factors            |
|----------------|--------------------|----------------------|--------------------|----------------------|---------------------------|
| Turnover ratio | EPS                | Asset turnover ratio | Capitalization     | Market value         | Growth rate of net profit |
|                | Asset profit ratio |                      |                    | Marketability        |                           |
|                |                    |                      |                    | Market current ratio |                           |
|                |                    |                      |                    | PB                   |                           |
|                |                    |                      |                    | P/E                  |                           |

## 3 Results & Discussion

### 3.1 Exposure & Correlation Analysis of Factor

By analyzing the factor exposure in the constituent stocks of the index, This article can determine whether the factors are universal and represent significant characteristics of the benchmark [8]. This article selected 11 factors and calculated the average exposure of each factor in 2020-2023. The average deviation of *market\_cap* and *capitalization* is the highest, which is almost around 0.3. This also indicates that the current price of these stocks has increased by approximately 30% relative to the corresponding benchmark price. Secondly, based on the results, the deviation of *inc\_net\_profit\_year\_on\_year* and *pcf\_ratio* factors is almost 0 between 2020 and 2023, indicating that the effectiveness of these two factors is low.

Highly correlated factors often lead to significant errors in linear models, resulting in a decline in the model's predictive ability. To obtain a better combination of factors, this paper analyzes the correlation between factors to avoid selecting highly correlated factors in the factor combination. In this paper, the mean of the correlation and standard deviation of the correlation were calculated. According to the results, index for judging the strength of the correlation between factors was obtained by dividing the two values. The larger the absolute value of the index, the stronger the correlation between the factors. According to the results, it is intuitive that the correlation strength between these five groups of factors is much greater than that of the other groups, namely, *capitalization*, *pb\_ratio*, *ps\_ratio*, *volume*, and *turnover\_ratio*. In this paper, by studying the IC value of factors, the explanatory power of factors is analyzed to determine the ability of factors to predict the future returns of individual stocks [9]. Except for *turnover\_ratio* and *volume*, which have relatively high

IC values, and pcf\_ratio, which has a relatively low IC value, the other factors have similar IC values.

### 3.2 Regression & Backtesting

In selecting a combination of investment factors, it is advisable to choose those factors that have high deviation, low correlation, and strong stock selection ability. This paper selected turnover\_ratio, capitalization, and ROE as independent variables, and performed multiple regression analysis with the daily rate of return as the dependent variable. The regression equation obtained is:

$$Y = 0.041433 \times \text{turnover\_ratio} - 0.000005 \times \text{capitalization} - 0.094065 \times \text{ROE} + 0.468433. \quad (1)$$

According to the William-F-Sharp market model, beta is a measure used in fundamental analysis to determine the volatility of an asset or portfolio relative to the overall market [10]. In this paper, the beta coefficient is chosen as the metric used to measure systematic risk and is back-tested by entering the selected factors on the Ricequant Rice Basket quantitative platform. There is a relationship between

the return of an individual security or a portfolio and the return of a market index, and its correlation coefficient is defined as the  $\beta$  value.

When the  $\beta$  coefficient of a security or portfolio is equal to 1, its systematic risk is exactly the same as that of the market portfolio; when  $\beta > 1$ , its systematic risk is greater than that of the market portfolio, and it can be called an active security; when  $\beta < 1$ , its systematic risk is less than that of the market portfolio, and it is a defensive security; when  $\beta = 0$ , it means there is no systematic risk. Similarly, in this paper, all sectors, all industries, and all stocks after filtering ST in "CSI 500" are considered as a portfolio, and the sensitivity of its return to the market index can be expressed by  $\beta$ . As shown in Table 2, based on the results of the regression analysis, this paper has selected the following indicators. Subsequently, the annual stock returns for the period 2020/1/1 to 2023/1/1 were back-tested and the following results were obtained. As presented in Table 3. In the backtest results,  $\alpha = 0.0933$ , indicating that the strategy achieved excess returns relative to the risk.  $\beta = 0.6339$ , indicating that the investment portfolio moves in the same direction as the benchmark. In addition, the backtest results show good Sharpe ratio, Sortino ratio, and Information ratio.

**Table 2.** Selection of backtest indicators in this paper.

| Stock Selection Indicators    | Sorting Indicators(Ascending)                | Trading Models  | Risk Control                                                   |
|-------------------------------|----------------------------------------------|-----------------|----------------------------------------------------------------|
| Transaction amount >1 billion | Total Market Capitalization=-0.000005        | Regular rolling | Take Profit Stop<br>Loss:<br>Individual Stock Stop<br>Loss 20% |
|                               | Average turnover rate for the year =0.041433 |                 | Broad Market<br>Timing:<br>Transaction amount >10 billion      |
|                               | ROE=-0.094065                                |                 |                                                                |

**Table 3.** Backtesting Summary

|                                        |             |
|----------------------------------------|-------------|
| Start time                             | 2020/1/1    |
| Ending time                            | 2023/1/1    |
| Stock initial capital                  | 1000000     |
| Benchmark                              | 000300.XSHG |
| Backtest Return                        | 28.37%      |
| Backtest annualized return             | 9.03%       |
| Benchmark Return                       | -5.49%      |
| Benchmark Annualized Return            | -1.94%      |
| Accumulated Net Value                  | 1.284       |
| Excess Return Maximum retracement rate | 62.016      |
| Annual Win Rate                        | 0.6667      |
| Quarterly Win Rate                     | 0.6667      |
| Monthly Win Rate                       | 0.5833      |
| Daily Win Rate                         | 0.4835      |
| Alpha                                  | 0.0933      |
| Beta                                   | 0.6339      |
| Sharpe ratio                           | 0.3809      |

|                                |        |
|--------------------------------|--------|
| Sortino ratio                  | 0.5238 |
| Information ratio              | 0.5335 |
| Annualized volatility          | 0.2327 |
| Maximum retracement            | 23.89% |
| Annualized Tracking Error      | 0.2066 |
| Annualized Downside Volatility | 0.1692 |

## 4 Conclusion

In summary, this article selected the data of "CSI 500" for the three-year period of 2020-2023, and conducted experiments with multiple regression model, in which this article showed that ROE, Capitalization, Turnover ratio, and three combination factors play a significant role in stock selection, and updated the time of the model. The validity of the factors is fully ensured by the factor exposure analysis and IC value correlation analysis. Nevertheless, there are still shortcomings in the experiment. For example, the use of a large number of factors and data to select stocks may lead to over-fitting risk, resulting in poor performance of the factor model in the future, as well as failing to respond to changes that occur in the market due to the inability to represent the current market situation, and other shortcomings that need to continue to be explored in depth.

## References

1. J. Xu, Exp. in Fin. Theory, **173(03)**, 30-38 (2017)
2. L. Zou, China Busi. J., **847(24)**, 100-103 (2021)
3. Y. Wu, Z. Zhao, Ent. Tech. & Dev., **478(08)**, 64-66 (2021).
4. Y. Geng, G. Wang, T. Gu, et al. J. of Heilongjiang Ins. of Tech. (Comprehensive Edition), **21(06)**, 97-104 (2021)
5. Y. Lin, Y. Shen, Y. Li, et al. Infor. Tech. & Stan., **438(06)**, 44-50 (2021).
6. Y. Liu, W. Jin, L. Yuan L. Price Theory & Prac., **454(04)**, 141-145 (2022)
7. E. F. Fama, K. R. French, J. of Fin. Eco., **116** (2015).
8. POLYWIDE: Factor analysis article series (IX): a multi-factor research framework, Retrieved from: <https://www.joinquant.com/view/community/detail/96e60785423795aecfc99aba1957c7a9?type=1>
9. China Galaxy Securities Research Institute, P. Lai, Multi-factor series: Determination of factor combinations of multi-factor model systems Retrieved from: <https://www.doc88.com/p-47347328581053.html>
10. STEVEN NICKOLAS, Beta Formula: How to Calculate the Beta of a Stock, Retrieved from: <https://www.investopedia.com/ask/answers/070615/what-formula-calculating-beta.asp>