The Optimization and Test Method of Quantitative Stock Selection Strategy Based on Factor Model in Emerging Markets

Zhengrui Lv

Pingdu No.1 High School of Shandong Province, Pingdu, China

13061452860@163.com

Keywords: Quantitative stock selection; Multi-factor model; Emerging markets; Model optimization

Abstract: This study aims to investigate and optimize the quantitative stock selection strategy based on the multi-factor model and conduct empirical tests in the context of emerging markets. As emerging markets evolve and expand, traditional investment strategies encounter increased challenges and risks. Meanwhile, quantitative stock selection has gained investors' attention due to its scientific and systematic approach. This study begins by defining and constructing a multi-factor stock selection model with six basic factors: market, scale, value, momentum, quality, and volatility. It uses advanced statistical methods and machine learning techniques to tune the model parameters and optimizes the factor weights through genetic algorithms. This study backtracks historical data for backtesting and uses forward-looking out-of-sample tests to verify the stock selection effect of the model. The research indicates that the optimized multi-factor model can reliably generate positive excess returns in multiple emerging markets, taking into account transaction costs and liquidity constraints. This finding validates the model's applicability and reliability in a diversified market environment. By empirical tests, this study also reveals the differences in factor returns in emerging markets and provides a powerful reference for investors to formulate more accurate investment strategies. The findings of this research are significant for understanding investment opportunities in emerging markets and enhancing the theoretical application of quantitative investment models.

1. Quantitative Stock Selection Strategy Optimization

As an important branch of modern financial engineering, quantitative stock selection uses mathematical models and computer technology to screen out stocks with investment value from many stocks. The traditional quantitative stock selection strategy depends on historical data and statistical analysis. However, due to the rapid change in market structure and the explosive growth of information, it is essential to optimize the quantitative stock selection strategy [1].

1.1 Factor Screening and Optimization

Factor screening is the key step to quantify investment, and reasonable factors provide accurate market trend guidance for formulating investment strategies. When screening factors in emerging markets, it is important to consider not only the traditional factors such as market, scale, value, momentum, quality, and volatility but also those specifically relevant to emerging markets, including country risk, liquidity, and policy impact. Through historical regression analysis, the factors with statistical significance and stable prediction ability can be screened out. For instance, it has been demonstrated that the market value factor is more significant in small-cap companies operating within emerging markets. These findings necessitate empirical validation through data analysis. To illustrate this point, let us consider a hypothetical scenario. We posit that the average excess rate of return for small-cap companies within emerging markets is 15% over five years, as evidenced by a back test of an emerging market.

1.2 Model Parameters Optimization

After the stock selection factor is determined, parameter optimization is important in improving the strategy's performance. Machine learning algorithms, especially support vector machine (SVM),

  DOI: 10.25236/gemmsd.2025.023

ensemble learning, and artificial neural networks, are widely used to adjust model parameters. They optimize decision boundaries and parameter settings by automatically learning complex patterns in historical data. For example, when optimizing the factor weights, a genetic algorithm (GA) can be used to iteratively update and find the optimal weight combination [2]. Through testing, it has been determined that the model performs best when the market factor weight is set at 20%, the scale factor weight at 25%, and other factors share the remaining weights accordingly. The test results provide clear guidance for the research.

1.3 Dynamic Adjustment of Strategy

A quantitative stock selection strategy must adapt to the dynamic changes in the market. In empirical testing, we should evaluate the strategy's performance using static historical data and recognize the need for dynamic adjustments to the strategy. It involves many factors, such as analyzing market status, risk preference, and capital flow, and adjusting stock selection strategies or factor weights according to these factors to adapt to the instability of emerging markets. For example, in a certain period, it is found that the responsiveness of emerging markets to value investment has increased, so the weight of value factors can be increased promptly.

1.4 Risk Control of Optimization Strategy

An effective risk control system must be established for the optimization strategies to ensure that the possible losses are reduced while pursuing the benefits. In emerging markets, managing risk is crucial due to limited market liquidity and a lack of transparency. Effective risk management strategies encompass controlling market risks (such as leveraging sound investment insights), credit, liquidity, and operational risks. Analysts often employ the Value at Risk (VaR) model to gauge the potential maximum loss within a portfolio, enabling them to establish stop-loss thresholds to mitigate significant financial setbacks [3]. In addition, the model needs to incorporate stress tests to test the strategy's performance in extreme market conditions. For instance, a simulation test discovered that when sudden political events cause market volatility to spike, implementing an optimization strategy can cut losses by trimming positions and making timely adjustments.

To sum up, optimizing a quantitative stock selection strategy based on a factor model is not static but a multidimensional dynamic adjustment process involving data analysis, model construction, back-test verification, and risk control. Different market characteristics and uncertainties lead to additional challenges in emerging markets. Ongoing research, empirical testing, and timely strategy adjustments significantly enhance the performance and adaptability of quantitative stock selection strategies in emerging markets.

2. Emerging Markets Data Collection and Processing

2.1 Definition and Choice of Emerging Markets

Emerging markets usually refer to those countries or regions with high economic development levels, but their capital markets are not as perfect as mature markets. For investors, emerging markets offer substantial growth potential and promising return prospects. However, they also come hand in hand with elevated risks. When defining emerging markets, we rely on the emerging market index of Morgan Stanley Capital International (MSCI). We pick four typical emerging market countries, China, India, Brazil, and Russia, to conduct empirical research.

2.2 Data Sources and Collection

Regarding data sources, we mainly collect data through multiple channels: first, some data are official data, such as public data of stock exchanges and central statistical offices in various countries. Second, some data originated from international databases, such as Bloomberg, Datastream, CRSP, Compustat, and other financial and economic databases. Finally, some data come from professional research reports and financial analysis articles. We collected financial statements, transaction, and market macroeconomic data, and the data directly related to the investment of all listed companies in these emerging markets from 2010 to 2020.

2.3 Data Preprocessing

In the data preprocessing, we cleaned the data and dealt with missing values and abnormal values. For instance, missing values are addressed by utilizing the industry median for replacement or data from the previous period to maintain continuity and integrity. In the outlier detection, the box chart method is adopted, and the data points exceeding 1.5 times the interquartile distance (IQR) are regarded as outliers to be eliminated or adjusted. Additionally, we have standardized different accounting standards for financial data to make them conform to International Financial Reporting Standards (IFRS) [4].

2.4 Analysis of the Characteristics of Emerging Market Data

The characteristics of data from emerging markets are analyzed in detail. Unlike mature markets, emerging markets are characterized by higher volatility and skewness of stock returns. Using China as a case study, we have observed that the average daily fluctuation for small to mid-cap stocks has surged to 2.5% over the past half-decade. This figure dwarfs the 1.2% volatility of the S&P 500 index during the same stretch. The distribution of stock returns skews heavily toward the positive, with a high occurrence of outlier gains. This trend underscores investors' penchant for high-stakes ventures and hints at potential market irrationalities. In addition, the capital control and foreign exchange policy in emerging markets has also caused the particularity of research, which has a potential impact on capital liquidity and return on investment.

The above data collection and processing work laid a solid foundation for the subsequent factor model construction and stock selection strategy optimization. The processed data enhances the accuracy of subsequent analyses and allows for the identification of unique investment patterns in emerging markets.

3. A Test of Emerging Markets' Strategies

3.1 Test Indicators and Methods

3.1.1 Test Indicators

This study selected several key indicators to comprehensively evaluate the effectiveness of quantitative stock selection strategy based on factor model in emerging markets.

Sharpe Ratio: It is evident that the strategy in question yields supplementary benefits that surpass the risk-free return when undertaking unit risk. The calculation formula is:

$$Sharpe\ Ratio = Rp - Rf/\sigma p$$

In the formula, Rp represents the average rate of return of the strategy; Rf represents the risk-free rate of return; σp represents the standard deviation of the rate of return. A higher Sharpe ratio means that the strategy can achieve better returns at the same risk. [5]

Information Ratio: It measures the excess return ability and stability of the strategy relative to the benchmark index. The formula is express as:

$$\text{Information Ratio=Rp-Rb/}\sigma\text{p-b}$$

Rb is the return rate of the benchmark index, and σp-b indicates the standard deviation of the excess return rate using the strategy.

Max Drawdown: It is used to describe the maximum loss that may be suffered during a specific period of time using this strategy. It reflects the risk tolerance of the strategy. The calculation formula is:

$$\text{Max Drawdown} = \max_t \left(1 - \frac{V_t}{V_{\text{peak},t}}\right)$$

In the formula, V_t represents the net value of the strategy at the time of t, and V_{peak} represents the highest net value before the time of t. [6]

3.1.2 Test Method

In historical data backtesting, stock data from emerging markets is divided into in-sample and out-of-sample data based on a specific time interval. Strategy optimization and parameter adjustment are performed on the in-sample data, and strategy verification is performed on the out-of-sample data. Additionally, the Monte Carlo simulation method is used to analyze the strategy's sensitivity and evaluate its stability in different market environments.

3.2 In Sample Test

3.2.1 Data Selection

In our study of quantitative stock selection strategies in emerging markets, we have chosen several representative stocks to demonstrate the overall investment opportunities and risks. The sample design is specifically intended to model the constituent stocks of major stock indexes in emerging markets from 2010 to 2018. The selected time period encompasses the full economic cycle experienced by the market, including periods of both bull and bear markets and the rapid change phase, thereby ensuring the wide applicability and credibility of the obtained analysis results.

Considering the need for security and benchmark comparison, it is reasonable to assume that the risk-free rate of return is 3% per year, which is close to the long-term national debt yield of major economies worldwide in the same period. Moreover, using the composite index of emerging markets as the benchmark index enables us to assess the advantages and disadvantages of quantitative strategies from the perspective of relative performance.

3.2.2 Strategy Implementation

When constructing the portfolio, we use the approach based on the factor model according to the predefined stock selection criteria, such as market value, price-earnings ratio, price-to-book ratio, historical volatility, and trading volume. Score and weight each factor quantitatively, and select the stock with the highest score to be included in the monthly portfolio. It enables regular adjustments to stock holdings, aligns the portfolio's activity with market trends, and accurately reflects market dynamics and changes in individual stocks.

In this study, the daily rate of return of selected stocks and the whole portfolio during the sample period is calculated to analyze and compare the performance of a fixed portfolio and an adjusted portfolio according to a factor model. Through the calculation of the daily rate of return, we can determine the cumulative rate of return, risk-adjusted rate of return, and potential maximum drawdown of the portfolio. These key indicators not only assess the success of the quantitative strategy but also provide essential data for optimizing and refining the strategy.

3.2.3 Results of the Test

From the results of the in-sample test, the average return rate of the strategy reaches 12.5 %, and the Sharpe ratio is 0.8, indicating that the strategy can obtain certain additional benefits when undertaking unit risk. The information ratio is 0.6, suggesting that the strategy has a moderate ability to achieve excess returns compared to the benchmark index. However, the maximum drawdown is 25%, indicating that there is a significant risk of loss during the sample period. As shown in Table 1:

Table 1 Results of the in-sample test

Indicators	Value
Average rate of return	12.5%
Sharpe ratio	0.8
Information ratio	0.6
Max drawdown	25%

3.3 Out-of-Sample Test

3.3.1 Data Selection

The out-of-sample test is an important step to verify the performance of quantitative investment strategies in the market environment. Therefore, the selected data cycle has a significant impact on the accuracy and reliability of the test results. This study's out-of-sample data includes the representative three-year period from 2019 to 2021. It encompasses a range of emerging market conditions, such as the effects of the US-China trade war, the volatility of the global economy due to the COVID-19 pandemic, and the market's reactions to various political and economic events. Choosing this time frame allows us to assess the strength and flexibility of quantitative investment strategies during market turmoil and uncertainty. [7].

3.3.2 Strategy Implementation

Based on the strategies and parameters optimized from the sample data, we apply the same quantitative model to select stocks and invest under market conditions beyond the sample. This requires us to follow the same frequency and factor weight as in the sample to ensure the consistency and rationality of the strategy. During the out-of-sample period, the portfolio is adjusted once a month according to the strategy's guidance, and the daily portfolio yield is calculated based on the results of these adjustments.

Throughout this process, we meticulously monitor the market landscape and our portfolio's performance, gauging whether our strategy is adept at capitalizing on market uptrends while mitigating or delaying the potential negative effects on our investments. Furthermore, by calculating real-time performance indicators, such as excess return, Sharp ratio, and maximum drawdown, we can comprehensively evaluate the sustained performance of the strategy in influencing unknown market data and determine whether it needs further adjustment and optimization. The findings reveal the strategy's efficacy during the historical era and furnish a robust foundation for anticipating future market changes.

3.3.3 Results of the Test

The results of the out-of-sample test show that the average rate of return of the strategy has dropped to 9%, and the Sharp ratio and information ratio have also decreased, which are 0.6 and 0.4, respectively. The maximum drawdown was expanded to 30%. It shows that the strategy's performance outside the sample is not as good as that inside the sample, and there may be over-fitting. Therefore, the strategy needs to be further optimized. As shown in Table 2:

Table 2 The results of the out-of-sample test

Indicators	Value
Average rate of return	9%
Sharpe ratio	0.6
Information ratio	0.4
Max drawdown	30%

3.4 Sensitivity Analysis

3.4.1 Analysis Method

The Monte Carlo simulation method simulates a strategy's performance in various market conditions by randomly altering the key parameters that influence it. The key parameters selected include factor exposure, warehouse adjustment frequency, etc.

3.4.2 Analysis Results

After conducting many simulations, the distribution of the Sharp ratio and maximum drawdown of the strategy under different parameter combinations is obtained.

From the results of sensitivity analysis, different parameter combinations have significant effects on the performance of the strategy. The average Sharp ratio of group 3 is the highest, while the average maximum drawdown is relatively low. It indicates that the risk-return characteristics of the strategy under this parameter combination are better. In summary, in practical application, suitable parameter combinations can be chosen based on market conditions to enhance the stability and effectiveness of the strategy. As shown in Table 3:

Table 3 Test results

Parameter combinations	Mean Sharpe Ratio	The mean value of the maximum drawdown
Group 1	0.7	28%
Group 2	0.5	32%
Group 3	0.8	26%

To sum up, the quantitative stock selection strategy based on factor model performs well in the in-sample test of emerging markets, but there is a certain degree of attenuation in the out-of-sample test. We can find a more suitable parameter combination using sensitivity analysis to further optimize the strategy and improve its adaptability and profitability in emerging markets.

4. Conclusion

The in-depth analysis and empirical test of this study provide a comprehensive evaluation framework for applying a quantitative stock selection strategy based on the factor model in emerging markets. The results show that the well-designed and optimized multi-factor model has potential and value in emerging markets. Regarding factor screening and optimization, the research shows that traditional factors like scale, value, and momentum can possess a certain degree of forecasting prowess in emerging markets. Moreover, country risk, liquidity, and policy influence, which are closely tied to the characteristics of emerging markets, are also of particular significance in the stock selection process. By systematically analyzing and combining various factors, we can more precisely get a handle on the market dynamics.

In model parameter optimization, machine learning technology shows excellent data fitting and pattern recognition ability. The strategy's performance has significantly improved after applying the algorithm to optimize factor weights and the decision boundary. Moreover, the adjusting strategies dynamically ensure that it maintains good adaptability and responsiveness in different stages of the market. As one of the important links of quantitative stock selection strategy, risk control can ensure that potential losses can be avoided while pursuing benefits. By applying the VaR model and market stress test, the researchers precisely forecasted the strategy's performance under extreme conditions and secured the investment safety margin.

The strategy has achieved relatively positive back-test results in the in-sample and out-sample tests, showing its effectiveness in historical data. However, the performance in the out-of-sample test is slightly unsatisfactory, which reminds researchers to be alert to the phenomenon of over-fitting and to constantly adjust their strategies to adapt to the actual market. Sensitivity analysis further strengthens this view, indicating that fine-tuning strategy parameters are crucial to improving their adaptability.

References

[1] Hu Y. Multi-Factor Stock Selection Model Based on Categorical Prediction Model[J]. Open Access Library Journal, 2023, 10(7): 1-8.

[2] Alhijawi B, Awajan A. Genetic algorithms: Theory, genetic operators, solutions, and applications[J]. Evolutionary Intelligence, 2024, 17(3): 1245-1256.

[3] Kilian L, Zhou X. The econometrics of oil market VAR models[M]//Essays in honor of joon y. park: Econometric methodology in empirical applications. Emerald Publishing Limited, 2023: 65-95.

[4] Tsalavoutas I, Tsoligkas F, Evans L. Compliance with IFRS mandatory disclosure requirements: a structured literature review[J]. Journal of International Accounting, Auditing and Taxation, 2020, 40: 100338.

[5] Kan R, Wang X, Zheng X. In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models[J]. Journal of Financial Economics, 2024, 155: 103837.

[6] Riley T, Yan Q. Maximum drawdown as predictor of mutual fund performance and flows[J]. Financial Analysts Journal, 2022, 78(4): 59-76.

[7] Gordeev D, Singer P, Michailidis M, et al. Backtesting the predictability of COVID-19[J]. arXiv e-prints, 2020.