Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/47/20230374

Analysis of the Effectiveness of Alpha Strategy for Optimized Multi-Factor Models

Guanzhen Lu[1,a,*]

[1]*School of Finance, Jiangxi University of Finance and Economics, No. 169 Shuanggang East Street, Nanchang, China*
a. 2201904949@stu.jxufe.edu.cn
**corresponding author*

Abstract: The emergence of ChatGPT signifies the advent of a new artificial intelligence revolution. Generative AI technologies, represented by ChatGPT, will profoundly afferct the research paradigm in finance and the systemic ecosystem of the financial industry. This study primarily focuses on the stocks within the ChatGPT concept sector of China's A-share market and investigates the effectiveness of a multi-factor Alpha stock selection strategy in this sector. Firstly, the research optimizes the Fama-French three-factor model and establishes a four-factor model. Using this optimized four-factor model, stock selection is performed within the ChatGPT concept sector, resulting in the formation of an investment portfolio. Based on backtesting, the strategy is found to be highly effective, though it requires a certain level of risk tolerance from investors. These results represent a proactive attempt to apply the Alpha strategy to the ChatGPT concept sector of China's A-share market, providing a preliminary examination of its performance characteristics and offering valuable insights for further research by subsequent scholars.

Keywords: ChatGPT concept sector, Alpha strategy, multi-factor model

1. Introduction

ChatGPT, a large-scale language model developed by OpenAI, has attracted significant attention across various fields in recent times. Unlike traditional language models, ChatGPT departs from the approach of designing independent model structures for specific and singular language tasks. Instead, it capitalizes on extensive data to capture the interrelationships between language tokens. It primarily adopts the Decoder structure of the Transformer model, while accomplishing various language tasks by adjusting the input-output formats specific to each task. Currently, the model has evolved to the GPT 4.0 version. Zhang evaluates this model as possessing significant characteristics such as big data, high computing power, advanced algorithms, and human-like intelligence. It has found widespread applications in industries or domains such as education, research, e-commerce, social media, finance, media, and manufacturing [1]. Concerning the field of finance, Wang et al. analyze the application and prospects of ChatGPT in finance. They propose eight research directions and issues worthy of attention, including ChatGPT's relationship with financial data security, sentiment analysis in financial markets, and personalized investment advisory [2]. Qiu and Lan highlight that the development of ChatGPT and similar AI technologies presents opportunities for the finance industry to enhance risk control capabilities, market insight, and competitive advantage. However, they also acknowledge the challenges it brings in terms of systemic financial risks, regulatory complexity, and technological limitations [3].

The theoretical foundation of the multi-factor stock selection model is established upon the mean-variance model proposed by Markowitz. This model quantifies the expected return and risk of risky assets by measuring the mean and variance of their returns, thereby examining the issues of portfolio construction and selection [4]. The original multi-factor model initially included only one factor, namely the CAPM (Capital Asset Pricing Model) proposed by Sharpe as an improvement and extension of the mean-variance model. The CAPM asserts that the expected return of an asset or portfolio is linearly related to the market risk premium [5]. Due to the inability of the CAPM to explain anomalies other than systematic risk, subsequent research by Fama and French incorporated two new factors into the CAPM: value factor and size factor, leading to the development of the renowned Fama-French three-factor model. This model enhanced the explanatory power of portfolio returns and provided better estimation of potential returns on asset portfolios in the market [6]. Following the introduction of the Fama-French three-factor model, various scholars have made improvements and advancements, proposing additional multi-factor models, e.g., the four-factor model, five-factor model. These models have been tailored and optimized by considering the development of capital markets in different countries and regions, leading to the inclusion and refinement of specific factors. Tian et al. conducted a comparative analysis of the application of the Fama-French three-factor model in the stock markets of China and the United States. They found that the Chinese market exhibited prominent systematic risk and demonstrated the presence of a size effect. However, the B-M ratio effect was insignificant. Compared to the US stock market, Chinese stock market was not fully realized its resource allocation function [7]. Considering the unique issue of shell value in the Chinese A-share market, Liu et al. reconstructed the size and value factors, proposing the Chinese version of the three-factor model. This model effectively explained most of the cross-sectional anomalies observed by the academic community in the Chinese market [8].

The Alpha strategy originated from the CAPM, which suggests a positive linearly relationship between portfolio return and the risk of it. According to the CAPM, asset or portfolio return equals to the risk-free rate plus the compensation for systematic risk. However, empirical studies have found differences between the expected returns calculated by the CAPM and the actual returns of portfolios, there exist excess returns or alpha. Because the CAPM relies on several restrictive assumptions and the securities market is not completely efficient. The Alpha strategy involves selecting a stock portfolio expected to generate high returns in the future and constructing a hedge fund to mitigate systematic risk by incorporating financial derivatives. Kang examined the mean reversion characteristics of the single-factor model Alpha using the A-share market in China as an example. He found that combining the Alpha strategy with quantitative stock selection and hedging the spot stock portfolio with stock index futures can achieve higher returns [9]. Li also verified the effectiveness of the Alpha strategy in the A-share market through a multi-factor model [10]. Ji analyzed the sources of returns and reasons for drawdowns in the Alpha strategy. Considering the characteristics of the Chinese A-share market, a specific strategy was designed that could achieve relatively stable absolute returns in historical market conditions and adapt to the current market conditions as well [11].

This study aims to optimize the traditional Fama-French three-factor model and primarily investigate the effectiveness of the optimized multi-factor model-based Alpha strategy in the Chinese A-share market, specifically focusing on the ChatGPT concept sector. The structure of this paper is as follows. The first section is the introduction about related research and background. The second section focuses on the optimization of the three-factor model. The third section involves stock selection and backtesting based on the optimized model, analyzing the effectiveness of the strategy. The fourth section provides a discussion on limitations and concludes the paper.

2. Model Construction

2.1. Sample Selection and Data Sources

The stocks selected in this paper are sourced from the ChatGPT concept sector of the China A-share market. The time window spans from January 1, 2018, to May 31, 2023, with the testing period covering January 1, 2018, to December 31, 2021. This period is utilized for factor correlation analysis and stock selection. The back testing period, on the other hand, encompasses January 1, 2022, to May 31, 2023, and is employed to analyze the effectiveness of the strategy. For the sake of data stability and completeness, further screening of the sample stocks is necessary. Stocks with the "ST" or "*ST" notation are excluded. This category of listed companies has significant uncertainties in their prospects, and the extreme price fluctuations of their issued stocks in the short term do not reflect the true market conditions. It can lead to empirical results that diverge greatly from the actual situation. Stocks with a listing period of less than one year are also excluded. Such stocks are in the early stages of listing and may be subject to IPO premium pricing issues. Additionally, due to their short-listing period, relevant financial information may not have been disclosed yet or the disclosed financial information may be incomplete. Stocks with necessary data missing and those that have not been in existence for at least 5 years and 5 months are also eliminated to ensure the completeness of data and an adequate sample size within the research period. After applying the screening and processing to all the stocks in the ChatGPT concept sector, a total of 40 listed company stocks remained as the final sample. This sample will serve as the data foundation for this study, and all data are sourced from the CSMAR and RESSET database. The empirical analysis was conducted by using Stata and Python.

2.2. Variable Description

This paper improves upon the Fama-French three-factor (henceforth referred to as FF3) model and the original mathematical expression of the FF3 model is as follows:

$$E\left(R_{i,t}\right) - R_{f,t} = \beta_{i,MKT}(MKT_t) + \beta_{i,HML}(HML_t) + \beta_{i,SMB}(SMB_t) \tag{1}$$

The MKT factor (means market) measures the risk premium of the market and is calculated as the expected return of the market portfolio minus the risk-free rate. The HML factor (means high minus low) measures the value effect of stocks and is calculated by the difference between the expected return of value stocks and growth stocks. The SMB factor (means small minus big) represents the simulated portfolio return of the book-to-market (B-M) factor at time t and is used to measure the size effect of stocks. This study primarily considers optimizing the basic FF3 model by incorporating liquidity-related factors. The candidate factors include the turnover ratio (Turn), the Amihud indicator (Amihud), the Roll indicator (Roll), and the Zeros indicator (Zeros). The factor from the candidate pool with the highest absolute correlation with stock returns (Return) will be included in the FF3 model, thus forming the four-factor model used in this study. The data frequency for each factor and stock returns is monthly.

2.3. Correlation Analysis

Descriptive statistical analysis was conducted on the data used in this study, and the results are demonstrated in Table 1. The monthly average return for the sample stocks during the testing period is 0.86%, with a standard deviation of 14.16%, showing significant volatility in the returns of ChatGPT concept sector stocks. In terms of the four liquidity indicators, the average turnover rate is 64.55%, and the average values for the other three indicators are close to zero, suggesting overall good liquidity in these stocks. The correlation analysis between the four factors in the candidate factor pool, Turnover Ratio, Amihud Indicator, Roll Indicator, Zeros Indicator, and monthly stock return is demonstrated in Table 2. Seen from Table 2, it can be observed that the monthly stock returns are positively correlated with turnover rate, and negatively correlated with Amihud, Roll and Zeros. This indicates that stocks with higher liquidity tend to generate higher returns. Comparing the absolute correlation coefficients, the correlation between monthly stock returns and turnover rate is the highest. Therefore, this indicator will be included in the FF3 model to form the fourth factor.

Table 1: Descriptive statistics for the entire sample.

Variable	N	Max	Min	Mean	Std
Return	1920	1.0766	-0.4766	0.0086	0.1416
Turn	1920	7.8043	0.0190	0.6455	0.6170
Amihud	1920	0.2288	0.0007	0.0230	0.0223
Roll	1920	0.0120	0.0000	0.0006	0.0006
Zeros	1920	0.0138	0.0000	0.0004	0.0010
MKT	1920	0.1452	-0.0818	0.0048	0.0454
HML	1920	0.1140	-0.0808	0.0014	0.0397
SMB	1920	0.0724	-0.0859	-0.0051	0.0340

Table 2: Correlation coefficient matrix of candidate factors.

	Return	Turn	Amihud	Roll	Zeros
Return	1.0000				
Turn	0.3181***	1.0000			
Amihud	-0.0736***	-0.1767***	1.0000		
Roll	-0.1500***	-0.2439***	0.7992***	1.0000	
Zeros	-0.0684***	-0.2290***	0.2870***	0.2932***	1.0000

Note: *** indicates significance at the 1% level, ** indicates significance at the 5% level, * indicates significance at the 10% level. The definitions for all other tables in this paper remain the same.

2.4. Model Optimization

To ensure the accuracy of the optimized multi-factor model, it is necessary to test whether there are redundant factors among the four factors to avoid multicollinearity issues. The correlation analysis of the MKT, HML, SMB, and Turn factors is conducted, and the results are presented in Table 3.

Table 3: Four-factor correlation coefficient matrix.

	MKT	HML	SMB	Turn
MKT	1.0000			
HML	-0.3363***	1.0000		
SMB	0.1002***	-0.4454***	1.0000	
Turn	0.0837***	-0.0922***	0.1267***	1.0000

The results of the correlation analysis indicate that the absolute correlation coefficients between variables are not high, suggesting the absence of severe multicollinearity among the variables. Therefore, the mathematical expression of the four-factor model formed by including the turnover rate factor in the FF3 model is as follows:

$$E(R_{i,t}) - R_{f,t} = \beta_{i,MKT}(MKT_t) + \beta_{i,HML}(HML_t) + \beta_{i,SMB}(SMB_t) + \beta_{i,Turn}(Turn_{i,t}) \quad (2)$$

The following equation will be used in the empirical analysis part in the next chapter:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MKT}(MKT_t) + \beta_{i,HML}(HML_t) + \beta_{i,SMB}(SMB_t) + \beta_{i,Turn}(Turn_{i,t}) + \varepsilon_{i,t} \quad (3)$$

3. Empirical Analysis

3.1. Portfolio Construction

Using the regression model given by Equation (3), a multiple linear regression analysis is conducted on the 40 sample stocks from the ChatGPT concept sector. The individual stock alpha returns are recorded and ranked in descending order. The period considered for this analysis is from January 2018 to December 2021, and the results are presented in Table 4. According to the principle of the Alpha strategy, which aims to achieve excess returns, all Alpha returns are sorted in descending order, and the top 5 stocks are selected. These stocks are Thunder Software Technology Co., Ltd. (300496), Business-Intelligence of Oriental Nations Corporation Ltd. (300166), Shanghai Fengyuzhu Culture Science and Technology Co., Ltd. (603466), LingNan Eco and Culture-Tourism Co., Ltd. (002717), and Shenzhen Tianyuan Dic Information Technology Co., Ltd. (300047). The stocks are treated with equal weights, assuming that the stocks are divisible, and an equally weighted portfolio is formed. Each stock is assigned the weight of 20%.

Table 4: Summary of individual stock alpha returns in the chatGPT concept sector.

Stock Code	Alpha Return	Stock Code	Alpha Return
300496	0.0252	600415	-0.0445
300166	-0.0006	002292	-0.0469
603466	-0.0038	002362	-0.0470
002717	-0.0118	300133	-0.0486
300047	-0.0136	300002	-0.0510
002642	-0.0137	600718	-0.0568
300479	-0.0171	000676	-0.0581
002230	-0.0185	300520	-0.0590
002649	-0.0189	300315	-0.0596
600100	-0.0215	300188	-0.0604
002878	-0.0264	300663	-0.0615
000851	-0.0279	600449	-0.0616
300245	-0.0286	300459	-0.0620
300033	-0.0294	300291	-0.0715
600797	-0.0304	300465	-0.0801
300352	-0.0317	300058	-0.0864
002123	-0.0321	300612	-0.0920
002354	-0.0323	300418	-0.0995
002803	-0.0354	300182	-0.1026
002229	-0.0419	300364	-0.1084

3.2. Analysis of Strategy Effectiveness

Backtesting was conducted on the equally weighted portfolio, and its performance was compared against the Shanghai Shenzhen 300 Index (CSI 300). The backtesting period was from January 1, 2022, to May 31, 2023. The cumulative return curve during the backtesting period is illustrated in Figure 1. Figure 1 provides a visual representation of the cumulative return's comparison between the equally weighted portfolio and the CSI 300 Index. Overall, the cumulative returns curve of the portfolio constructed using the Alpha strategy is above that of the CSI 300 Index. During the entire backtesting period, the cumulative return of the equally weighted portfolio is -2.69%, while the CSI 300 Index has a cumulative return of -23.11%. The annualized returns are -1.90% and -16.93% respectively. The equally weighted portfolio generates higher excess returns but exhibits higher volatility. The maximum drawdown during the backtesting period for the portfolio is -36.43%, with an annualized standard deviation of 38.35%, compared to -23.42% and 20.42% for the CSI 300 Index. Therefore, investors need to have a higher risk tolerance. Additionally, the Sharpe ratio of the equally weighted portfolio is 3.37%, while the CSI 300 Index has a Sharpe ratio of -30.22%. This indicates that the equally weighted portfolio enjoys higher risk-adjusted returns despite the higher risk. Investors who can bear the associated risks can potentially achieve higher returns from the equally weighted portfolio. Based on the analysis above, the equally weighted portfolio constructed using the Alpha strategy demonstrates the ability to generate higher excess returns, but it also requires a higher risk tolerance from investors. The optimized multi-factor model's Alpha selection strategy shows effectiveness when applied to the ChatGPT concept sector.

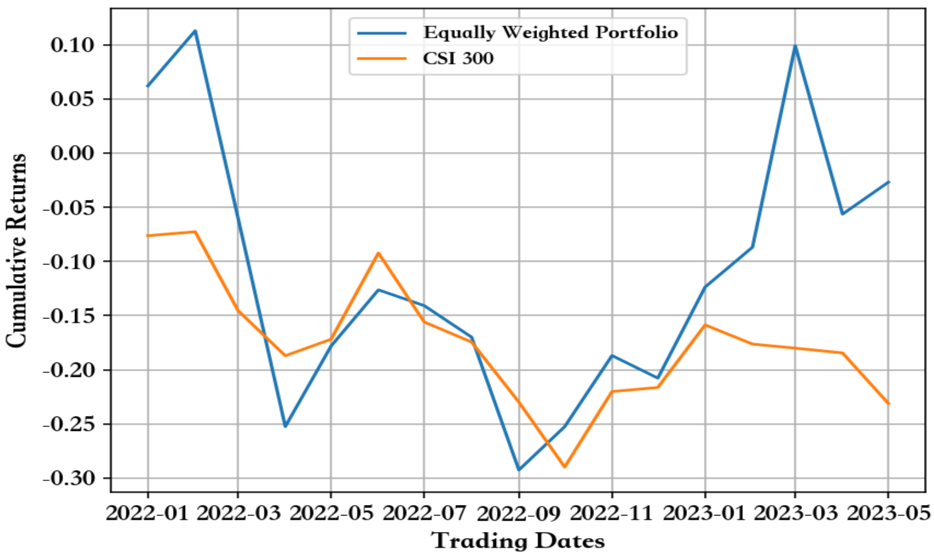

Figure 1: Cumulative returns comparison.

4. Limitations

The limitations of this paper are as follows. Firstly, the stock selection and backtesting phases did not consider rebalancing operations, which led to a lack of flexibility in the strategy and potential biases in the selection of portfolio holdings, resulting in a decrease in overall accuracy. Secondly, the study only focused on the ChatGPT concept sector in the Chinese A-share market, and the effectiveness of the strategy in other sectors or the entire A-share market remains to be further analyzed.

5. Conclusion

To sum up, this study constructs a multi-factor Alpha strategy for stock selection and portfolio construction based on the FF3 model. Firstly, the three-factor model is optimized by incorporating the turnover rate factor. Then, 40 stocks from the ChatGPT concept sector in the Chinese A-share market are selected. The four-factor model regression is performed using the stock's monthly returns

DOI: 10.54254/2754-1169/47/20230374

data from January 2018 to December 2021. The top five stocks with the highest Alpha returns are chosen to construct an equally weighted portfolio. The effectiveness of this strategy is empirically studied using the stock returns data and market information from January 2022 to May 2023. The main findings of this study are as follows. The four-factor model Alpha strategy, incorporating the turnover rate factor, demonstrates high effectiveness in the ChatGPT concept sector of the Chinese market. It can generate excess returns and outperforming the benchmark index, the Shanghai Shenzhen 300 Index. However, investors need to have a higher risk tolerance to implement this strategy. This study represents a proactive attempt to apply the Alpha strategy in the ChatGPT concept sector of the Chinese A-share market, providing initial insights into the performance characteristics of the strategy and offering valuable references for future researchers in this area.

References

[1] Zhang, X.H., (2023) Potential Social Changes Triggered by the New Generation of Artificial Intelligence Technology (ChatGPT). Review of Industrial Economics, 3, 22-30.

[2] Wang, S.Y., Li, M.C., Yang, K., Lin, W.C., Jiang, S.R. and Wei, Y.J. (2023) ChatGPT+Finance: Eight Noteworthy Research Directions and Issues. Management Review, 4, 3-11.

[3] Qiu, D.Y. and Lan, Y. (2023) Opportunities, Challenges, and Issues Brought by ChatGPT to the Financial Industry. Southwest Finance, 6, 1-12.

[4] Markowitz, H. (1952) Portfolio Selection. The Journal of Finance, 1, 77.

[5] Sharpe, W.F. (1964) Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. The Journal of Finance, 3, 425-442.*

[6] Fama, E.F. and French, K.R. (1992) The Cross-Section of Expected Stock Returns. The Journal of Finance, 2, 427-465.

[7] Tian, L.H., Wang, G.Y. and Zhang, W. (2014) Three-Factor Model Pricing: How Does it Differ Between China and the United States?. Studies of International Finance, 7, 37-45.

[8] Liu, J.N., Stambaugh, R.F. and Yuan, Y. (2018) Size and value in China. Journal of Financial Economics, 1.

[9] Kang, Z.Y. (2018) Quantitative Stock Selection Strategy Based on Alpha Mean Regression Characteristics. Shanghai Normal University.

[10] Li, K.L. (2021) Empirical Research on Multiple-factor Alpha Strategy Based on A-share Market. Harbin Institute of Technology.

[11] Ji, Y. (2017) A Design of Alpha Strategy in the Domestic Stock Market. University of International Business and Economics.