Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 4, Issue 6: 10-14, DOI: 10.25236/AJCIS.2021.040603

# Quantified Investment Strategies and Excess Returns: Stock Price Forecasting Based on Machine Learning

**Zihe Tang[1,a]*, Yanqi Cheng[2,b], Ziyao Wang[3,c]**

*[1]Department of Mathematics, University of California, Los Angeles, California, United States*
*[2]Department of Economics, University of California, San Diego, California, United States*
*[3]Department of Corporate Finance, University of International Business and Economics, Beijing, China*
*[a]lillytangzihe0915@foxmail.com, [b]evacheng1211@163.com, [c]201806194@uibe.edu.cn*
**Corresponding author: lillytangzihe0915@foxmail.com*
*These authors contributed equally to this work.*

***Abstract:*** *Under frequent trading strategy, this paper assumes that rational investors will adopt a long-term conservative frequent trading strategy. By studying historical data by machine learning, we predict the future trend of stock prices and compare the returns of frequent trading strategy and long-term holding outcomes. The results show that the trend of stock prices is the key factor affecting quantitative investment strategy. This study enriches the relevant literature in the field of quantitative investment and focuses on the comparison of returns between frequent trading strategies and long-term holding strategies. It provides standards for selecting investment strategies in the securities market. In the practical sense, this paper proposes investment suggestions for investors to benefit from market fluctuations.*

***Keywords:*** *Quantitative Investment, Linear Regression Machine Learning Model, Long-Term Conservative Frequent Trading Strategy*

## 1. Introduction

Quantitative trading is a marketing strategy that relies on mathematics, statistics, and computer programming to model an integrated trading condition and analyze historical data based on excluding subjective factors (market opinion, investor inexperience, and investment irrationality). Therefore, quantitative trading has several compelling advantages. First, the investment results with quantitative trading are more stable, i.e., investments are made through a unique trading strategy established through a range of trading conditions. The trading strategy used in this paper is the long-term conservative frequent trading strategy, which if we predict the stock price will increase greater than 0.1% of the closing price after 3 days, we will sell the stock on the fourth day; if the price continues to rise in those 3 days, and the increase is greater than 0.1% of the forecasted price, we will not sell the stock until a downward trend occurs. After that, a new round of buying begins. While strict adherence to trading strategies may not guarantee returns, using the right model has a higher possibility to yield profits than relying solely on investor experience. Second, prospect theory suggests that many people are loss aversion, and they trigger the disposition effect, which means that investors are more sensitive to gains than losses, resulting in different criteria for disposing of profitable and loss-making stocks (Camerer, 2000). Thus, quantitative trading avoids irrational investments caused by risk-seeking shifts; yet, the stability of quantitative trading also makes it challenging for investors to make huge gains in a short time because of the modest profits per trade.

The quantitative financial strategy has varied applications in financial markets. Zhang, Shan, and Su used time theory in [6] to analyze the macroeconomic and share price movements. They predicted short-term stock prices using ARCH and ARIMA models and compared them with real data. They found that ARCH fits better because the ARIMA model assumes the stock market has the same variance and zero residual on average. Huang and Liu in [4] added the human emotion index to improve the accuracy of predicting the stock price. Fang, Cheng, and Xue built the LSTM-SVR model in the study [2] based on the depth learning model LSTM to eliminate the predictive deviation, improve the accuracy of the forecast, and introduce hedging concepts into the quantitative investment strategy. They found that hedging improved the profitability and stability of the quantitative investment strategy. Experiments show that compared with the traditional quantitative investment strategy, applying a deep learning strategy brings a higher rate of return, a more stable yield curve, and better resistance to decline. Ta, Liu, and Tadesse (2020) used deep learning techniques to predict stock movements based on historical data and build portfolios for higher returns. Based on previous research and the efficient market hypothesis, our research points out that in an incompletely weak effective market, technical analysis based on historical data can obtain considerable income in the securities market. We built our model based on depth learning to determine when the best time is for making trades. Compared to traditional forecasting methods, machine learning methods can measure the risk premium of a single stock and the entire market more accurately, so it can provide significant returns for investors (S. Gu, B. T. Kelly, and D. Xiu, 2018). Through machine learning, we used eight forecasting models to study the historical stock price and select the model with the best prediction effect to make further predictions. We selected a stock that maximizes returns while not considering the situation of multiple stock combinations. We maximized the return by selecting the right buying and selling time and compared the returns of frequent trading strategies with long-term holding strategies of quantitative investments to provide investors with investment analysis and advice.

## 2. Proposed Method

### *2.1 The Overall Analysis*

This article conducts the research and corresponding analysis through the following steps. First, the study takes the historical data of representative stocks in various industries. Second, applying eight machine learning models to forecast future stock price changes, the research compares and selects the machine learning model with the best prediction effect. Third, the study sets up functions to simulate stock trading conditions, and, last, compares stock returns under different quantitative trading strategies.

### *2.2 Data Acquisition and Preprocessing*

This research uses the Yfinance package provided by Yahoo!Finance to obtain historical stock data. Controlling the three parameters of the stock code, start date, and end date, the research uses the yfinance.download function to set up the time interval of the specific stock and download its data. The data includes the stock's daily opening price, highest price, lowest price, closing price, and trading volume of the stock within the periods. Then, the data is preprocessed by using the dropna() function from the Pandas package to delete the data with a closing price of zero.

### *2.3 Machine Learning Models*

This research uses the Linear Regression Model from the supervised learning algorithm to predict stock price changes. The Linear Regression Model is one of the regression analyses in machine learning algorithms, which can perform linear analysis and modeling of the relationship between predictor variables and response variables.

One-Variable Linear Regression Model: Use the linear function of the predictor variable $x$ to build a model for the response variable $y$ :

$$y = w \cdot x + b \tag{1}$$

where $w$ and $b$ are regression coefficients.

General Multiple Linear Regression Model: Use multiple predictor variables $x_1$ , $x_2$ , ..., $x_n$ or attribute linear functions of sample $X$ to model the response variable $y$ :

$$y = w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_n \cdot x_n + b \tag{2}$$

where $w_i$ and $b$ are regression coefficients.

### *2.4 Quantitative Investment Trading Strategy*

The study tries to maximize the selected stocks returns without considering the situation of multiple stocks combinations. There are mainly six types of quantitative investment strategies in the modern stock trading. This study uses the long-term conservative frequent trading strategy and the long-term holding strategy to compare the returns of the same stock.

ISSN 2616-5775 Vol. 4, Issue 6: 10-14, DOI: 10.25236/AJCIS.2021.040603

Long-Term Conservative Frequent Trading Strategy: Predict the stock price in 3 days. If the price increase is greater than 0.1% of the closing price of the day, hold the stock for 3 days and then sell it. Within those 3 days, if the price is predicted to has a continuous upward trend, and the increase is greater than 0.1% of the forecasted price, keep holding the stock until the price shows a downward trend. After that, a new round of trading began.

Long-Term Holding Strategy: Within a certain period, buy the stock at the beginning and sell it at the end, with no other operations in between.

## 3. Strategy Analysis Based On Real Data

### *3.1 Dataset*

In the study, 6 stocks of Zhejiang University Netnews (600797.SS), SINOPEC (600028.SS), Amazon (AMZN), McDonald's (MCD), Google (GOOGL), and Apple (AAPL) are selected. Applying the yfinance package, the study obtains the historical price data from the start date 01/01/2000 to the end date 01/03/2021, provided by yahoo! Finance, and removes data with zero closing price.

### *3.2 Learning*

The study splits the processed data into two sets: the data from the start date 01/01/2000 to 241 days prior to the end date as the training set, and the data of the last 241 days as the trading stimulation set. After splitting data sets, the study applies eight machine learning regression models to study the historical stock data. The 8 regression models are the Decision Tree Regressor Model, the Linear Regression Model, the K Neighbors Regressor Model, the AdaBoost Regressor Model, the Support Vector Machine Model, the Gradient Boosting Regressor Model, the Bagging Regressor Model, and the Extra Tree Regressor Model. The study selects the most optimal model by declaring a function to compare each prediction's mean, variance, and mean square error regression loss. Since the comparison shows the prediction of the eight models are very similar, this study selects the Linear Regression Model for prediction, using the date as the predictor variable x and the stock price as the response variable y. The forecast stock price results are visualized as follows, where the blue line is the closing price, and the green line is the predicted price:

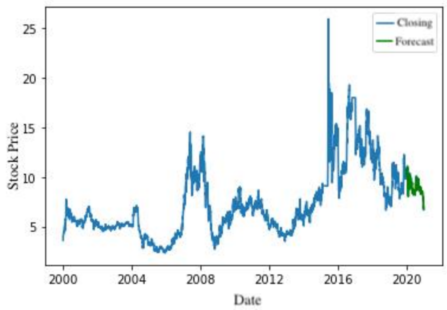


*Figure 1: Zhejiang University Netnews (600797.SS.)*

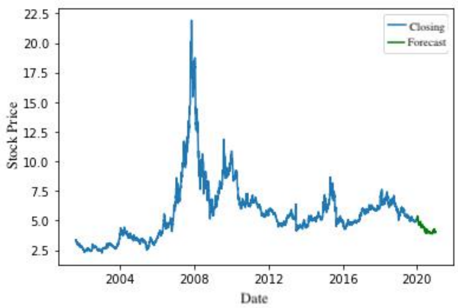


*Figure 2: SINOPEC (600028.SS).*

ISSN 2616-5775 Vol. 4, Issue 6: 10-14, DOI: 10.25236/AJCIS.2021.040603

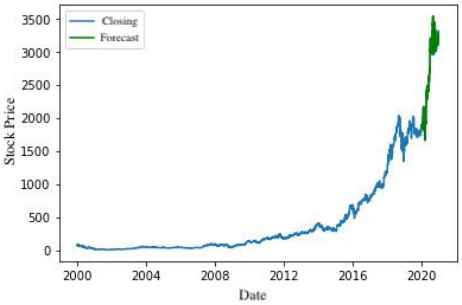


*Figure 3: Amazon (AMZN).*

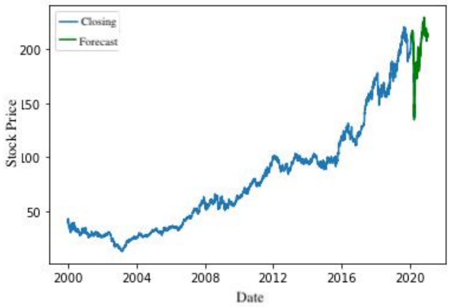


*Figure 4: McDonald's (MCD).*

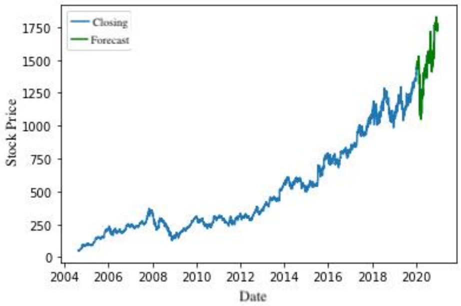


*Figure 5: Google (GOOGL).*

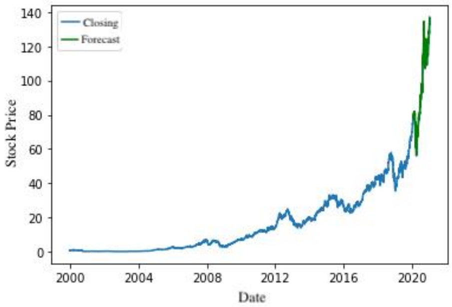


*Figure 6: Apple (AAPL).*

### *3.3 Trading Simulation and Evaluation*

Assuming that the initial principal is 100,000 yuan, the results of using a long-term conservative frequent trading strategy and a long-term holding strategy in the period of 241 days are compared. In the long-term conservative frequent trading, the study tries to maximize the return by selecting the right buying and selling time. Using the three days later forecasted stock price as the reference price, the study sets up the function get_income_circle_3 to calculate the specific dates of the buying and selling operations, based on the price increase and decrease trends. Next, the study sets up a profit-calculating function, cal_income, which returns the holding revenue during the period by entering the stock buying date and selling date. Thus, applying the obtained operating dates in the frequent trading strategy to the function, the study calculates the total revenue under this strategy. Last, taking the 241 days of the simulated trading period as the total holding time, the study uses the function cal_income again to calculate the revenue under the long-term holding strategy. The simulated trading results of the two strategies for each stock are as follows:

*Table 1 Trading Simulation Results.*

| Stock Price Changing Trend | Stock | Long-Term Conservative Frequent Trading Strategy | Long-Term Holding Strategy |
|---|---|---|---|
| Upward | Amazon | 172142.17(+72.14%) | 149756.42(+49.76%) |
| | McDonald's | 101619.62(+1.62%) | 81179.83(-18.82%) |
| | Google | 118241.86(+18.24%) | 111048.93(+11.05%) |
| | Apple | 118241.86(+67.66%) | 147941.15(+47.94%) |
| Fluctuated | Zhejiang University Netnews | 66222.64(-33.78%) | 77566.12(-22.44%) |
| | SINOPEC | 75186.57(-24.81%) | 82480.73(-17.52%) |

In summary, the two stocks of Zhejiang Netnews and Sinopec, which are predicted to have fluctuating stock prices, have better returns under the long-term holding strategy than under the frequent trading strategy; and the four stocks that are predicted to have greater changes in their stock prices, Amazon, McDonald's, and Google, and Apple, all have better returns when using frequent trading strategy than using long-term holding strategy.

## 4. Conclusions

The results show that the trend of the stock price is an important factor affecting investment decisions. Facing stocks with significant upward trends, the return rate of frequent trading strategy is higher than that of long-term holding strategy. For volatile stocks, the long-term holding strategy yields higher than the frequent trading strategy. The results of this paper enrich the research literature related to quantitative investment and extend the research results. In many historical documents, the trading strategies of quantitative investment mainly focus on the transformation among multiple trading strategies, and there is little discussion on specific practices of a single trading strategy —— frequent short-term trading around the market. This paper compares the income differences via frequent trading strategy and long-term strategy for investors to choose the right kind of trading criteria. Meanwhile, suggestions are put forward and specific business operations are provided according to real historical data, which helps investors understand the meaning of the actual market value.

## References

*[1] C. F. Camerer, "Prospect theory in the wild: Evidence from the field. Social Science Working Paper," in California Institute of Technology, Pasadena, CA,1998.*
*[2] Y. Fang, J. Chen and Z. Xue, "Research on Quantitative Investment Strategies Based on Deep Learning," in Algorithms,12(2),p.35, 2019.*
*[3] S. Gu, B. T. Kelly, and D. Xiu, "Empirical Asset Pricing via Machine Learning," in SSRN Electronic Journal, 2018.*
*[4] J. Huang and J. Liu, "Using social media mining technology to improve stock price forecast accuracy," in Journal of Forecasting, 39(1), pp.104–116, 2019.*
*[5] V. D. Ta, C. M. Liu and D. A. Tadesse, "Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading," in Applied Sciences, 10(2), p.437, 2020.*
*[6] J. Zhang, R. Shan and W. Su, "Applying Time Series Analysis Builds Stock Price Forecast Model," in Modern Applied Science, 3(5), 2009.*