Journal of Industrial Information Technology and Application Vol.5 No.1

ISSN(Online): 2586-0852

A Novel Method For Stock Selection Using Weight Threshold

Thien Nguyen [1,*], Trang Nguyen [2]

[1] Natural Language Processing and Knowledge Discovery Laboratory, Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[2] Faculty of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia

Abstract. This paper studies the portfolio with the optimal in-sample Sharpe ratio, constructed from a large stock pool. Basing on the portfolio, this paper proposes to select only stocks which weights surpass a certain weight threshold in order to avoid over-fitting. The weights of the selected stocks will be normalized to 100%. The new portfolio has better out-of-sample Sharpe ratio than out-of-sample Sharpe ratio of the original portfolio, as the same time, has significantly smaller size, which is more manageable for investors.

The first part of the paper explains the goal of our work. The second part reviews related works on building an investment portfolio. The third section presents the proposed method. The fourth part describes the data in this study. The fifth and sixth parts of the paper discusses experimental results and conclusion, respectively.

With regard to the focus of this conference, an experiment was performed on 6-year historic price data of stocks listed on the Ho Chi Minh City Stock Exchange (HSX), the largest stock market in Vietnam.

Experimental results show that the proposed method of building a portfolio from the HSX market obtains the better out-of-sample Sharpe ratio, reducing the size of portfolio.

Keywords: Portfolio construction, Sharpe ratio, Ho Chi Minh City Stock Exchange

---

* Corresponding author: nguyenchithien@tdtu.edu.vn

Received: Feb 8, 2021; Revised: Feb 17, 2021; Accepted: Mar. 31, 2021

1. Introduction

A primary concern of investors on stock exchange markets is stock selection in portfolio management. Stock selection has been studied my many researchers using both fundamental analysis and technical analysis.

Fundamental analysis evaluates stocks and companies represented by the stocks. It determines the intrinsic values of stocks by analyzing financial factors, management of the companies. Fundamental analysis also considers the microeconomic conditions. The aim of fundamental analysis is to produce the expected future price of stocks that an investor can compare with their current market price. Then, fundamental analysts decide which stocks to select to build a portfolio. Taken together, the complicated process of fundamental analysis requires specialists and costs heavily, especially, when a large stock pool and related companies need to be studied.

On the other hand, technical analysis only considers the market price series of stocks. It assumes that the market price can be disconnected from business fundamentals and is govern solely by the demand and supply from market participants. Technical analysis assumes that price is stationary in short enough period of time despite of its non-stationary characteristics in general. Using such assumption we can apply statistical methods to time series of stock market prices. Therefore, the process of technical analysis can be computerized, utilized by a lot of people, hence, its cost is divided, and reasonable.

Investors may choose to combine both fundamental analysis and technical analysis in stock selection. This study aims to help such investors by proposing a technical method to construct a portfolio with small size, which is more manageable for fundamental analysis, and high technical score in terms of Sharpe ratio, calculation of which can be computerized. The method proposes to select stocks using threshold of weights of stocks in the portfolio with the optimal in-sample Sharpe ratio.

According to Harding [1], Sharpe ratio can be define as a statistic which aims to sum up the desirability of a risky investment strategy or instrument by dividing the average period return in excess of the risk-free rate by the standard deviation of the return generating process. Sharpe [2] distinguishes between ex-ante Sharpe ratio and ex-post Sharpe ratio. Ex-post Sharpe ratio is computed from historic data, while ex-ante indicates the expected value in the forthcoming period. In this paper we use another terminology from the statistics and machine learning community. Throughout this paper, the term in-sample Sharpe ratio will refer to the value computed from data sample in a certain period of time. The out-of-sample Sharpe ratio is the Sharpe ratio computed from another data sample in the following period of time.

Portfolio, giving optimal in-sample Sharpe ratio, does not guarantee good out-of-sample Sharpe ratio due to the problem of over-fitting. The problem of over-fitting is very popular in the statistics and machine learning community. According to Cawley [3], an excessively complex model with too many parameters leads to over-fitting. The over fitted model has poor predictive performance, because it overreacts to the training data sample. We believe that the problem could be happen here in finance in constructing a portfolio, so we suggest to select only stocks which weights surpass a certain weight threshold in order to avoid over-fitting. By this way, we reduce in-sample Sharpe ratio, but improve out-of-sample Sharpe ratio, which represents the forecasting ability.

2. Literature Review

There is a large volume of published studies which focus on computer-based methods of constructing an investment portfolio. Machine learning techniques in computer science were widely applied to data from stock markets. Chavarnakul and Enke [4] proposed a hybrid stock selection and trading system for intelligent technical analysis-based equivolume charting. Esfahanipour and Mousavi [5] found a genetic programming model to select stocks and generate risk-adjusted technical trading rules in stock markets. A hybrid stock selection model using genetic algorithms and support vector regression is proposed by Huang [6]. Support vector machine is used by Yu et al. [7] for stock selection. Surveys such as that conducted by Hu et al. [8] have shown application of evolutionary computation for rule discovery in stock selection and trading. Yang et al. [9] combined stock selection with stock prediction based on computational intelligence. Tan et al. [10] used random forest to select stocks. Yuan et al. [11] also deployed random forest in feature selection and stock price trend prediction, then used the results to select stocks.

Despite the rich literature on the topic, it was difficult to apply these different methods to emerging markets, i.e. Vietnam. Compared to developed markets emerging markets are not complete. For example, Vietnam lacks tools like short selling, regulations and enforcement in the Vietnam market are different from those in developed market, restriction on foreign ownership leads to disproportionate effect of foreign fund-flows on market return and volatility in the relatively small Vietnam market [12], making fund flow information critical in decision making. Due to the large size of institutional investors and small stock market size in Vietnam institutional ownership hugely effects volatility of stock returns [13].

There are numerous quantitative criteria for selecting a set of stocks from a large, diverse pool, such as dividend yield, volatility, trend, momentum. Assuming that a trend tends to continue on the same direction, Covel [14] proposed to follow trend where instead of trying to predict the future, selecting stocks are made reactively to the price movement. A similar approach is taken by Peachavanish [15] where two types of technical indicators were used: trend and momentum. Tola [16] demonstrated that use of clustering algorithms can improve the reliability of the portfolio in terms of the ratio between predicted and realized volatility.

In this study we utilize Sharpe ratio as a quantitative criteria for selecting a set of stocks from a large pool. We focus on constructing a portfolio with improved out-of-sample Sharpe ratio. It has been argued that for certain applications the Sharpe ratio is not the most appropriate performance measure, when the returns are far from normally distributed. On the other hand, there is recent evidence that the Sharpe ratio can result in almost identical fund rankings compared to alternative performance measures in Eling et al. [17]. We do not enter this debate. Instead, we believe that the task of choosing the appropriate performance measure is up to investors. Our aim is to provide a reliable method of stock selection in case the Sharpe ratio is deemed of interest.

3. Methodology

A. Data preparation

As of Jun 2017, there were over 300 listed companies on the Ho Chi Minh City stock exchange market (HSX), Vietnam with total market capitalization of over VND 500 trillion. Among them a set of $N = 114$ stocks from HSX market is used in this paper. These stocks are chosen because they are constantly available on the market from January 2011 to December 2016, therefore, most likely to be investment-grade and least likely to be maliciously manipulated. Stock prices change on a daily basis, altering the value of investments. Daily stock returns could be calculated to monitor the magnitude of this change. The daily return measures the change in a stock's price as a percentage of the previous day's closing price. A positive return means the stock has grown in value, while a negative return means it has lost value. A stock with lower positive and negative daily returns is typically less risky than a stock with higher daily returns, which create larger swings in value. From daily closing price of stocks the daily return is calculated on the base of the Equation 1:

$$r_j = \frac{(v_j - v_{j-1})}{v_j} \times 100\% \quad (1)$$

where r_j is the daily return at day j, v_j is the closing price on the current trading day j, v_{j-1} is the closing price on the previous trading day $j - 1$.

For each stock i, where $i = 1, \dots, N$, time series R_i of daily returns is formed (Equation 2)

$$R_i = r_{i1}, r_{i2}, \dots \quad)2)$$

where r_{ij} is the daily return of stock i at day j.

B. Constructing portfolio with optimal Sharpe ratio

In finance, the Sharpe ratio (also known as the Sharpe index, the Sharpe measure, and the reward-to-variability ratio) is a way to examine the performance of an investment by adjusting for its risk. The ratio measures the excess return (or risk premium) per unit of deviation in an investment asset or a trading strategy, typically referred to as risk (and is a deviation risk measure), named after William F. Sharpe [2] (see Equation 3):

$$S = \frac{E[R] - R_f}{\sigma_R} \quad (3)$$

where S is Sharpe ratio, $E[R]$ is the expected value of return, R_f is constant risk-free return, σ_R is standard deviation of return representing risk. Risk-free return could be savings interest rate.

Time series of daily returns of all stocks are used to construct the portfolio with optimal in-sample Sharpe ratio. The weight of each stock is calculated using method SLSQP implemented by Jones et al. [18]. The method uses Sequential Least Squares Programming to minimize a Sharpe ratio function of stocks weights with bounds from 0 to 1 (Equation 4):

$$S(W) = \frac{W^T \times R - R_f}{W \times \Sigma_R \times W^T} \tag{4}$$

where $S(W)$ is Sharpe ratio function, W^T is row-vector of stock weights, W is column-vector of stock weights, R is column-vector of expected stock daily return, Σ_R is their covariance matrix between stocks

Stock weight is its portion in a portfolio. The sum of all stocks weights is equal to 100%. Stock weight has bounds from 0 to 1. It means investor cannot do short selling. They can buy and only sell what they have already had. Until now (2017), Vietnam market does not allow short selling, so the limit on bounds of stock weights is reasonable.

Jones et al. wrap the SLSQP Optimization subroutine originally implemented by Dieter Kraft [19].

C. Selecting stocks by weight threshold

When portfolio with optimal in-sample Sharpe ratio is constructed using a large stock pool, in the result each stock has some weight. By that way the portfolio with optimal in-sample Sharpe ratio consists of many stocks. Portfolio of a lot of stocks is not a problem for traders who use technical techniques, but a problem for those who use fundamental analysis. In addition, for traders in emerging market with high transaction tax and fee, portfolio with a lot of stocks is not profitable, since it needs more stock selling/ buying to maintain the portfolio.

We observed that in the portfolio many stocks possess insignificant weights. Inspired by election threshold in party-list proportional representation systems, we suggest to keep only stocks with weights which surpass some threshold. The electoral threshold is the minimum share of the primary vote which a candidate or political party requires to achieve before they become entitled to any representation in a legislature. This limit can operate in various ways. For example, in party-list proportional representation systems, an election threshold requires that a party must receive a specified minimum percentage of votes (e.g., 5%). In addition, reducing the number of stocks can reduce over-fitting problem, therefore, improve out-of-sample Sharpe ratio.

Full algorithm to select stocks is follows:

Input:

A set of N stocks with their daily returns

Weight threshold $h > 0$ to select stocks

Output:

Size-reduced portfolio with integer weight for each stock. The new portfolio has significantly small size ($M \ll N$)

Step 1: Calculate the weight w_i for each stock i, where $i = 1, \dots, N$, in the input set to achieve the maximum of in-sample Sharpe ratio.

Step 2: Select stocks with weight greater than weight threshold $w_i > h$. Such way we create new portfolio with size smaller than original one w_i, $i = 1, \dots, M$, $M \ll N$.

Step 3: Normalize the weight of selected stocks to 100%

Step 4: Round the weights to integer to make them easier for fundamental analysis, keeping their sum equal to 100%

1) Sort weights decently by fractional part
2) Calculate summation of integer part of all weights $U = \sum \lfloor w_i \rfloor$, where U is the sum of the integer part of all stock weight, $\lfloor w_i \rfloor$ is the integer part of weight of stock i in the size-reduced portfolio
3) Update the weight of i-th stock by assigning their corresponding integer part $w_i = \lfloor w_i \rfloor$
4) Add one to the first $(100 - U)$ stock weights sorted decently by fractional part $w_i = w_i + 1,\ i = 1,2, \dots, (100 - U)$

D. Evaluating the proposed method of constructing portfolio

To evaluate the proposed method of constructing portfolio we use cross-validation procedure for time series, which theoretical background is provided by Bergmeir et al. [20]. Step by step procedure is provided below:

1) Divide time interval from January 2011 to December 2016 into sequence of K separate fragments with some equal time length T.
2) Use daily returns in k-th time fragment as the training sample to calculate weight w_i^k, $i = 1, \dots, N$ for each stock to maximize Sharpe ratio. In such way, we utilize already known daily returns in k-th time fragment as historic data to predict the future. Therefore, we calculated the in-sample Sharpe ratio.
3) Use steps 2-4 of the proposed algorithm for selecting stocks by weight threshold to select $M \ll N$ stocks, which weights w_m^k, $m = i_1, \dots, i_M$ are higher than weight threshold h.
4) Use selected M daily returns in $(k + 1)$-th time fragment to calculate vector of expected daily returns and covariance matrix between stocks of size-reduced (M) portfolio
5) Use all daily returns in $(k + 1)$-th time fragment to calculate vector of expected daily returns and covariance matrix between stocks of full-size (N) portfolio
6) Use weights, expected returns, covariance matrix from steps 2-5 to calculate Sharpe ratios of full-size (N) portfolio and of size-reduced (M) portfolio,

applying Sharpe ratio function $S(W)$. We calculated out-of-sample Sharpe ratios.

7) Repeat steps 2-6 for time fragment $k = 1, \ldots, K - 1$
8) Compare the calculated out-of-sample Sharpe ratios.

4. Data

For this study a set of $N = 114$ stocks from HSX market is prepared according to the data preparation procedure in chapter 3. These stocks are listed by companies, categorized into 10 different economic sectors: Materials, Financial, Consumer Staples, Real Estate, Energy, Industrials, Utilities, Information Technology, Consumer Discretionary, Health Care. The classification is provided by Ho Chi Minh City Stock Exchange, Ministry of Finance, Vietnam at its official website www.hsx.vn.

The pie chart below (Figure 1) shows the shares of stocks in economic sectors.

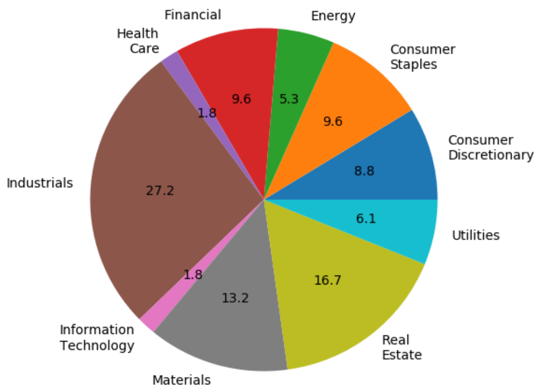

Fig.1 - Shares of stocks in economic sectors

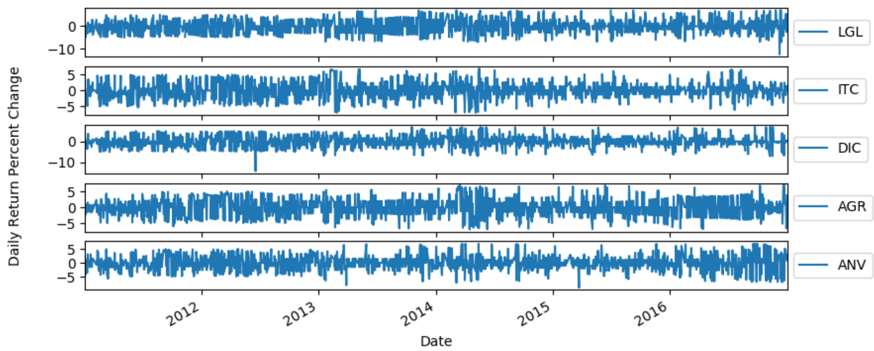

Fig. 2 - The daily stock returns

Daily stock returns span continuously on the market from January 2011 to December 2016. Figure 2 shows the daily returns of five stocks. They represent the most popular economic sectors. The stock LGL is from Industrials, ITC from Real Estate, DIC from Materials, ARG from Financial, ANV from Consumer Staples.

5. Results and Discussion

According to the procedure evaluating the proposed method of constructing portfolio in chapter 3 we divide time interval from January 2011 to December 2016 into sequence of $K = 36$ separate fragments with equal time length $T = 2$ months. In this analysis we apply risk-free rate $R_f = 0.025$. The value was chosen because it is closed to average interest rate on a savings account in Vietnam. For example, Joint Stock Commercial Bank for Foreign Trade of Vietnam (Vietcombank), one of the biggest bank in Vietnam, applies one year interest rate of 6.3%, which is equivalent to $\frac{6.3\%}{250\ trading\ days} = 0.0252\%$ per day.

In Vietnam there has not been any default of banks. Therefore, interest rate of savings may be counted as rick-free.

In this experiment we used stock weight threshold of 3%, which is inspired by the fact, that The Parliamentary Assembly of the Council of Europe recommends for parliamentary elections a threshold not higher than 3%.

The results of the procedure evaluating the proposed method of constructing portfolio are presented in Figure 3.

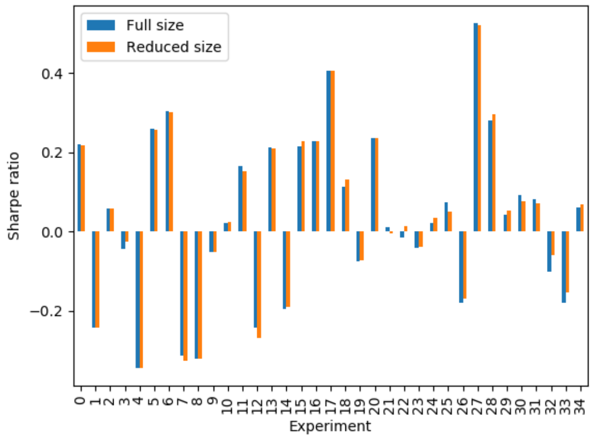

Fig. 3 - Sharpe ratios of full-size and reduced-size portfolios

As shown in Figure 3, out-of-sample Sharpe ratios of reduced-size portfolios are comparable with that of full-size portfolios. The mean score of reduced-size was 0.039, meanwhile, the mean score of full-size was 0.037. Interestingly, variation of the score of reduced-size is slightly better than variation of the score of full-size. Variation of the score of reduced-size was 0.042 and variation of the score of full-size was 0.043.

Distributions of out-of-sample Sharpe ratios of full-size and reduced-size portfolios are shown in Figure 4.

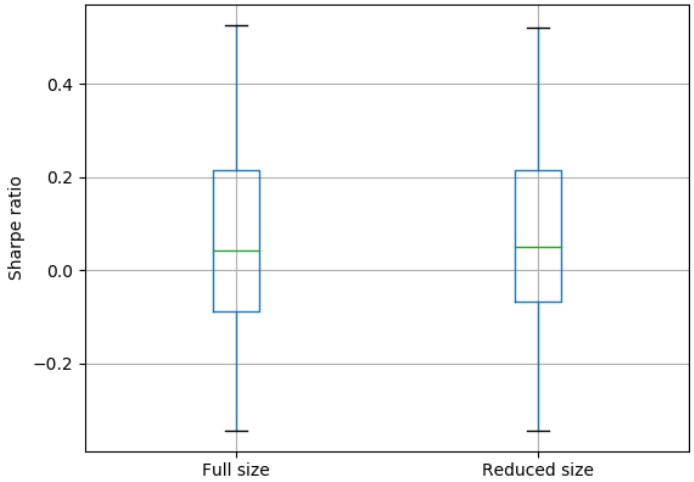

Fig. 4 - Distributions of Sharpe ratios

Overall, these results indicate that the method of selecting stocks constructs reduced-size portfolio which performed not worse than portfolio with full-size. It means that the proposed method is effective in picking out good stocks. When the number of stocks in full-size portfolio is always 114, the number of stocks in reduced-size portfolio is on average 9, which is greatly lower. Detail information of the number of stocks in reduced-size portfolio through experiments is presented in Figure 5.

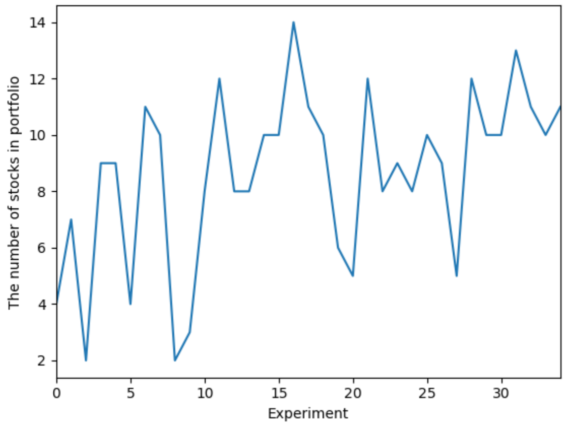

Fig. 5 – The number of stocks in portfolios with reduced-size

Figure 6 shows the number of economic sectors in portfolios with reduced-size. Its mean value is 6.

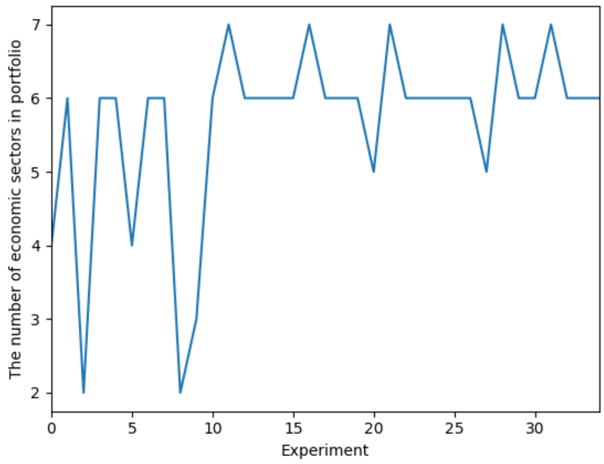

Fig. 6 – The number of economic sectors in portfolios with reduced-size

Portfolio with full-size consists of all 114 stocks listed by companies, operating in 10 different economic sectors. It means that the ratio of economics sectors to stocks is equal to $\frac{10}{114} = 0.09$. As the same time, portfolio with reduced-size consists of, on average, 9 stocks listed by companies, operating in 6 different economic sectors. Hence, the ratio of economics sectors to stocks is $\frac{6}{9} = 0.67$. The observed increase in the diversity could be attributed to the better performance of the proposed method, which leads to higher out-of-sample Sharpe ratio.

6. Conclusions

This paper contributes a method that uses weight threshold to identify a small group of stocks that are most likely to outperform a larger stock pool. The results show that the proposed method using a small group of stocks can on average outperform a large stock pool.

To improve, additional works need to be done to identify weight threshold for selecting stocks. In addition, it is likely that positive correlation exists between the diversity of economic sectors in portfolio and the value of out-of-sample Sharpe ratio. However, more research on this topic needs to be undertaken before the association between the diversity of economic sectors in portfolio and the value of out-of-sample Sharpe ratio is more clearly understood.

References

[1] D. Harding, “A critique of the Sharpe ratio,” *Wilt. Cap. Manag.*, 2002.

[2] W. F. Sharpe, “The sharpe ratio,” *J. Portf. Manag.*, vol. 21, no. 1, pp. 49–58, 1994.

[3] G. C. Cawley and N. L. C. Talbot, “On over-fitting in model selection and subsequent selection bias in performance evaluation,” *J. Mach. Learn. Res.*, vol. 11, pp. 2079–2107, 2010.

[4] T. Chavarnakul and D. Enke, “A hybrid stock trading system for intelligent technical analysis-based equivolume charting,” *Neurocomputing*, vol. 72, no. 16–18, pp. 3517–3528, 2009.

[5] A. Esfahanipour and S. Mousavi, “A genetic programming model to generate risk-adjusted technical trading rules in stock markets,” *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8438–8445, 2011.

[6] C.-F. Huang, “A hybrid stock selection model using genetic algorithms and support vector regression,” *Appl. Soft Comput.*, vol. 12, no. 2, pp. 807–818, 2012.

[7] H. Yu, R. Chen, and G. Zhang, “A SVM stock selection model within PCA,” *Procedia Comput. Sci.*, vol. 31, pp. 406–412, 2014.

[8] Y. Hu, K. Liu, X. Zhang, L. Su, E. W. T. Ngai, and M. Liu, “Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review,” *Appl. Soft Comput.*, vol. 36, pp. 534–551, 2015.

[9] F. Yang, Z. Chen, J. Li, and L. Tang, “A novel hybrid stock selection method with stock prediction,” *Appl. Soft Comput.*, vol. 80, pp. 820–831, 2019.

[10] Z. Tan, Z. Yan, and G. Zhu, “Stock selection with random forest: An exploitation of excess return in the Chinese stock market,” *Heliyon*, vol. 5, no. 8, p. e02310, 2019.

[11] X. Yuan, J. Yuan, T. Jiang, and Q. U. Ain, “Integrated long-term stock selection models based on feature selection and machine learning algorithms for China stock market,” *IEEE Access*, vol. 8, pp. 22672–22685, 2020.

[12] X. V. Vo, “Foreign ownership and stock return volatility--Evidence from Vietnam,” *J. Multinatl. Financ. Manag.*, vol. 30, pp. 101–109, 2015.

[13] X. V. Vo, “Does institutional ownership increase stock return volatility? Evidence from Vietnam,” *Int. Rev. Financ. Anal.*, vol. 45, pp. 54–61, 2016.

[14] M. Covel and B. Ritholtz, *Trend following*. PH Professional Business, 2009.

[15] R. Peachavanish, “Stock selection and trading based on cluster analysis of trend and momentum indicators,” in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2016, vol. 1, pp. 317–321.

[16] V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna, “Cluster analysis for portfolio optimization,” *J. Econ. Dyn. Control*, vol. 32, no. 1, pp. 235–258, 2008.

[17] M. Eling and F. Schuhmacher, “Does the choice of performance measure influence the evaluation of hedge funds?,” *J. Bank. Financ.*, vol. 31, no. 9, pp. 2632–2647, 2007.

[18] E. Jones, T. Oliphant, P. Peterson, and others, “SciPy: Open source scientific tools for Python,” 2001.

[19] D. Kraft and others, “A software package for sequential quadratic programming,” 1988.

[20] C. Bergmeir, R. J. Hyndman, and B. Koo, “A note on the validity of cross-validation for evaluating autoregressive time series prediction,” *Comput. Stat. Data Anal.*, vol. 120, pp. 70–83, 2018.