Stability of Price Prediction Rule for High-Frequency Trading

Kenichi Yoshida* and Shigeki Kohda

***Abstract*—When using the data mining method for stock trading, it is essential to confirm the future stability of the rule found by the data mining method. Even if a rule can explain past data, in situations where the model behind the data changes, there is a risk of loss if an investment is made based on the rule learned using past data. In this study, we analyzed the mechanism behind the high-frequency trading prediction rule proposed in past research and demonstrated its stability. By demonstrating stability, this paper shows that investment using the rule can continue to be the basis for low-risk investment methods in the future.**

***Index Terms*—Data mining, high-frequency trading, stock price prediction, algorithmic trading**

I. Introduction

The efficient market hypothesis (EMH [1]) was widely accepted in financial market studies. It asserts that financial markets are "informationally efficient." As a consequence, future stock prices change randomly. In other words, they are unpredictable [2]. However, recent studies reported the predictability of high-frequency trading (HFT), e.g., [3] and [4]. Here, HFT is a trading method that uses an algorithm for stock trades and is executed by machines without human decision-making. Since high-frequency traders are the leading participants in today's exchanges, we believe EMH is denied as far as HFT is concerned.

However, when using the data mining method for stock trading, it is essential to confirm the future stability of the rule found by the data mining method. Even if a rule can explain past data, in situations where the mechanism behind the data changes, there is a risk of loss if an investment is made based on the rule learned using past data. Unfortunately, previous studies only reported experimental results based on past data. Even if they follow an out-sampled framework, the out-sampled framework only shows the predictability of the past data. The out-of-sample framework cannot guarantee that the mechanism that brings the predictability will not change in the future. It does not guarantee the eternity of found rules.

In this study, we analyzed the mechanism behind the HFT's prediction rule proposed in past research [3] and demonstrated its stability. To be precise, we found a mechanism behind the rule in [3]. Since the mechanism seems to be the essential (i.e., eternal) mechanism of HFT, we believe the rule found in [3] is also eternal and stable. By demonstrating its stability, this paper shows that investment using the rule can continue to be the basis for low-risk investment methods in the future.

The remainder of this paper is organized as follows. First, Section II discusses the related studies in this domain. Subsequently, Section III explains the prediction rule found in [3]. Then, Section IV discusses its stability. Finally, Section 5 summarizes our findings.

Manuscript received October 11, 2022; revised December 23, 2022; accepted March 1, 2023.

Kenichi Yoshida is with Graduate School of Business Science, University of Tsukuba, Otsuka 3-29-1, Bunkyo, Tokyo 112-0012, Japan.

Shigeki Kohda is a Doctoral Candidate of University of Tsukuba, Japan.

*Correspondence: yoshida.kenichi.ka@u.tsukuba.ac.jp (K.Y.)

II. Research on HFT and Its Forecast

The characteristics of HFT have been actively studied in recent years. By analyzing NASDAQ HFT transaction data, Brogaard *et al*. [5] found a correlation between price changes and buy/sell orders. Salighehdar *et al.* [6] analyzed price liquidity and found a cluster structure among transactions. Manahov *et al*. [7] analyzed the behavior of HFT in the Australian stock market. Consequently, it was found that numerous limit orders are cancelled within 50 milliseconds. They also showed a high tendency of cancellation for recently placed orders near the best market price. Matteo *et al.* [8] also reported similar findings, i.e., short-time reaction in HFT trading. Saito *et al.* [9] classified the top ten stocks on the Tokyo stock exchange (TSE) market in terms of the number of stock orders. According to [9], when stock prices rise, the execution volume and order volume shares increase with the market's direction. To be precise, they arranged trading information in 10-second units and reported that the HFT would place orders in the direction of market prices. This is different from what Niederhoffer *et al*. [10] reported as a characteristic of the traditional market. Niederhoffer et al. [10] analyzed the trend of future price fluctuations and reported that the direction of price fluctuations would reverse. Kohda *et al*. [11] perform different analyses by changing time scales and supports [10]. Lakshmi *et al.* [12] showed that HFT improves transaction costs, volatility, and trading imbalances in Indian stock exchanges. Anagnostdis *et al.* [13] investigated the liquidity risk in the French HFT stock market and found a relationship between the order size and the price fluctuations.

Forecasting of stock price fluctuations also has been studied for a long time, e.g., [14-18]. Even if the topic is limited to HFT, Yoshida *et al.* [3] reported a simple price prediction rule for HFT. Zhang *et al.* [4] also proposed a

doi: 10.18178/ijfcc.2023.12.2.599

prediction method with an accuracy of 80%.

III. Prediction Rule of HFT Price

In [3], Yoshida et al. show that a simple rule can classify short-term stock prices with an 82.9% accuracy. They count the number of buy and sell orders.

Their analysis uses order book information from the Tokyo stock exchange (TSE). The TSE has been distributing real-time stock exchange information through a high-speed trading network named "Arrowhead" [19]. When the TSE processes an order, it sends a UDP packet that contains information about the order, i.e., the stock ID, size of the order, and ask/bid price. In particular, each UDP packet contains the eight lowest ask orders and the eight highest bid orders (See Fig. 1). The packet can therefore be described as a list of orders that record the interest of buyers and sellers on a particular stock, and is essentially an order book for the stock. A comparison of the current packet with its predecessor for the same stock enables the extraction of the information of the corresponding order. e.g., the number of buy and sell orders.

```
e14 <= -1
|   e13 <= -1: n (111.0/24.0)
|   e13 > -1: p (878.0/166.0)
e14 > -1
|   e13 <= 0: n (845.0/135.0)
|   e13 > 0: p (125.0/26.0)

Number of Leaves :      4
Size of the tree :      7
...
=== Error on test data ===

Correctly Classified Instances    942  81.9843 %
Incorrectly Classified Instances  207  18.0157 %
```

Fig. 1. Information in UDP packets (in [3] Fig. 2).

Fig. 2 shows an example of a rule they learned from the TSE order book information. Here, "e14" is the difference between the recently quoted price and its predecessor. "e13" is "e11 - e12" where "e11" is number of UDP packets for ask (i.e., buy) with the same price as the most recent quoted price after the corresponding trade and "e12" is the number of UDP packets for bid (i.e., sell) with the same price as the most recent quoted price after the corresponding trade. "*n*" and "*p*" show the direction of price fluctuation.

In their experiments with the out-of-sample framework, [3] showed that this simple rule explains price fluctuations. However, it does not consider why such a simple rule holds. If such a rule is discovered, the structure of the market may change, and the discovered rule would no longer hold. The out-of-sample framework does not consider such a situation. Therefore, it is important to analyze the reason why this rule holds. However, [3] does not go into that much discussion. Thus, this paper analyses the mechanism behind this simple rule to show its stability.

IV. Mechanism Behind the Rule

We believe Fig. 3 shows the mechanism behind the rule shown in Fig. 2. It shows a simple trading strategy and branches an ordering process depending on whether the price rises or drops. First, if a stock price drops, the algorithm buys at the changed price (line 5). When the stock price returns to the original price, the algorithm sells the bought stock (line 6). In this case, Algorithm 1 can achieve a profit because it can sell higher than the original stock price. If it does not return to its original price within 5 minutes, it sells the stock (line 8). In this case, Algorithm 1 will lead to a loss. When a stock price rises, the algorithm sells the stocks at the changed price (line 10). It buys the stock when the stock price returns to the original price (line 11). If the stock price does not return to the original stock price within 5 minutes, Algorithm 1 also executes a short covering (line 13). In this case, it will lead to a loss.

```
J48 pruned tree
------------------

e14 <= -1
|   e13 <= -1: n (111.0/24.0)
|   e13 > -1: p (878.0/166.0)
e14 > -1
|   e13 <= 0: n (845.0/135.0)
|   e13 > 0: p (125.0/26.0)

Number of Leaves :      4
Size of the tree :      7

Time taken to build model: 0.14 seconds
Time taken to test model on training data: 0.11 seconds

=== Error on training data ===

Correctly Classified Instances        1608         82.0827 %
Incorrectly Classified Instances       351         17.9173 %
Kappa statistic                          0.6417
Mean absolute error                      0.1956
Root mean squared error                  0.3128
Relative absolute error                 58.6688 %
Root relative squared error             76.615  %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances             1959

=== Error on test data ===

Correctly Classified Instances         942         81.9843 %
Incorrectly Classified Instances       207         18.0157 %
Kappa statistic                          0.6393
Mean absolute error                      0.1978
Root mean squared error                  0.3145
Relative absolute error                 59.3132 %
Root relative squared error             77.014  %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances             1149
```

Fig. 2. j48 pruned tree (in [3] Fig. 12).

Note that similar trading strategies have been studied to exploit short-term price moves, e.g., [20] and [21]. We believe many high-frequency traders commonly use similar strategies, and the phenomenon reported in [11] are the side effects of this strategy. This strategy and similar strategies can make a profit if the price returns within 5 minutes. Kohda *et al.* [11] reported the probability of price return. According to their observation, the price will return with a probability of 90%. Many high-frequency traders seem to know this probability and use the strategy shown in Fig. 3.

Fig. 4 and 5 show the side effects of this strategy. Fig. 4 shows the number of orders per millisecond after each execution. As shown in Fig. 4, the number of orders per unit of time is high immediately after executions. There exists no apparent difference in the number of orders for the direction of price movement. Fig. 5 shows the ratio of buy orders after a stock price change. The dashed line is the percentage of buy orders after a price drops, and the solid line is after a price rises. The ratio of a buy order is significantly higher after its price drops and lower after its price rises (i.e., the ratio of a sell order is high).

```
procedure Simple Trading Strategy(P, p)
    P : Most recent execution price of Stock
    p : Previous execution price of Stock

    if P < p then
        Buy Stock at price P
        Try to Sell Stock at price p within 5min
        if Fail then
            Sell Stock at any price
    else if P > p then
        Sell Stock at price P
        Try to Buy Stock at price p within 5min
        if Fail then
            Buy Stock at any price
```

Fig. 3. Trading strategy.

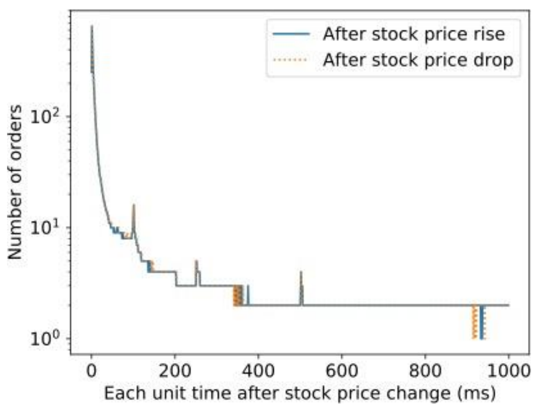

Fig. 4. Number of orders after price change (in [11] Fig. 5).

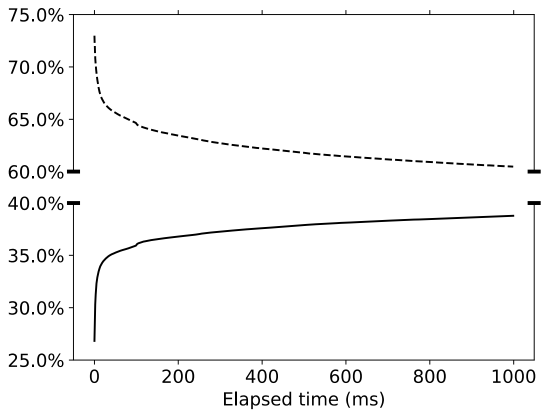

Fig. 5. Orders' ratio after stock price change (in [11] Fig. 7).

These two figures show a significant bias in buy and sell orders immediately after each execution. Thus, we conclude that the simple rule reported in [3] is a natural result caused by the trading strategy shown in Fig. 3. We cannot deny the possibility that there are other causes of this bias. Still, the existence of this bias strongly suggests the presence of traders using strategies similar to those shown in Fig. 3. In addition, we believe small price fluctuations are unavoidable characteristics of stock prices. This price fluctuation is behind the price return probabilities shown in Fig. 6. Thus, the probability of price return shown in Fig. 6 is an eternal characteristic of HFT, and high-frequency traders will not discard the trading strategy shown in Fig. 3. The rule found in [3] is eternal and stable.

The characteristics of HFT have been actively studied in recent years, and many studies attempt to predict future stock prices by data mining methods. Although it is essential to confirm the future stability of the rules found by the data mining method, few of them discuss the stability of the found rules. However, the above discussion using Fig.5,5, and 6 shows the stability of the rules proposed in [3]. Furthermore, though the data used in this study is only about the TSE stock market [19], we believe the discussion of this paper can apply to other stock markets.

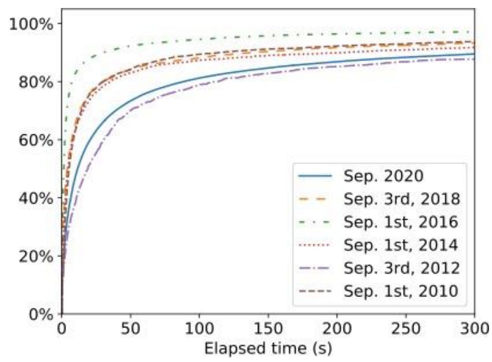

Fig. 6. Probability of price return after price change (in [11] Fig. 9).

V. Conclusion

This study analyses the model behind the high-frequency trading prediction rule proposed in [3] and demonstrates its stability. To be precise, the prediction rule results from the characteristics of HFT-specific trading strategies. Therefore, the rule would continue to hold because the trading strategies seem to hold in the future. By demonstrating stability, this paper shows that investment using the rule can continue to be the basis for low-risk investment methods in the future.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Kenichi YOSHIDA conducted the research and wrote the paper; Shigeki KOHDA analyzed the data; all authors had approved the final version.

Acknowledgment

This study was partly supported by JSPS KAKENHI Grant Number 19H04165.

References

[1] S. Basu, "Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis," *The Journal of Finance*, vol. 32, no. 3, pp. 663-682, 1977.

[2] B. G. Malkiel, "A random walk down Wall Street: Including a life-cycle guide to personal investing," WW Norton & Company, 1999.

[3] K. Yoshida and A. Sakurai, "Short-term stock price analysis based on order book information," *Information and Media Technologies*, vol. 10, no. 4, pp. 521-530, 2015.

[4] Z. H. Zhang, S. Zohren, and S. Roberts, "Deeplob: Deep convolutional neural networks for limit order books," *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 3001-3012, 2019.

[5] J. Brogaard, T. Hendershott, and R. Riordan, "High-frequency trading and price discovery," *The Review of Financial Studies*, vol. 27, no. 8, pp. 2267--2306, 2014.

[6] A. Salighehdar, Y. Liu, D. Bozdog, and I. Florescu, "Cluster analysis of liquidity measures in a stock market using high frequency data," *Journal of Management Science and Business Intelligence*, vol. 2, no. 2, pp. 1-8, 2017.

[7] V. Manahov, "High-frequency trading order cancellations and market quality: Is stricter regulation the answer?" *International Journal of Finance and Economics*, 2020.

[8] M. Aquilina, E. Budish, and P. O. Neill, "Quantifying the high-frequency trading # arms race #," *The Quarterly Journal of Economics*, vol. 137, no. 1, pp. 493-564, 2022.

[9] T. Saito, T. Adachi, T. Nakatsuma, A. Takahashi, H. Tsuda, and N. Yoshino, "Trading and ordering patterns of market participants in high frequency trading environment: Empirical study in the japanese stock market," *Asia-Pacific Financial Markets*, vol. 25, no. 3, pp. 179-220, 2018.

[10] V. Niederhoffer and M. F. M. Osborne, "Market making and reversal on the stock exchange," *Journal of the American Statistical Association*, vol. 61, no. 316, pp. 897-916, 1966.

[11] S. Kohda and K. Yoshida, "Characteristics and forecast of high-frequency trading," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 37, no. 5, pp. 1-9, 2022.

[12] A. Lakshmi and S. Vedala, "A study on low latency trading in indian stock markets," *International Journal of Civil Engineering and Technology*, vol. 8, pp. 733-743, 2017.

[13] P. Anagnostidis and P. Fontaine, "Liquidity commonality and high frequency trading: Evidence from the french stock market," *International Review of Financial Analysis*, vol. 69, p. 101428, 2020.

[14] A. Ang and G. Bekaert, "Stock return predictability: Is it there?" *The Review of Financial Studies*, vol. 20, no. 3, pp. 651-707, 2007.

[15] X. Ding, Y. Zhang, T. Liu, and J. W. Duan, "Deep learning for event-driven stock prediction," 2015.

[16] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using lstm, rnn and cnn-sliding window model," in *Proc. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1643-1647, 2017.

[17] R. Singh and S. Srivastava, "Stock prediction using deep learning," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 18569-18584, 2017.

[18] Z. G. Jin, Y. Yang, and Y. H. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9713-9729, 2020.

[19] Tokyo Stock Exchange, Arrownet, 2012.

[20] W. J. Duncan, *Stocks*, Lulu. com, 2016.

[21] W. L. Silber, "Marketmaker behavior in an auction market: An analysis of scalpers in futures markets," *The Journal of Finance*, vol. 39, no. 4, pp. 937--953, 1984.