Application of Computer Science in Risk Management of Financial Investment

Andy Ruichen Lu[1, †], Yuhan Pan[2, †], Yu Yang[3, †, *], Zhiyi Ying[4, †]
arl633@stern.nyu.edu, jacobpan@nbt.edu.cn, [*]Corresponding author: yangyu8@msu.edu, sgzying@liverpool.ac.uk

[1]Stern School of Business New York University New York, US, 10012

[2]Department of Finance Ningbo Tech University Ningbo, China, 315100

[3]Department of Accounting Michigan state University, US east Lansing, MI, 48823

[4]Department of Mathematical Sciences University of Liverpool Liverpool, England, L69 7ZX

[†]These authors contributed equally.

Abstract— Forecasting for the stock market is a popular domain that many researchers have focused on in recent years. However, accurate prediction for different types of stocks is still challenging. In this research, by using the computing power of a programming tool called Python, financial mathematical models were transferred into a computer model, and this study innovatively realized the use of computers to complete the preliminary stock market forecast, to make the forecast and asset distribution more accurate, which is very helpful for practitioners in the stock market. Investment in the field has played an important role in avoiding risks. This study obtained data from Yahoo Finance and made relevant predictions using mainly Fama-French model, Linear Regression model, Decision Tree Regressor model and Portfolio Optimization with Monte Carlo Simulations experiment. The result indicated that the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of the multi-factor model, Linear Regression model and Decision Tree Regressor model are all less than 1, which means the prediction effect of the experiment is satisfactory. This demonstrates the feasibility of emerging computer algorithms to replace traditional methods.

Keywords— Python, Fama-French model, Linear Regression model, Decision tree Regressor model, Portfolio Optimization with Monte Carlo Simulations

1 INTRODUCTION

Prediction of stock prices has always been a classic and very important subject. Since the birth of the stock market, experts, and scholars in the fields of finance, mathematics, and economics have conducted in-depth and systematic research on the prediction of stock prices and the analysis of the stock market and have established many mathematical models. These models have had a very positive impact on stock market forecasts. As the research of Academic abilities, education, and performance in the stock market in 2018, mathematics and quantitively education, in the stock market, the level of mathematical ability and quantitative data analysis plays one of the most critical roles in improving returns [1]. People with higher intelligence,

FFIT 2022, October 28-30, Shenzhen, People's Republic of China

DOI 10.4108/eai.28-10-2022.2328463

test scores and education perform better in stock market investments. This proves the validity of the mathematical model to a certain extent.

However, it can be found that in many cases based on the current open literature, many mathematical models have not been effectively applied in stock market trading. People still prefer to predict stock market growth and decline based on their feelings and experience. Usually, the reason is that these models assume of economic man, but people are not economic people, and various social information will affect people's subjective judgments, causing people to become impulsive and irrational.

Some related studies are summarized. For example, Naseem et al. indicated that COVID-19 has successfully disrupted investors' psychology for investment decisions [2]. Public health crises and psychological barriers affect the economic and financial well-being of individuals and global investors. Valle-Cruz et al. found that Twitter posts affected financial indices during both outbreaks [3]. The impact of social media publications was more pronounced during the COVID-19 pandemic than during the H1N1 pandemic, where stock prices fell more sharply because of more speculation, rumors, and negative news. Sentiment on Twitter has a significant impact on financial indices, and this effect was observed days after information was posted on Twitter. Over a nearly 11-year period, social media spread (more Twitter accounts) had a direct impact on the stock market's index. According to an analysis of the article Intraday News Trading: The Reciprocal Relationships Between the Stock Market and Economic News, the number of tweets by Reuters and Bloomberg is positively correlated with volatility in stock market prices, especially when considering 1-hour intervals.

To solve the problem of uncontrollable human emotions for stock market forecasts, this study believes that after systematic research, computers should be introduced to replace humans to make more accurate stock market forecasts. To be more specific, machine learning models can be considered as the effective alternative method in this case since they achieved success in other domains [4-6]. To achieve this goal, this study has conducted research on some representative and typical financial forecasting models. By using the computing power of python, financial mathematical models were transformed into computer models, and this study innovatively realized the use of computers to complete preliminary Stock market forecast, to make the forecast and asset distribution more accurate and play a certain role in risk aversion for practitioners' investment in the stock market.

2 METHODS

2.1 Dataset Description

We retrieved mkt, smb, hml, adj close price, close price, high price, low price, open price, and volume price from 2012 to 2021 as our research sample data from Yahoo Finance and put them into an Excel. Thereinto, we used 80% of the data as the training set of the model and 20% as the testing set. To be more specific, we used mkt, smb and hml as the factors of Multi-Factor model and adj close price, close price, high price, low price, open price and volume price as the samples of Linear Regression model and Decision Tree Regressor model.

During our research, we selected 3 industries as experimental samples, namely Energy Industry, Food & Beverage Industry, and Software/Consumer Electronics Industry. For each industry, we selected the stocks of the two most representative companies, so our sample has a total of 6 stocks finally. For Energy Industry, we selected Exxon Mobil and Chevron. For Food & Beverage Industry, we picked Pepsi and Nestle. For Software/Consumer Electronics Industry, we picked Microsoft and Apple. These stocks have high market capitalization, and they are the leading companies in their respective industries. Some of these industries are highly cyclical, and others are non-cyclical. We believe by choosing the stocks in this way, the result would be worth optimizing, because if all the stocks are selected from the same industry, especially if it is a cyclical industry, the covariance between the stocks would be high. Thus, it would be hard to obtain a satisfactory result. Exxon Mobil and Chevron are leading US companies in the energy industry. Their market capitalizations are higher than $300 billion. The energy industry can be considered cyclical because gasoline demand tends to increase during summer and decrease during winter. Although many countries are investing heavily in renewable energy, fossil fuels are still the most widespread source of energy generation currently. Thus, the two companies we chose are both oil companies. On the other hand, Pepsi and Nestle are the leaders in the food and beverage industry, a non-cyclical sector. Pepsi’s market capitalization is higher than $200 billion, and Nestle’s is higher than $300 billion, while Pepsi’s primary focus is on the beverage industry and Nestle’s is on the food industry. Apple and Microsoft are leading companies in the consumer electronics and software development industries. Both companies also provide different online services, and both companies have a market capitalization of more than $2 trillion.

Table 1 Sample data in the collected dataset.

Date	**Adj CloseAAPL**	**Adj CloseCVX**	**Adj CloseMSFT**	**Adj CloseNSRGY**	**Adj ClosePEP**
2012/1/3	12.57591724	72.38985443	21.62313652	43.46416473	49.25931931
2012/1/4	12.64349937	72.26524353	22.13200951	43.15871429	49.51155472
2012/1/5	12.78386784	71.55686951	22.35817528	42.8309021	49.1257782
2012/1/6	12.91750336	71.03871155	22.70550156	42.24234772	48.51003647
2012/1/9	12.89701843	71.81266785	22.40663719	42.77130127	48.76226807

Table 1 is only a partial example of the data source. In the collected dataset, all six stocks show a generally increasing trend. Among them, AAPL and MSFT perform a great growth while XOM and CVX perform a relatively small increase. Second, all stocks have shown periodic fluctuations during the whole-time range, which provide several sets of extreme data which can be viewed as peak and bottom. Despite of this huge change, all stocks grow according to a long-time range as mentioned in the first point. Third, regarding the changes pointed out above, the fluctuations of the six stocks barely happened similarly or at nearby times. This proves the uncorrelation of the chosen data. In addition, these trends and patterns of the past data can be used to roughly verify the result and trend of the prediction. They can also be used to negate extreme errors.

2.2 Adj Open price prediction

2.2.1 Linear Regression and Decision Tree Regressor model

Linear Regression model is the mathematical modeling of the correlation between independent variables and dependent variables in statistics [7]. It is often used to explain quantitative variables and can predict the future based on known data. The advantage of the linear model is that it is easier to fit, and the prediction results are easier to determine. In addition, the decision tree as a famous algorithm that is good at finding the complex nonlinear relationship is also considered [8].

For the Linear Regression and Decision Tree Regressor model, we import some libraries, such as sklearn, which is about machine learning and train_test_split, which is used for dividing the training set and test set, training the model on the training set, and testing the performance of the trained model on the test set. We carried out the following data processing work on the selected 6 representative stocks respectively. First, we retrieved the data in Excel prepared earlier. Secondly, we trained the data model using the training set. X and Y values were imported for training. Thirdly, we made the prediction with the test value set, and compared our results with the actual value by figuring out the Root Mean-Square Error (RMSE) and Mean Absolute Error (MAE). Finally, we visualized the results and made the fitted curves of the prediction and actual value.

2.2.2 Multi-Factor model

As an extension of Capital Asset Pricing Model (CAPM) [9], the multi-factor model is an important financial forecasting model. It is generally used to describe the returns of stocks. According to the theory of Fama and French [10], mkt, hml and smb are the three major factors that affect the return of a company's stock.

For the multi-factor model, we imported some libraries, such as pandas, which is used for working with tabular data, numpy, which is used for matrix operations, matplotlib, which is used for display image, and yfinance, which is used to get data set. We used the S&P 500 index as the standard of the stock market benchmark, because it includes more companies. Firstly, by using the capital assets pricing model, we figured out the beta, which means covariances of the 6 assets that reflects the changes in the return on assets. Using the data above, we defined and fit the regression model. The model above does not show the risk-free rate, but this is an important factor for our investment, so we did also calculate the risk-free rate and draw the images of the risk-free rate of the 6 assets. Then, we started implementing the multi-factor model. By using the data from yahoo finance, we figured out the monthly returns on the risky assets by using this formula. Then, we merged the datasets and calculated excess returns, and used it to estimate the three-factor model.

2.2.3 Portfolio Optimization with Monte Carlo Simulations

Monte Carlo Simulation uses random sampling of historical stock information, including close, high and low stock prices, data of annualized average returns, corresponding standard deviation from each stock repeatedly to simulate random portfolio weights and metrics. Through creating a joint data Frame with all data, locating the numerous data points to generate the efficient frontier plot of six stocks, Maximum Sharpe ratio, Minimum Volatility ratio portfolio respectively, and forming the efficient frontier plot for visualization.

3 RESULT AND DISCUSSION

3.1 Linear Regression and Decision Tree Regressor model

Due to the obvious repeatability of the experimental results, we selected one of the typical characteristic results with representative significance. The experimental results show that in the linear regression model, rmse is approximately equal to 0.43, and mae is approximately equal to 0.33. In the Decision Tree Regressor model, rmse is approximately equal to 0.46, and mae is approximately equal to 0.29. The above data are kept to two decimal places. Figure 1-4 present our visualizations.

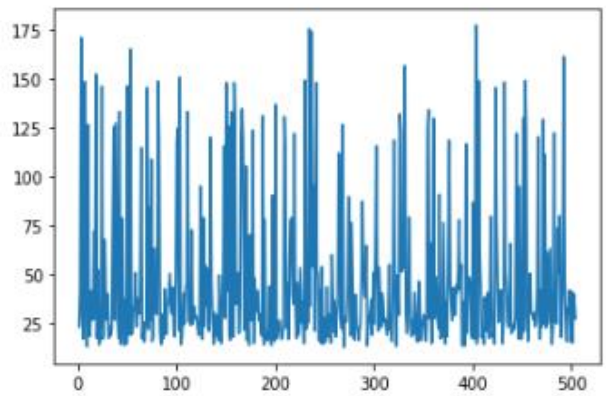

Figure 1. The Fitted curve of test value of Linear Regression model of APPL.

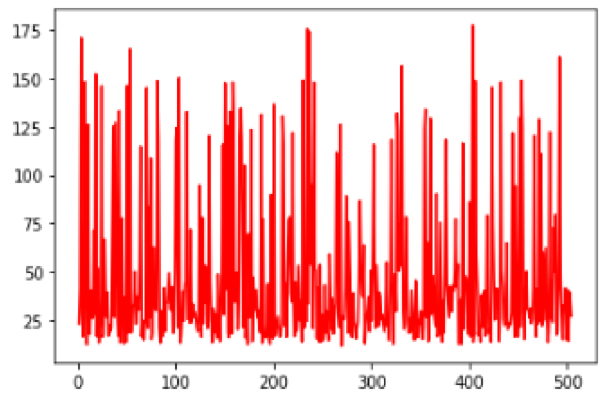

Figure 2. The Fitted curve of predict value of Linear Regression model of APPL.

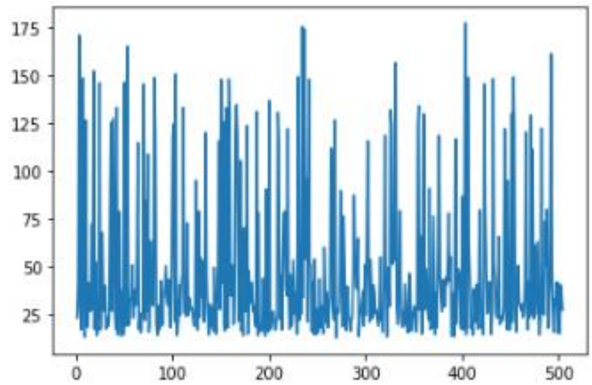

Figure 3. The Fitted curve of Test value of Decision Tree Regressor model of APPL.

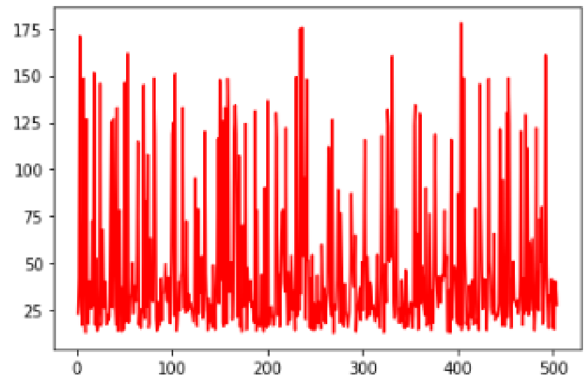

Figure 4. The Fitted curve of test value of Decision Tree Regressor model of APPL.

Some repeatable and accurate experimental results have proved that the rmse and mae of the two models are both less than 1 in both models, so it can be proved that the predicted price of the model is very similar to the price of the actual stock market, and the prediction result of the Linear Regression and Decision Tree Regressor model is also relatively similar. It is known that our visual fitting curve can also prove that the fitting effect of the financial models of these two computers is very good. In the same way, if we follow the mathematical model and only change the meaning of the representation of X and Y of the model without changing the basic operating mode of the model, the model can still be used normally. Obviously, in the laboratory, the effect of using computer models instead of traditional methods to make stock market forecasts is very significant.

3.2 Fama-French three Factor model (Multi-factor model)

Obviously, according to the above visual chart results, the returns, and risks of stocks of companies in the energy industry are the most unstable, while the returns and risks of stocks of companies in the technology industry and food industry are relatively stable, and the predicted income is also relatively considerable. But due to a series of unexpected events such as

COVID-19 and the Russian-Ukrainian war, financial markets have been hit. According to the article on financial markets under the global pandemic of COVID-19, due to the nervousness of the market and the introduction of some policies, such as unlimited quantitative easing, zero-percent interest rate, etc., financial markets have become more unstable, financial markets in all countries risk levels have risen sharply. This has led to significant volatility in the market. According to the article COVID-19 and Corporate Performance in the Energy Industry, market results show that, in this crisis, companies in the energy industry have received a relatively large impact, there has been relatively large market volatility, and stock prices have declined. According to the article on The Changes of Apple's Stock Price During the Pandemic, companies in the high-tech industry, represented by Apple and Microsoft, suffered some losses in the early stages of the crisis, but managed to recover with their own internal adjustments. Its share price has continued to grow during the crisis. According to the article on Measuring the impact of COVID-19 on stock prices and profits in the food supply chain, companies in the food industry experience greater stock price volatility and investment risks during the crisis, and the cumulative return of company stocks will be lower. It is easy to see that the prediction results of the model are correct for companies in the energy and high-tech industries, but not accurate for companies in the food industry. This prediction is correct to a certain extent. Theoretically speaking, the cumulative returns of stock investments of companies in all industries can be predicted more accurately. However, this is not the case, and the error may come from the following aspects. Because the training data of the model comes from January 2011 to January 2022, and during most of this period, the data of the financial market has not been affected by the epidemic and war, so the model may fluctuations are not sensitive enough. If we want to solve this special problem, we still need to invest a lot of work in the future to carry out more in-depth research.

3.3 Portfolio Optimization with Monte Carlo Simulations

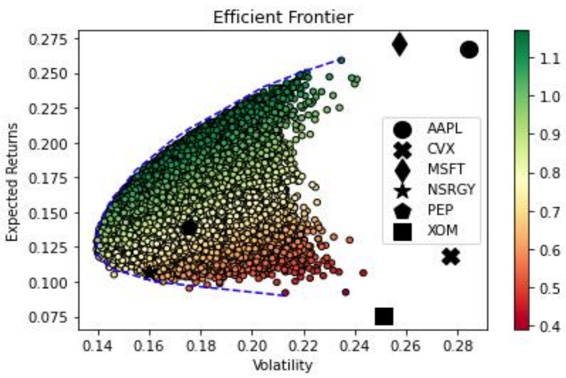

Figure 5. The Efficient Frontier of Six Stocks

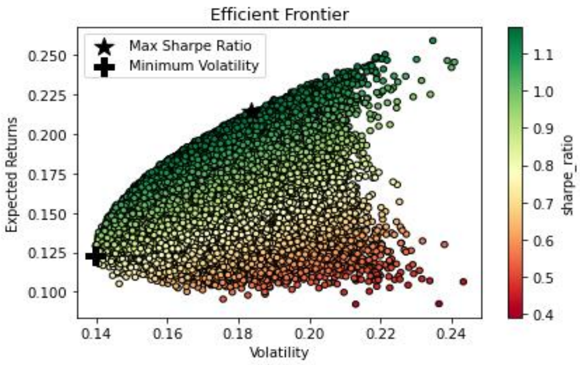

Figure 6. The Efficient Frontier of The Maximum Sharpe Ratio and The Minimum Volatility Ratio

After generating the two types of portfolios, the efficient frontier of six stocks is represented by the blue dotted line in Figure 5. The six different shaped icons represent the predicted returns of the underlying stocks with favorable Sharpe and Volatility Ratio in Figure 6. Portfolios with various random Sharpe Ratio and Volatility Ratio are stippled. And these two kinds of portfolios are invested in the following proportions shown in Table 2 and Table 3.

Table 2. Performances and Weights of Maximum Sharpe Ratio portfolio

Maximum Sharpe Ratio Portfolio	
Performance	Returns: 21.45% Volatility: 18.34% Sharpe ratio: 116.98%
Weights	AAPL: 26.33% CVX: 0.05% MSFT: 36.11% NSRGY: 16.99% PEP: 19.69% XOM: 0.83%

Table 3. Performances and Weights of Minimum Volatility portfolio

Minimum Volatility Portfolio	
Performance	Returns: 12.34% Volatility: 13.97% Sharpe ratio: 88.31%
Weights	AAPL: 0.31% CVX: 2.22% MSFT: 5.59% NSRGY: 46.60% PEP: 32.42% XOM: 12.87%

The yields of Maximum Sharpe Ratio portfolio achieved to 21.45% with higher volatility rates and per unit of risk, especially the two tech stocks MSFT and AAPL occupy more than 50% of the portfolio. CVX and XOM, which are in the energy sector, hold minor proportion due to their relatively smaller yields and relatively larger risk fluctuations. Overall, it is more suitable for investors who prefer high risk and high yield.

In the conservative case with the Minimum Volatility Ratio, the yield only attains to 12.34% of which the consumption stocks NSRGY and PEP, account for more than 70% of the investment.

This type of low-risk stock investment has a certain relationship with the consumer industry's stable status all year round, and is more suitable for steady investors.

4 CONCLUSION

In this research study, the financial mathematical models were transferred into a computer model using the computing power of python, The use of computers to complete the preliminary stock market forecast is innovatively realized, so that the forecast and asset distribution are more accurate. It can be clearly seen that the MAE and RMSE of the multi-factor model, Linear Regression model and Decision Tree Regressor model are all less than 1, which means the prediction effect of the experiment is very significant. This demonstrates the feasibility of emerging computer algorithms to replace traditional methods. The results in this study can improve efficiency in the field of stock investment, simplify the operation process, and replace the traditional method. Although it is still in the preliminary stage and cannot cover the impact of some emergencies, such as epidemics, wars, etc., it has shown great application potential in many aspects, such as stock and futures investment. In the future, further studies may conduct more in-depth research on the problem of unexpected events affecting the accuracy of computer models' return forecasts for stocks.

REFERENCES

[1] T. Talpsepp, L. Kristjan, and V. Tarvo. "Academic abilities, education and performance in the stock market." Journal of Banking & Finance, 117, 2020.

[2] S. Naseem, et al. "The investor psychology and stock market behavior during the initial era of COVID-19: a study of China, Japan, and the United States." Frontiers in Psychology, vol. 12, 2021.

[3] D. Valle-Cruz, et al. "Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods." Cognitive computation, vol. 14, 2022, pp. 372-387.

[4] M. Syakur, et al. "Integration k-means clustering method and elbow method for identification of the best customer profile cluster." IOP conference series: materials science and engineering. vol. 336, no. 1. IOP Publishing, 2018.

[5] P. Smith, G. Siva, and L. Ping. "A comparison of random forest regression and multiple linear regression for prediction in neuroscience." Journal of neuroscience methods, vol. 220, 2013, pp. 85-91.

[6] Y. Qiu, et al. "Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training." Biomedical Signal Processing and Control, vol. 72, 2022.

[7] G. Seber, and J. Alan. "Linear regression analysis. " John Wiley & Sons, 2012.

[8] G. Biau, and S. Erwan. "A random forest guided tour." Test, vol. 25, 2016, vol. 197-227.

[9] A. Perold. "The capital asset pricing model." Journal of economic perspectives, vol. 18, 2004, pp. 3-24.

[10] J. Griffin, "Are the Fama and French factors global or country specific?" The Review of Financial Studies, vol. 15, 2002, pp. 783-803.