Analysis of the Effectiveness of Stock Selection Strategy Based on Quality Indicators in Stock Market Volatility

Kangrui Lu [a]*, Yongxin Zhao[b]

[a]18438589811@163.com, [b]13261028029@163.com

Hebei Finance University, BaoDing,China

Abstract. This study aims to explore the effectiveness of a quality-based stock selection strategy during periods of stock market volatility. We utilize the real data from the CSI 300 Index from January 2014 to April 2023 and divide this period into stable and volatile periods using the historical volatility threshold method. Through machine learning-based multiple regression analysis, we construct a quality-based stock selection strategy based on profitability, growth, and safety factors, and conduct strategy backtesting. The results show that the performance of this strategy is better during volatile periods compared to stable periods. Therefore, we conclude that the quality-based stock selection strategy is effective during periods of stock market volatility and provides useful guidance for investors' decision-making.
CCS CONCEPTS • Applied computing~Law, social and behavioral sciences~Economics • Computing methodologies~Machine learning~Machine learning approaches~Logical and relational learning~Statistical relational learning • Mathematics of computing~Probability and statistics~Multivariate statistics

Keywords: Quantitative investment; Machine learning; Quality indicators; Stock market volatility

1 Introduction

In recent years, substantial changes in the global macro environment have resulted in an increasingly intricate volatility landscape in the Chinese stock market. During periods of stock market oscillation, investors grapple with the challenge of discerning market trends and the genuine value of stocks, complicating the formulation and execution of stock selection strategies. As market fluctuations become more intricate, investors urgently seek a relatively robust investment strategy to maintain the stability and growth of their assets. In this context, quantitative stock selection strategies based on quality indicators have garnered considerable attention. In contrast to traditional indicators, quality indicators focus more on a company's intrinsic value and long-term profitability, thereby mitigating the impact of market volatility on stock selection strategies to a certain extent. Therefore, investigating the effectiveness of stock selection strategies based on quality indicators during periods of stock market oscillation holds significant theoretical and practical importance.

Researchers globally have extensively explored quality indicators. For example, Zaremba A [1] discovered that profitability indicators contribute to explaining stock returns. Lalwani V et al.

MSIEID 2023, December 08-10, Guangzhou, People's Republic of China

DOI 10.4108/eai.8-12-2023.2344712

[2] conducted a study on stock selection strategies based on four fundamental quality indicators, namely market, size, value, and momentum, and identified that quality-based investment portfolios in the Indian stock market can achieve excess returns. Asness C S et al. [3] defined quality as characteristics for which investors are willing to pay a higher price, and established quality indicators based on profitability, growth, and safety. Wu W [4] constructed a quantitative investment strategy using quality indicators to balance returns and risks. Bradrania R et al. [5] observed that high-quality stocks exhibit no beta anomaly.

Beyond quality indicators, this study delves into the classification of stock market volatility and stability periods. The Historical Volatility Threshold method serves as a common classification technique, enabling researchers to analyze and predict stock market volatility using Bollerslev T's Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. Bhowmik R et al. [6] found that, due to the lack of similar characteristics in the stock market, only one GARCH model is highly sensitive to the analysis of market volatility and returns. Consequently, the selection of both the stock market and the model becomes somewhat challenging. Yao Y et al. [7] utilized the HAR-LSTM model to study the volatility of the Chinese stock market and found it to have strong predictive ability.

Moreover, machine learning is extensively applied in quantitative investment. Khan W et al. [8] utilized machine learning algorithms to predict stock market volatility, highlighting the impact of external factors through social media and financial news data. Kumar G et al. [9] conducted a comprehensive survey on stock market predictions, emphasizing computational intelligence methods. Ayala J et al. [10] integrated technical indicators with machine learning for trading decisions, enhancing competitiveness in trading signals and proposed rules, as evidenced in tests on indices like Ibex35 (IBEX), DAX, and Dow Jones Industrial Average (DJI).

In summary, this study addresses a notable gap in the existing literature by delving into the limited research on constructing quality indicators using machine learning methods and the insufficient exploration of their performance during stock market oscillation periods. Consequently, this paper strategically selects profitability, safety, and growth as the pivotal construction variables for quality indicators. Building upon this conceptual framework, a quantitative stock selection strategy is formulated. To effectively classify distinct market phases, the research employs the historical volatility threshold method. This approach not only contributes to the theoretical understanding of the subject but also provides valuable practical insights for investors navigating the complexities of the stock market.

2 Research Methods and Theory

2.1 Historical Volatility

Historical volatility of stocks is an indicator that measures the level of price fluctuations in the stock market. It reflects the extent of price fluctuations of a stock over a certain period of time. Generally, periods with high volatility can be considered as volatile periods, while periods with low volatility can be considered as stable periods. The historical volatility of a stock can be calculated using the following formula[11]:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{a}^{b}(R_t - \mu)^2} \tag{1}$$

where σ represents daily volatility, N is the time window for calculating historical volatility, R_t represents daily returns, and μ is the mean of N-day returns. In practical calculations of historical volatility, a sliding time window N can be used to obtain a historical volatility curve. Then, by analyzing the historical volatility curve and selecting an appropriate threshold, the stock market can be divided into volatile periods and stable periods.

2.2 Selection of Quality Indicators

Based on literature research and rational investment requirements for good company performance, stable growth, and low risk, this paper constructs quality indicators from three dimensions: profitability, growth, and safety. The calculation methods for the three-dimensional factors are shown in Table 1.

The performance of profitability reflects the results of a company's performance. This study selects three factors, namely gross margin ratio (GMAR), return on assets (ROA), and return on equity (ROE), to measure the profitability of a company.

Growth represents the potential for future development of a company. This paper selects three factors, namely net asset growth rate (net_asset_growth), net profit growth rate (net_profit_growth), and total asset growth rate (total_asset_growth), to measure the growth of a company.

Safety represents a company's ability to resist risks. A higher level of safety indicates lower risk, corresponding to lower beta coefficient and debt-to-asset ratio for stocks. This paper selects three factors, namely 21-day beta (beta21), 126-day beta (beta126), and debt-to-asset ratio (debt_asset_ratio), to measure the safety of a company. Among them, Beta coefficient is the estimated coefficient obtained by regressing the excess returns of individual stocks against the excess returns of the market during the same period, using weighted least squares regression. The formula is as follows:

$$r_t - r_f = \alpha + \beta\left(r_m - r_f\right) + \varepsilon \tag{2}$$

where r_t represents the return rate of a stock, r_f represents the risk-free rate, and r_m represents the return rate of the market portfolio, which is represented by the CSI 300 Index. α is the constant term, and ε is the error term. β is the regression beta coefficient.

Table 1: Calculation Methods for the Three Dimensions of Factors

Dimension	Calculation Method
Profitability	Gross margin = (Revenue - Cost of Goods Sold) / Revenue
	Return on assets = Net profit / Average total assets for the latest four quarters
	Return on equity = Net profit / Average shareholder's equity for the latest four quarters
Growth	Net asset growth rate = (Current shareholder's equity / Previous year's shareholder's equity) - 1
	Net profit growth rate = (Current year's net profit / Previous year's net profit) - 1
	Total asset growth rate = (Current year's total assets / Previous year's total assets) - 1

Dimension	Calculation Method
Safety	Beta21 represents the volatility of the stock relative to the market in the past 21 trading days
	Beta126 represents the volatility of the stock relative to the market in the past 126 trading days
	Debt-to-asset ratio = Total liabilities / Total assets

3 Data Selection and Processing

3.1 Research Object and Data Source

The data used in this study is derived from the constituent stocks of the CSI 300 Index in China. The research period spans from January 1, 2014, to April 30, 2023, covering a total of 10 years of sample data. Within this research period, multiple bull and bear market cycles are included to ensure the robustness of the results. The quantitative backtesting software employed in this study is the Auto Trader Quantitative Research Platform, and the programming software used is Python.

3.2 Distinguishing Between Periods of Stock Market Volatility and Stability

This study calculates historical volatility using the closing prices of the CSI 300 Index on the last trading day of each month, from January 2014 to April 2023. The upper part of Figure 1 displays the monthly closing price information, while the lower part depicts the historical volatility curve and its average value.

Figure 1: Trading Information and Historical Volatility Curve of the CSI 300 Index

To differentiate between periods of stock market volatility and stability, the average historical volatility is used as a threshold. Periods where the historical volatility curve fluctuates above and below the average are considered as volatile periods, while periods with historical volatility below the average are seen as stable periods. Since this study aims to examine the effectiveness of quality strategies during volatile periods, and violent fluctuations are influenced by various factors and have a low correlation with quality strategies, data from periods of violent fluctuations are excluded. Therefore, only the effectiveness of quality strategies during stable and volatile periods is compared and analyzed.

From Figure 1, we can see that the average historical volatility is 0.003. From June 2016 to December 2018, the historical volatility is below the average, indicating a stable period. From January 2019 to April 2023, the historical volatility fluctuates around the average, indicating a volatile period. The subsequent research in this study uses this division as a benchmark to compare and analyze the performance of quality strategies during periods of stock market stability and volatility.

3.3 Construction of Quality Indicators

Based on the selection of quality factors in section 2.2, quality indicators are constructed based on three dimensions: profitability, growth, and safety. Multiple regression analysis from machine learning is employed to regress the three factors within each dimension, and factors with larger regression coefficients are selected as representatives for each dimension to participate in the construction of the quality indicators.

3.3.1 Multiple Regression for Selecting Representative Factors

The first step is to use multiple regression analysis in machine learning to select the most representative factors. Regression analysis refers to using various factors of stocks (company characteristics) as independent variables and historical returns as the dependent variable for multiple regression analysis. The multiple regression model constructed in this study is as follows:

$$y = w_1x_1 + w_2x_2 + w_3x_3 + b \tag{3}$$

Here, y represents historical returns; x_1, x_2, x_3 represent the three factors to be regressed, taking profitability as an example, x_1, x_2, x_3 represent roa, roe, and gross_income_ratio respectively; w_1, w_2, w_3 are the regression coefficients corresponding to the three factors, and b is the bias term in multiple regression.

During the process of machine learning, the regression coefficients and bias term are continuously adjusted based on the principle of gradient descent, aiming to minimize the loss value in each iteration. The multiple regression model is considered to have a good fit when the loss decreases continuously and the regression accuracy reaches above 85%[12].

In the specific implementation, a random selection of 20% of stocks from the CSI 300 Index was made, and multiple regression analysis was conducted for each of the three dimensions - profitability, growth, and safety - using the constructed multiple regression model. The regression results are shown in Table 2.

Table 2: Multiple Regression Results

Dimension	Factor 1	Factor 2	Factor 3
Profitability	gross_income_ratio	roa	roe
	0.17304328	0.1390492	0.27459017

Dimension	Factor 1	Factor 2	Factor 3
Growth	net_asset_growth	net_profit_growth	total_asset_growth
	0.24098279	0.34245178	0.10796858
Safety	debt_asset_ratio	beta21	beta126
	-0.05812929	-0.11940734	-0.2260211

In Table 2, each dimension has three factors, and each factor is associated with a regression coefficient. The largest coefficient in each dimension was chosen as the representative factor for that dimension: return on equity for profitability, net profit growth rate for growth, and 126-day beta for safety.

3.3.2 Construction of Quality Indicators based on the Selected Factors

After the regression analysis, the selected three factors are used to construct the quality indicators. Since the factors have different units and magnitudes, simply adding the factor values lacks practical significance. Therefore, it is necessary to standardize each factor using Z-score normalization. Additionally, considering that extreme values can affect the effectiveness of the factors, this study employs the method of trimming extreme values to process the factor data.

It should be noted that the impact direction of 126-day beta is different from that of return on equity and net profit growth rate on cumulative returns. Therefore, for the construction of the quality indicators, the negative sign is added to 126-day beta. The construction method for the quality indicators is as follows:

Profitability is represented by ROE and processed as follows:

$$\text{Profitability} = Z(\text{roe})$$

Growth is represented by net profit growth and processed as follows:

$$\text{Growth} = Z(\text{net\_profit\_growth})$$

Safety is represented by -126-day beta and processed as follows:

$$\text{Safety} = Z(-\text{beta126})$$

Here, Z represents Z-score normalization.

The construction method for the quality indicator is as follows:

$$\text{Quality} = \text{Profitability} + \text{Growth} + \text{Safety}$$

3.4 Validation of Quality Indicator

After constructing the quality indicators, a stratified backtesting approach was used to verify the effectiveness of the indicators. The steps for stratified backtesting are as follows:

Initially, all sample stocks were ranked in ascending order based on their quality indicator values. The stocks were then divided into 5 groups of equal size based on their rankings. Each group was equally weighted, resulting in 5 investment portfolios labeled as 1 (lowest), 2, …, 5 (highest).

Next, the returns of the quality factor were calculated, and the composition of the portfolios was adjusted weekly. Through a retrospective test, the return series of the 5 portfolios was obtained, showing a trend in stratification.

Figure 2 displays the cumulative returns of each group based on the quality indicators. The purple line represents the cumulative returns of the highest quality group (Group 5), while the red line represents the cumulative returns of the lowest quality group (Group 1). Throughout the backtesting period, although the cumulative returns curves of Group 4 and Group 5 intersected at times, the overall cumulative returns of the 5 groups diverged significantly, demonstrating clear differentiation. The results of stratified backtesting indicate that stock selection based on this quality indicator is effective.

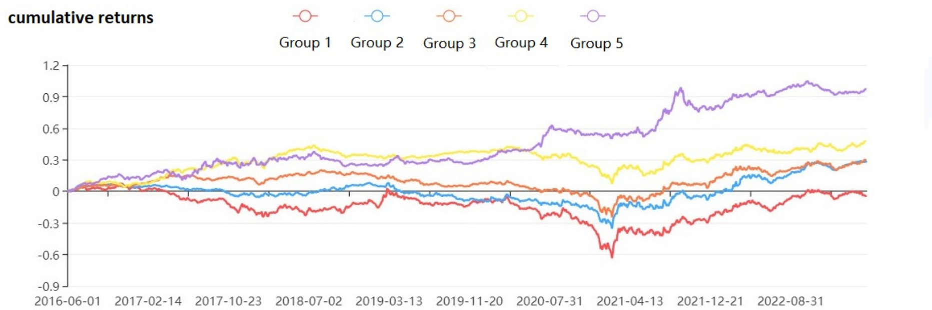

Figure 2: Cumulative Returns of the Quality Indicators by Group

4 Empirical Analysis of the Quality Strategy

4.1 Strategy Construction and Backtesting

The quality strategy constructed in this study is based on the quality indicators mentioned earlier. The specific steps for constructing the strategy are as follows: firstly, the sample stocks are ranked in ascending order based on the quality indicators, and the top 20% ranked stocks are selected as the investment targets. The strategy is rebalanced monthly, where the sample stocks are sorted again according to the quality indicators. Stocks that no longer meet the selection criteria are sold, and new stocks that meet the criteria are bought, while retaining previously purchased stocks.

To validate the effectiveness of this strategy, backtesting is conducted. The backtesting period is from June 1, 2016, to April 30, 2023, covering both periods of market stability and volatility. The results of the backtesting are as follows: the cumulative return is 192.07%, compared to the benchmark return of 27.48% for the same period of the CSI 300 Index. The annualized return of the strategy is 13.72%. The backtesting results demonstrate that the stock selection strategy based on the quality indicators achieves excess returns throughout the testing period, confirming its effectiveness. Detailed backtesting results are shown in Table 3.

Table 3: Results for Different Backtesting Periods

	Cumulative Return (%)	Benchmark Return (%)	Annualized Return(%)	Information Ratio (%)	Annual Turnover Ratio (%)
Stable Period	21.35	-4.74	9.96	256.30	154.32
Volatile Period	165.75	35.68	18.84	682.73	130.85
Both	192.07	27.48	13.72	851.32	138.83

4.2 Analysis of Strategy Backtesting Results

Based on the backtesting results during different periods shown in Table 3, we can draw the following conclusions: the quality-based stock selection strategy performs differently in different market conditions, with better performance during periods of significant market volatility compared to relatively stable periods. The specific analysis is as follows:

4.2.1 Stable Period (January 1, 2016, to December 31, 2018)

The backtesting results of the strategy during stable periods are as follows: the cumulative return is 21.35%, compared to a benchmark return of -4.74%, indicating significant excess returns during stable periods. The annualized return rate is 9.96%, significantly higher than the benchmark return rate, indicating that the strategy can consistently generate stable returns in relatively calm market conditions. The information ratio is 256.30%, further emphasizing the stability and superiority of the strategy during stable periods. A high information ratio suggests that the strategy's excess return relative to risk is reasonable. The one-sided annualized turnover rate is 154.32%, indicating that the strategy's trading activity is moderate during stable markets, helping to reduce trading costs and maintain the relative stability of the investment portfolio.

4.2.2.Volatile Period (January 1, 2019, to April 30, 2023)

The backtesting results of the strategy during volatile periods (January 1, 2019, to April 30, 2023) are as follows: the cumulative return reaches 165.75%, compared to a benchmark return of 35.68%. The strategy achieves significant excess returns during this period as well. The annualized return rate is 18.84%, demonstrating stronger investment performance compared to the benchmark during this period of market volatility. This highlights the strategy's flexibility and its ability to achieve significant profits during market fluctuations. The information ratio is as high as 682.73%, further confirming the superiority of the strategy during this period. A high information ratio indicates a good balance between risk control and excess returns. The one-sided annualized turnover rate is 130.85%, slightly lower than during the stable period but still at a reasonable level. This indicates that the strategy can still adjust flexibly in volatile markets to adapt to market changes.

In summary, the performance of the quality-based stock selection strategy differs significantly in different market environments, especially outperforming during periods of significant market volatility compared to relatively stable periods.

Figure 3 Performance of the Strategy during Stable Period(a) and Volatile Period(b)

5 Conclusion

This study explores the effectiveness of a quality-based stock selection strategy during periods of stock market volatility. Based on profitability, growth, and safety dimensions, a quality-based stock selection strategy was constructed and analyzed using the real data from the constituents of the CSI 300 Index from January 2014 to April 2023. The performance of this strategy during periods of market volatility and stability was compared.

Empirical results demonstrate that the quality-based stock selection strategy achieved significant outperformance during periods of market volatility. In contrast, during stable periods, the strategy's performance did not significantly differ from the overall market. Therefore, this study suggests that the quality-based stock selection strategy has some effectiveness during periods of market volatility. These findings provide investors with a viable stock selection strategy to mitigate market risks and potentially achieve excess returns during periods of market volatility.

In conclusion, this study finds that the quality-based stock selection strategy has some effectiveness during periods of market volatility and provides investors with some guidance for investment decisions during stock market turbulence.

References

[1] Zaremba A. Quality investing and the cross-section of country returns[J]. Studies in Economics and Finance， 2016.

[2] Lalwani V, Chakraborty M. Quality investing in the Indian stock market[J]. Managerial finance, 2018, 44(2): 127-141.

[3] Asness C S，Frazzini A，Pedersen L H. Quality minus junk[J]. Review of Accounting Studies，2019，24(1): 34-112.

[4] Wu W. Risk Management of Enterprise Quantitative Investment Strategies through Data Modeling[J]. Journal of Engineering, Project & Production Management, 2023, 13(1).

[5] Bradrania R, Veron J F, Wu W. The beta anomaly and the quality effect in international stock markets[J]. Journal of Behavioral and Experimental Finance, 2023, 38: 100808.

[6] Bhowmik R, Wang S. Stock market volatility and return analysis: A systematic literature review[J]. Entropy, 2020, 22(5): 522.
[7] Yao Y, Li Y. Modeling and Forecasting of Fractal Volatility of Highly Volatile Markets[J]. IAENG International Journal of Applied Mathematics, 2022, 52(4): 1-9.
[8] Khan W, Ghazanfar M A, Azam M A, et al. Stock market prediction using machine learning classifiers and social media, news[J]. Journal of Ambient Intelligence and Humanized Computing, 2020: 1-24.
[9] Kumar G, Jain S, Singh U P. Stock market forecasting using computational intelligence: A survey[J]. Archives of computational methods in engineering, 2021, 28: 1069-1101.
[10] Ayala J, García-Torres M, Noguera J L V, et al. Technical analysis strategy optimization using a machine learning approach in stock market indices[J]. Knowledge-Based Systems, 2021, 225: 107119.
[11] Garman M B, Klass M J. On the estimation of security price volatilities from historical data[J]. Journal of business, 1980: 67-78.
[12] Hastie T, Tibshirani R, Friedman J H, et al. The elements of statistical learning: data mining, inference, and prediction[M]. New York: springer, 2009.