Proceedings of CONF-MLA 2024 Workshop: Semantic Communication Based Complexity Scalable Image Transmission System for Resource Constrained Devices
DOI: 10.54254/2755-2721/83/2024GLG0068

Optimizing Short and Long Term Investment Returns Using Multi-Armed Slot Machine Algorithms

Qiaojia Liu

Northeastern University, 360 Huntington Avenue,200 Kerr Hall, Boston, MA 02125, USA

liu.qiaoj@northeastern.edu

Abstract. This study focuses on comparing the effectiveness of UCB, Thompson sampling and Epsilon-Greedy algorithms in multi-armed slot machine algorithms for short-term and long-term investment return optimization in financial markets. This analysis examines stock performance data from Tesla, General Motors, and Ford over specific periods: five years of weekly data (2019-2024) and six months of daily data (February to August 2024). From the results, it is shown that for long-term investments, Over the five-year period, Thompson Sampling outperformed UCB and Epsilon-Greedy algorithms in terms of stability and overall return, demonstrating a consistent upward trend with fewer fluctuations. For short-term investments, although the UCB algorithm performs similarly to the Thompson sampling algorithm, the returns in the UCB algorithm grow progressively slower as the timeline is lengthened. In the six-month daily data analysis, the Epsilon-Greedy algorithm excelled in capturing short-term gains due to its aggressive exploration mechanism, though it also showed higher volatility in response to market fluctuations, and the algorithm should be chosen to be able to keep up with the fluctuations in the stock market in real time, and the investor should take into account the cost of time as well as changes in the market in order to formulate an optimal investment strategy.The findings of this study will provide a practical type of approach to investors when developing investment strategies. Future research will explore other classes of new algorithms and consider more market factors so that they can be better adapted to the complex investment environment in the future.

Keywords: Multi-Armed Bandit Algorithms, Algorithm Performance Comparison, Statistical Analysis and Results, Investment Strategy Optimization.

1. Introduction

With globalization and the rapid development of technology, financial markets have become increasingly complex and uncertain. In this environment, traditional investment strategies have gradually become insufficient to effectively respond to rapid market changes. Such limitations pose a number of challenges to investors, including difficulties in making timely portfolio adjustments, coping with market volatility, and achieving desired returns in an uncertain market environment. Therefore, it becomes particularly important to explore novel algorithms that can adapt to high volatility and uncertainty environments.

The theme of this research focuses on the application of multi-armed slot machine algorithms in short-term and long-term investments in financial markets. The study adopts an empirical approach,

DOI: 10.54254/2755-2721/83/2024GLG0068

combining historical stock market data and Monte Carlo simulation, to compare the performance of the three algorithms, UCB, Thompson sampling and ε-Greedy, in different market situations. Through quantitative analysis, the effectiveness of each algorithm in balancing exploration and exploitation and its impact on investment returns are assessed.

2. Literature Review

As an advanced decision optimization tool, Multi-Armed Bandit (MAB) algorithms have demonstrated their effectiveness in a wide range of domains such as media content optimization, advertising strategy, and healthcare [1-3]. However, despite the excellent performance of MAB algorithms in these areas, their potential and effectiveness in financial investment strategies, especially in applications that balance short-term and long-term investments, have not been fully investigated.

Existing studies have mainly focused on the performance of MAB algorithms in deterministic environments, while relatively little research has been conducted on their adaptability and efficiency in highly uncertain and dynamically changing environments like financial markets. Nevertheless, it has been shown that these algorithms are able to maximize returns by dynamically adapting their strategies, thus demonstrating the potential to optimize complex decisions [4, 5]. In addition, it has been shown in the literature that insufficient data sharing in the field of financial research is an important factor limiting the application of these algorithms, a research gap that further highlights the need for exploring efficient algorithms in financial markets [6, 7].

In addition, a study discusses the limitations of upper confidence interval-based strategies, noting that research on non-parametric methods is gaining favor [8]. Other studies have addressed the application of market data processing in portfolio optimization [9], as well as the shortcomings of traditional MAB models in coping with complex scenarios and proposed to optimize the decision-making efficiency by dynamically adjusting the number of selection arms [10]. There are also studies exploring the application of Bayesian methods for learning and decision making in uncertain environments and pointing out the main challenges of greedy algorithms in parameter setting [11], as well as an in-depth analysis of the sampling behavior of the UCB strategy in the worst case scenario [12]. In addition, the application of the Epsilon-Greedy algorithm to multimedia search is described and two algorithms for improving search efficiency in the presence of poor initial indexing are proposed [12].

The current research is insufficient in that, despite the abundance of studies on the application of multi-armed bandit algorithms, there is still a lack of empirical analysis regarding their performance in long-term investment strategies, particularly in exploring their effectiveness across different market environments. The significance of this study is to fill this research gap by deeply analyzing and comparing the performance of the three algorithms, UCB, Thompson Sampling, and Epsilon-Greedy, in the stock market with the aim of providing investors with a more accurate investment strategy, especially when the market is unstable. By combining theoretical research with actual market data, this study hopes to improve the adaptability and decision-making efficiency of these algorithms in practical applications.

The goal of the study is to validate the effectiveness of these algorithms in complex financial environments and to provide practical strategy recommendations for investors and decision makers to optimize short-term and long-term investment returns. In addition, this study aims to explore how these algorithms are implemented in real financial applications and how they perform in response to market volatility.

3. Methodology

For the UCB algorithm we have implemented it using a classical formula in which each arm is chosen based on the average return therein plus an upper bound on the confidence interval and based on this formula.$UCB = \overline{X_i} + \sqrt{\frac{2 \ln n}{n_i}}$. 'n' represents the total number of experiments and n_i is for the number of times a particular arm was selected. By being able to automatically adjust how often each arm is explored in this way, the experiment prioritizes those arms with high uncertainty or high return potential.

DOI: 10.54254/2755-2721/83/2024GLG0068

The Thompson sampling algorithm is implemented by random probabilistic sampling through a Beta distribution, where the parameters of the Beta distribution for each arm are determined by (α, β). The number of times the arm achieved success and failure were recorded, respectively. If success α plus one, if failure β plus one. In each round, we take a random sample from the Beta distribution for each arm and we choose the arm with the largest sample size to invest in based on the results. This random sampling method is more adaptable to real-time changes in the stock market and allows us to adjust the balance between exploration and utilization based on actual returns.

Epsilon-Greedy is a relatively simple and convenient algorithm, which usually sets a fixed probability of ε at the very beginning, and the algorithm randomly selects an arm in each selection with the fixed probability set at the very beginning and selects the arm with the highest probability of the current estimated return with a probability of $1 - \varepsilon$. So the key to this algorithm is the fixed probability set at the very beginning of the experiment, since This fixed probability will directly affect the explorability and utilization of the next algorithm.

3.1. Performance Evaluation

In this study, three algorithms are used, and the Sharpe ratio is included as the main performance metric in order to get a complete picture of the usefulness of these three algorithms. After these algorithms have calculated the average returns in the data, the Sharpe ratio is able to weigh the volatility of the average returns and is able to adjust the algorithm's risk efficiency based on the volatility of the returns. ANOVA and T-test are also used in this study; ANOVA identifies if there are differences in multiple sets of data, while T-test is used to further compare the differences between two algorithms, which further confirms that one algorithm is more appropriate for the market. Through these methods, the performance of different multi-armed slot machine algorithms in the real investment environment can be comprehensively assessed, thus providing a basis for the adjustment of stock investment strategies.

4. Experimental Design

4.1. Data description

The datasets used in this study are all publicly available from yahoo finance. The datasets include weekly stock price data for Tesla Inc (TSLA), General Motors (GM), and Ford (F) from 2019 through 2024, as well as semiannual daily data for 2024. These companies were chosen because they are all part of the automotive industry and their stock price volatility is more pronounced, making them suitable for testing investment strategies.

Data pre-processing was conducted to ensure consistency and reliability of the analysis. This included removing all entries from non-trading days to avoid data distortions, as financial markets are closed on these days and provide no price changes. Additionally, all date formats were standardized to YYYY-MM-DD across all datasets to eliminate any discrepancies in temporal data handling.

4.2. Rationale for the choice of experimental parameters

This study will use weekly closing prices for five years and daily closing prices for the last six months for Tesla (TSLA), General Motors (GM) and Ford (F) as data. The three multi-armed slot machine algorithms, UCB (Upper Confidence Bound Algorithm), Thompson Sampling, and ε-Greedy, were chosen because of their ability to optimize long-term and short-term returns on investment during the exploration and exploitation phases. All three algorithms are implemented using Python: the UCB algorithm is implemented by calculating the upper confidence bound level parameter, Thompson sampling samples and selects again by randomly drawing samples from a Beta distribution and updating the results after each analysis, and the ε-Greedy algorithm combines random and optimal choices to balance exploration and exploitation.

DOI: 10.54254/2755-2721/83/2024GLG0068

4.3. Key metrics for evaluating algorithm performance

In the model, buy, hold, and sell actions are defined for each "arm", which represent the core trading strategies in the stock market, allowing the algorithm to choose the most beneficial strategy at each decision point to optimize investment returns. The reward mechanism is based on stock price returns, especially for short-term datasets, calculated as the percentage difference between the daily opening and closing prices. This approach allows a direct assessment of the impact of each trade, facilitating the evaluation of the effectiveness of different trading strategies. During the experimental phase, each algorithm was executed 100 times on each dataset to ensure the reliability and fairness of the results. In addition, the starting point of each experiment is randomly selected to ensure the independence of the experiment and to minimize the effect of random variation.

5. Results

In this chapter, the performance of the three algorithms, UCB, Thompson Sampling, and ε-Greedy, on Tesla (TSLA) stock will be analyzed in two dimensions: short-term and long-term. Firstly, we will reveal the adaptability and effectiveness of the algorithms in the financial market by comparing the cumulative return performance of each algorithm over different time periods. Then, we will further analyze the differences among the algorithms through statistical methods to explore their relative advantages in optimizing investment returns. Ultimately, these results are combined to provide investors with investment strategy recommendations for different market conditions.

5.1. TSLA

Figure 1 shows the cumulative return performance of the three multi-armed slot machine algorithms on long-term and short-term data for Tesla stock, with the image on the left representing the short-term data and the image on the right representing the long-term data. The results of the short-term data analysis show that all three algorithms produced an increasing trend in cumulative returns, which indicates that all three algorithms produced positive returns over the observation period. The performance of UCB and Thompson sampling are very close to each other, and the cumulative return curves are almost overlapping, which indicates that these two algorithms have similar performance in the short term. And the cumulative return of Epsilon-Greedy is higher than the other two algorithms, which suggests that the Epsilon-Greedy algorithm does a better job in exploring and exploiting in the short term. In the long term data, again Epsilon-Greedy performs better than the other two algorithms and grows faster than the other two algorithms in the mid to late term. In this case, UCB has a smoother performance and slower growth rate, but Thompson sampling is performing better than UCB.

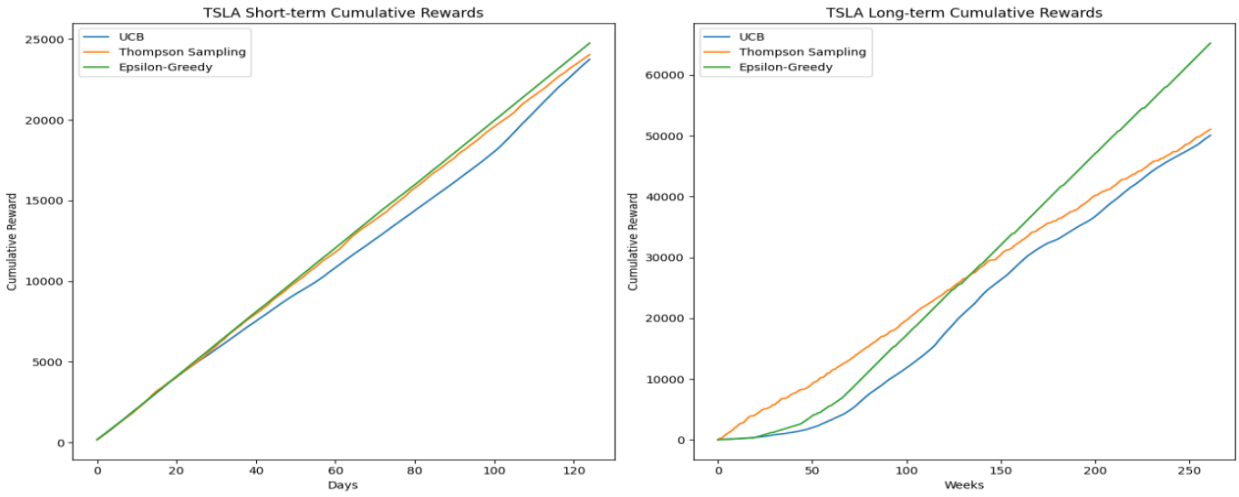

Figure 1. algorithms on Tesla stock in the long and short term

DOI: 10.54254/2755-2721/83/2024GLG0068

Table 1. Three algorithms in Tesla's long- and short-term stock market through Sharpe ratios, t-tests, and ANOVA.

Statistic	Short-Term	Long-Term
Sharpe Ratio (UCB)	6.9965	2.0764
Sharpe Ratio (Thompson)	7.3833	2.0869
Sharpe Ratio (Epsilon)	8.3934	4.7910
ANOVA F-value	35.6127(p=7.06e^{-15})	105.5008(p=2.69e^{-41})
t-test UCB vs Thompso	t=-0.8635, p=0.0638	t=-0.1386, p=0.8898
t-test UCB vs Epsilon	t=-6.0905, p=<1e^{-24}	t=-13.3414, p=<1e^{-35}
t-test Thompson vs Epsilon	t=-3.0026, p=7.63e^{-7}	t=-3.1315, p=7.50e^{-7}

This table 1 contains the Sharpe ratio as well as the results of the ANOVA and T-test. The results based on the Sharpe Ratio show that all three algorithms outperform in the short term. Epsilon-Greedy has the highest risk-adjusted returns in both the short and long term, at 8.3934 and 4.7910, respectively.

Based on the ANOVA results for both short and long term there is a very significant difference in the performance of all the three algorithms, especially in the long term as the F-value is very high and the P-value is very small. By comparing the algorithms with each other, from the results of the T-test, the T-value and P-value of UCB and Thompson sampling do not show any significant difference and their performance is basically the same. And Epsilon-Greedy has a very significant difference compared to both other algorithms and is superior to the other two algorithms both in the long and short term.

5.2. GM

Figure 2 illustrates the performance of these three algorithms on the short-term stock of General Motors. The Epsilon-Greedy algorithm shows a growing trend but with some volatility, which indicates that the algorithm is able to utilize known information to a certain extent to make investment decisions, which can be adapted to market changes based on changes in the investment strategy, which suggests that this algorithm can be used in market environments with a high degree of predictability.The UCB algorithm shows more pronounced fluctuations in its initial stages, and after exploring the complete Arm will gradually smooth out. While Thompson sampling has the highest volatility, which indicates that it is most active in the exploration and utilization phase.

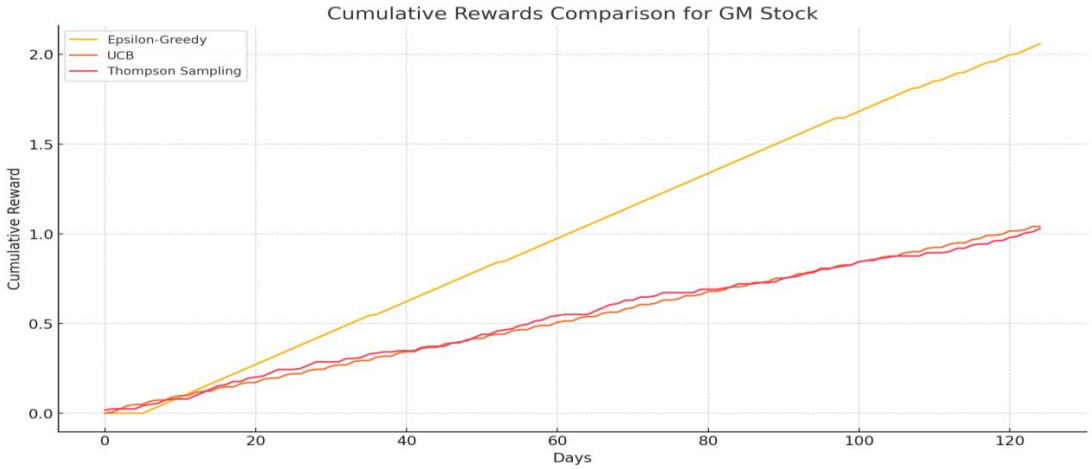

Figure 2. This chart shows how the three algorithms performed in the stock market over a six-month period at General Motors

DOI: 10.54254/2755-2721/83/2024GLG0068

This figure 3 shows the performance of the three algorithms over five years of weekly data from General Motors, the graph shows that the cumulative rewards have been declining this shows that the stock of General Motors as a whole has been trending downward over the five year period, for the investment strategies, Epsilon-Greedy performs relatively flat because you stay flat in the face of long term investing by defining an investment strategy first in the exploratory phase, while the remaining The rapid decline of the two algorithms suggests that there may be over-exploration in both calculations.

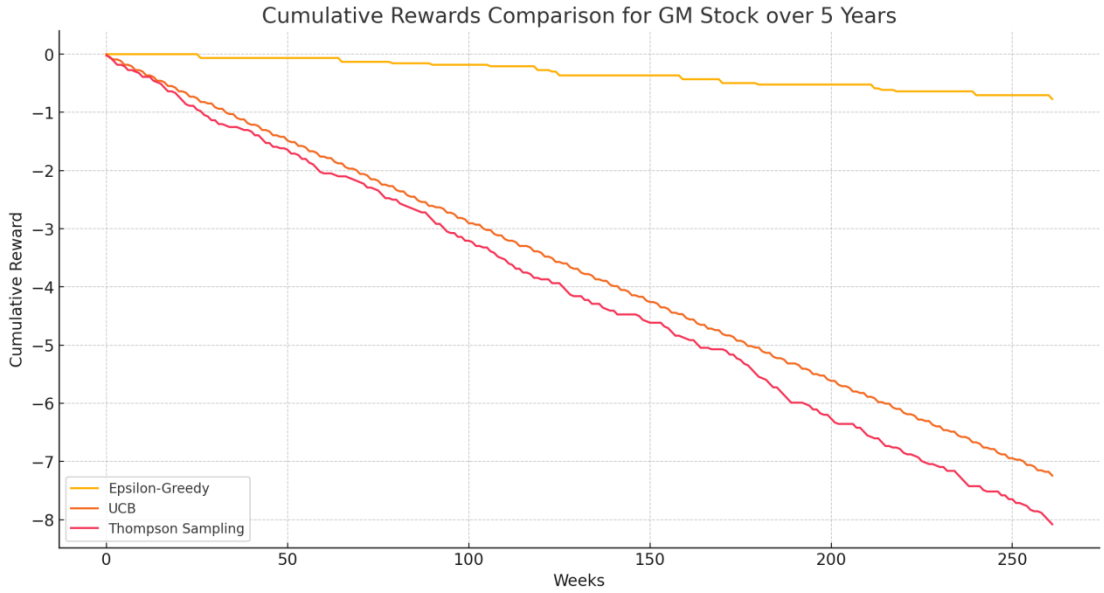

Figure 3. Three algorithms performed in the stock market over a five-year period at General Motors

Table 2. Three algorithms in General Motors long- and short-term stock market through Sharpe ratios, t-tests, and ANOVA.

Statistic	Short-Term	Long-Term
Sharpe Ratio (UCB)	1.1009	-1.0362
Sharpe Ratio (Thompson)	1.0956	-1.1188
Sharpe Ratio (Epsilon)	3.4033	-0.28244
ANOVA F-value	0.0	0.0
t-test UCB vs Thompson	t=0.0986, p=0.9216	t=1.0477, p=0.2953
t-test UCB vs Epsilon	t=-10.1006, p=8.0826e^{-20}	t=-12.5421, p=1.1013e^{-30}
t-test Thompson vs Epsilon	t=-10.2655, p=2.5186e^{-20}	t=-13.7597, p=1.5204e^{-35}

This table 2 were analyzed by subpopulation ratios as well as ANOVA and T-test tests to analyze the performance of the three algorithms. In the short term, the Epsilon-Greedy algorithm achieves a Sharpe ratio of 3.4033, which realizes a higher return compared to the other two algorithms. The Sharpe ratios for the UCB and Thompson sampling are lower due to the fact that the two algorithms have more explorations within the short term investments and this exploration leads to an increase in risk. In the long term, all the algorithms have negative Sharpe ratios, indicating that the stock has not outperformed the increase in risk within the long term investment strategy, this is due to the fact that this stock has not realized well over the past five years and is due to the market conditions. the results of the ANOVA are all close to 0, which shows that the differences between the three algorithms are very significant in both the long term and the short term data. the t-test reveals that difference between the two algorithms in market performance, within the short-term market, the p-value of UCB and Thompson sampling is higher, which indicates that there is no significant difference between the two algorithms. However,

DOI: 10.54254/2755-2721/83/2024GLG0068

when comparing both algorithms with Epsilon-Greedy at the same time, there are very low p-values, which indicates that the latter algorithm performs better. In the long run also the Epsilon-Greedy algorithm has a very low P-value, but since the Sharpe ratio is negative, it means that although there is a significant difference in the performance of the three algorithms, the overall performance is still unsatisfactory.

5.3. F

This figure 4 depicts the cumulative reward trend graphs of the three algorithms, from which it can be seen that the cumulative reward fluctuation of the Epsilon-Greedy algorithm is relatively large, which indicates that the optimal strategy was not chosen in the exploratory phase, which is also his flaw, and can not be updated according to the real-time changes in the market.The cumulative reward line of the UCB algorithm shows a gradual upward trend, and although there are some small fluctuations, but the overall realization is relatively robust. Thompson's sampling has more ups and downs which shows the randomness in the selection of the arms during the exploration phase.

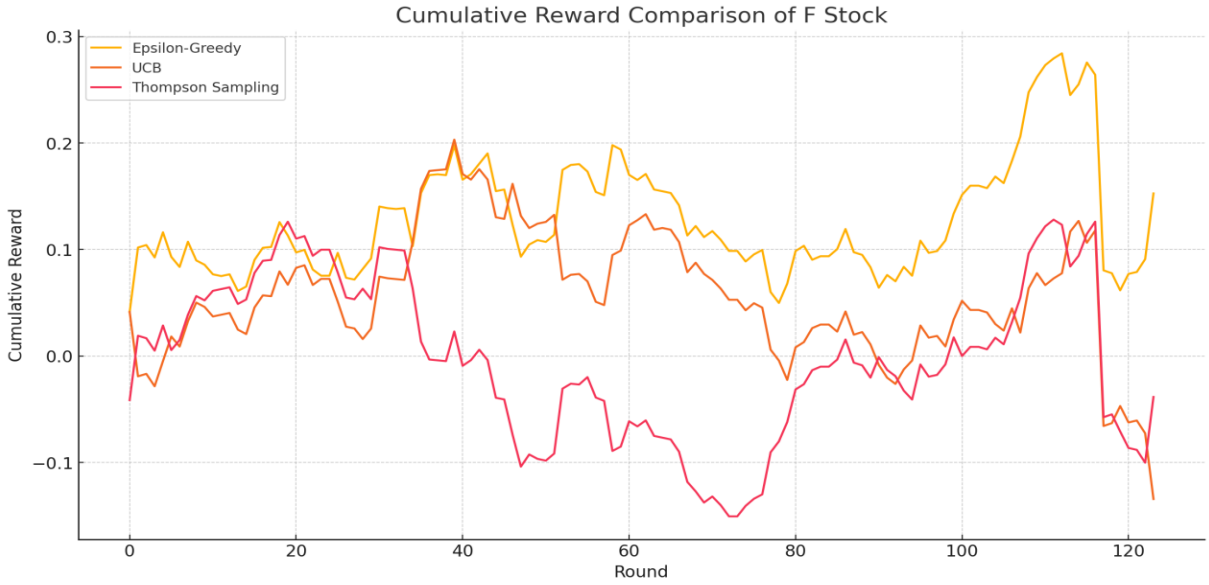

Figure 4. This graph shows the performance of the three algorithms in Ford's stock market over a six-month period

This figure 5 shows the trend of the three algorithms in terms of accumulated rewards. From the graph, we can see that Epsilon-Greedy's cumulative reward rises relatively quickly in the beginning phase, which demonstrates the ability of the Epsilon-Greedy algorithm to grow at a fluctuating rate at the very beginning through the ability to utilize known information or initial gains, followed by the subsequent growth of the cumulative reward. Whereas the UCB algorithm has very little overall movement and shows a steady upward trend, Thompson Sampling, because it selects strategies by random sampling, can be seen to have more pronounced ups and downs in the trend, which clearly reflects the uncertainty associated with frequent adjustments to the strategy when the market changes.

DOI: 10.54254/2755-2721/83/2024GLG0068

Figure 5. Three algorithms performed in Ford's stock market over a five-year period

6. Discussion

This part will delve into the performance of the three algorithms, UCB, Thompson Sampling and ε-Greedy, under different market conditions

6.1. Short-term investments

For short-term investing, the Epsilon-Greedy algorithm performs well across all datasets (Tesla, GM, and Ford). This is attributed to its aggressive exploration strategy, which allows it to quickly adapt to short-term market fluctuations and capture quick gains. However, this aggressive exploration also leads to high volatility in returns, as evidenced by fluctuations in cumulative returns. Nonetheless, the algorithm's ability to quickly adjust to market conditions has allowed it to perform best in short-term scenarios, earning the highest Sharpe ratio.

6.2. Long-term investing

In contrast, the performance of the algorithms varies significantly in long-term investing.Thompson sampling demonstrates greater stability and consistency in the long-term data for Tesla, with total returns outperforming those of UCB and Epsilon-Greedy.This is in line with the algorithm's design, which is able to balance exploration and utilization more efficiently over a longer period of time, allowing it to adapt to market changes without over-adjusting for short-term volatility. to market changes.The UCB algorithm, while also performing well, is growing at a slower rate, suggesting that it may be better suited to scenarios that prioritize stability over quick gains.

6.3. GM and Ford

The results for GM and Ford stocks are less favorable, especially in long-term investments. The negative Sharpe ratios for all algorithms suggest that these stocks underperformed over the observation period, leading to less than optimal investment results. This highlights the importance of considering underlying market conditions and stock-specific factors when applying these algorithms. In these cases, all algorithms struggled to achieve positive returns in the long term, although Epsilon-Greedy still showed the strongest performance in the short term.

6.4. Algorithm suitability

The significant differences in algorithm performance are further validated by the results of ANOVA and t-tests. The dominance of Epsilon-Greedy in short-term scenarios is statistically significant, whereas the differences between UCB and Thompson sampling are less pronounced, especially in the short-term.

DOI: 10.54254/2755-2721/83/2024GLG0068

This suggests that while Epsilon-Greedy may be the preferred choice for short-term returns, Thompson sampling may provide a more balanced approach for long-term investing, especially in markets where stability and consistency of returns are required.

6.5. Implications for investors

For investors, these findings provide important insights into how different algorithms can be utilized depending on the investment cycle and market conditions. Short-term investors may benefit from the aggressive nature of the Epsilon-Greedy algorithm, while investors with a long-term outlook may prefer the stability provided by Thompson sampling. However, performance differences between stocks (e.g., Tesla vs. General Motors) underscore the need to tailor algorithmic choices to stock-specific characteristics and the broader market environment.

7. Conclusion

This study provides insights into the effectiveness of three algorithms - UCB, Thompson Sampling and ε-Greedy - in practice in the financial markets, particularly their performance in short- and long-term investments. The results show that the ε-Greedy algorithm performs well in responding to rapid short-term changes in the market and is able to capture opportunities effectively, and that although the returns of this approach may be volatile, overall the risk and reward are relatively reasonable.

For long-term investing, the Thompson Sampling algorithm demonstrates more stable returns, especially when the market changes, and the algorithm is better able to maintain the balance between risk and return. In addition, the UCB algorithm demonstrates a unique advantage in balancing exploration and utilization, especially in the face of uncertain market environments, by calculating upper confidence bounds, which ensures the continued utilization of known optimal choices without ignoring potentially high-return options. This allows the UCB algorithm to provide more stable returns in long-term investments. This study focused on these three algorithms and did not cover more other types of algorithms, which may limit our ability to fully understand the applicability of these algorithms. In addition, while these algorithms are theoretically applicable to a wide range of market conditions, their actual performance is heavily influenced by the complexity of the market itself. Our empirical tests show that these algorithms perform inconsistently in different market environments, exposing the inadequacy of existing algorithms to adapt to dynamic markets. It is possible that the single data and the very variable market environment cause these three that algorithms do not perform very well in the market.

To address these issues, future research can be improved in several ways. First, more kinds of algorithms can be introduced, such as using advanced deep learning techniques, which may perform better when dealing with complex market data. Second, research should further test the effectiveness of these algorithms under different market conditions, such as in different market trends (e.g., bull or bear market), and in different types of financial products, in order to develop more flexible and stable investment strategies.

In addition, future research should consider how to combine algorithms with the individual needs of investors, such as risk preferences, investment cycles, and capital requirements. Research should develop algorithms that can adjust and optimize decisions in real time to help investors make better investment decisions as markets change. This will help us understand more fully how algorithms perform in practice and provide investors with more scientific and personalized advice.

References

[1] Chan H P 2020 THE MULTI-ARMED BANDIT PROBLEM. The Annals of Statistics. 48(1), 346-373.

[2] Slivkins A 2019 Introduction to multi-armed bandits. Foundations and Trends® in Machine Learning. 12(1-2), 1-286.

[3] Fouché E, Komiyama J, and Böhm K 2019 Scaling multi-armed bandit algorithms. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. p1449-1459

DOI: 10.54254/2755-2721/83/2024GLG0068

[4] Marti G, Nielsen F, Bińkowski M, and Donnat P 2021 A review of two decades of correlations, hierarchies, networks and clustering in financial markets. Progress in information geometry: Theory and applications. p245-274.
[5] Rebentrost P, and Lloyd S 2024 Quantum computational finance: quantum algorithm for portfolio optimization. KI-Künstliche Intelligenz. p1-12.
[6] Bouneffouf D, and Rish I 2019 A survey on practical applications of multi-armed and contextual bandits. arXiv - CS - Machine Learning.1904.10040
[7] Kar D, Lyu Z, Ororbia A G, Desell T, and Krutz D 2024 Enabling An Informed Contextual Multi-Armed Bandit Framework For Stock Trading With Neuroevolution. In Proceedings of the Genetic and Evolutionary Computation Conference Companion. p1924-1933
[8] Kuang N L, and Leung C H 2019. Performance effectiveness of multimedia Information search using the epsilon-greedy algorithm. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). p929-936
[9] Gimelfarb M, Sanner S, and Lee C G 2020 {\epsilon}-bmc: A bayesian ensembleapproach to epsilon-greedy exploration in model-free reinforcement learning. arXiv preprint arXiv. 2007.00869.
[10] Khurshid S, Abdulla M S, and Ghatak G. 2024 Optimizing Sharpe Ratio: RiskAdjusted Decision-Making in Multi-Armed Bandits. arXiv preprint arXiv.2406.06552.
[11] Kalvit A, and Zeevi A 2021 A closer look at the worst-case behavior of multi-armed bandit algorithms. Advances in Neural Information Processing Systems. 34, 8807-8
[12] Zhang Z, Zohren S, and Roberts S 2020 Deep learning for portfolio optimization. arXiv preprint arXiv. 2005.13665.