Construction and Empirical Study of Quantitative Investment Models in High Frequency Trading Environment

Qirui Lou

Tianjin Chengjian University, Tianjin, 300000, China

L1285653957@163.com

Keywords: High frequency trading environment; Quantitative investment; Model construction and empirical analysis

Abstract: Quantitative investment, as a method of using mathematics, statistics, and computer technology to guide investment decisions, has become an important component of the modern financial investment field. It establishes a sustainable mathematical model by deeply analyzing market environment, macroeconomic factors, and diversified information such as the fundamentals and technical aspects of stocks, in order to seek profit opportunities in the market. High frequency trading refers to the behavior of completing a large number of transactions in a very short time using high-performance computer systems and fast data transmission technologies. It requires investors to quickly obtain, process, and analyze market data, and make quick and accurate trading decisions based on it. In the high-frequency trading environment, how to construct appropriate quantitative investment models and adopt appropriate methods to measure risks has become a hot and cutting-edge issue in the field of financial research. With the continuous progress of technology and the continuous growth of the market, investors need to constantly learn and explore new methods and technologies to improve investment efficiency and reduce risks. This article delves into the construction and empirical analysis of quantitative investment models in high-frequency trading environments.

1. Introduction

In traditional financial markets, low-frequency stock trading data with daily, weekly, and monthly sampling periods dominate[1]. Investors often rely on these low-frequency data, combined with fundamental analysis and technical analysis methods, to predict and judge the future trend of stocks[2]. However, due to the inherent problems of information lag and data sparsity in low-frequency data, investors often face significant uncertainty and risk in their decisions[3]. Especially in modern financial markets with high market volatility and rapid information updates, the limitations of low-frequency data are becoming increasingly prominent. With the arrival of the big data era, the amount of data in the financial market is showing explosive growth. Especially for intraday high-frequency trading data, due to its high sampling frequency, large data volume, and rich information, it has gradually become active in the financial market[4]. High frequency trading data can reflect the dynamic changes of the market in real time, provide more accurate and comprehensive market information, and provide investors with more decision-making basis[5].

Meanwhile, with the continuous growth of computer technology and optimization of algorithms, investors can process and analyze high-frequency data more efficiently, thereby more accurately grasping market trends and investment opportunities[6]. In the high-frequency trading environment, how to construct appropriate quantitative investment models and adopt appropriate methods to measure risks has become a hot and cutting-edge issue in the field of financial research[7]. Quantitative investment is a method that uses mathematics, statistics, and computer technology to guide investment decisions. It excavates and analyzes historical data to identify the patterns and trends of stock price changes, thereby constructing a mathematical model that can predict future stock prices[8]. In a high-frequency trading environment, quantitative investment models need to be able to process and analyze high-frequency data in real-time to quickly respond to market changes

  DOI: 10.25236/etmhs.2024.064

and make accurate investment decisions. Building a suitable quantitative investment model requires addressing multiple key issues. Firstly, it is necessary to select appropriate data sources and indicators. Although high-frequency trading data is rich in information, there are also issues such as data redundancy and noise interference. Therefore, it is necessary to clean, integrate, and standardize the data to extract key features related to changes in stock prices.

At the same time, it is necessary to select appropriate data indicators as input variables for the model based on factors such as investment goals and risk preferences. Secondly, it is necessary to choose appropriate algorithms and models. The construction of quantitative investment models involves the selection of various algorithms and models, such as linear regression, logistic regression, deep learning (DL), etc. Different algorithms and models have different characteristics and applicability, and need to be selected and optimized according to actual situations. At the same time, it is necessary to train and validate the model to ensure its stability and reliability in practical applications. Finally, appropriate methods need to be adopted to measure and manage risks. Although quantitative investment can improve the accuracy and efficiency of investment decisions, there are also certain risks involved. In a high-frequency trading environment, with high market volatility and rapid information updates, investors need to pay more attention to risk management and control. Therefore, multiple methods are needed to measure and evaluate risks, such as volatility, maximum drawdown, Sharpe ratio, etc., in order to adjust investment strategies in a timely manner and control risks. In summary, the construction and empirical analysis of quantitative investment models in high-frequency trading environments is a complex and important research field. This article aims to explore in depth the construction and empirical issues of quantitative investment models in high-frequency trading environments, providing useful references and guidance for investors.

2. The Application of DL in Quantitative Investment

In the context of artificial intelligence (AI), DL, with its powerful data-driven and adaptive learning capabilities, is gradually penetrating into various fields and demonstrating enormous potential and value[9]. Quantitative investment, as an important branch of the financial field, relies on a large amount of data analysis and model prediction in its decision-making process[10]. The introduction of DL technology has brought new ideas and methods for quantitative investment, which is expected to improve the accuracy and efficiency of investment decisions. DL is a branch of machine learning (ML) that simulates the working mode of human brain neural networks to achieve automatic feature extraction and model training of data. Compared with traditional ML algorithms, DL has the following significant characteristics. DL can automatically learn the complex structure and features of data without the need for manual feature selection and extraction. The DL model has strong non-linear processing ability and can capture non-linear relationships in data. DL improves the model's generalization ability by constructing a multi-level neural network structure to achieve layer by layer abstraction and representation of data. Feature extraction and selection are crucial steps in quantitative investment.

Traditional quantitative methods often rely on manual experience and professional knowledge for feature selection, which has subjectivity and limitations. DL technology can automatically learn the complex structure and features of data, thereby achieving automatic feature selection and extraction. For example, the DL model can be used to learn from historical stock prices, trading volumes, financial indicators, and other data, extracting key features related to stock price changes. The DL model has strong predictive ability. In quantitative investment, DL models can be used to construct predictive models for predicting the future prices of stocks. Compared with traditional statistical models, DL models can capture nonlinear relationships in data and improve prediction accuracy. For example, recurrent neural networks (RNN) or long short-term memory networks (LSTM) can be used to model time series data of stocks and predict the future trend of stock prices. Figure 1 shows the LSTM structure.

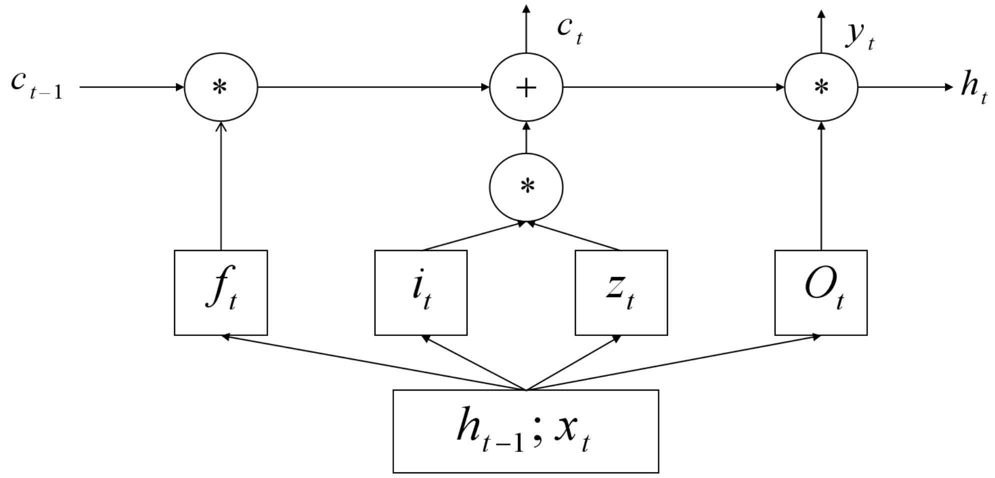

Figure 1 LSTM structure

Risk management is an essential part of quantitative investment that cannot be ignored. DL technology can be used to quantify risk management in investment. For example, DL models can be used to evaluate and predict the risk of investment portfolios, guiding investors in risk control and asset allocation. In addition, DL technology can also be used to build risk warning systems, monitor market changes in real time, identify potential risks in a timely manner, and take corresponding measures. In order to overcome the limitation that single perspective features cannot fully represent the characteristics of things, this paper combines multi perspective features with DL methods for research. Multi perspective features can more comprehensively represent different aspects and dimensions of things, thereby improving the predictive ability and robustness of the model. In the process of multi view feature processing, the feature data of each perspective can be represented by a matrix, so multi view features can be represented by multiple single view feature data matrices. By inputting these feature data matrices into the DL model for training and learning, a quantitative investment model that can simultaneously process multiple perspective features can be obtained.

3. Construction and Empirical Analysis of Quantitative Investment Models

3.1. Model Building

The comprehensive understanding of things is achieved through in-depth understanding and analysis of things from multiple perspectives. Each perspective provides partial information about things, and the superposition and complementarity of this information helps us to have a more comprehensive understanding of the essence and characteristics of things. The multi perspective understanding method has the characteristics of comprehensiveness, accuracy, and complexity, which are particularly evident in the description of short-term timing classification features. In short-term timing classification feature description, the multi perspective approach requires us to combine multiple single perspective features to form a more complete and accurate feature expression.

$$MVF = \left[VF_1, VF_2, \ldots, VF_n\right] \tag{1}$$

$$VF_n = \left[D_1^{VF_n}, D_2^{VF_n}, \ldots, D_m^{VF_n}\right]^T \tag{2}$$

$$D_m^{VF_n} = \left[D_{m1}^{VF_n}, D_{m2}^{VF_n}, \ldots, D_{mj}^{VF_n}\right]^T \tag{3}$$

MVF represents the multi view short-term timing classification feature, VF_n represents the result value of the n th single view feature, and $D_m^{VF_i}$ represents the m th feature representation in the i th single view.

When generating training set data, standardizing the data is an important step that helps eliminate differences in dimensionality and value limits between indicators, making different features numerically comparable. A common method of standardization processing is Z-score standardization, which is achieved by converting data into a distribution with a mean of 0 and a standard deviation of 1. The data preprocessing formula is shown as follows:

$$X = \frac{x - \overline{x}}{\sigma} \tag{4}$$

Among them, $\overline{x}$ is the data mean, σ is the data standard deviation, and x, X is the original data value and the value after Z-score, respectively.

Activation functions play a crucial role in neural networks. They introduce nonlinearity, enabling neural networks to learn and represent complex nonlinear relationships. Without an activation function, the neural network will only be a linear model, and its expressive power will be very limited. The sigmoid function is one of the commonly used activation functions in neural networks, and its definition is as follows:

$$S(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

The value range of the Sigmoid function is (0,1).

3.2. Experimental Result

The display in Figure 2 fully verifies the effectiveness of combining multi view features with DL methods. By comparing the profit performance before and after using the model in this article, it is evident that the profit performance has significantly improved after using a model that combines multi perspective features with DL for prediction.

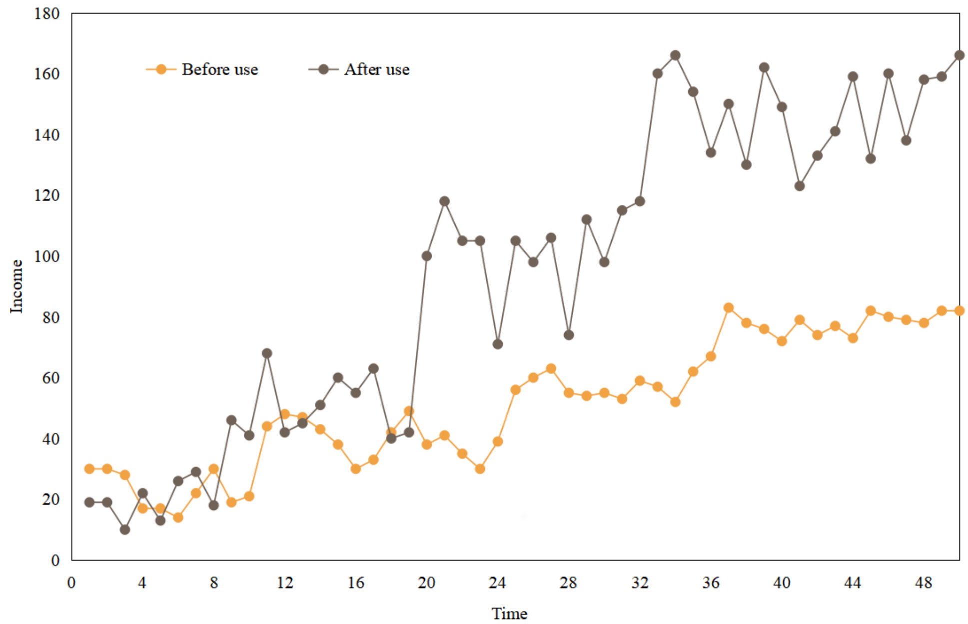

Figure 2 Revenue comparison

This improvement is mainly due to the following aspects. Multi perspective features can cover multiple aspects and dimensions of things, providing more comprehensive and rich information. In quantitative investment, this means that we can analyze investment targets from multiple perspectives (such as market trends, technical indicators, fundamental data, market sentiment, etc.) to make more accurate investment decisions. The DL method has strong representation learning and non-linear processing capabilities, which can automatically learn the complex structure and features of data, and capture non-linear relationships in the data. This enables the DL model to fully utilize the information in multi view features and discover potential patterns and patterns hidden in the data. By combining multi perspective features with DL methods, we can construct a comprehensive and powerful quantitative investment model. This model not only utilizes the information provided by multi perspective features, but also utilizes the powerful capabilities of DL methods to discover potential patterns and patterns in data, making more accurate predictions and decisions. Therefore, the comparison of returns shown in Figure 2 fully demonstrates the effectiveness and advantages of combining multi perspective features with DL methods in quantitative investment. This combination approach can not only improve investment returns, but also provide investors with more comprehensive and accurate decision support.

4. Conclusions

With the vigorous growth of AI technology, especially the continuous progress in the DL field, and the increasing demand for financial data analysis by individuals and enterprises, DL technology has been widely penetrated into the forefront of research in the financial field. Especially in financial market forecasting, text information processing, and trading strategy optimization, DL technology has shown tremendous potential and value. However, despite significant achievements in the application of DL in the financial field, with the continuous changes in the market and advances in technology, investors still need to maintain a keen insight and exploratory spirit, constantly learn and master new methods and technologies, in order to improve investment efficiency and reduce risks. This article delves into the construction and empirical analysis of quantitative investment models in high-frequency trading environments, and proposes a new quantitative investment strategy by combining multi perspective features with DL technology. This strategy can effectively utilize the multidimensional information of financial markets and perform complex pattern recognition and prediction through the powerful capabilities of DL. The experimental results show that the quantitative investment model constructed using the method proposed in this paper has better performance in high-frequency trading environments. However, the method proposed in this article also has some shortcomings. The performance of DL models highly depends on the quality and quantity of training data. In the financial field, high-quality data is often difficult to obtain, and there are problems such as noise and outliers. This may lead to overfitting or underfitting during the training process, affecting the model's generalization ability.

References

[1] Virgilio G P M. A theory of very short-time price change: security price drivers in times of high-frequency trading[J]. Financial Innovation, 2022, 8(1): 66.

[2] Brigida M, Pratt W R. High-frequency trading and the weekly natural gas storage report[J]. Studies in Economics and Finance, 2019, 36(3): 547-566.

[3] Lei B, Zhang B, Song Y. Volatility forecasting for high-frequency financial data based on web search index and deep learning model[J]. Mathematics, 2021, 9(4): 320.

[4] Manahov V, Hudson R, Urquhart A. High-frequency trading from an evolutionary perspective: Financial markets as adaptive systems[J]. International Journal of Finance & Economics, 2019, 24(2): 943-962.

[5] José Cerda, Nicolás Rojas-Morales, Minutolo M C, et al. High Frequency and Dynamic Pairs Trading with Ant Colony Optimization[J]. Computational Economics, 2022, 59(3): 1251-1275.

[6] Cerda J, Rojas-Morales N, Minutolo M C, et al. High frequency and dynamic pairs trading with ant colony optimization[J]. Computational Economics, 2022, 59(3): 1251-1275.

[7] Bongaerts D, Van Achter M. Competition among liquidity providers with access to high-frequency trading technology[J]. Journal of Financial Economics, 2021, 140(1): 220-249.

[8] Zhuang X N, Chen Z Y, Wu Y C, et al. Quantum computational quantitative trading:

high-frequency statistical arbitrage algorithm[J]. New Journal of Physics, 2022, 24(7): 073036.

[9] Santosa P W. Determinants of price reversal in high-frequency trading: empirical evidence from Indonesia[J]. Investment Management and Financial Innovations, 2020, 17(1):175-187.

[10] Lenczewski M C J. Hidden and Fast Liquidity - Hidden Orders and High-Frequency Trading[J]. e-Finanse, 2020, 16(1):27-35.