CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Significance of Data Decimation in Stock Market Forecasting and Stock Selection

by

Nosherwan Khan

A dissertation submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Management & Social Sciences
Department of Management Sciences

2024

Significance of Data Decimation in Stock Market Forecasting and Stock Selection

By

Nosherwan Khan

(PM113011)

Dr. Abdoul Aziz Ndoye, Associate Professor
Université d' Orléans, Rue De Blois, Orléans, France
(Foreign Evaluator 1)

Dr. Tahir Suleman, Senior Lecturer
University of Otago, New Zealand
(Foreign Evaluator 2)

Dr. Corrado Crocetta, Professor
University of Bari Aldo Moro, Italy
(Foreign Evaluator 3)

Dr. Arshad Hassan
(Research Supervisor)

Dr. Lakhi Muhammad
(Head, Department of Management Sciences)

Dr. Arshad Hassan
(Dean, Faculty of Management & Social Sciences)

DEPARTMENT OF MANAGEMENT SCIENCES
CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY
ISLAMABAD
2024

My gratitude goes to my parents and teachers.
Thank you for keeping the interest rates on
everything I owe you so low.

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD
Expressway, Kahuta Road, Zone-V, Islamabad
Phone:+92-51-111-555-666 Fax: +92-51-4486705
Email: info@cust.edu.pk Website: https://www.cust.edu.pk

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in the dissertation, entitled "**Significance of Data Decimation in Stock Market Forecasting and Stock Selection**" was conducted under the supervision of **Dr. Arshad Hassan**. No part of this dissertation has been submitted anywhere else for any other degree. This dissertation is submitted to the **Department of Management Sciences, Capital University of Science and Technology** in partial fulfillment of the requirements for the degree of Doctor in Philosophy in the field of **Management Sciences.** The open defence of the dissertation was conducted on **January 09, 2024**.

Student Name : Nosherwan Khan (PM113011)

The Examination Committee unanimously agrees to award PhD degree in the mentioned field.

Examination Committee :

(a)	External Examiner 1:	Dr. Qaisar Ali Malik Professor Foundation University, Islamabad	
(b)	External Examiner 2:	Dr. Nadeem Iqbal Professor IIU, Islamabad	
(c)	Internal Examiner :	Dr. Imran Riaz Malik Associate Professor CUST, Islamabad	
Supervisor Name :		Dr. Arshad Hassan Professor CUST, Islamabad	
Name of HoD :		Dr. Lakhi Muhammad Associate Professor CUST, Islamabad	
Name of Dean :		Dr. Arshad Hassan Professor CUST, Islamabad	

AUTHOR'S DECLARATION

I, **Nosherwan Khan** (Registration No. **PM113011)**, hereby state that my dissertation titled, "**Significance of Data Decimation in Stock Market Forecasting and Stock Selection**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/ world.

At any time, if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my PhD Degree.

(**Nosherwan Khan**)

Dated: 09 January, 2024

Registration No : PM113011

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the dissertation titled "**Significance of Data Decimation in Stock Market Forecasting and Stock Selection**" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete dissertation has been written by me.

I understand the zero-tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled dissertation declare that no portion of my dissertation has been plagiarized and any material used as reference is properly referred/ cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled dissertation even after award of PhD Degree, the University reserves the right to withdraw/ revoke my PhD degree and that HEC and the University have the right to publish my name on the HEC/ University Website on which names of students are placed who submitted plagiarized dissertation.

(Nosherwan Khan)

Registration No : PM113011

Dated: 09 January, 2024

List of Publications

It is certified that following publication(s) have been made out of the research work that has been carried out for this dissertation:-

1. Nosherwan Khan, Arshad Hassan and Amir Iqbal Bhatti (2023). "Quick Decision Making To Achieve The Competitive Edge: Significance Of Decimated Data For Forecasting The Stock Market". Journal of Positive School Psychology, Vol. 7, No. 4, 1341-1353

Nosherwan Khan
Registration No: PM113011

Acknowledgement

First and foremost, I thank Allah for providing me with the power, expertise, and knowledge to conduct this research. I am quite grateful to the institution for assigning him the responsibility of creating the Research Dissertation in the level of doctoral degree program with a specialization in Management Sciences (major in Finance). This study met the university's requirements, and I am grateful to my friends and colleagues for their assistance and time in completing this research. I am grateful to those individuals for their efforts, time, knowledge, and resources in assisting me to finish my proposal. I am very fortunate to have his heart blessed, and I am grateful to our adviser (Dr. Fazal ur Rehman) and supervisors (Dr. Arshad Hassan and Dr. Amir Iqbal Bhatti) for providing me with suggestions, directions, and knowledge sharing to create this research. Furthermore, I am grateful to all of my friends, class fellows, and seniors for their time and recommendations in helping me write the research.

Nosherwan Khan

Abstract

Online portfolio management algorithms have largely replaced human bookmakers with the advent of computerized trading, which has transformed today's stock markets. These algorithms expedite the process of buying stocks. The computing requirements of these algorithms, however, present difficulties in terms of both speed and profitability. Slowing the rate at which stock data is supplied to the decision-making computers is one way to address this issue by reducing the computational load.

In order to overcome the computing difficulties inherent in performing online portfolio optimization, this study presents the Decimated method. The approach uses decimated data in an effort to keep performance steady while decreasing computational demands. Decimation requires the application of strict theoretical requirements generated from data analysis, including the selection of appropriate statistical criteria and decimation rates.

This study uses the Decimated algorithm to the numbers from five well-known businesses: Apple, Amazon.com, Alphabet, Meta Platforms, and Netflix. The efficiency of the algorithm is measured by how well it handles investment decisions in the allotted time frame. Results are shown using real-world stock market data to verify the effectiveness of the suggested method.

The results show that the Decimated algorithm has promise as a method for optimizing the stock return in real time. The method exhibits stable performance because it uses well-established ideas from information theory and makes calculated decisions about decimation rates. This study adds to our understanding of the role of portfolio management methods in the context of automated trading and sheds light on how to achieve optimal performance in this area while minimizing computational costs.

Key words: Portfolio, OLMAR, SMAR, Stock, Online Trading, D-OLMAR

Contents

List of Figures

List of Tables

Abbreviations

AA	Aggregating Algorithm
AI	Artificial Intelligence
Anticor	Anti Correlation
BAH	Buy and Hold
CRP	Constant Re-Balanced Portfolio
CW	Confidence Weighted
CWMR	Confidence-Weighted Mean Reversion
DR	Direct Reinforcement
DWMR	Dense Weighted Mean Reversion
D-OLMAR	Decimated Online Moving Average Reversion
EP	Exponential Gradient
EEMD-FFH	Empirical Mode Decomposition−Fractional frequency hybrid
FTL	Follow the Leader
MV	Mean Variance
NLP	Natural Language Processing
OIM	Optimal Investment Management
OLPs	Online Portfolio Selection
OLMAR	Online Moving Average Reversion
ONS	Online Newton Step
OLC	Online Complex Optimization
PA	Passive Aggressive
PMAR	Passive–Aggressive Mean Reversion
RMR	Robust Median Reversion
RS	Risk Seeker

SF	Safety First
SCRP	Semi Constant Rebalanced Portfolios
UP	Universal Portfolio

Symbols

$x_{t,i}$	The ratio of the closing price
$\mathbf{b}_t$	Combination of price vector
S_0	Sum of wealth
$\bigodot$	Element wise Product
$\prod$	Factor
ω	Window size
ρ	Price vector
C	Index set of similar price relatives
S_n	Cumulative Wealth
x_1^n	Historical Price Movement
Z	Normalization Term
ω	Window / Weight
η	Learning Rate
β	Trade off Parameter
x	Price vector
$\tilde{x}_{t+1}$	Next price movement
t	th period
L	Lagrangian function

Chapter 1

Introduction

An interesting story of technology progress and financial democratization is the historical path of online trading. The ultimate peak occurred in the 1980s with the rise of personal computers and financial networks, even if the foundations started growing in the 1960s with the introduction of Electronic Data Interchange (EDI) for data transfer. The internet grew rapidly in the 1990s, which led to the creation of online brokerage services like E*Trade and Charles Schwab. By enabling everyone with an internet connection to easily trade stocks, bonds, and other assets from the comfort of their homes, these platforms transformed financial access. Online trading is significant in many ways, since it democratizes access to financial markets and makes investment possible for those who would not otherwise be able to use traditional brokerage services. Price transparency, efficiency, and market liquidity have all grown as a result of this democratization. Furthermore, new financial services and products including fractional shares, trading algorithms, and mobile apps have all emerged as a result of internet trading. Due to these advancements, investing has become more accessible and inexpensive for a wider range of people, democratizing the process.

However, it is crucial to understand that there are risks associated with internet trading. Online transactions' convenience and quickness might lead to quick decisions and excessive trade. Vigilance and precaution are also necessary due to the increasing number of online scams and fraudulent activities. The following are a few of the major changes:

Democratization of Finance Online trading has shattered barriers related to wealth and knowledge, broadening access to investing Han (2019).

Increased Market Liquidity The surge in participants has significantly enhanced market liquidity, streamlining trade execution Malceniece et al. (2019).

Enhanced Transparency Real-time quotes and accessible market data empower investors, reducing the information gap favoring professionals Urman and Makhortykh (2023).

Innovation and Disruption Online platforms have been pivotal in introducing new financial products and services, from fractional shares to automated investment tools Li et al. (2020).

Challenges and Risks The ease of online trading can lead to impulsive decisions and overtrading, especially for inexperienced investors. Risks also include online scams and cybersecurity threats El Hajj and Hammoud (2023).

A recent study by El Hajj and Hammoud (2023) highlights the ongoing relevance and influence of online trading on the financial landscape by demonstrating how the democratization of finance through online trading has enhanced market participation and been linked to favorable economic outcomes.

Online trading has significantly improved the financial landscape despite these obstacles, empowering people, increasing market efficiency, and encouraging innovation. Investors expect more revolutionary developments as technology develops, such as the introduction of blockchain-based trading and the incorporation of AI-powered investing algorithms.

1.1 The Evolution of Online Trading and Learning in Financial Markets

The Internet's widespread reach and the revolutionary potential of digital technology have significantly changed the financial environment in recent decades. The dynamic interaction between online learning and online commerce, which is changing how people access, comprehend, and engage with the complex environment of markets, is at the center of this revolution.

1.1.1 Increasing Access and Choice

In the past, financial markets were dominated by experienced experts and had significant entry barriers. However, the introduction of online trading platforms has broken down these boundaries, ushering in a period of democratization and inclusivity. Platforms such as Interactive Brokers, E*Trade, and mobile apps like Robinhood have played critical roles in democratizing financial participation by making it available to a wide range of users, from seasoned experts to total beginners Sathye (2004). This increased accessibility, aided by user-friendly interfaces and lower middleman fees, has allowed people from all walks of life to traverse markets and potentially accumulate wealth Stephens (2013).

1.1.2 Greater Access, Greater Accountability

Having more access to the financial markets is beneficial, but it also entails greater accountability. The complex world of investing requires ongoing education and skill development. This educational component is crucial in providing people with the information and resources needed for online learning and in assisting them in navigating the complex world of financial strategies and goods. Such information makes it possible for people to make wise decisions, lower the risks involved in online transactions, and foster an environment where participation is done responsibly Christensen et al. (2016).

1.1.3 Online Transactions and Learning Convergence

The integration of online learning and transactions is expected to support the ongoing growth of the financial ecosystem, going beyond democratization and empowerment. The financial market will probably see the development of new trading methods and the exploration of unexplored territory due to the arrive in of new participants who are equipped with a variety of knowledge and analytical tools Michael Spector (2017).

1.2 The Effect of Algorithmic Trading on the Financial Environment

The financial environment has undergone a fundamental transformation due to the disruptive influence of computerized trading, commonly referred to as algorithmic trading. Algorithmic trading, which uses pre-programmed algorithms to execute transactions based on predetermined criteria, analyzes enormous volumes of data in milliseconds and takes advantage of short-lived market opportunities Fan et al. (2012). Both praise and criticism have been directed towards this technological advancement, which has brought up important issues of transparency, equity, and possible risk amplification.

1.2.1 Algorithmic Trading's Supporters and Detractors

Proponents of algorithmic trading emphasize the effectiveness and efficiency of the practice, praising its quickness (which outpaces human comprehension) and capacity to seize momentary arbitrage possibilities Brogaard et al. (2010). Moreover, its ability to analyze large amounts of data and spot intricate patterns is thought to help traders make well-informed decisions and maybe lessen market volatility Lattemann et al. (2012). It is hypothesized that removing the emotional biases and impulsive choices that human traders frequently make will increase market stability and create a more logical trading environment Sonkin and Johnson (2017).

1.2.2 Issues & Difficulties

Nonetheless, there are drawbacks to algorithmic trading's quick rise. Since the intricate algorithms function as "black boxes" shrouded in secret, critics express concerns about their opacity Hansen (2020). Due to linked algorithms that magnify market collapses through cascading feedback loops, this lack of transparency might increase suspicion and systemic risk Bardoscia et al. (2021). Furthermore, market fragmentation and liquidity problems may be made worse by the high-frequency nature of automated trading, making it more difficult for traditional investors to engage in the market successfully Siering et al. (2017).

Demand for a Comprehensive Review of the Financial Ecosystem The financial ecosystem needs to be thoroughly reevaluated in light of the rise of algorithmic trading. In order to maintain openness and reduce potential dangers, regulatory frameworks need to be modified

The findings of this study were intended to help financial professionals better manage the complexities that arise when dealing with a large variety of products. Our goal was to create an algorithm that could efficiently process and assess large data sets. We can improve computer performance and speed up decision making by increasing the frequency with which stock data is sent to decision-making systems. So, picking the finest stocks and precisely forecasting their performance over several time periods was the primary emphasis of this study. Our primary goals were to identify the most effective window size and frequency, and to draw attention to the unique contributions this research made to the study of online learning.

1.3 Theoretical Framework

The optimal allocation of wealth between different best stock assets is a key research subject in the fields of portfolio selection and applied financial engineering. Both the Mean Variance Theory, made famous by Markowitz, and the Capital Growth Theory, developed by Kelly, have been investigated as potential solutions to this issue. When trying to optimize the portfolio's predicted growth rate, the

Mean Variance Theory looks just at a single time period, while the Capital Growth Theory considers a range of time periods.

Optimal portfolio selection decisions can be gleaned from both theories, but the "Online" approach that considers a number of time periods is the focus of this study. The advent of high-frequency trading is just one example of how the trading sector has been impacted by the development of information technology. Trading within short time frames (seconds to 24 hours) requires buying and selling often. Massive volumes of intraday data are produced by high-frequency trading, necessitating lightning-fast processing power. Timely reactions are vital to grasp chances in today's market, but human stockholders often have trouble keeping up with the market's rapid pace.

Complex difficulties and greater hazards are brought forth by high-frequency trading and the prediction of future price unpredictability. Although price patterns are often interpreted subjectively, indicators can be used to make more objective market predictions. Even if indications from indicators are there, these methods may not provide explanations for some market swings. In order to deal with the complexities and uncertainties of high-frequency trading, this study suggests employing machine learning approaches.

This study incorporates the Capital Growth Theory into a discussion of how to maximize the equitable distribution of wealth for optimal investment decision making. The "Online" method, which takes into consideration numerous time periods in investment decision-making, is highlighted. The study recognizes the value of IT in high-frequency trading and emphasizes the importance of quick and efficient tools and procedures. The study also suggests incorporating machine learning strategies to deal with the difficulties of high-frequency trading and future price forecasting uncertainty using the decimated data with varying sample size and frequency.

1.4 Research Gap

A significant challenge in the financial market pertains to the handling of big and volatile data for stock market prediction and stock selection. With the advancement of technology, there has been an exponential increase in the availability and accessibility of financial data. This includes not only historical price and volume data but also a wide range of alternative data sources such as news sentiment, social media trends, and macroeconomic indicators.

However, the sheer volume and volatility of this data pose considerable challenges. Traditional statistical models and analysis techniques often struggle to effectively process and extract meaningful insights from such vast amounts of data. Moreover, the stock market is known for its inherent volatility, making it even more challenging to accurately predict future price movements and optimize investment portfolios.

Existing literature on stock market forecasting and stock selection frequently emphasizes the use of conventional data sets and techniques, such as daily or high-frequency stock price data. However, there is a dearth of research examining the efficacy and prospective benefits of using decimated data in these prediction models.

A subset of the original data that has been systematically reduced or sampled at specific intervals is referred to as decimated data. This reduction in data granularity may have potential benefits, such as reducing noise, enhancing computational efficiency, and capturing long-term trends as opposed to short-term fluctuations.

Consequently, decimated data can provide trustworthy insights and competitive advantages in terms of precision, speed, and profitability in comparison to traditional data sets.

In this study, a custom-tailored algorithm was developed to utilize decimated data for stock market forecasting and stock selection. This research investigates the efficacy and performance of this approach through empirical analysis and evaluation, thereby contributing significantly to the fields of financial forecasting and investment strategies.

1.5 Problem Statement

Dealing with an enormous quantity of financial data can be difficult in the area of stock prediction and selection. The challenge is that because markets generate massive amounts of data in real time, it is critical to filter through the vital information in order to make appropriate decisions. This is where "data decimation" comes in - it's a method of reducing down the data using various ways, and it could be a solution to deal with the complexities of dealing with large amounts of financial data.

Despite the increased interest in data decimation, we don't fully grasp how significant it is and how it affects the accuracy of projecting stocks and making financial decisions. This dissertation seeks to fill this knowledge gap by investigating how data reduction approaches such as data decimation affect the efficacy of prediction models in financial markets. We'll look at the pros and downsides, such as the danger of losing some information vs the benefits of faster calculations. In addition, this study looks at how data decimation works in real-world stock market scenarios. This research determines if it helps prevent overly specific predictions (overfitting), if it makes models easier to grasp, and if it improves the dependability of forecasting algorithms. By delving into these crucial components, the study seeks to provide a clear grasp of how critical data decimation is when it comes to stock analyses.

1.6 Research Questions

This dissertation aims to investigate the following core questions:

1. Can the decimated data be used to trade on the basis of algorithm?

2. Does algorithm-based trading outperform market-based performance?

3. Is the long time frame better than the short time frame for forecasting and devising?

4. Does the algorithm-based trading strategy face more risk in comparison to the passive investment strategy?

5. Does algorithm based strategy offer batter risk adjusted return against downside risk?

1.7 Research Objectives

The proposed method, as well as the incorporation of the decimated aspects, will be thoroughly examined using the most recent empirical data to evaluate its effectiveness across a variety of statistical circumstances. The study will dive into critical aspects such as sample size and data collecting frequency, with a focus on the "Online" approach. This technique seeks to tackle the intricacies of stock selection and prediction efficiently and quickly, while eliminating additional responsibilities.

This research aims to improve traders' decision-making capabilities in today's fast-paced financial markets by addressing these difficulties head-on and employing cutting-edge algorithms and approaches. The incorporation of the decimated algorithm increases the potential for making informed and sensible decisions, allowing traders to negotiate the high-frequency terrain with enhanced precision and agility while reducing operational expenses. The following is a list of core objectives of this study:

1. To propose an algorithm for investment decision.

2. To compare the the performance of proposed algorithm with market based performance of stock.

3. To compare the risk of algorithm based trading with risk of the stock.

4. To examine the long horizon forecasting outperform the short horizon prediction.

5. Is the algorithm-based trading strategy less risky than the passive investment strategy?

1.8 Significance of the Study

The suggested decimated algorithm for Online Stock Selection is developed with the intention of improving upon existing methods. The algorithm's goal is to maximize the ability of stock prediction and investment decision making by making use of optimally decimated data and frequency. This algorithm places more emphasis on the trade mechanism of "Online" algorithms for resolving the stock forecasting and selection concerns than do other existing techniques, which tend to rely on theory or single stock trading.

In order to deal with the difficulties presented by large and unpredictable data sets, the method uses decimated data. The method makes data processing quicker and more efficient by lowering the amount of data without losing any relevant information. As a result, businesses are better equipped to react quickly to shifts in the market and seize opportunities that may otherwise go unnoticed.

The decimated algorithm has many potential applications, including quantitative trading, automated wealth management, and the management of hedge funds. By offering more precise stock market projections, it can aid in better decision making for investment.

In conclusion, the suggested decimated algorithm is a feasible approach to addressing the challenges of enormous and changeable data in financial investment management, which present obstacles to the complexities of stock selection and prediction.

The proposed decimated algorithm for online stock selection and prediction is useful for both academia and industry. The effects can be seen in a variety of ways:

1. An important step forward, the decimated algorithm for online stock selection allows researchers to focus on the technical components of the problem. It offers researchers a fresh viewpoint on the problems of stock selection & prediction and a novel approach to resolving them.

2. This research stands out because it tackles the challenging issue of stock selection across many asset types, including equities, fixed income instruments, and derivatives. Beyond the more limited fields of single-stock trading and the theory of portfolio management, this research broadens the scope of previous work.

3. The proposed algorithm uses computational intelligence techniques to forecast financial time series. This integration improves the accuracy and reliability of Online Stock Selection and Predictions.

4. This investigation into the "Online" algorithmic trading mechanism provides insight into the development of efficient and successful portfolio management strategies for use in live market conditions.

5. The proposed algorithm improves upon existing techniques for optimizing investment decision. It enlarges the possible avenues of inquiry, examination, and refinement for scholars and students.

6. This research makes a significant impact by introducing AI-based techniques to the field of computational finance. It allows academics to create innovative computational models and conduct ground-breaking theoretical research.

7. This study helps to close the gap between the theoretical underpinnings of portfolio management and its actual application in the real world. This opens the door for academics to examine the practical implications of algorithmic portfolio selection methods.

8. The suggested algorithm has applications in quantitative trading firms, hedge funds, and automated wealth management services. It aids these businesses in better managing their investment portfolios, resulting in higher returns and lower losses.

9. Technological Advances This research contributes to the improvement of cutting-edge high-frequency trading and data-processing methods. This

technological development allows the company to take full use of the benefits of online trading procedures and stay competitive in the face of the ever-changing nature of financial markets.

Overall, the proposed decimation algorithm helps researchers, academics, and businesses make greater strides in their application of Online Stock Selection for portfolio optimization. It has far-reaching consequences, from helping to close knowledge gaps and advancing academic progress to enhancing the quality of investment management decisions and increasing output.

1.9 Structure of Dissertation

The research study is thoroughly explored throughout this dissertation's multiple chapters.

The first chapter introduces the subject and then analyzes it in depth. It provides a synopsis of the field of study, stressing its relevance and justification. This section, describe the theoretical and conceptual foundations of the study and highlight the contributions we want to make with this research.

The second chapter is a literature review that examines the history of online portfolio selection strategies and provides a summary of the relevant studies. By examining the relevant research and outlining the gaps and limits of current techniques, this chapter lays the groundwork for the succeeding chapters.

The theoretical foundations of the algorithms employed in online portfolio selection are discussed in Chapter 3. Famous methods like online linear regression and online passive regression are discussed. Following an overview of the methods currently in use for online portfolio selection, the dissertation narrows in on the topic at hand: online moving average methods and its modeling.

The findings of the research are summarized in Chapter 4. The results for each method are presented separately, with the most important analyses highlighted. This section's goal is to give readers a thorough grounding in the outcomes of using online moving average algorithms and its modeling.

In the last chapter, reviewed the results and analyze their significance. It highlighted the importance of the findings and addresses their caveats. Suggestions for potential future study topics and directions are provided in this section as well.

This dissertation is organized as follows: introduction; topic analysis; literature review; theoretical framework; findings analysis; discussion and conclusion. This structure helps researchers conduct in-depth studies and adds to the body of knowledge in the domain of online portfolio management.

Chapter 5 serves as a concluding chapter, summarizing the discussion and implications of the results. It highlights the significance of the findings and discusses their limitations. This chapter also offers suggestions for future research directions and areas that could benefit from further investigation.

Overall, this dissertation follows a structured approach, beginning with an in-depth analysis of the topic, followed by a literature survey, theoretical background, analysis of results, and finally, a discussion and conclusion. This organization ensures a comprehensive examination of the research area and contributes to the existing knowledge in the field of online portfolio selection.

Chapter 2

Literature Review

The emergence of algorithmic trading, often known as "Algo-Trading," has had a profound impact on regulatory frameworks and investment strategies in the financial markets. The objective of this dissertation research is to investigate this complex field in further detail, looking at its various techniques, efficacy, drawbacks, and possible directions going forward.

2.1 Algorithmic Approaches

2.1.1 Algorithms Based on Fundamental Analysis

These algorithms evaluate the inherent value of a firm based on financial documents, economic statistics, and news sentiment. Although their long-term efficacy is still up for question, studies by Battiston et al. (2012) and Fama and French (2015) demonstrate their potential for alpha production.

2.1.2 Algorithms Based on Technical Analysis

These algorithms use past volume and price data to spot trends and patterns that can be used to forecast future movements in the market. Relative Strength Index (RSI) and Moving Averages are two well-known examples. Although beneficial for short-term trading, market efficiency has limitations, as noted by Bessembinder and Chan (1998) and Lo et al. (2000).

2.1.3 Algorithms Based on Statistical Arbitrage

These algorithms take advantage of mean reversion occurrences by utilizing price differences between comparable assets across markets. Russakovsky et al. (2015) and Reddaf et al. (2020) both show that they are successful in identifying market inefficiencies, mostly in short-term settings.

2.1.4 Algorithms Based on Machine Learning and Deep Learning Concepts

By utilizing AI's enormous potential, these algorithms examine a wide range of data sets, such as news stories, social media posts, and even satellite images, in order to spot possible trading opportunities. While highlighting their potential, Guidotti et al. (2018) and Dong et al. (2023) warn about their "black-box" character and overfitting vulnerability.

Existing methods for online portfolio selection (OLPS) generate explicit portfolio updating strategies by describing previous portfolio optimization techniques. Table: 3.1 identifies the major ideas and a few exemplary algorithms (Li and Hoi, 2014). Literature introduced a number of benchmark algorithms. Then, based on the directions of weight transfer, offered three categories of principles or algorithms with explicit portfolio updating strategies. The first strategy, "follow the winner," gives more weight to experts or stocks that have performed better in the past. The second strategy, called "follow the loser," on the other hand, gives less successful experts or stocks more weight, or moves weight from winners to losers. The pattern matching-based technique, the third category, builds portfolios based on comparable historical patterns and lacks clear instructions. Finally, this review described several similar meta-algorithms that are applied to a group of experts, each of whom has any algorithm from the first three categories.

2.2 Historical Development of Online Portfolio Strategies

One important question is whether financial data from Mathematical point of view is random or not. Like any other signal or data that may be acoustical, electrical or financial data series, they should be which can be electrical, acoustical, or otherwise, financial data series should be exposed to the same signal analysis procedures. A lot of literature claims data not to be random. In trading there would have been no point it had been so. In financial market analysis, on chart market price data versus time is often noted by traders (Neftci, 1991). Due to bulls and bears market, each pattern is understood in a different way. These patterns lead the market head. Peterson (1997a) suggested that through complete randomness such patterns can take birth. Is there any reason for variation in price of certain market if the market is a random process? The core reason is that more traders use the technical analysis tool, i.e., prediction based on the past data. If we can accept this so each trader has its own rule and past would give the same future especially in Futures and Bond market. But individual markets may not be assessed with these phenomena where investor and fund managers use fundamental analysis tools, i.e. decision making also include the P/E ratio, MTB ratio, etc. In these markets, the past data are not playing a role of the future price function. So, in these markets, the non-randomness may cause other reason. If it is correct that market is not random would we be able to predict the market based on the mathematical model which based on the probabilistic rather than deterministic.

The "online" option emphasizes the significance of prompt and well-informed decision making in portfolio selection. Using this method, investors would get data on the market in order and allocate their funds right away. Multi-period portfolio selection research is essential because of the sequential structure of the process.

According to Cover (1991), in such a situation, the portfolio is re-balanced to an allocation determined by the next available trading period. The goal is to make the most of the consistent logarithmic yield throughout a variety of time frames for making trades. Kelly (1956b) discusses the Capital Growth Theory, which emphasizes the need of developing a sound betting strategy. It stresses the futility of constantly risking one's entire fortune on an advantageous but uncertain bet. Kelly proposes a new method of betting that accelerates the rate at which one's wealth grows.

Researchers like Breiman (1960); Hakansson (1975); Bell and Cover (1980) have expanded on these ideas. While Markowits (1952) work is a good example of a single-period study, this investigation goes beyond that. Instead, it emphasizes portfolio choices that maximize the present value of expected future returns, since it is impossible to know for sure what will happen in the future and must instead be approximated. There are many possible refinements to this rule.

The evaluation of the incorporation of "anticipated" returns in investment decisions relies heavily on the analysis of risk and reward tradeoffs. If you have any reservations about the market's efficiency, you shouldn't base your investment decisions exclusively on the discounted return you can achieve. Existing recommendations do not universally favor an undiversified portfolio over a diversified one, but there are exceptions.

Li and Ng (2000) broke with the common practice of creating short-term investment portfolios. To solve the portfolio selection issue over multiple time periods, they used Markowitz's mean-variance approach. The multi-period variance formulation was kept tractable in its original settings, and they gave an analytical solution for it. Investors were given newfound flexibility in their decision-making thanks to this solution's illumination of the dynamic tension between risk and expected return.

Using stochastic control theory, Dai et al. (2010) presented a continuous-time portfolio selection model that incorporated transaction costs. They used tools like Lagrangian multipliers and partial differential equations to figure it out. We classified the practical constraints, obtained the function value from the PDE, and applied the Lagrange multiplier to solve the resulting algebraic problem. This method provided a thorough structure for evaluating portfolio asset replacement and selection.

In a study of the market, El-Yaniv (1998) zeroed in on the selection and replacement of portfolio assets as points of differentiation. Their study suggested exciting new algorithms that improved upon prior approaches. Uncertain demand projections associated with stock buy and sell decisions are the primary focus of the Optimal Investment Management (OIM) theory.

In addition, the study covers a wide range of interconnected issues that can be broken down into their own subfields. There's a class of investing strategies known as "Follow the Winner," which includes putting greater weight on profitable stocks than the market as a whole or the top-performing stock. Cover (1991) Universal Portfolio Strategy is an example of such an approach. They developed algorithms to boost the efficiency of investment portfolios. Cover and Ordentlich (1996) elaborated on this strategy by including side news and a polynomial factor, showing that the capital achieved by the -weighted universal portfolio exceeds the capital of the best state-constant re-balanced portfolio at any time x^n based on price dynamics (xn) and side news (yn). This discovery is not limited to averages or probabilities, but rather applies to each series independently.

Blum and Kalai (1997)) suggested a universal portfolio algorithm after investigating the effects of transaction costs on online portfolio management. The idea of deep exterior regret was also considered. It is possible to modify these algorithms to reduce swap remorse as well. Additionally, a "sleeping expert" can play an important part in real-time selection (Blum and Mansour, 2007).

Vovk and Watkins (1998) proposed a new algorithm for picking arbitrarily good investing strategies. Vovk (1990) also created algorithms to count this type of strategy. Similarly aggregating algorithm (AA) proposed by Vovk (1990) was called for a constant number of Constant Re-balanced Portfolios (CRPs).

Ordentlich and Cover (1998) wrote about how different price structures for "m" assets across "n" periods are reflected in the market. The proportion of total capital gained by a non-anticipating asset was compared to the wealth attained by a leading constant adjustable portfolio, with the latter figure determined with reference to the reflection portion of the pricing arrangement.

Cross and Barron (2003) improved upon Cover's research by creating a new universal portfolio approach. This method yields results that are comparable to those obtained by knowing stock prices in advance. Specifically, within the predetermined portfolio categories, it performs better than the conventional constant re-balanced portfolio strategy. These target groups are based on historical stock prices and other data that is tangentially related to stock prices, making for a more trustworthy portfolio classification. The ability to compute with simple closed-form expressions in a polynomial number of steps is a significant benefit of this method. However, other universal processes typically include numerical estimation and exponential calculations.

Akcoglu et al. (2002) developed a common online algorithm for all investing strategies. Methods for optimal offline determinations of daily average performance were discussed. These methods take into account the distribution of wealth among various shares and provide a fundamental foundation for universalizing investing strategies. Their algorithms were designed with consideration for the exponential time limits that they impose. In order to efficiently compute universal portfolios, it is necessary to sample long concave functions, which opens up the method to a wider range of investment strategies.

The Online Newton Step (ONS) approach was developed by Hazan et al. (2007) to solve a rounded online optimization issue. A decision-maker in this scenario will select plausible decision points in Euclidean space from a fixed set of options. The decision-maker is presented with a set of convex cost functions at the outset and is then given a concave payoff function associated with the selected spots. Additive regret of 0(T), where T is an arbitrary time horizon specified for the cost function, was introduced by Zinkevich (2003) in his straightforward online gradient technique. These algorithms shed light on the follow-the-leader strategy by utilizing Newton methods from offline optimization.

Kozat et al. (2008) created a competing approach for building a piece-wise constant re-balanced portfolio using context trees and sequential probability assignments. The optimum piece-wise constant re-balancing portfolio can be chosen with an exponentiation complexity of 0(ln(n)). Subsequently, Kozat and Singer (2011) tackled a sequential-time investment problem. They investigated several investment strategies and thought about practical concerns, such as when and how to re-balance a portfolio when dealing with transaction expenses.

Evidence theory, learning theory, and signal processing are just a few of the fields that have looked into sequential portfolio investment strategies. While Helmbold et al. (1998); Vovk and Watkins (1998); Borodin et al. (2003) dove into learning theory facets, Cover and Ordentlich (1996) investigated sequential universal portfolio techniques. Semi-constant re-balanced portfolios were first introduced by (Kozat et al., 2008), who developed this sequential strategy with an emphasis on signal processing. By making adjustments to wealth depending on specific price ratios, these portfolios strive to generate asymptotically high wealth equivalent to the best Semi-Constant Re-balanced Portfolios (SCRP). Usually, decisions on such portfolios are made retroactively. The authors also proposed a universal sequential portfolio that maximizes wealth relative to the optimal SCRP while accounting for fixed operation costs. These algorithms can only hope to outperform the best SCRP recitation by a factor of O(ln(n)) in terms of wealth. These methods follow a sequential structure and can be applied independently of the values of k and n. The authors provide a thorough explanation of the methods' usefulness, demonstrating the methods' linear complexity with respect to the n-length of the data, and relating the methods to other well-known approaches.

Retrospectively based on real market outcomes, Helmbold et al. (1998) proposed algorithms that reach the same wealth as the best constant rebalanced portfolio. The multiplicative updating rule, developed by Kivinen and Warmuth (1997), uses continuous storage and calculates the time per stock at each time period. The data they collected backed up the assumption and shed light on the relatively minor role that transaction fees play. For the purpose of minimizing exp-concave loss, Hazan and Kale (2015) devised a highly effective approach. Portfolio selection makes use of limitations to study the impact of return variability. Following the seminal work of Cover (1991), this paper makes substantial contributions to the study of regret boundaries for universal portfolio selection. According to data, in a standard Geometric Brownian Motion model, more frequent trading does not inevitably result in more regret. The results of this study are the first to come close to the GBM standard.

With this research, we hope to bridge the gap between traditional portfolio theory and stochastic portfolio theory. The strong connection between regime switching and competing models was demonstrated by a model provided (Hardy, 2001).

During the devastating economic collapse and slowdown in commerce that hit in 1987, their efforts were widely recognized. The transition to a high volatility regime can be seen in the increased volatility of equities returns during times of economic turmoil. The unsuitability of Gaussian and Geometric distributions for the market presents difficulties for this method in producing an accurate estimate of the distribution. Stylized empirical data from a numerical analysis of pricing discrepancies in several financial markets were reported (Cont, 2001).

These characteristics are not tied to any particular model of the profit-making procedure but rather rest on broad qualitative notions. For stochastic models to properly capture the underlying statistical features of profits, they must serve as restrictions. However, modern models have had trouble successfully recreating all these statistical properties, revealing some limitations. The "Follow the Loser" strategy involves switching money from successful businesses to unsuccessful ones. It proposes re-allocating capital from high performing assets to those that are under-performing.

Mean reversion, as investigated by De Bondt and Thaler (1985), is the foundation of the "Follow the Loser" strategy. The results of their psychological tests disproved the efficiency of markets by showing that people frequently exaggerate the impact of recent developments on stock prices. The overreaction hypothesis is supported by these results, which were derived from an examination of CRSP monthly gain data. Previous losers tend to outperform previous winners in weak-form markets.

Li et al. (2012) provide a novel online portfolio selection approach they call "Passive Aggressive Mean Reversion" (PAMR) to better comprehend mean reversion. Mean reversion trading is elucidated, and how PAMR takes advantage of the mean reversion features of financial markets to strike an optimal balance between return and risk.

Binary classification, regression, uni-class prediction, multiclass problems, cost-sensitive multiclass classification, and learning with structured output are only some of the online learning issues that Shalev-Shwartz et al. (2003) examine through a unified algorithmic lens. For the purpose of performing binary classifications online, they introduce a straightforward technique they name Passive-Aggressive (PA). Several algorithms with bounded losses are derived, and the scope of this framework extends all the way from optimal layout to system prediction. All of the boundaries' proofs rest on a common lemma.

The work gives a single lemma that forms the basis for proofs of all bounds, and it concentrates on online environments. It emphasizes the importance of online algorithms in the creation of efficient batch algorithms and draws attention to the prospective extensions of this work.

Borodin et al. (2003) proposed the Anti-Correlation approach, which takes use of statistically expected relationships between groups of stocks. While correlation and other simple statistical relationships might provide substantial profits, it is expected that more advanced machine learning approaches could produce even better portfolio selection algorithms capable of bigger returns, notwithstanding the existence of increased commission fees.

In addition, Duchi et al. (2008) provide two forecasting strategies and efficient methodologies for projecting a vector onto the l1-ball. Both methods provide correct predictions, although the first does it in O(n) time (where n is the dimension of the space) and the second in O(klog(n)) time (where k is the number of observable) by accommodating k vectors with factors outside the l1-ball. Useful in text categorization and other online learning contexts where feature space is limited.

Li et al. (2011b) present a method called Confidence Weighted Mean Reversion that uses machine learning to build financial market portfolios into smart business activities. The study's primary goals are (1) creating trading algorithms for efficient market behavior and (2) establishing non-parametric learning methodologies for effective policy learning. Confidence Weighted online learning is consistent with the idea of mean reversion trading, and empirical evidence reveals that portfolio changes, rather than absolute prices or price relative, play a key effect.

Dredze et al. (2008) offer a related work called Confidence-weighted linear classifiers, which are linear classifiers that are guided by confidence evidence. This method provides a fresh way of learning that can be used to the problem of sluggish performance in NLP applications. Parameter sensitivity, decision making, and distribution variance are all improved by the technique. It also streamlines classifier combination techniques, which boosts NLP applications' precision.

Using ideas from Gaussian distribution, the Confidence-weighted (CW) learning techniques introduced by Crammer et al. (2012) present a new online portfolio learning method for linear classifiers. In this method, the uncertainty in weights and correlation measurements is represented by a covariance matrix. The scalability of these methods is stressed, and an error-bound analysis is performed using data from both synthetic and Natural Language Processing (NLP) studies.

The Confidence Weighted Mean Reversion, introduced by Li et al. (2013), is an online portfolio selection policy that extends the functionality of the preceding method. Mean reversion and confidence weighted online learning are brought together in this method. Using the mean reversion trading method, the portfolio weights are assumed to follow a Gaussian distribution and their distribution is subsequently updated over time. The goal of this fresh strategy is to improve portfolio management processes.

Mean reversion theory can be efficiently implemented in portfolio selection with the help of the user-friendly framework provided by the Dense Weighted Mean Reversion (DW-MR) model. Robust Median Reversion (RMR) is a novel method for online trading that incorporates data from several time periods. It was introduced by Li et al. (2013) Using the L1 median estimator in a mean reversion strategy over a single time period, this method investigates the phenomenon of stock price reversal.

Li and Hoi (2014) highlight the significance of online portfolio selection in the realm of competitive finance, machine learning, and statistics. They offer a complete explanation and survey of numerous online portfolio selection procedures in a number of distinct fields of study.

A third method, Pattern-Matching, is explored for non-parametric sequential investing strategies. This method ensures optimal development and uniform consistency in both deterministic and ergodic markets. It provides a trustworthy means of adjusting financial projections to match market trends.

Cover and Gluss (1986) suggest picking numerous consecutive portfolios according to the market's past performance. These portfolios not only detect differences, but also shed light on potential outcomes. The method takes into account the experimental distribution of the market across n periods in advance to construct optimal portfolios.

Gyorfi and Schafer (2003) examine the use of loss functions such squared loss, 0-1 loss, and log utility for forecasting stationary time series. To this end, they stress the need for developing universal prediction rules that hold true for all imaginable stationary processes.

George and Hwang (2004) compared the results of three distinct investment strategies. The first method invests broadly in the top 30% of high-return equities based on historical data of individual stock performances. The second method uses past industry return data to identify the top 30% of performing sectors. The third strategy is an innovative method that considers how close a stock's current price is to its 52-week high. This duration-based strategy implies that closeness to the 52-week high predicts individual returns, regardless of whether or not the highest return was previously obtained. Traders rely on this time frame frequently, utilizing it as a reference point to correct for biases and factor in the potential impact of news. In the context of the market, it has been noted that stock values tend to climb and reach their highs during periods of positive news or good news. Despite the fact that the available data suggests that the stock price may increase further, surprisingly, traders typically refrain from bidding on these equities. Traders' tendency toward caution has helped fuel the market's recent price rise.

When unfavorable news hits the markets, stock prices typically fall from their 52-week highs. When investors hear bad news, they immediately begin selling the affected stocks. Because of this influx of buyers, prices in the market have fallen.

These regularities call attention to the impact of investor behavior and market psychology on stock prices. Market participants have an impact on the dynamics of stock prices, as seen by the cautious approach during positive news periods and the subsequent selling pressure during bad news periods.

When a stock's price drops significantly from its 52-week high as a result of bad news, however, investors are initially reluctant to sell. Investors' reluctance to sell at what they see as a bargain price is a factor in the market's current bearish trend.

Algorithms described in Györfi et al. (2006) have been subjected to both empirical and theoretical testing, and they have been shown to be effective at tackling the difficulties of sequential investing and deriving maximum returns with minimal assumptions. These strategies use an in-depth familiarity with market dynamics to consistently provide an asymptotic rate of return. There is empirical evidence that demonstrates the usefulness of these tactics.

A kernel-based logarithmic estimating strategy for optimal investment was presented (Györfi et al., 2007). This method has the benefit of maximizing return with the fewest necessary assumptions about the market. Even under the simplified assumptions of market stationarity and ergodicity, empirical evidence support the usefulness of this strategy. It's interesting to see that when tested with NYSE data from the past, both semi-log-optimal and log-optimal methods fare similarly.

Ottucsák and Vajda (2007) asymptotic proposes a non-parametric Markowitz-type technique, which is a streamlined version of the semi-log-optimal method. Portfolio pickers based on closest neighbor approaches were subsequently proposed by Györfi et al. (2008), which offer log-optimal solutions for a wide variety of stationary and ergodic random processes. Empirical research verifies the methods' durability, demonstrating their utility in a range of contexts.

Biau et al. (2010) created nonparametric strategies for sequential prediction by employing an ensemble method called "experts." Their work involves a thorough examination of real-world datasets, proving the approaches' universal reliability under modest assumptions. The findings demonstrate that nonparametric methods frequently beat ARMA techniques in terms of standardized cumulative estimated error and are more versatile and faster to implement.

The goal of the log-optimal investment strategy described by Ormos and Urbán (2013) is to produce a growth rate for investments that is nearly ideal. The relevance of separate investment periods is emphasized, and the effect of reordering costs on portfolio restructuring is analyzed. When experts forecast a negative price shift, they recommend looking into a risk-free investment solution to protect portfolios. The model also includes the potential for short selling. Expert predictions can be made even more accurate by tracking the market during bull and bear times. A trades and quotes catalog is used to circumvent the problem of the required security not being readily available in the market at closing prices.

Using iterative optimization techniques, the Meta-Learning Algorithms developed by Das and Banerjee (2011) learn predictive models in data mining. Up front optimization can be difficult with conventional methods. Online trading is dominated by base algorithms, however Meta Algorithms demonstrate competitiveness through their adaptable combination of strategies. These algorithms offer superior performance to earlier portfolio algorithms. Meta-algorithms provide assurances by adhering to a set of rules and limiting the optimal performance that can be achieved by the best heuristics.

According to Agarwal et al. (2006), the online Newton Stages algorithm is introduced as part of the Online Gradient & Newton Updates strategy for online trading. This algorithm has theoretical guarantees and empirical data suggests that it achieves those promises with remarkable speed in practice. In addition, it outperforms other algorithms in keeping tabs on the best stocks.

Hazan and Seshadhri (2009) delves into the idea of education in a changing world. They introduce techniques for online convex optimization that display adaptability by frequently and rapidly locating local optima. They also suggest a different performance metric to use in conjunction with the common regret meter. In these algorithms, performance is measured against an evolving benchmark and characterized by a continuous flow of data. Using this method, regret-based algorithms can be converted into adaptive algorithms, boosting their performance in dynamic settings.

To implicitly or explicitly predict future price changes, many existing algorithms incorporate specialized trading methods on top of the ideas of Capital Growth Theory. The momentum technique, discussed in Lee and Swaminathan (2000), is one example of such a trading strategy. Their studies center on how the Momentum and Values approaches relate to one another. High turnover companies are studied for their ability to attract additional interest and to sustain financial gains in the future. They do, however, note that price momentum typically reverses within the following five years. In conclusion, the results of this study indicate that predictions about the future can be made with some accuracy using only information about the past.

Cooper et al. (2004)'s momentum trading tactics place a premium on paying attention to the current market environment. Overreaction models studies by Daniel et al. (1998) and Hong and Stein (1999) lend credence to the momentum idea, indicating that a momentum portfolio with a holding period of six months is more likely to see positive returns during bull markets. As the market lags behind, the momentum return rises, but even significant lags in yields are no guarantee of future gains; yields might fall without being rejected. Overreaction theory predicts that gains from momentum will be lost in the long run, and this is what actually happens (Lee and Swaminathan, 2000; Jegadeesh and Titman, 2001) . In addition, they discover that negative market conditions can reverse significantly over the long run even when they lack initial velocity. These results suggest that long-term declines can't be ascribed exclusively to the reversal of preceding momentum but instead reflect the impact of conditional information on lagged market returns, which are intrinsically linked to the forecasting of momentum trends.

Furthermore, progress has been made using cutting-edge methods of prediction.

Llorente et al. (2002)) used a straightforward strategy to investigate the link between stock returns and trading activity. Their research showed that speculative trade gains persist while risk-sharing trade returns reverse. The study also looked at the correlation between volume and returns for specific stocks and found that volume significantly affects correlation on a daily basis.

In addition, this investigation does not account for a number of additional concerns or articles that deal with similar topics. Both probabilistic and deterministic settings have devoted considerable effort to studying the universal forecasting problem, and there are numerous commonalities between the two methods. In particular, self-information loss is important in both contexts for a number of reasons.

To begin, there are various reasons why self-information loss functions are significant in their own right. The self-information loss function appears naturally when viewing the forecasting problem as a problem of probability assignment, which is a major factor. Second, the self-information loss function has been studied extensively and provides a firm theoretical grounding. Finally, the self-information loss function is commonly used as a starting point for deriving or relating conclusions for other loss functions.

There is a noteworthy similarity between the probabilistic and deterministic models of universal forecasting (Dhar, 2011). The goal of many forecasting models in regression and classification issues is to identify a single model that is capable of reliably addressing the vast majority of the cases seen in the training data. The hunt for a model that can generalize well across multiple examples is a cornerstone of the study of forecasting.

The opposing extreme consists of a collection of copies called "small disjoints," each of which is limited to a tiny fraction of the overall decision space. The advantages and disadvantages of these two sorts of replicas have been well-documented and explored. Single models, especially linear ones, are straightforward and easy to understand. But if their complexity isn't managed correctly, they might have overfitting problems that cause big mistakes when used on out-of-sample data.

Direct reinforcement (DR) methods, such as reinforcement learning, have been investigated for their potential use in the context of predicting financial time series and engaging in single stock trading (Moody and Saffell, 2001). Stochastic control issues form the basis of optimization strategies for trading systems and asset allocations. In this article, we introduce recurrent reinforcement learning (RRL), a specialized adaptive learning approach for portfolio selection. RRL uses techniques like TD learning and Q-Learning, as opposed to classic dynamic programming and reinforcement algorithms. To identify optimal solutions without the dimensional complexity of Bellman's method, RRL evaluates the value function through control issues.

The accounting-based measurement of operational expenses and the optimization of risk-adjusted return are both determined (Dempster et al., 2001). They use a time-honored method of financial forecasting called technical analysis. Artificial intelligence (AI) is often cited as a leading method in the present active research domain. In the past, researchers have utilized approaches like out-of-sample testing and genetic algorithms to create trading rules in the hope that they will provide positive out-of-sample gains after accounting for operational costs.

The authors recommend utilizing well-known technical indicators as a starting point for finding success. Common computational learning methods include reinforcement learning and genetic programming. The Marko decision problem can be solved precisely with these methods. All strategies are equally important in producing in-sample and out-of-sample returns, the authors discover when operational costs are eliminated. Empirical data also show that the evolutionary algorithm method excels when there are no operating expenses, but the other approaches all perform poorly.

Overfitting to the sample data is another issue the authors bring out, noting that if in-sample learning is not limited (Mahfoud and Mani, 1996). Genetic algorithm is another method used in stock analysis. This method can be used to resolve problems in databases, however it may be incompatible with more conventional approaches. Results from financial forecasting utilizing the genetic algorithm and a neural network system are compared for roughly 1600 equities. The genetic algorithm is shown to produce valid and trustworthy outcomes, according to the study's findings.

In contrast to the exogenously set rules used by most trading algorithms, the evolutionary algorithms proposed by Allen and Karjalainen (1999) aim to optimize trading performance. The rules that do not pass the out-of-sample test do not outperform the buy-and-hold strategy if transaction costs are factored in. During times of low volatility in commercial transactions, traders can make large returns by adopting long positions, as do the guidelines. In contrast, negative earnings are accompanied by excessive volatility, prompting the rules to switch to short positions.

Volatility forecasting using Genetic Algorithms has been shown to be successful. The empirical evidence shows that the volatility of inputs increases with the degree of their interconnectedness, but variables that satisfy certain requirements can be used as predictors over a specified time frame. Technical analysis has shown more accurate at predicting currency market outcomes than other algorithms, including those tried and tested (Tsang et al., 2004). In the field of machine learning, trading strategies based on Artificial Intelligence (AI) have been increasingly popular over the past two decades. When it comes to technical analysis indicators, AI-based algorithms encompass a wider range, leading to well-defined and profitable trading rules, than prior studies that concentrated on standalone trading rules or genetic algorithms.

The algorithm Support Vector Machine (SVM) was developed by Cao and Tay (2003) for the purpose of forecasting financial time series. In terms of generalization performance, SVM is superior to both multilayer backpropagation neural networks and regularized radial basis function neural networks. To reduce the maximum amount of generalization error, SVM employs the SRM (Structural Risk Minimization) approach. To account for the inevitable structural shifts in financial data, SVM employs adaptive parameters.

Creamer (2007) presented a method for decreasing bias in algorithms by giving more consideration to training data and less to outliers. It should be highlighted, however, that these algorithms have high computational and financial requirements. Companies that use effective strategies have a better chance of succeeding in the marketplace.

For a long time, investors have argued about whether or not market behavior is random. When chartists plot market price data against time, they look for shapes like rectangles and triangles, which they then use to infer the relative strength of buyers and sellers in the market (Pring, 1991). Skeptics, however, point out that these sorts of patterns can also result from chance occurrences (Peterson, 1997b). Imagine the night sky crammed with stars to demonstrate this. It's common knowledge that celestial patterns like lions and bears are purely coincidental, despite our ability to see them.

Computer programs can replicate market behavior by treating pricing as a random walk, providing a way to test the theory of randomness in the market. According to Peter J. Brockwell (1996), random walks are the sum of i.i.d. variables. Different curves are produced every time the program is executed, and regular shapes like triangles and rectangles are typically seen. Traders can use these patterns to help them determine when to enter and quit a deal.

Ramsey theory, a field of mathematics, has particular situations that correlate to this observed occurrence of seeming regularity emerging from randomness (Peterson, 1997a). The existence of regular patterns among a large set of randomly selected objects is explained by Ramsey theory (Mandelbrot, 1982; Avnir, 1998).

The concept of a one-dimensional Brownian motion to simulate price changes was first suggested (Mandelbrot and Mandelbrot, 1982; Avnir et al., 1998). Small particles suspended in a liquid exhibit Brownian motion, an irregular motion with a normal (Gaussian) distribution. Bachelier admitted, however, that his Gaussian random walk model did not perfectly fit the financial data.

Different viewpoints and insights from the fields of mathematics, statistics, and finance theory are presented here to shed light on the ongoing dispute over the existence of patterns and randomness in the market.

There are two major observations that cast doubt on the idea that market price fluctuations follow a normal distribution with a finite variance. In the first place, there is time-varying behavior in the sample variance of price changes. This statistic measures the typical squared dispersion around the mean for a given set of data. The second issue is that not all instances of price fluctuations follow a normal distribution.

Price fluctuations, according to Mandelbrot's 1963 proposal Mandelbrot and Mandelbrot (1982); Mandelbrot (1983), Mantegna and Stanley (1995), follow a Levy stable distribution with infinite variance and larger tails than those expected from a normal process. This suggests that extreme price swings are more common than would be expected from a normal distribution.

Mantegna and Stanley (1995) conducted additional research on the subject of S&P 500 index fluctuations. From one minute to one thousand minutes, they discovered that the random process was best described by a Levy distribution. They proposed the idea of a truncated Levy flight distribution, which has a finite variance, as a solution to the problem of infinite variance. This distribution takes into account the constraints of actual monetary systems.

These findings cast doubt on the commonplace belief that price fluctuations in the financial sector are normally distributed and have a finite variance, stressing the need to investigate other distributions and models that more closely match market reality.

The efficient market hypothesis maintains that the market is inherently unpredictable, regardless of the existence of different price distributions. However, the hypothesis is incorrect since it presumes that all investors are aware of the same piece of news at the same time. In practice, insiders know things before the general public does, experienced traders keep a close eye on news updates in real time, and mutual fund investors usually find out about news the same day or the day after. Despite the ease with which information may now be shared, there will always be discrepancies in how and when certain pieces of news reach individual traders and investors.

Palmer and Rock (1994) raised doubts about the efficient market hypothesis and its close cousin, the rational expectation theory. Market participants, they said, lack perfect information, aren't completely logical, and may have trouble figuring out what their best moves should be. They also have no way of guaranteeing that others will make the same choices they do. The data and strategies available to and used by various agents may vary. Indicators such as these point to the existence of prevailing market tendencies, suggesting that monetary fluctuations are not completely at random.

Changes in signals over time are key to grasping their meaning. Continuous-time signals, also known as analog signals, are unbroken in time and can be further subdivided into discrete-time signals, which are functions of an integer-valued time variable. It is common practice to sample a continuous-time signal with an analog-to-digital converter in order to generate a discrete-time signal. Tick data from the financial markets is one example of a signal that is intrinsically discrete-time.

A "tick" is an up or down change in the price of a security's trading on the stock market. Many tick data points can occur in a short period of time, with ten ticks possible in only one minute of trade. Tick data is frequently collected at regular intervals in order to simplify and consolidate this information. In a 10-minute chart, for instance, the greatest and lowest ticks of each interval as well as the opening and closing ticks are depicted. This approach of financial market data visualization is distinct from the more common practice of using the period's closing value.

Market participants put a lot of stock on the day's high, low, open, and close prices as indicators of future movement. Each hourly interval in a chart displays these four prices again. In most cases, the last reported prices are used as the basis for analysis, while the median of the day's range may also be taken into account. It is more practical and consistent with the mathematical idea of sampling to use the previous day's closing prices. If you have a 10-minute chart, you can extrapolate the closing price for the hour by taking a sample every sixth point.

In 180 7, French mathematician Joseph Fourier introduced the concept of the Fourier series, which expresses any practical signal as the sum of numerous sine waves. The field of digital signal processing has benefited greatly from this mathematical ideas (Broesch, 1997; Hubbard, 1998).

Financial data can be modeled as a sum of sine waves using the Fourier theorem. Since a sine wave only has one frequency, it can be sampled at 32 times every cycle. The frequency of the sine wave is equal to 1/32 of its period, as 1 period equals 32 points. Since 2 radians is equal to 360 degrees, the sampling frequency (in degrees per second) is 2/32, or 7/16 radians per cycle.

The Nyquist theorem Broesch (1997) states that in order to faithfully duplicate a signal, its samples must be taken at a rate that is more than twice the frequency of the signal's highest frequency component. In other words, the maximum frequency requires more than two samples every cycle. This criterion is the same as saying the greatest frequency component must have a circular frequency of fewer than radians. The primary signal in the financial market is price, which is plotted as a series of bars on a chart over a certain period of time. When attempting to predict market moves, technical analysts place a premium on pricing. Various manipulations and indicators have been created to extract hidden information from this raw price data in order to predict the direction of the market. One issue with these algorithms is that they don't all use the same procedure for figuring out what the best sample frequency should be.

The purpose of this research is to use mathematics to investigate the non-random character of financial markets and to analyze various scientific theories that shed light on market dynamics. Since the state of the art in mathematics and science is always expanding, we will be using scholarly articles to back up our claims as we investigate. Understanding the market is still in its infancy, as its behavior is driven by the aggregate activities of individuals, which are little understood.

The ability to quickly absorb large amounts of varied information and make investment judgments is a critical component in the creation of effective algorithms for E-trading markets. Human intervention should be minimized, bias should be removed, resources should be conserved, and time should be saved if this goal is to be realized. The use of decimated data in algorithms is a viable strategy for tackling these issues.

To "decimate" means to select a smaller selection of data that is indicative of the whole. Algorithms based on decimated data allow for efficient management of big datasets, with the retained information and complexity reduced. This method improves the performance of algorithms by allowing them to zero in on the most important features and patterns while ignoring the less relevant information.

Creamer and Freund (2010) trading system is a good illustration of this trend; it uses machine learning techniques to produce positive anomalous returns over a wide range of stock markets. The system's weighting algorithm takes into account both expert opinion and a risk-management mechanism to determine which forecasts are the most reliable and to prevent trading during volatile market periods. This method's use of decimated data facilitates rapid processing and analysis of market data, which in turn enhances the precision of forecasts.

Creamer (2012) also suggests logtiboost approaches, which automatically normalize trading models by utilizing several forms of technical analysis. This strategy uses decimation to find the best possible set of technical indicators and to propose novel trading rules from the perspective of a smaller dataset. This shows how decimated data can improve trading system efficiency, as the algorithm surpasses a collection of typical technical analysis tools.

When it comes to analyzing financial markets, the value of Big data-based algorithms resides in their capacity to analyse massive amounts of data with minimal human bias, expense, and time investment. These algorithms increase the accuracy of predictions and the quality of investment choices by applying decimation techniques into the examination of complex market data.

The weekly buying and selling trend of TWSE stocks was studied by Chang et al. (2018), who then defined it as an example of irrational conduct. The Risk Seeker Indicator (RSI) was used to classify the weekly market situation into one of six categories. The study's authors zeroed in on the top 150 equities to determine where irrational trading had the most noticeable effects.

All 150 stock securities were included in the study's stock pool for analysis. The first data period began on April 1, 2002, and ended on June 15, 2012; the second data period, which was used for backtesting, began on June 15, 2012, and ended on May 9, 2014. The study's authors created 5,000 hypothetical weekly stock return scenarios to model extreme irrationality. The detected unreasonable behavior's consequences were measured and evaluated over the course of a 100-week backtesting period.

Chang et al. (2018) investigated the many ways in which people behave differently while making choices. In particular, they compared SF1's (Safety First) results to those of its rivals. Importantly, SF1 exceeded the Market in all but one indicator (standard deviation), when it lagged behind. This research shows that RS1 (Risk Seeker) and SF1 are superior to other methods.

RS1 (SF1) also fared better than RS2 (SF2), MV (Mean Variance), and the Market when evaluating total return over a 100-week time horizon. The investigation also uncovered an essential defect in the aforementioned algorithms, which originated in the choosing of sample frequency. It was found that the optimum sample frequency parameter cannot be determined by a lone method. For machine learning algorithms to work as intended, precise sample sizes and frequencies are required.

Chang et al. (2018) research provided insight into individual differences in decision-making style and showed that SF1 and RS1 were superior to their rivals. It was stressed that the success of machine learning algorithms is heavily dependent on picking suitable sample sizes and frequencies.

Li and Lin (2020) exploring found evidence of significant entropy even when considering larger time periods in their investigation. Twenty stocks from the SSE 50 index were chosen for further analysis, including the five with the highest and lowest entropies based on daily data and the same five stocks chosen using high frequency financial data collected every five minutes.

Researchers used a coarse-graining approach with amplification ratios ranging from 0 to 20 to account for the impact of multi-scale impacts on entropy computation and stock selection. Using this approach, they were able to construct a coarse-grained dataset from which they could determine the median and mean values, thereby allowing them to distinguish between stocks with high and low entropy. The median and the average were used to make the final cut.

Plz, an information-theoretic predictability estimator derived from the Lempel-Ziv estimator, was developed by Li and Lin (2020) for use in evaluating the predictability of financial time series. Plz attempts to put a numerical value on the role that past values played on the time series, with the ultimate goal of lowering the degree of uncertainty surrounding future values.

2.3 Theoretical/Methodological Background of Algorithms

This study focuses on online learning algorithms and their theoretical and empirical gratifications.

2.4 Online Linear Regression

Learning that takes place online is based on a series of data samples that are timestamped. The researcher is given an instance $x_t \in \mathbb{R}^t$ at each and every step t in the process. It begins by making an attempt to create a prediction for the output of the incoming instance. This prediction is made using a formula that, in the case of linear regression, has the form $\hat{y}_t = \omega_t^T x_t$, where $\omega_t \in \mathbb{R}^n$ is the weight vector that has been incrementally learned. Following the making of the prediction, the true output $y_t \in \mathbb{R}$ is revealed, and the learner then computes the incurred loss $\ell\left(y_t\right), \left(\hat{y}\right) \in \geq_0$, depending on some criterion to assess the difference between the learner's forecast and the revealed true output. At the end of each learning step, the learner will ultimately determine, using this loss, whether or not to update the regression model, as well as how to do so.

2.4.1 Statistical Learning Theory

Understanding and foresight are the desired outcomes of the learning process. One of the most important foundations for the theoretical analysis of machine learning problems is statistical learning theory, which was initially presented in the late 1960s. This theory is particularly useful for supervised learning. There are many different types of learning, such as supervised learning, unsupervised learning, learning through online platforms, and learning through reinforcement. Understanding supervised learning is easiest when approaching it from the point of view of statistical learning theory proposed by Poggio et al. (2012), learning from an existing training set of data is what supervised learning entails. At each and every point in the training, there is a pair of input and output points, with each input mapping to one of the outputs. The learning issue consists of deducing the function that maps between the input and the output in such a way that the learnt function can be used to predict the output based on the input that will be provided in the future.

Problems in supervised learning can either be regression problems or classification problems Osisanwo et al. (2017), depending on the type of output that is being analysed. This indicates that there is a regression problem because the output takes on a continuous range of values. A regression could be carried out with voltage as the input variable and current as the output variable, using Ohm's Law as an illustration. The functional connection between voltage and current would be determined by the regression to be $\mathbb{R}$ which would mean that $V = I\mathbb{R}$ Problems that fall under the category of classification are ones in which the answer will be one label from among a distinct collection of options. Applications of machine learning often involve classification in some form or another. For the purpose of facial recognition, for instance, the input would be a photograph of a person's face, and the output label would be that person's given name. The input would be represented as a very large multidimensional vector, and each member of that vector would stand in for a pixel in the picture.

After learning a function based on the information contained in the training set data, the function is then validated using a test set of data, which contains information that was absent from the training set.

2.4.2 Convex Optimization Theory

The use of convex sets and functions in optimization models Bertsekas (2015) is extremely beneficial because of their rich structure, which makes it easy to conduct analysis and develop algorithms. A significant portion of this structure can be traced back to a small number of its fundamental features. As an illustration, each point on the boundary of a convex set can be approached through the relative interior of the set, and each halfline belonging to a closed convex set still belongs to the set when translated to start at any point in the set. Closed convex sets can be described in terms of the hyperplanes that support the set, and each point on the boundary of a convex set can be approached through the relative interior of the set. Nevertheless, in spite of their advantageous structure, convex sets and the analysis of them are not devoid of anomalies and unique behaviour, both of which lead to significant challenges in theoretical work and practical implementations. For instance, in contrast to affine and compact sets, the degree to which the closedness of closure convex sets is preserved by fundamental operations such as linear translation and vector sum is not guaranteed to be constant. The discussion of several fundamental optimization concerns, such as the existence of optimal solutions and duality, is thus made more difficult as a result of this.

2.5 Online Passive-Aggressive Regression

The online passive-aggressive (PA) regression approach was initially described (Crammer et al., 2006). It has been implemented in a variety of fields, such as non-negative matrix factorisation and completion Blondel et al. (2008), online portfolio selection Li et al. (2012), and statistical machine translation (Turchi et al., 2014). Upon receiving an instance x^t at the beginning of each cycle, the PA regression algorithm produces a prediction for the target value $\hat{y}_t = \omega_t^T x_t$ using its internal regression function. This forecast might be considered an approximation. Following the conclusion of a prediction, the algorithm is presented with the actual value of the target, at which point it immediately incurs a loss due to its faulty prediction. PA regression depends on the insensitive loss function (ILF) (Vapnik, 1998).

2.5.1 Online Gradient Descent

Numerous issues with online learning can be re-framed as "online convex optimisation" (OCO) tasks, as was previously stated, and then solved using the "online gradient descent" (OGD) algorithm.

In the broadest sense imaginable, it is pertinent to the OCO setting. Zinkevich (2003) was the first to deliver it in an online setting. Its development was based on the offline optimization technique known as standard gradient descent.

Beginning where it left off in the previous iteration, the algorithm moves linearly, taking one step at a time.

The direction that the gradient of the prior cost would go in. It is feasible to reach a point that does not belong to the underlying convex set if you use this strategy in its entirety. The approach "projects" the point back onto the convex set in situations like these. This means that it locates the convex point that is most near the target point. Even while the next cost function may be completely different from the costs that have been seen up to this point, the regret that the algorithm ultimately produces is still sublinear (Hazan et al., 2016).

2.6 Second-Order Online Learning

In recent years, significant progress has been made toward the development of second-order online learning algorithms with the goal of improving the efficiency of first-order learning strategies. When using these approaches, the weights are almost always given a Gaussian prior distribution. This distribution has a mean vector of $\mu \epsilon \mathbb{R}^n$ and a covariance matrix of $\Sigma \varepsilon \mathbb{R}^{n \times n}$. Keeping the model parameters up to date is an integral part of the online learning process. Examples of second-order algorithms for linear regression include adaptive regularisation of weight vectors Crammer et al. (2009), which was motivated by confidence-weighted learning Dredze et al. (2008), and a more recently introduced Bayesian treatment of online PA learning, namely online Bayesian passive-aggressive learning (Liu et al., 2017).

2.6.1 Confidence-Weighted Learning

Confidence-weighted (CW) learning was introduced by Dredze et al. (2008) as a probabilistic extension of online passive-aggressive learning. CW learning accounts for model/parameter uncertainty by retaining a probabilistic confidence measure. As a probabilistic extension of online passive-aggressive learning, CW learning was created. In each weight with a lower confidence level are altered more frequently than those with a higher confidence level. A Gaussian distribution on weight vectors is used to formalise weight confidence. This distribution is adjusted for each new training instance so that the probability of successful classification for that instance under the updated distribution meets a specified confidence threshold.

2.6.2 Adaptive Regularisation of Weights

It has been shown that confidence-weighted algorithms perform well in practise (see, for example, Dredze et al. (2008), however these algorithms also have a number of limitations. To begin, the update takes a rather proactive stance, since it requires that the chance of accurately forecasting each example be at least more than half, despite the fact that this increases the cost to the aim. When labels are noisy, this can lead to significant over-fitting, which is especially problematic when considering the assumption made by the CW framework that data can be linearly separated. Second, they are designed to be used for classification, and it is not entirely clear how they may be applied in other situations, such as regression. This is due, in part, to the fact that the constraint is stated in discrete terms, meaning that the forecast will either be accurate or inaccurate.

Adaptive regularisation of weights, also known as AROW, is a variation of CW that was proposed by Crammer et al. (2009) to address both of the aforementioned challenges, specifically dealing with label noise in a more efficient manner and generalising the benefits of CW learning in an extendable manner.

In conclusion of the literature review, this study focuses on testing the following research hypotheses:

1. Algorithmic trading algorithms can effectively use decimated data.
2. In terms of overall results, algorithm-based trading performs better than market-based trading.
3. Longer forecasting and decision-making time horizons produce superior outcomes over shorter ones.
4. Compared to passive investment techniques, algorithm-based trading strategies have more long-term risks.
5. Algorithm-based methods provide a more robust approach to market volatility by providing greater risk-adjusted returns against downside risks.

Chapter 3

Research Methodology

3.1 Online Strategies-Background

Online portfolio selection is an important part of computational finance, therefore it has attracted a lot of attention from academics in domains as diverse as finance, statistics, AI, ML, and data mining.

Based on Li and Hoi (2014) and similar studies, summarized in the Table 3.1, the most popular machine learning methodologies currently being employed in this field of study. The goal is to compile and categorize existing research in this field.

The algorithms' advantages and disadvantages are laid forth in a systematic manner after a thorough literature analysis (Li et al., 2013, 2015; Hoi et al., 2021). The Follow the Winner strategy, which has its origins in the Regret Bound Theory, is one of the most popular approaches. Successful equities would theoretically be given a greater share of the portfolio under this approach. However, data reveals that in practice, it does not function particularly well. More importantly, real market data does not corroborate the theory's central tenet, which is that relative prices follow an independent and identically distributed (i.i.d.) pattern. The Follow the Loser method is another option, and it's based on the idea that price fluctuations will eventually return to their averages. For short-term data, such daily price data, however, the strategy's claimed high cumulative returns have been debunked by academics.

Pattern-matching methods incorporate winner-and-loser algorithms and nonparametric sequential investing strategies. Pattern matching and optimization are the two main phases of these methods. However, they gloss over the existence of cyclical trends in the industry.

The Fund of Funds (FoF) theory provides a foundation for many of the expert tactics used in Meta-Learning Algorithms. These methods may share a common ancestor or originate from entirely separate classes. This method incorporates intricate mathematical models, which is a huge plus in terms of sophistication.

In conclusion, scholars in the subject of online portfolio selection have investigated a wide range of tactics, taking into account a wide variety of theoretical frameworks and methods. The i.i.d. assumption, inconsistent performance, and failing to capture recurring patterns are some of the criticisms leveled at these methods. There is hope in the development of meta-learning algorithms because of their potential to mix different approaches taken by experts and make use of mathematical simulations. The efficiency and reliability of online portfolio selection algorithms could benefit from more study and development in this area.

3.2 Mathematical Prospective of Problem

Suppose there is a financial market with m different assets; in this market, we distribute our funds throughout n different trading sessions among all of the assets. The fluctuation in market price is characterised by an m-dimensional price relative vector denoted by $\mathbf{x}_t \varepsilon \mathbb{R}^m_+, where t = 1...., n$, and the i^{th} component of the i^{th} price relative vector, denoted by $x_{t,i}$, is the ratio of the i^{th} closing price to the i^{th} asset's most recent closing price.

As a consequence of this, the value of an investment made during period t in asset i grows by a factor of $x_{t,i}$. A market window, which consists of a sequence of price relative vectors $\mathbf{x}_{t_1}^{t_2}$, is another way that we refer to the changes in market price that occur between periods t_1 and t_2 where t_2 is longer than t_1. t_1 equals $\left\{\mathbf{x_{t_1,....}}, ...\mathbf{x_{t_2}}\right\}$, where t_1 denotes the beginning of the period and t_2 denotes the

TABLE 3.1: Classification of Algorithms: Reproduced from (Li and Hoi, 2014)

CLASSIFICATIONS	**ALGORITHMS**	**REFERENCES REPRESENTATION**
Benchmarks	Buy and Hold, Best stocks	(Kelly, 1956a)
	Constant Rebalance Portfolio	(Cover, 1991)
Follow the Winner	Universal Portfolios	(Cover, 1991; Cover and Ordentlich, 1996)
	Exponential Gradient	(Helmbold et al., 1998)
	Follow the Leader	(Gaivoronski and S., 2000)
	Follow the Regularized Leader	(Agarwal et al., 2006)
	Aggregating-type Algorithms	(Vovk and Watkins, 1998)
Follow-the-Loser	Anti Correlation	(Borodin et al., 2003, 2004)
	Passive Aggressive Mean Reversion	(Li et al., 2012)
	Confidence Weighted Mean Reversion	(Li et al., 2011b, 2013)
	Online Moving Average Reversion	(Li et al., 2012)
	Robust Median Reversion	(Huang et al., 2012)
Pattern-Matching Approaches	Nonparametric Histogram Log optimal Strategy	(Györfi et al., 2006)
	Nonparametric Nearest Neighbor Log-optimal Strategy	(Györfi et al., 2006)
	Nonparametric Kernel-based Log-optimal Strategy	(Györfi et al., 2008)
	Correlation-driven Nonparametric Learning Strategy	(Li et al., 2011a)
	Nonparametric Kernel-based Semi-log-optimal Strategy	(GYÖRFI et al., 2007)
	Nonparametric Kernel-based Markowitz-type Strategy	(Ottucsák and Vajda, 2007)
	Nonparametric Kernel-based GV-type Strategy	(Györfi et al., 2008)

Table 3.1: **(Continued)**

CLASSIFICATIONS	**ALGORITHMS**	**REFERENCES REPRESENTATION**
Meta-Learning Algorithm's	Aggregating Algorithm	(Vovk, 1990; Vovk and Watkins, 1998)
	Fast Universalization Algorithm	(Akcoglu et al., 2002, 2004)
	Online Gradient Updates	(Das and Banerjee, 2011)
	Online Newton Updates	(Das and Banerjee, 2011)
	Follow the Leading History	(Hazan and Seshadhri, 2009)

conclusion of the period. $\mathbf{x_1^n} = \{\mathbf{x_1}, ...\mathbf{x_n}\}$ signifies a one-of-a-kind market window that extends from period 1 all the way up to period n.

An investment that will be made at the start of the t^{th} term is denoted by the portfolio vector $\mathbf{b}_t$, where $t = 1, ...n$. The proportion of total portfolio capital that is allocated to the i^{th} asset is represented by the value of the i^{th} portfolio component, $b_{t,i}$. In general, we operate on the assumption that a portfolio is self-financed and that neither margin trading nor short sells are allowed. Therefore, if a portfolio meets the conditions $\mathbf{b}_t \, \varepsilon \, \Delta_m \, , where \; \Delta_m = \left\{\mathbf{b} : \mathbf{b} \succeq \mathbf{0}, \mathbf{b^T 1} = \mathbf{1}\right\}$, where each input is positive and the sum of all entries equals one, then the portfolio is said to be in good standing. One represents a vector of length m that is composed entirely of ones, while $\mathbf{b^T 1}$ represents the inner product of one and another vector, 1. A portfolio strategy that depicts the investment operation from period 1 to period n is comprised of the mappings described in the following lines:

$$\mathbf{b}_t = \frac{1}{m}\mathbf{1}, \; \mathbf{b}_t : \mathbb{R}_+^{m(t-1))} \rightarrow \Delta_m, \; t = 2, 3, ..., n \tag{3.1}$$

where $\mathbf{b}_t = \mathbf{b}_t \left(\mathbf{x}_1^{t-1}\right)$ is the portfolio from the market window of the time before the current time period i.e. $\mathbf{x}_1^{t-1}$. The method for managing a portfolio over n

different time periods can be expressed as $\mathbf{b}_1^n = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$.

At the beginning of time t, a portfolio manager invests money into the portfolio in accordance with portfolio $\mathbf{b}_t$ and keeps it there until the end of time t. Accordingly, the portfolio's wealth will grow by a factor of $\mathbf{b}_t^T \mathbf{x}_t = \Sigma_{i=1}^m \ b_{t,i} x_{t,i}$.

The portfolio's value will increase at an exponential rate due to the model's use of price relationships and subsequent reinvestment of profits. After n iterations, the cumulative wealth is equal to S_0, assuming that the portfolio strategy $\mathbf{b}_1^n$ raises the initial wealth S_0 by a factor of $\prod_{t=1}^n \mathbf{b}_t^\top \mathbf{x}_t$ from period 1 to period n i.e.

$$S_n\left(\mathbf{b}_1^n\right) = \ S_0 \prod_{t=1}^{n} \mathbf{b}_t^\top \mathbf{x}_t = S_0 \prod_{t=1}^{n} \Sigma_{i=1}^m b_{t,i} x_{t,i} \tag{3.2}$$

We define the exponential growth rate for a strategy $\mathbf{b}_1^n$ as follows because the model implies multi-period investment:

$$W_n\left(\mathbf{b}_1^n\right) = \frac{1}{n} \, log \; S_n\left(\mathbf{b}_1^n\right) = \frac{1}{n} \, \Sigma_{t=1}^n \; log \; \mathbf{b}_t . \mathbf{x}_t. \tag{3.3}$$

Let's put the finishing touches on the online portfolio selection approach by putting all the components in place. A portfolio manager is the person who makes the final decision in a portfolio selection task. A portfolio manager's job is to devise a portfolio strategy $\mathbf{b}_1^n$ in order to achieve a set of predetermined goals. The objective is to get a cumulative value for the portfolio that is as high as possible, S_n, while maintaining the same general principles that are used by the algorithms outlined in **Algorithm** 1. Portfolio managers learn a new portfolio vector $\mathbf{b}_t$ at the beginning of period t based on the previous market window $\mathbf{x}_1^{t-1}$ for the subsequent price relative vector $\mathbf{x}_t$, where the decision criterion differs between managers/strategies. The portfolio $\mathbf{b}_t$ is determined by the portfolio period return $\mathbf{b}_t \cdot \mathbf{x}_t$. This process is repeated until period n, at which point the portfolio's total wealth S_n is used to evaluate the strategy's success.

It is possible to back-test any online Stock selection algorithm by following the steps outlined in **Algorithm** 1, which presents the overall architecture of online stock selection. This popular model makes the following assumptions:

Algorithm 1 Framework for Stock Prediction

Input:$\mathbf{x}_1^n$: Historical market sequence;
Output:S_n: Cumulative wealth (end)
Initialize $S_0 = 1$; $\mathbf{b}_1 = \left(\frac{1}{m}, ..., \frac{1}{m}\right)$
for $t = 1, 2, ..., n$ **do**
  Stock $\mathbf{b}_t$ computes;
  Relative Price $\mathbf{x}_t$ reveals;
  Over time period retun $\mathbf{b}^\top\mathbf{x}$, updates $S_t = S_t - 1 \ \times \left(\mathbf{b}^\top\mathbf{x}\right)$;
  Stock selection rules updates by decision manage;r
end

3.2.1 Model Assumptions

First, the approach relies on the absence of transaction fees and taxes. Secondly, the model assume that every asset can be bought or sold in any quantity at its closing price, which suggests that the market is liquid. Thirdly, as far as we're aware, there is no impact-costing approach of stock selection that also assumes no change in market behaviour.

3.2.2 Dealing with Transaction Costs Issue

This research follow the widely adopted model "Proportional Transaction Cost" proposed by Vovk and Watkins (1998), wherein the transaction cost is proportionally associated during the re-balancing mechanism of transferring of wealth among the different set of assets.

In this approach the investor manager plans to re-balance the stock from the close-price adjusted returns $\hat{\mathbf{b}}_{t-1}$ to a new combination of stock $\mathbf{b}_t$ at the start of the t^{th} period. Here, they determine each component of $\hat{\mathbf{b}}_{t-1}$ by solving $\hat{\mathbf{b}}_{t-i} = \frac{b_{t-1,i}x_{t-1,i}}{\mathbf{b}_{t-1}^T\mathbf{x}_{t-1}}$, *for i=1,...,*m However, it is crucial to keep in mind that this transnational solution has nothing to do with those strategies that specifically deal with the issue of transaction costs (Györfi et al., 2008; Ormos and Urbán, 2013).

3.3 A Brief of Online Learning Algorithms and Strategies

3.3.1 Benchmarks

This online algorithm responds to each request in the game of requests and answers before serving the next one. Such actions each result in a certain gain. In competitive analysis, an algorithm's performance is determined by dividing the sum of the gains from the best off-line algorithm and the best online algorithm, both of which service the same series of requests across the worst-case inputs (Chen et al., 1999). The competitive ratio is the ratio that the online algorithm aims to minimise.

3.3.1.1 Buy and Hold Strategy

The buy-and-hold (BAH) strategy, which invests money in the market with an initial portfolio of $\mathbf{b_1}$, and retains the portfolio until the end, is the most popular baseline. The portfolio holdings are altered implicitly in reaction to market movements, with the manager only making asset acquisitions at the beginning of the first period. In this way the holding of portfolio becomes $\frac{\mathbf{b}_1 \odot \mathbf{x}_1}{\mathbf{b}_1^\top \mathbf{x}_1}$, where $\odot$ used for element wise product and refraining from re-balancing in future periods. The final product under BAH strategy becomes;

$$S_n\left(BAH\left(\mathbf{b}_1\right)\right) = (b)_1 \cdot \left(\bigodot_{t=1}^{n} \mathbf{x}_t\right) \tag{3.4}$$

The uniform BAH strategy is the BAH strategy with an initial uniform portfolio of $\mathbf{b}_1 = (\frac{1}{m}, ..., \frac{1}{m})$, and it is commonly employed as a market strategy to generate a market index.

3.3.1.2 Best Stock (Best) Strategy

The Best Stock (Best) strategy, a unique BAH method that invests all capital on the best stock in retrospect, is another popular benchmark. It has a hindsight approach since its starting portfolio can be estimated as

$$\mathbf{b}^{\circ} = \underset{\mathbf{b}\,\varepsilon\,\Delta_m}{\arg\max}\ \mathbf{b} \cdot \left(\bigodot_{t=1}^{n} \mathbf{x}_t \right) \tag{3.5}$$

This allows us to derive an expression for the cumulative wealth attained by the Best approach over time as:

$$S_n(Best) = \max_{\mathbf{b}\,\varepsilon\,\Delta_m} \mathbf{b} \cdot \left(\bigodot_{t=1}^{n} \mathbf{x}_t \right) = S_n(BAH(\mathbf{b}^{\circ})) \tag{3.6}$$

3.3.1.3 Constant Re-balanced Portfolio

Constant Re-balanced portfolio maintains the same wealth distribution among a group of assets throughout time. These algorithms are built on the premise that a fixed asset allocation approach can produce high returns. The wealth produced by a continuously re-balanced portfolio, however, is frequently significantly less than the wealth produced by an ad hoc investing strategy that responds to market fluctuations because stock markets are far from being stationary. The portfolio strategy can be written as $\mathbf{b}_1^n = \{\mathbf{b}, \mathbf{b}, ...\}$ So, the total portfolio value achieved by a CRP approach after n periods is defined as,

$$S_n(CRP(\mathbf{b})) = \prod_{t=1}^{n} \mathbf{b}^{\top}\mathbf{x}_t. \tag{3.7}$$

3.3.2 Follow the Winner Approaches

The "Follow-the-Winner" technique gives more weight to the advice of experts/-stocks who have been correct in the past. Generally these algorithms follow the the BCRP technique, which has been proved to be the optimal strategy in an i.i.d. market (Cover, 1991). This optimality promotes "universal portfolio selection" algorithms to come close to the performance of the hindsight BCRP for any sequence of price relative vectors, commonly known as individual sequences.

3.3.2.1 Universal Portfolios

Algorithms of the Universal Portfolio variety work on the principle of pooling the profits earned by a group of experts in a specific field. The Buy and Hold (BAH) strategy is an analogue of this tactic. The approach of investing in a single stock, on the other hand, is known as base BAH expert, and the number of experts corresponds to the number of stocks. This means that the BAH technique comprises the acquisition of stocks, the subsequent sale of those stocks, and the final pooling of those stocks' capital. In contrast, the foundational expert in the Follow-the-Winner grouping can be any market-investing strategy class. Furthermore, algorithms in this class are related to Meta-Learning Algorithms (MLA), which are discussed in later section, however MLA often applies to experts in more than one class.

3.3.2.2 Exponential Gradient

Techniques of the Exponential Gradient type frequently centre their attention on the following optimization issue:

$$\mathbf{b}_{t+1} = \underset{\mathbf{b}\,\varepsilon\,\Delta_m}{\arg\max} \quad \eta \log \mathbf{b} \cdot \mathbf{x}_t - R\left(\mathbf{b}, \mathbf{b_t}\right) \tag{3.8}$$

where $\eta > 0$ represent the learning rate and $R\ (\mathbf{b}, \mathbf{b_t})$ for regularisation term. Following the stock with the best return during the prior time period while maintaining a degree of resemblance between the new and old portfolios is one straightforward interpretation of optimization.

3.3.2.3 Follow the Leader

The Follow the Leader (FTL) approach makes an effort to resemble the Best Constant Rebalanced Portfolio (BCRP) until time t i.e.

$$\mathbf{b}_{t+1} = \mathbf{b}_t^{\star} = \underset{\mathbf{b}\,\varepsilon\,\Delta_m}{\arg\max} \sum_{j=1}^{t} \log\left(\mathbf{b} \cdot \mathbf{x}_j\right). \tag{3.9}$$

There is no doubt that the BCRP is the category leader and, indeed, the ultimate leader across all time periods. A method for obtaining portfolios by combining the BCRP up to time t with a uniform portfolio was briefly discussed by Ordentlich Ordentlich and Cover (1998), which is as follows;

$$\mathbf{b}_{t+1} = \frac{t}{t+1}\,\mathbf{b}_t^{\star} \;+\; \frac{1}{t+1}\frac{1}{m}1. \tag{3.10}$$

3.3.2.4 Category of Follow the Regularized Leader

Another family of approaches, called Follow the Regularized Leader (FTRL), takes the FTL concept and applies a regularisation term to it. As a whole, the following constitutes FTRL methods:

$$\mathbf{b}_{t+1} = \underset{\mathbf{b}\,\varepsilon\,\Delta_m}{\arg\;\max} \sum_{\tau=1}^{t} \log\left(\mathbf{b}\;\cdot\;\mathbf{x}_{\tau}\right) \;-\; \frac{\beta}{2}\,R\left(\mathbf{b}\right) \tag{3.11}$$

where $R\ (\mathbf{b})$ is a regularisation term on $\mathbf{b}$ and β is the trade-off parameter. In contrast to the EG method, the regularisation term here only applies to the subsequent portfolio because all relevant past information has already been collected in the first term. In L2 norm, where $R(\mathbf{b}) = \|\mathbf{b}\|^2$, is an example of a common regularizer.

3.3.2.5 Category of Aggregating-type Algorithms

However, in real markets, the i.i.d. assumption is debatable, therefore the optimal portfolio may not be a CRP or fixed fraction portfolio even though BCRP is the optimal strategy for such a market. It's possible to use some algorithms that were developed specifically for the purpose of following a certain type of subject matter experts. These algorithms are conceptually related to the Meta-Learning Algorithms discussed later in later paragraph. Whereas Meta-Learning Algorithms typically apply to more complicated experts from several classes, in this case the basis experts are of a specific class: individual experts that invest just in a single stock.

Cover's UP is a special example of the online portfolio selection task, which was addressed by Vovk and Watkins (1998) using the Aggregating Algorithm (AA) (Vovk, 1990; Gaivoronski and S., 2000). In order to obtain good performance that is not poorer than any predetermined combination of underlying experts, the typical context for AA involves defining a countable or finite collection of base experts and sequentially allocating the resource among many base experts. The mathematical formulation for its online portfolio selection update is:

$$\mathbf{b}_{t+1} = \frac{\int_{\Delta_m} \mathbf{b} \prod_{i=1}^{t-1} (\mathbf{b} \cdot \mathbf{x}_t)^{\eta} P_0 \, (d\mathbf{b})}{\int_{\Delta_m} \prod_{i=1}^{t-1} (\mathbf{b} \cdot \mathbf{x}_t)^{\eta} P_0 \, (d\mathbf{b})} \tag{3.12}$$

in above the experts' initial weights, $P_0 \, (d\mathbf{b})$, are appended. Cover's Universal Portfolios is a particular instance of AA when the prior distribution is uniform and $\eta = 1$.

3.3.3 Follow-the-Loser Approaches

In order for BCRP to work optimally, it assumes the market is i.i.d., which is not necessarily true in practise and can lead to subpar empirical performance, as has been seen in a number of literatures. The Follow-the-Loser strategy, in contrast to the Winner-Take-All approach, typically involves redistribution of money from the winners to the losers. Mean reversion is the premise upon which this strategy is built De Bondt and Thaler (1985); Poterba and Summers (1988); Lo and MacKinlay (1990), which states that high-performing assets will continue to perform at high levels in subsequent periods.

3.3.3.1 Anti Correlation

A Follow-the-Loser portfolio approach called Anti Correlation (Anticor) (Borodin et al., 2003, 2004). Unlike Cover's UP, which makes no distributional assumptions, Anticor's approach to the market operates under the premise that it will tend toward its mean. Specifically, it uses a statistical wager on the reliability of positive lagged cross-correlation and negative auto-correlation in order to take advantage of the mean reversion feature.

Anticor utilises logarithmic price relatives Hull (2008) in two distinct market windows $\mathbf{y}_1 = \log\left(\mathbf{x}_{t-2w+1}^{t-w}\right)$ and $\mathbf{y}_2 = \log\left(\mathbf{x}_{t-w+1}^{t}\right)$ to generate a portfolio for the *t+1*st period. The matrix of cross-correlation between $\mathbf{y}_1$ and $\mathbf{y}_2$ is then determined;

$$M_{cov}\,(i,j) = \frac{1}{w-1}\left(\mathbf{y}_{1,i} - \bar{y}_1\right)^{\top}\left(\mathbf{y}_{2,j} - \bar{y}_2\right) \tag{3.13}$$

$$M_{cov}\,(i,j) = \begin{cases} \frac{M_{cov}(i,j)}{\sigma_1(i)\,\star\,\sigma_2(j)} & \sigma_1(i), \sigma_2(j) \neq 0 \\ 0 & otherwise \end{cases} \tag{3.14}$$

Anticor's algorithm then adjusts the matching quantities in the cross-correlation matrix to redistribute wealth in accordance with the mean reversion trading idea, shifting resources from outperforming to underperforming companies.

In instance, Anticor claims a transfer from asset i to j with the amount equal to the cross correlation value $M_{cori,j}$ minus their negative auto correlation values ($min\ \left\{0, M_{cor}\,(i,j)\right\}$) and ($min\ \left\{0, M_{cor}\,(j,j)\right\}$). In order to maintain the portfolio in the simplex domain, these transfer claims are normalised.

3.3.3.2 Passive–Aggressive Mean Reversion

Either every stock falls at the same time or only some stocks fall, but both scenarios are equally undesirable. Because it concentrates an excessive amount of wealth on "mine" stocks, such as Bear Stearns during the subprime crisis, actively rebalancing may not be the best strategy when confronted with such circumstances. This is due to the fact that it diversifies the portfolio less. It is in an investor's best interest to maintain the same portfolio they had before investing in "mine" stocks in order to reduce the likelihood of incurring losses as a result of the potential risks associated with those stocks. A novel online portfolio selection (OLPS) method called passive-aggressive mean reversion (PAMR) is based on mean reversion (Li et al., 2012). This algorithm derive the initial loss function of the mean reversion principle. This method, when combined with passive-aggressive online learning, can successfully take advantage of mean reversion (Crammer et al., 2006). When we look at PAMR's updating strategy, we see that it effectively reflects the mean reversion concept and strikes a balance between portfolio return and volatility risk. The PAMR method outperforms all benchmarks and nearly all cutting-edge algorithms on the majority of performance metrics.

3.3.3.3 Confidence-Weighted Mean Reversion

The confidence-weighted mean reversion (CWMR) algorithm modifies the distribution in a step-by-step manner so that it is consistent with the notion of mean reversion. This is accomplished by modelling the portfolio vector as a Gaussian distribution. This model was conceptualised by Dredze et al. and Crammer et al. with the use of the financial mean reversion principle and the online confidence-weighted (CW) machine learning method (Dredze et al., 2008; Crammer et al., 2008). Both of these studies were published that same year (2008). The closed form updates provided by CWMR provide unmistakable proof of the mean reversion trading idea as well as the interaction between first-order and second-order data.

This strategy in learns portfolios effectively apply the mean-reversion feature that is present in financial markets as well as the second-order knowledge that is associated with a portfolio. The CWMR update techniques are accomplished by finding solutions to two optimization problems that take into consideration first- and second-order information of a portfolio vector. This method is an improvement above any of the preceding methods, which solely considered first-order information.

3.3.3.4 Online Moving Average Reversion

Li et al. (2012) defined a multiple-period mean reversion called Moving Average Reversion, and proposed OnLine Moving Average Reversion (OLMAR) to take advantage of the multiple-period mean reversion, whereas PAMR and CWMR implicitly assume single-period mean reversion, resulting in one failure case on real data set.

The notion behind OLMAR is that the price vector associated with an $\mathbf{x}$ can be used to make implicit predictions about future prices: $\hat{\mathbf{p}}_{t-1} = \mathbf{p}_{t-1}$, where $\mathbf{x}$ has a price vector of $\mathbf{p}$. The severe single-period projections made in Li et al. (2012) may have been incorrect in certain cases because to the inherent problems of such predictions.

The authors instead presented a multiple period mean reversion, which explicitly predicts the future price vector as the moving average inside a window. They used the simple moving average, which is calculated as $MA_t = \frac{1}{w}\sum_{i=t-w+1}^{t} \mathbf{p}_i$. The following price related Li et al. (2012) becomes;

$$\hat{\mathbf{x}}_{t+1}(w) = \frac{MA_t(w)}{\mathbf{p}_t} = \frac{1}{w}\left(1 + \frac{1}{x_t} + ... + \frac{1}{\bigodot_{i=0}^{w-2} \mathbf{x}_{t-i}}\right) \tag{3.15}$$

for which w is the window size and $\bigodot$ is the element-wise product symbol. Next, they started using PAMR-like passive aggressive online learning Crammer et al. (2006) to acquire knowledge about a portfolio.

$$\mathbf{b}_{t+1} = \arg\min_{\mathbf{b}\,\varepsilon\,\Delta_m} \frac{1}{2}\ \|\mathbf{b}\ -\ \mathbf{b_t}\|^2 \ \text{s.t.}\ \mathbf{b}\ \cdot\ \mathbf{x}_{t+1}\hat{}\ \geq \varepsilon \tag{3.16}$$

3.3.3.5 Robust Median Reversion.

Existing mean reversion algorithms generally suffer from estimate errors, leading to non-optimal portfolios and poor performance in practise since they do not account for disturbances and outliers in the data. Huang et al. (2012) introduced a unique portfolio selection technique dubbed Robust Median Reversion for dealing with disturbances and outliers by exploiting mean reversion via a robust L1-median estimator (RMR).

3.3.4 Pattern-Matching based Approaches

The methods that fall under the Follow-the-Winner and Follow-the-Loser categories aren't the only ones that can use winners and losers; pattern matching strategies can, too. Specifically, this class includes non-parametric sequential investment strategies that are guaranteed to be consistent everywhere at all times, with trading rules that are growth optimal for every stationary and ergodic market process. The Follow-The-Winner strategy is not based on the optimality of BCRP like it is for the i.i.d. market.

Pattern-matching based approaches maximise the conditional expectation of log-return given prior observations by taking into account the non i.i.d. market (Algoet and Cover, 1988). In addition, the sequential prediction problem is also addressed by employing some of these methods (Biau et al., 2010).

3.3.4.1 Correlation-Driven Nonparametric Learning

At this point in time, the pattern-matching method, which is naturally straightforward, is capable of reaching its maximum level of performance. Nevertheless, identifying the days of the week that provide the most competitive rates is one of the aspects of this strategy that presents the greatest challenge. The Euclidean distance is frequently utilised by existing solutions in order to ascertain the degree of resemblance that exists between two preceding market periods. Euclidean distance, on the other hand, simply looks at the area around the most recent market windows and ignores whether or not there is a linear or nonlinear relationship between the two market windows. This relationship is essential for creating an accurate price relative estimate.

Euclidean distance, on the other hand, simply analyses the neighbourhood of the most recent market windows and ignores the linear or nonlinear relationship between two market windows, despite the fact that this relationship is essential for the assessment of price relativeness. Li et al. (2011a) proposed exploiting comparable patterns through a correlation coefficient, which effectively evaluates the linear relationship, and offered a novel pattern matching-based online portfolio selection technique called "CORrelation-driven Nonparametric learning." Li et al. (2011a) (CORN). At the moment, the proposed CORN is able to capture the linear relationship that exists between two market windows, and it is even possible to capture the nonlinear relationship that exists between them. This is an improvement over the previous implementation of the CORN, which was unable to capture either of these types of relationships. This new implementation of the CORN is an improvement over the one that came before it. It would be more attractive to develop a strategy that can manage the risk appropriately without cutting too much return than it would be to develop a strategy that can generate high returns but is frequently associated with high risk. Although high return strategies are frequently associated with high risk, it would be more attractive to develop a strategy that can manage the risk appropriately without cutting too much return. This is due to the fact that the process of building a strategy that can effectively control risk without sacrificing too much return would be more difficult than the process of developing a strategy that can create high returns despite frequently being linked with high risk.

3.3.4.2 Selection of Samples

The main goal of sample selection is to compare the previous market windows of the two price relatives in order to find historical price relatives that have samples that are similar. Assume that one of our objectives is to find the price relatives that are comparable to the next price relative, $\mathbf{x}_{t+1}$. The basic method is to go through all relative price vectors from the past $\mathbf{b}_i, i = \omega + 1, .., t$ and count $\mathbf{x}_i$ as comparable if the previous market window $\mathbf{x}_{i-\omega}^{i-1}$ is similar to the most recent market window $\mathbf{x}_{t-\omega+1}^{t}$.

In collection C, we keep track of the indices of close relatives based on how much they cost to buy. When comparing two market windows, which are both ω matrices, the concatenated $\omega \times m$-vector is a common measure of their similarity. Sample Selection $(C\ (\mathbf{x}_1^t, \omega))$ is elaborated upon in the **Algorithm** 2

Algorithm 2 Framework for sample selection ($C\ (\mathbf{x}_1^t,\ w)$)

```
Input: x_1^t: Historical market sequence; ω: window size;
Output: C: Index set of similar price relatives.
Initialize C = Ø;
if t ≤ ω + 1 then
    return;
end
for i = ω + 1, ω + 2, ..., t do
    if x_{i-ω}^{i-1} is similar to x_{t-ω+1}^t then
        C = C ∪ {i};
    end
end
```

3.3.5 Category of Meta-Learning Algorithms

Another topic of research in the realm of online portfolio selection is the MetaLearning Algorithm (MLA) (Das and Banerjee, 2011). In the field of machine learning, there is a close connection between it and the concept of expert learning (Cesa-Bianchi and Lugosi, 2006). This is especially important to keep in mind for "Funds of Funds," which are investment vehicles that sell portions of their portfolio to other funds. In most cases, MLA takes into account the opinions of a number of base experts, who may or may not come from the same strategy class. In preparation for the future period, each specialist generates a portfolio vector, which MLA combines in order to form a complete portfolio.

However, in situations in which it is difficult to draw a theoretical bound, the universal quality can be given to the entire MLA system by combining a universal strategy with a heuristic algorithm like Anticor, etc. In conclusion, MLA has the potential to incorporate every existing approach, providing a far broader scope of applicability. The majority of the online portfolio selection algorithms discussed above have a fascinating connection to the Capital Growth Theory. The Table 3.2 describe a brief summary, shown the connection of existing algorithms wit the Capital Growth theory.

Table 3.2: Algorithms Connection with Theory

Algorithms	$\hat{\mathbf{x}}_{t+1}$	**Prob.**	**Forms of Capital Growth**
BCRP	$\mathbf{x}_i, i = 1, .., n$	*1/n*	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} \frac{1}{n} \sum_{i=1}^{n} \log \mathbf{b} \cdot \mathbf{x}_i$
EG	$\mathbf{x}_t$	100%	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} \log \mathbf{b} \cdot \mathbf{x}_t - \lambda R(\mathbf{b}, \mathbf{b}_t)$
PAMR	$\frac{1}{\mathbf{x}_t}$	100%	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} \mathbf{b} \cdot \mathbf{x}_t + \lambda R(\mathbf{b}, \mathbf{b}_t)$
CWMR	$\frac{1}{\mathbf{x}_t}$	100%	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} P(\mathbf{b} \cdot \mathbf{x}_t) + \lambda R(\mathbf{b}, \mathbf{b}_t)$
OLMAR/RMR	$\frac{1}{\mathbf{x}_t}$	Eq. 3.17 & 3.18	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} \mathbf{b} \cdot \hat{\mathbf{x}}_{t+1} - \lambda R(\mathbf{b}, \mathbf{b}_t)$
$B^H/B^K/B^{NN}/CORN$	$\mathbf{x}_i, i \,\varepsilon\, C_t$	$\frac{1}{\lvert C_t \rvert}$	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} \frac{1}{\lvert C_t \rvert} \sum_{i \,\varepsilon\, C_t} \log \mathbf{b} \cdot \mathbf{x}_i$
B^{GV}	$\mathbf{x}_i, i \,\varepsilon\, C_t$	$\frac{1}{\lvert C_t \rvert}$	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} \frac{1}{\lvert C_t \rvert} \sum_{i \,\varepsilon\, C_t} (\log \mathbf{b} \cdot \mathbf{x}_i + \log C(\cdot))$
FTL	$\mathbf{x}_t \; i, , , , t$	$\frac{1}{t}$	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} \frac{1}{t} \sum_{i=1}^{t} \log \mathbf{b} \cdot \mathbf{x}_i$
FTRL	$\mathbf{x}_t \; i, , , , t$	$\frac{1}{t}$	$\mathbf{b}_{t+1} = \arg\max_{b \,\varepsilon\, \Delta_m} \frac{1}{t} \sum_{i=1}^{t} \log \mathbf{b} \cdot \mathbf{x}_i - \lambda R(\mathbf{b})$

3.4 Focus of This Study: Online Moving Average Reversion

Empirical studies show that extremes in a stock's price range, both up and down, are transient and that the mean reversion phenomena tends to be followed by stock price relatives. Even though existing mean reversion algorithms can achieve acceptable empirical performance on a wide range of real data sets, they usually assume a single-period mean reversion assumption that is not always satisfied, leading to subpar performance on some real data sets. This is due to the fact that the assumption relies on mean reversion occurring periodically. The assumption is based on the finding that mean reversion always occurs over the same time frame. Multiple-period mean reversion, also known as the moving average reversion (MAR), was proposed by Li et al. (2015), along with a new online portfolio selection (OLPS) strategy exploiting MAR using potent online learning techniques and given the term online moving average reversion (OLMAR).

3.4.1 Modelling in Online Moving Average Reversion (OLMAR)

This study mainly focuses on Online Moving Average Reversion (OLMAR) and proposes a decimated version called the D-OLMAR algorithm. Classifying multi-period mean reversion as "Moving Average Reversion" (MAR) is the basic idea. MAR uses moving averages to explicitly estimate future price movements and uses online learning approaches to improve portfolios. OLMAR is significant for being the first algorithm to use moving averages in online portfolio selection. Extensive studies on real markets empirically support OLMAR, which, despite its simplicity, demonstrates efficient updates. The findings show that OLMAR performs better than current algorithms in terms of cumulative wealth, indicating a significant benefit, especially with regard to preventing performance degradation in certain datasets such as DJA Borodin et al. (2004); **?**(Borodin et al., 2004; Li et al., 2012). OLMAR is also faster to execute than state-of-the-art techniques nowadays, which makes it ideal for large-scale applications.

The vast majority of formulations that are now on the market conform to the conventional practise of Kelly-based stock selection (Kelly, 1956b; Thorp, 1971). To provide a greater level of detail, a investment manager will make a projection for $\bar{X}_{t-1}$ based on k probable values, which include $\bar{X}^1{}_{t-1}, ..., \bar{X}^t{}_{t+1}$ in addition to the probabilities that are associated with them $(\rho 1, ..., \rho k)$. It is essential to bear in mind that each x I t+1 represents a unique alternative combination vector of individual price relative estimates. It is also crucial to keep this fact in mind. Constructing a portfolio with the objective of getting the best feasible predicted log return is the next step for him or her to take.

Considering the limitations discussed by Li et al. (2013, 2015); Hoi et al. (2021) of mean reversion algorithms, : proposed multiple time period mean reversion algorithms, named Moving Average Reversion (MAR). Wherein, OLMAR assume that $\bar{p}_{t+1} = \bar{p}_{t-1}$ will revert to moving average (MA), where MA_t represent the moving average consistency till the end of period t. Further, it observes the long-term trend and thus overcomes the drawbacks of existing mean reversion algorithms. A brief mathematical introduction of Simple Moving Average strategy proposed by Li et al. (2015), is as follows;

3.4.1.1 Simple Moving Average Reversion

$$\begin{aligned}\bar{x}_{t+1}(\omega) = \quad & \tfrac{\text{SMA}_t(\omega)}{\rho_t} = \tfrac{1}{\omega}\left(\tfrac{\rho_t}{\rho_t} + \tfrac{\rho_{t-1}}{\rho_t} + \cdots + \tfrac{\rho_{t-\omega+1}}{\rho_t}\right) \\ & = \tfrac{1}{\omega}\left(1 + \tfrac{1}{x_t} + \cdots + \tfrac{1}{\odot_{i=0}^{\omega-2} x_{t-1}}\right) \end{aligned} \tag{3.17}$$

where SMA is Simple Moving Average, ω is the window size, ρ is the price vector and relative change in prices are shown by x and $\odot$ denotes the element-wise product (Li et al., 2015).

OLMAR capture the Creamer's basic idea of passive aggressive (PA), Creamer (2007) to explain the moving average reversion. In simple words, PA incorporate the previous output if the classification found correct and adopt aggressively approaches in case of incorrect classification.

OLMAR Algorithm

The OLMAR optimization problem presented by Li et al. (2015) is given as;

$$b_{t+1} = \arg_{b \in \Delta_m} \min \frac{1}{2} \|b - b_t\|^2 \quad \textbf{s.t.} \quad b.\tilde{x}_{t+1} \geq \varepsilon \tag{3.18}$$

where $\tilde{x}_{t+1}$ is the next price relative which is inversely proportional to last price relative x_t. In particular, they implicitly assume that next price $\tilde{p}_{t+1}$ will revert to last price p_{t-1}, as follows;

$$\mathbf{\tilde{x}_{t+1}} = \frac{\mathbf{1}}{\mathbf{x_t}} \Rightarrow \frac{\mathbf{\tilde{p}_{t+1}}}{\mathbf{p_t}} = \frac{\mathbf{p_{t-1}}}{\mathbf{p_t}} \Rightarrow \mathbf{\tilde{p}_{t+1}} = \mathbf{p_{t-1}} \tag{3.19}$$

and b_t represents a portfolio vector which is an investment in the market for the $t-th$ period, i.e $\mathbf{b_t} = \mathbf{b_t}\left(\mathbf{x_1^{t-1}}\right)$ and $\mathbf{b_1^n} = (\mathbf{b_1} \cdots \mathbf{b_n})$ As earlier discussed, the basic idea of this study is the decimation process explored in the of study of de Cheveigné and Nelken (2019) due to challenges faced in the today voluminous trading of high speed and high frequency instruments.The extended version of this model was proposed by Ingber (2020) using multiple and sequential time period to avoid or minimize the human biases errors. Like lowpass filtering a signal and produce some samples from the population (Soleymani and Paquet, 2020). If this process continue without the low pass filtering concept, its called down sampling. This is the core motivation behind this study for decimation is to reduce the cost of processing, time saving and prompt response to the quick opportunities, which create in today fast moving market. Therefore, the use of a lower sampling rate usually results in a cheaper and quick implementation. We modified the OLMAR algorithm for decimation process. However, it does not ensure aliasing degree and produces a shift invariant signal representation (Unser, 1995; Li et al., 2002).

3.4.2 A Brief About the Decimated Sampling

Decimation is a technique used in signal processing when a signal's sample rate is decreased. This procedure is frequently used to fit the signal to the capabilities of a particular system, lower the volume of data, or simplify processing needs. Within the framework of the suggested D-OLMAR algorithm, decimation is essential to improving the algorithm's effectiveness and performance.

3.4.2.1 The Decimation Process's Filtering

Filter for Anti-Aliasing

An anti-aliasing filter is applied to the signal prior to downsampling to remove high-frequency elements that may induce aliasing. By keeping only the frequency content below the Nyquist frequency, the anti-aliasing filter helps to eliminate distortions and inaccuracies during subsequent downsampling. Filtering methods such as Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) can be employed. The filter type chosen is determined by the unique needs of the application as well as the desired qualities of the filtered signal. After filtering, the signal is downsampled by selecting a subset of samples. This means picking every nth sample, where 'n' is the desired sampling rate reduction. Downsampling reduces the data volume while keeping the fundamental features of the signal within the new sample rate.

Integration of the D-OLMAR Algorithm

An important methodological contribution to the decimation process is made by the proposed D-OLMAR algorithm. It is probable that D-OLMAR will introduce innovative methods for enhancing the downsampling process, considering factors like computational effectiveness, preserving signal characteristics, and flexibility in handling different kinds of signals. Resampling the initial ("undecimated") data at a lower sampling rate is referred to as "decimated" data. Decimated data will therefore have a lower sample rate than undecimated data. This algorithm, known as D-OLMAR, is expressed as follows:

3.4.3 Undervaluation of Performance Metrics Using Decimated Data

The results of this thesis demonstrate the D-OLMAR algorithm's potential as a useful real-time stock return optimisation technique. Interestingly, the technique makes use of the Lagrangian concept, which is necessary for optimising risk-adjusted matrices. The D-OLMAR model addresses concerns about the possible undervaluation of risk-adjusted metrics associated with decimated data, specifically the Sharpe and Calmar ratios, by applying Lagrangian optimisation. By taking into account risk considerations in a more sophisticated manner, the Lagrangian technique improves the accuracy of risk-adjusted returns. Thus, the observed outperformance is due to the algorithm's capacity to strategically optimise returns while controlling downside risk, rather than being exclusively attributed to downsampling effects. The robustness and integrity of the risk-adjusted measures are enhanced by the Lagrangian notion, which adds an additional layer of depth.

3.4.4 Proposed: D-OLMAR Algorithm

$$b_{t+1} = \arg_{b \in \Delta_m} \min \frac{1}{2} \|b - b_t\|^2 \quad \textbf{s.t.} \quad b.\tilde{y}_{t+1} \geq \varepsilon \tag{3.20}$$

where y_{t+1} is the decimated version of x_{t+1} which is $y_{t+1} = x((t+1)M)$ and $\tilde{y}_{t+1}$ is the moving average of y_{t+1}.

3.4.4.1 Assumption

We need the following optimizations for Decimated OLMAR algorithm.

1. The given stock data preserve the ergodicity property for a given window size.

2. The down sampling factor is chosen such that the first order moment remains constant.

3.4.4.2 Proof

The Lagrangian for the optimization problem of D-OLMAR is;

$$L(b, \lambda, \eta) = \frac{1}{2} \|b - b\|^2 + \lambda(\varepsilon - b \ . \ \tilde{y}_{t+1}) + \eta(b \ . \ 1 - 1) \tag{3.21}$$

where $\lambda \geq 0$ and η are the Lagrangian multipliers. Taking the gradient with respect to b and setting it to zero, we get;

$$0 = \frac{\delta L}{\delta b} = (b - bt) - \lambda \tilde{y}_{t+1} + \eta_1 \Rightarrow b = b_t + \lambda \tilde{y}_{t+1} - \eta_1 \tag{3.22}$$

Multiplying both sides by 1^T , we get

$$1 = 1 + \tilde{y}_{t+1} \ . \ 1 - \eta_m \Rightarrow \eta = \lambda \bar{y}_{t+1} \tag{3.23}$$

where $\bar{y}_{t+1}$ denotes the average predicted price relative (market). Plugging the above equation to the update of b, we get the update of b;

$$b = b_t + \lambda(\tilde{y}_{t+1} - \bar{y}_{t+1} \ . \ 1) \tag{3.24}$$

For Lagrangian multiplier, let us plug the above equation to the Lagrangian;

$$L(\lambda) = \lambda(\varepsilon - b_t \ . \ \tilde{y}_{t+1}) - \frac{1}{2}\lambda^2 \|\tilde{y}_{t+1} - \bar{y}_{t+1}\|^2 \tag{3.25}$$

Taking derivative with respect to and setting to zero we get;

$$\begin{aligned} 0 = \frac{\delta L}{\delta \lambda} &= (\varepsilon - b_t \ . \ \tilde{y}_{t+1}) - \lambda \|\tilde{y}_{t+1} - \bar{y}_{t+1}\|^2 \\ \Rightarrow \lambda &= \frac{\varepsilon - b_t \ . \ \tilde{y}_{t+1}}{\|\tilde{y}_{t+1} - \bar{y}_{t+1}\|^2} \end{aligned} \tag{3.26}$$

Further projecting λ to $[0; \infty]$, we get

$$\lambda = max \left\{ 0, \frac{\varepsilon - b_t \ . \ \tilde{y}_{t+1}}{\|\tilde{y}_{t+1} - \bar{y}_{t+1}\|^2} \right\} \tag{3.27}$$

3.5 Sample Size and Frequency

Table 3.3: Sample Size and Frequency

Sample Size	**Frequency**
10	daily, weekly, and monthly
20	daily, weekly, and monthly
30	daily, weekly, and monthly
40	daily, weekly, and monthly
50	daily, weekly, and monthly

The NASDAQ stock exchange has chosen to highlight the following five rapidly growing companies: Apple Inc. (AAPL), Amazon.com, Inc. (AMZN), Alphabet

Inc. (GOOG), Meta Platforms, Inc. (META), and Netflix, Inc. (NFLX). The rationales that led to the selection of these companies as the primary partner in the market, where the data at a high frequency and a vast scale available. Additionally, these firms has the effects that are detrimental to the market and theses data are highly volatile. And they transform the market from traditional based trading into one for the new era.

For thorough testing of D-OLMAR, a total of 2739 days' worth of data from Yahoo Finance were obtained, starting on May 21, 2012 and ending on October 4, 2023. The applied window size and frequencies are mentioned in the table 3.3 .

Chapter 4

Results and Discussion

4.1 Understanding the Data

The sample and measurements that make up a study's data set are summed up in a Table 4.1. The mean, median,mode skewness and kurtosis of all five data sets indicate that we can classify the underlying companies based on their size and frequency of trading volume. The following comparison summarized the characteristics and behaviors of the data.

NFLX has the highest mean value, 228.447, indicates that its stock price is, on average, higher than that of the other companies taken into consideration. When compared to the other firms, AMZN has the second-highest mean, which is 70.867, which indicates that the company has a higher average stock price. The mean for META is 148.658, which is lower than the mean for NFLX but higher than the mean for AAPL, AMZN, and GOOG. This indicates that the price of META's stock is higher than the average price of AAPL, AMZN, and GOOG but lower than the price of NFLX shares. With a mean of 62.230, AAPL has the lowest average of all the companies that were shown. GOOG has the lowest average stock price of all the companies, coming in at 57.833 dollars per share on average.

The central tendency of stock price distributions can be discerned by comparison of the medians of such distributions. The fact that NFLX has the highest median value, 184.210, indicates that the middle value of its stock price distribution is the greatest among the companies that are being supplied. When compared to AAPL, AMZN, and GOOG's stock prices, META's stock prices have a bigger middle value, as indicated by META's median value of 144.960, which is the second-highest value. The median for AMZN is 54.711, which is a higher value than those for AAPL and GOOG but a lower value than those for META and NFLX. When contrasted with META, NFLX, and AMZN, GOOG's median value of 49.609 indicates that the company's stock prices are closer to the lower end of the price range. AAPL has the lowest median of the companies offered, which comes in at 39.570.

This result is consistent for the skewness. Where AAPL's positive skewness value of 1.013 indicates that its distribution is right-skewed. This means that the distribution may contain a tail of larger values. AMZN's positive skewness of 0.58 indicates a right-skewed distribution, albeit one that is less pronounced than AAPL's. The positive skewness of 0.977 for GOOG indicates a moderately right-skewed distribution. The skewness of META is 0.564, which indicates a minor right-skewed distribution. NFLX has a positive skewness of 0.575, indicating that its distribution is also slightly right-skewed. All companies exhibit positive skewness, indicating that the tails of their stock price distributions extend toward higher values. Nonetheless, the magnitudes of skewness differ among the companies, with AAPL possessing the greatest skewness and META the least.

In this particular situation, the values of kurtosis suggest the following:

The fact that AAPL's distribution has a negative kurtosis value of -0.533 implies that its tails are more narrow than those of the normal distribution. Because of this, it is less probable that values at the extreme end of the scale will be observed.The kurtosis score for AMZN is -0.967, which shows even smaller tails than those seen in AAPL's data.GOOG has a kurtosis value of -0.057, which is somewhat less than zero and indicates that its tail-heavy distribution is relatively close to the normal distribution.META has a negative kurtosis of -0.192, which indicates that its tails are thinner than the normal distribution. This characteristic is shared by AAPL and AMZN. The negative value of -0.795 for NFLX's kurtosis indicates that the company's tails are thinner than the normal distribution would predict.All of the companies have a negative kurtosis, which shows that the tails of their stock price distributions are smaller than the normal distribution would predict. Regarding the behavior of the tails, however, the magnitudes of kurtosis varied among the firms, with AMZN having the highest negative kurtosis and GOOG having a kurtosis near to zero, which indicates a distribution that is closer to the normal distribution. GOOG's kurtosis is close to zero because its distribution is closer to the normal distribution. The line graphs shown in Figure 4.6 indicate the historical price movements of Apple Inc. (AAPL), Amazon.com, Inc. (AMZN), Alphabet Inc. (GOOG), Meta Platforms, Inc. (META), and Netflix, Inc. (NFLX) individually. It is concluded that the distribution is flat in comparison to the typical distribution, and hence it is classified as platykurtic.

TABLE 4.1: Descriptive Statistics

Sample: May 21, 2012 to October 4, 2023, 2739 Observations

	AAPL	**AMZN**	**GOOG**	**META**	**NFLX**
Mean	62.230	70.867	57.833	148.658	228.447
Median	39.570	54.711	49.609	144.960	184.210
Mode	24.335	11.253	18.014	26.850	51.871
Skewness	1.013	0.580	0.977	0.564	0.575
Kurtosis	-0.533	-0.967	-0.057	-0.192	-0.795

4.1.1 Parameters Evaluation

The referred tables above presents a comprehensive investigation of the stock data for AAPL, AMZN, GOOG, META, and NFLX from May 21, 2012, through October 4, 2023. This investigation makes it possible to do comparisons and analysis across a wide range of sample sizes and frequencies.

Dimensions of the Sample: The dimensions of the samples in the table range from 10 to 50. This makes it possible to conduct an investigation of how statistical measures change in response to growing amounts of data.

FIGURE 4.1: AAPL

FIGURE 4.2: AMZN

FIGURE 4.3: GOOG

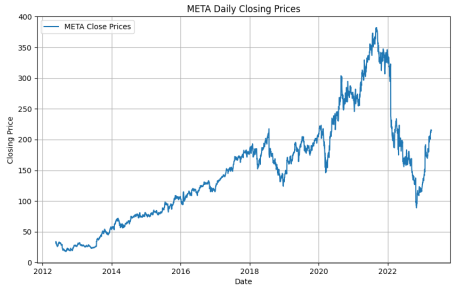

FIGURE 4.4: META

FIGURE 4.5: NFLX

FIGURE 4.6: Historical Price Movements of AAPL, AMZN, GOOG, META, and NFLX

Frequencies: The data are presented on a daily, a weekly, and a monthly basis, denoted by the letters D, W, and M respectively. This makes it possible to investigate how the statistics shift depending on the time intervals that are taken into consideration.

Mean Signal: The "Mean Signal" column provides a representation of the stock signal's typical value. When different sample sizes and frequencies are compared, the results provide information on the general trend and direction of the stock.

Decimated Mean Signal: Values that are included in the "Mean Signal" column are those that have been decimated. The method of reducing the number of data points while preserving the general characteristics is called decimation. In order to determine the influence that data reduction has on the typical behavior of stocks, it is helpful to compare the mean signal values before and after decimation.

Standard Deviation: The stock signal's standard deviation is presented in the "Std Signal" column of the table. It is a measure of how unpredictable or erratic the price movement of the item is likely to be. When standard deviations from different sample sizes and frequencies are compared with one another, patterns in the stock's stability or volatility may become apparent.

Standard Deviation with Decimal Placement: The "Std Signal" column, in addition to the "Mean Signal" column, includes numbers that have been decimated. By contrasting the standard deviation with and without decimation, one can determine how the decimation of data affects the volatility of stock prices.

Sharp Ratio: The "Sharp Ratio" column offers a risk-adjusted return measure. This is indicated by the column's name. It measures the additional return that was created for each additional point of risk. The performance of the stock can be evaluated in relation to its risk profile by comparing sharp ratios over a variety of sample sizes and frequencies. This makes it feasible to measure the stock's overall performance.

Decimated Sharp Ratio: Sharp Ratio After Decimation The "Sharp Ratio" column now also includes the values obtained after decimating the ratio. One can observe how the decimation of data impacts risk-adjusted returns by contrasting the results of sharp ratios calculated with and without decimation.

Calmar Ratio: The risk-adjusted return is computed in the "Calmar Ratio" column, which takes into account the largest drawdown experienced. It shows how well the stock performs in relation to its biggest loss, which is displayed as a percentage. Calmar ratios can be examined across a wide range of sample sizes and frequencies in order to evaluate the performance of the stock during times of significant decline.

Decimated Calmar Ratio: Calmar Ratio After Decimation The values that remain after the decimation process have been provided in the "Calmar Ratio" column. By contrasting Calmar ratios with and without decimation, one can determine the effect that data reduction has on the stock's risk-adjusted returns, which are determined by the greatest drawdown.

In general, this in-depth investigation of each stock provides important insights into its behavior across a range of sample sizes and frequencies, which can be used to make informed investment decisions. By making it possible to compare important statistical data, it makes it easier to investigate the performance of the stock, as well as its volatility, risk-adjusted returns, and level of stability.

4.2 Simulation and Results

Having proposed and derived the decimated version of OLMAR, i.e. D-OLMAR. It is appropriate to illustrate the effectiveness of the suggested algorithm using real-world stock market data. For this purpose, fast moving data set of Apple Inc. (AAPL), Amazon. com, Inc. (AMZN), Alphabet Inc. (GOOG), Meta Platforms, Inc. (META) and Netflix, Inc. (NFLX) are selected from NASDAQ stock exchange. For rigorous testing of D-OLMAR 2739 observations from May 21, 2012 to October 4, 2023 have been used. Various D-OLMAR results are shown in Table 4.2, 4.3, 4.4, 4.5 and 4.6 , decimated window size of daily, weekly and monthly with frequencies of 10, 20, 30, 40 and 50, details of the parameters are mentioned in Table 3.3.

4.2.1 Interpretation of Signal Strength

When using a window size of 50 and a monthly frequency, the greatest mean signal value is displayed, indicating a clear and significant overall rising tendency. When the size of the window is decreased, the mean signal values have a tendency to decrease as well. This suggests that the magnitude of the rising trend is diminished. The following expression exemplifies the standard deviation:

The values of the standard deviation that are the lowest are observed for window widths of 40 or 50 with a monthly frequency. This is indicative of lower price variability and volatility. When the size of the window is decreased, the standard deviation has a tendency to increase, which is indicative of larger price swings and volatility.

When using a window size of 50 and a monthly frequency, the Sharpe ratio values, which measure risk-adjusted returns, are at their maximum, which is an indication of good risk-adjusted performance. The risk-to-reward ratio, also known as the Calmar ratio, produces varying results depending on the window width and frequency being used. On the other hand, the results of applying a window size of 40 and a monthly frequency produce a Calmar ratio that is exceptionally high at 56.064. The following inferences and conclusions can be made on the basis of these findings:

In order to more accurately capture the overall trend, it is preferable to use a larger window size (40 or 50) in conjunction with a monthly frequency. This combination results in higher mean signal values and lower levels of volatility. Because it has better Sharpe ratio values, using a window size of 50 with a monthly frequency looks to be beneficial in terms of risk-adjusted metrics. This is because it has a higher Sharpe ratio. The Calmar ratio for a window size of 40 with a monthly frequency is outstanding, which indicates that there is the possibility for extremely favorable risk-adjusted returns. It is essential, however, to keep in mind that these conclusions were reached based on the information and analysis that were provided. The financial goals, level of comfort with risk, and length of time horizon are the factors that will determine whether or not a specific window size and frequency is appropriate.

4.2.2 Using AAPL Stock (Apple Inc.) to Make Predictions

The findings of the analysis may be seen in Table 4.2, which covers the time period from May 21, 2012 to October 4, 2023 and covers a variety of window widths and sample frequencies. In light of the criteria that have been presented, in the fallowing paragraph, a detailed analysis is conducted to determine which combination of sample frequency and window size is optimal for prediction and portfolio development. Figures 4.7 − 4.21 are displayed the price comparison between original and decimated signalling.

4.2.2.1 Measurements for Window 10

With a daily sample frequency, the mean decimated signal is 0.212, whereas the standard deviation of the decimated signal is 1.209. The sharp decimated ratio is 0.175, which indicates that the risk-adjusted return is moderate. Sharp decimated ratios are calculated using decimal places. The decimated ratio for Calmar is 0.079, which indicates that the company has a low risk-adjusted performance. When using a sample frequency of once per week, the standard deviation of the reduced signal drops to 1.185, while the mean signal drops to 0.367. The sharp decimated ratio goes up to 0.310, which indicates that the risk-return trade off has become more advantageous. The improvement in risk-adjusted performance is indicated by the fact that the Calmar decimated ratio has risen to 0.150. When using a monthly sample frequency, the mean decimated signal goes up to 0.679, but the standard deviation goes down to 1.100. The sharp decimated ratio goes up to 0.617, which indicates a higher return that is proportional to risk. The increased value of the Calmar decimated ratio, which now stands at 0.358, indicates improved performance relative to risk.

4.2.2.2 Measurements for Window 20

When using a sample frequency of once per day, the mean value of the decreased signal is 0.303, and the standard deviation is 1.311. It can be deduced from the fact that the sharp decimated ratio is 0.231 that the risk-adjusted return is not very high. The decimation ratio for Calmar is 0.094, which indicates that the company has a low risk-adjusted performance. The mean decimated signal goes up to 0.505 when the sample frequency is increased to once per week, while the standard deviation goes down to 1.237. The sharp decimated ratio improves to

0.409, which indicates that the risk-return trade off is in a more advantageous position. The increase in the Calmar decimated ratio to 0.165 is indicative of improved performance in relation to risk. When using a monthly sample frequency, the mean decimated signal goes up to 0.991, while the standard deviation goes down to 1.059. The sharp decimated ratio goes up to 0.936, which indicates a higher return that is proportional to risk. The increased value of the Calmar decimated ratio, which now stands at 0.547, indicates improved performance in relation to risk.

4.2.2.3 Measurements for Window 30

The mean value of the decreased signal while using a sampling frequency of once per day is 0.338, and the standard deviation is 1.345. It can be deduced from the fact that the sharp decimated ratio is 0.251 that the risk-adjusted return is not very high. The decimation ratio for Calmar is 0.095, which indicates that the company has a low risk-adjusted performance. With a weekly sample frequency, the mean decimated signal goes up to 0.677, while the standard deviation goes down to 1.252. The sharp decimated ratio goes up to 0.540, which indicates that the risk-return tradeoff has become more favorable. The increase in the Calmar decimated ratio to 0.197 is indicative of improved performance in relation to risk.

When using a monthly sample frequency, the mean decimated signal goes up to 1.309, while the standard deviation goes down to 0.887. The sharp decimated ratio improves to 1.476, which indicates an increased return that is proportional to the level of risk. The increase in the Calmar decimated ratio, which now stands at 2.133, is indicative of improved performance relative to risk.

4.2.2.4 Measurements for Window 40

The standard deviation of the decimated signal for a daily sample frequency is 1.350, while the mean decimated signal for the frequency is 0.370. It can be deduced from the fact that the sharp decimated ratio is 0.274 that the risk-adjusted

TABLE 4.2: AAPL-May 21, 2012 to Oct 4, 2023 (Total 2739 observations)

Sample Size	Frequency	Mean Signal	Decimated Mean Signal	Std Signal	Decimated Std Signal	Sharp Ratio	Decimated Sharp Ratio	Calmar Ratio	Decimated Calmar Ratio
10	D	0.212	0.212	1.209	1.210	0.175	0.175	0.079	0.079
10	W	0.212	0.367	1.209	1.185	0.175	0.310	0.079	0.150
10	M	0.212	0.679	1.209	1.100	0.175	0.617	0.079	0.358
20	D	0.309	0.303	1.311	1.311	0.236	0.231	0.096	0.094
20	W	0.309	0.505	1.311	1.237	0.236	0.409	0.096	0.165
20	M	0.309	0.991	1.311	1.059	0.236	0.936	0.096	0.547
30	D	0.347	0.338	1.345	1.346	0.258	0.251	0.098	0.095
30	W	0.347	0.677	1.345	1.252	0.258	0.540	0.098	0.197
30	M	0.347	1.309	1.345	0.887	0.258	1.476	0.098	2.133
40	D	0.385	0.370	1.350	1.351	0.285	0.274	0.103	0.099
40	W	0.385	0.797	1.350	1.225	0.285	0.650	0.103	0.314
40	M	0.385	1.526	1.350	0.825	0.285	1.850	0.103	56.064
50	D	0.431	0.411	1.350	1.351	0.319	0.304	0.114	0.109
50	W	0.431	0.874	1.350	1.206	0.319	0.725	0.114	0.312
50	M	0.431	1.793	1.350	0.710	0.319	2.526	0.114	-3.784

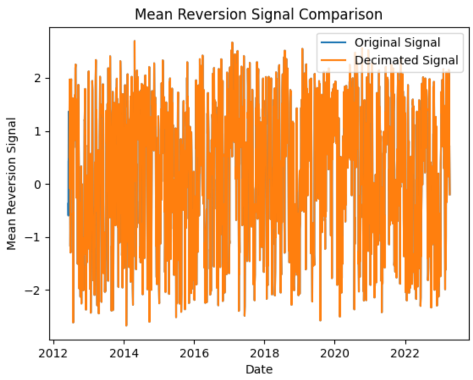

FIGURE 4.7: AAPL:10-Daily

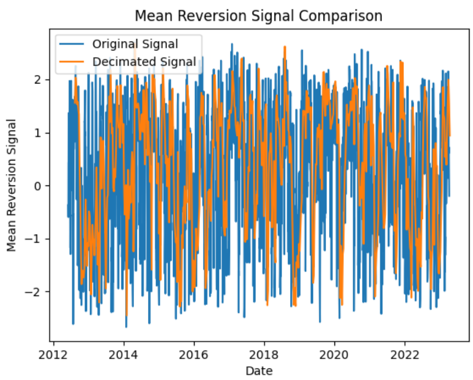

FIGURE 4.8: AAPL:10-Weekly

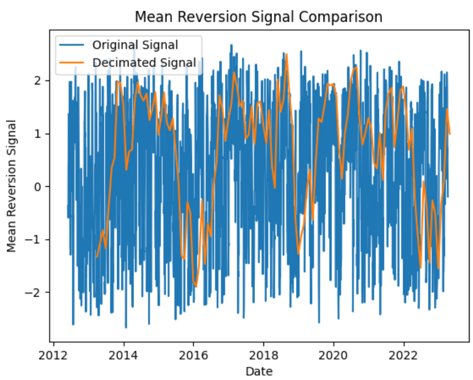

FIGURE 4.9: AAPL:10-Monthly

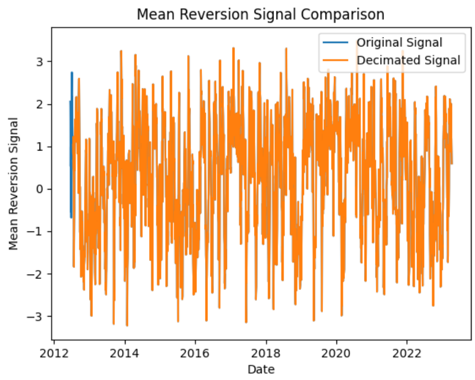

FIGURE 4.10: AAPL:20-Daily

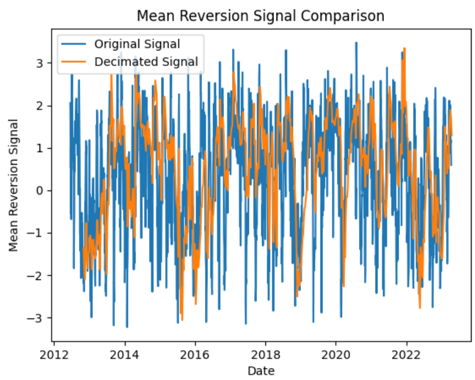

FIGURE 4.11: AAPL:20-Weekly

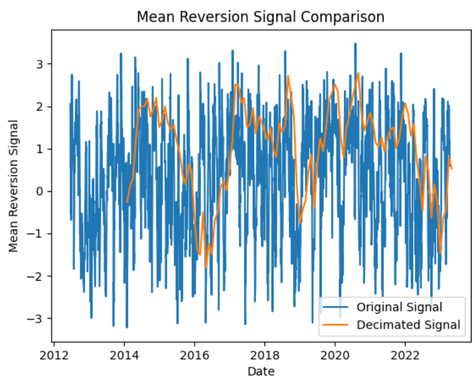

FIGURE 4.12: AAPL:20-Monthly

return is not very high. The decimation ratio for Calmar is 0.099, which indicates that the company has a low risk-adjusted performance.

The standard deviation of the decimated signal is reduced to 1.225, while the mean

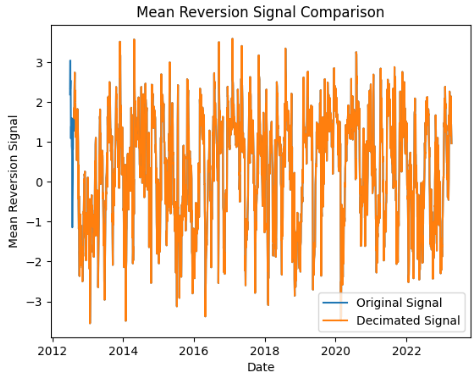

FIGURE 4.13: AAPL:30-Daily

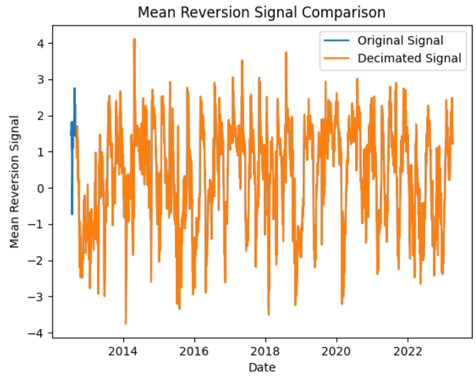

FIGURE 4.16: AAPL:40-Daily

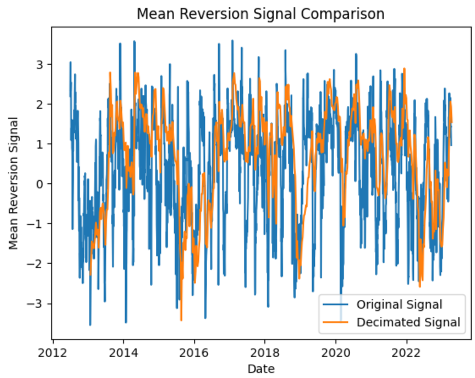

FIGURE 4.14: AAPL:30-Weekly

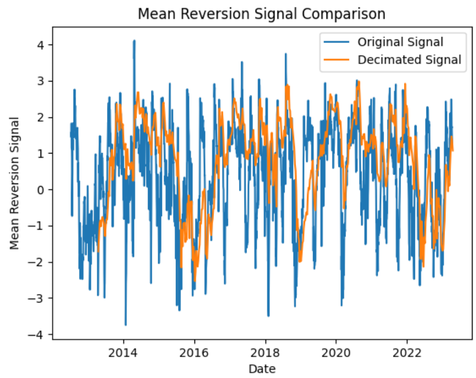

FIGURE 4.17: AAPL:40-Weekly

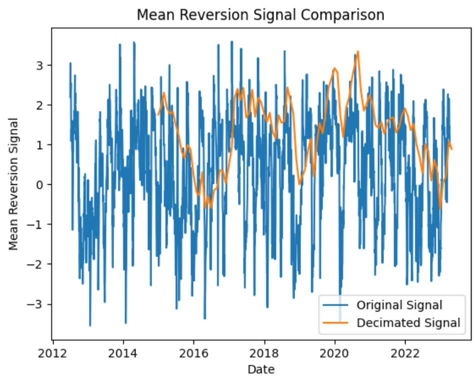

FIGURE 4.15: AAPL:30-Monthly

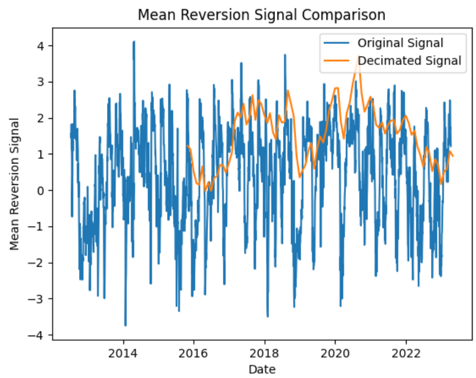

FIGURE 4.18: AAPL:40-Monthly

decimated signal rises to 0.797 when the sample frequency is weekly. The sharp decimated ratio goes up to 0.650, which indicates that the risk-return tradeoff has become more favorable. The increase in the Calmar decimated ratio to 0.314 is

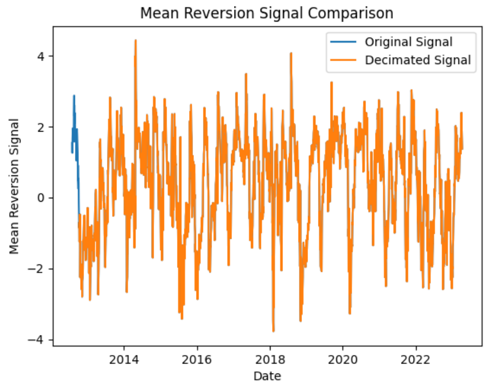

FIGURE 4.19: AAPL:50-Daily

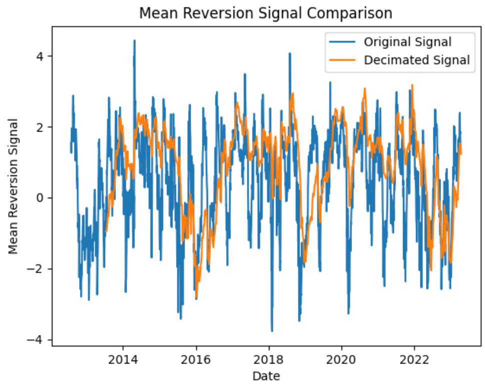

FIGURE 4.20: AAPL:50-Weekly

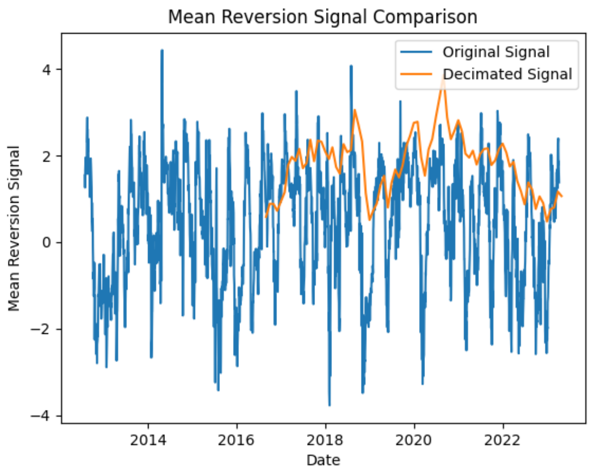

FIGURE 4.21: AAPL:50-Monthly

indicative of improved performance in relation to risk.

With a monthly sample frequency, the mean decimated signal rises to 1.526, while the standard deviation falls to 0.825. Both of these numbers are positive. A higher risk-adjusted return can be inferred from the sharp decimated ratio, which has increased to 1.850. A significant improvement in risk-adjusted performance is indicated by the steep jump in the Calmar decimated ratio to 56.064.

4.2.2.5 Measurements for Window 50

The standard deviation of the decimated signal with a daily sample frequency is 1.350, while the mean decimated signal is 0.411. The sharp decimated ratio comes in at 0.304, which indicates that the risk-adjusted return is on the moderate side. The decimated Calmar ratio is 0.109, which indicates poor performance while taking into account the level of risk. With a weekly sample frequency, the mean decimated signal goes up to 0.874, while the standard deviation goes down to 1.206. The sharp decimated ratio improves, reaching 0.725, which indicates that the risk-return tradeoff has become more advantageous. The increase in the Calmar decimated ratio to 0.312 is indicative of improved performance in relation to risk. In the case of a monthly sample frequency, the mean decimated signal goes up to 1.793, while the standard deviation goes down to 0.710. The sharp decimated ratio increases to 2.526, which indicates an increased return that is proportional to the level of risk. The decimation ratio for Calmar drops to -3.784, which is a sign of bad performance when risk is taken into account.

4.2.2.6 Summary

As a result of our investigation into the available variables, we have uncovered the following patterns:

When the window size is increased, there is a general trend toward an increase in the mean decimated signal, which indicates a stronger prediction signal. When the window size is increased, the standard deviation of the decimated signal typically decreases, which is suggestive of the possibility of a reduction in volatility. Increasing the window size typically results in improved Sharp decimated and Calmar decimated ratios, which indicates improved risk-adjusted performance. Monthly frequency has bigger mean decimated signals, lower standard deviations, and superior risk-adjusted performance as compared to daily and weekly frequencies. Additionally, monthly frequency has smaller standard deviations. According to these findings, it would appear that a larger window size (such as 40 or 50) combined with a monthly sample frequency would yield improved prediction accuracy and more opportunities for portfolio construction with AAPL stock.

4.2.3 Using AMZN Stock (Amazon.com, Inc.) to Make Predictions

The findings of the study are presented in Table 4.3 These findings contain data for a variety of window widths and sample rates that span the time period from May 21, 2012 to October 4, 2023. To see the results, you need only select the cell you're interested in from the table. In light of the requirements that were outlined, we are going to carry out research to ascertain the sample frequency and window size that provide the best opportunities for accurate forecasting and portfolio construction. Figures 4.22 − 4.36 are displayed the price comparison between original and decimated signalling.

4.2.3.1 Window Size 10

The following observations were made for an examination with a window size of 10:

The standard deviation of the decimated signal for the daily sample frequency is 1.184, while the mean signal for the daily frequency is 0.200. 0.169 is the acute decimated ratio, which represents a return that is proportionate to the amount of risk taken. The decimation ratio for Calmar is 0.073, which suggests that the company has poor performance when taking into account the risks involved. The average decimated signal goes up to 0.382, while the standard deviation goes down to 1.155 when the sample frequency increases to once every week. The ratio rises to 0.331, which indicates that the risk-reward balance has become more favorable. In addition, the Calmar decimated ratio goes up to 0.152, which is an indication of enhanced risk-adjusted performance. The mean decimated signal rises to 0.674, but the standard deviation falls to 1.117 as the monthly sample frequency increases. The severe decimated ratio improves to 0.603, which indicates an improvement in the return that is proportional to risk. Calmar's decimated ratio has increased to 0.279%, which indicates that the company's risk-adjusted performance has improved.

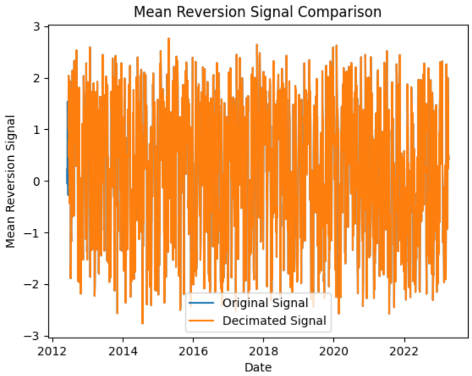

FIGURE 4.22: AMZN:10-Daily

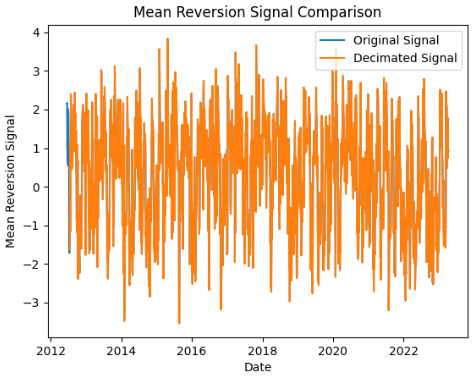

FIGURE 4.25: AMZN:20-Daily

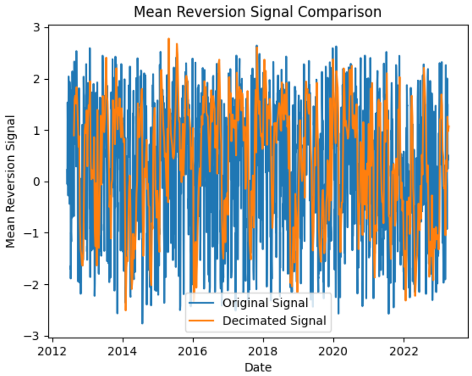

FIGURE 4.23: AMZN:10-Weekly

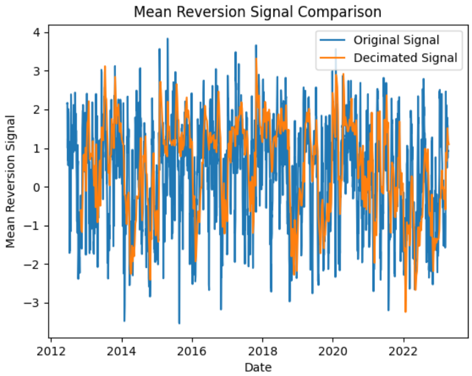

FIGURE 4.26: AMZN:20-Weekly

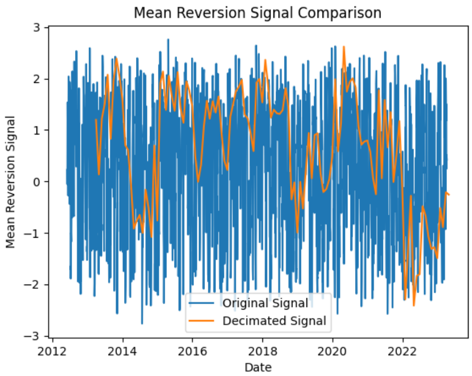

FIGURE 4.24: AMZN:10-Monthly

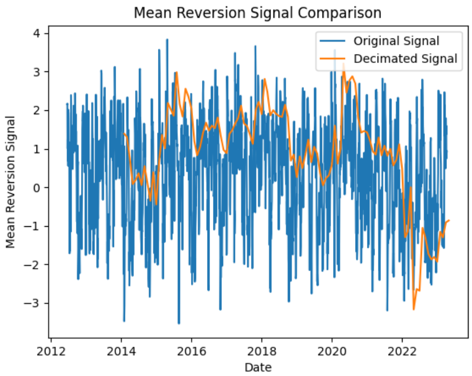

FIGURE 4.27: AMZN:20-Monthly

TABLE 4.3: AMZN-May 21, 2012 to Oct 4, 2023 (Total 2739 observations)

Sample Size	Freq-uency	Mean Signal	Decimated Mean Signal	Std Signal	Decimated Std Signal	Sharp Ratio	Decimated Sharp Ratio	Calmar Ratio	Decimated Calmar Ratio
10	D	0.200	0.199	1.184	1.185	0.169	0.168	0.073	0.072
10	W	0.200	0.382	1.184	1.155	0.169	0.331	0.073	0.152
10	M	0.200	0.674	1.184	1.117	0.169	0.603	0.073	0.279
20	D	0.308	0.304	1.294	1.294	0.238	0.235	0.087	0.086
20	W	0.308	0.495	1.294	1.226	0.238	0.404	0.087	0.153
20	M	0.308	0.899	1.294	1.271	0.238	0.707	0.087	0.284
30	D	0.372	0.368	1.320	1.322	0.282	0.278	0.088	0.087
30	W	0.372	0.614	1.320	1.270	0.282	0.484	0.088	0.185
30	M	0.372	1.238	1.320	1.275	0.282	0.971	0.088	0.506
40	D	0.418	0.403	1.325	1.326	0.315	0.304	0.105	0.102
40	W	0.418	0.731	1.325	1.300	0.315	0.562	0.105	0.215
40	M	0.418	1.412	1.325	1.106	0.315	1.277	0.105	0.829
50	D	0.461	0.441	1.328	1.330	0.347	0.331	0.129	0.123
50	W	0.461	0.837	1.328	1.330	0.347	0.629	0.129	0.238
50	M	0.461	1.481	1.328	1.016	0.347	1.458	0.129	1.174

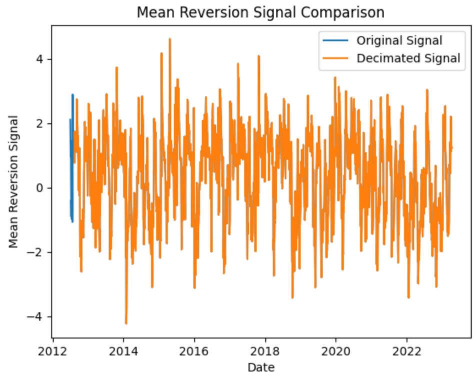

FIGURE 4.28: AMZN:30-Daily

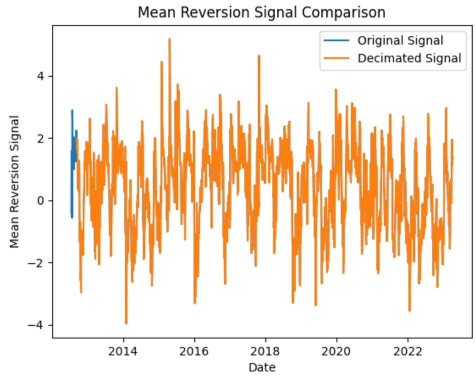

FIGURE 4.31: AMZN:40-Daily

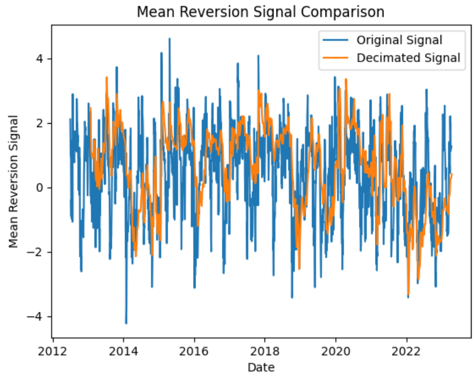

FIGURE 4.29: AMZN:30-Weekly

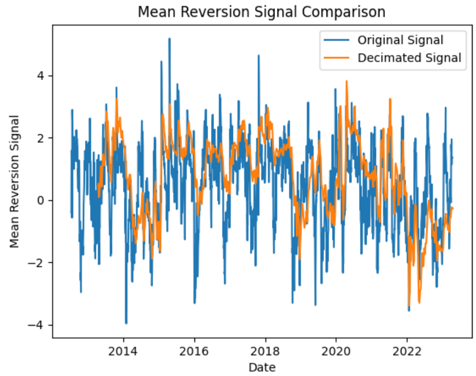

FIGURE 4.32: AMZN:40-Weekly

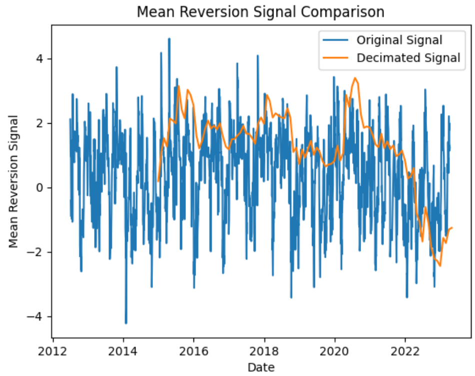

FIGURE 4.30: AMZN:30-Monthly

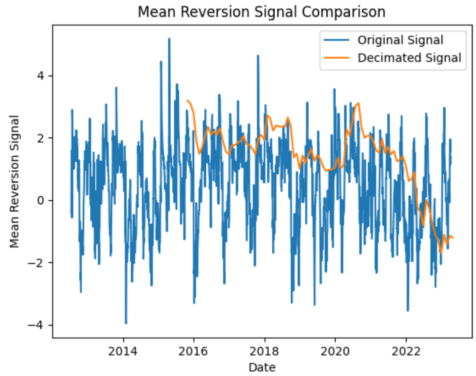

FIGURE 4.33: AMZN:40-Monthly

4.2.3.2 Window Size 20

Following are some of the observations that were made after conducting research on a window with 20 elements:

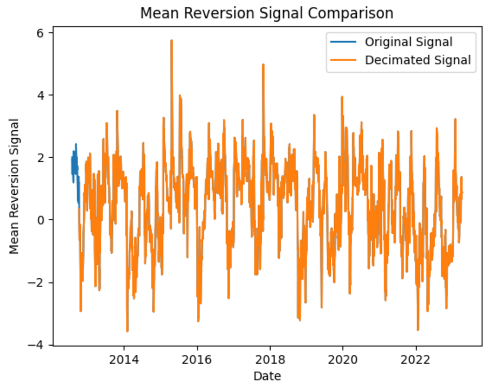

FIGURE 4.34: AMZN:50-Daily

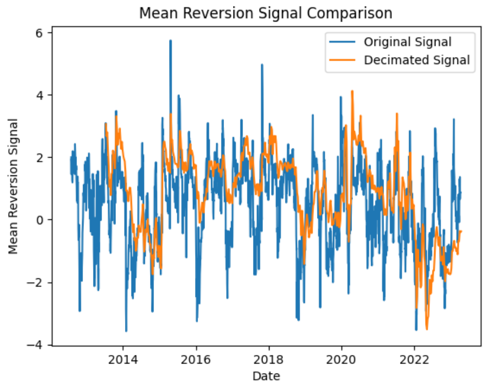

FIGURE 4.35: AMZN:50-Weekly

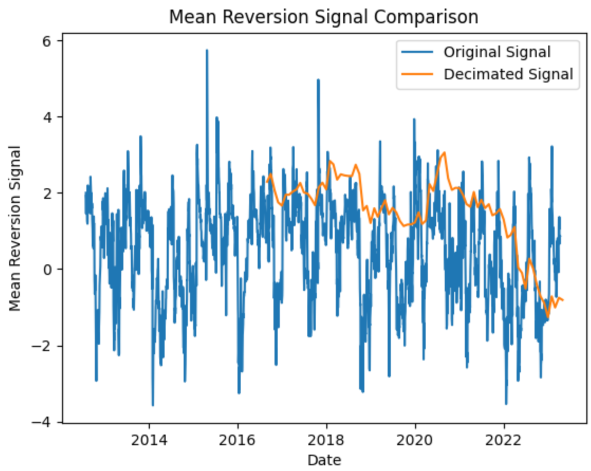

FIGURE 4.36: AMZN:50-Monthly

The standard deviation of the decimated signal for the daily sample frequency is 1.294, and the mean decimated signal for the daily sample frequency is 0.308. With a value of 0.238 for the severe decimated ratio, we can infer that the return will be moderate in comparison to the danger. The decimation ratio for Calmar is 0.086, which suggests that the company has poor performance when taking into account the risks involved. The standard deviation of the decimated signal is reduced to 1.226, while the mean decimated signal has increased to 0.495Frequency of Samples Taken Monthly: The increase in the decimated signal's mean brings it to 0.899, while the increase in the standard deviation brings it to 1.271. The severe decimated ratio goes up to 0.707

4.2.3.3 Window Size 30

When a window with a width of 30 pixels was investigated, the following findings were discovered:

The standard deviation of the decimated signal is 1.320, and the average decimated signal is 0.372. The daily sample frequency is once every day. With a value of 0.282 for the severe decimated ratio, we can deduce that the return will be moderate in comparison to the danger. The decimation ratio for Calmar is 0.087, which suggests that the company has poor performance when taking into account the risks involved. When using a weekly sample frequency, the mean decimated signal goes up to 0.614, while the standard deviation goes down to 1.270. The ratio rises to 0.484%, which indicates that the risk-return tradeoff has become more advantageous. In addition, the Calmar decimated ratio rises to 0.185, which is an indication of enhanced performance in relation to risk. The mean decimated signal has increased to 1.238, and the standard deviation has decreased to 1.275 as a result of the monthly sample frequency. The severe devastated ratio goes up to 0.971, which indicates a greater risk-adjusted yield on the investment. The decimated ratio for Calmar goes up to 0.506, which indicates that the company's risk-adjusted performance has improved.

4.2.3.4 Window Size 40

The subsequent observations were made for a window size of forty, and they are as follows:

The mean decimated signal for the daily sample frequency is 0.418, and the standard deviation is 1.325. The severe decimation ratio comes in at 0.315%, which indicates a moderate return in comparison to the risk. The decimation ratio for Calmar is 0.102, which suggests that the company has poor performance when risk is taken into account. The mean decimated signal has increased to 0.731, while the standard deviation has decreased to 1.300 as the weekly sample frequency increases. The ratio rises to 0.562, which indicates that the risk-reward balance has become more advantageous. In addition, the Calmar decimated ratio rises to 0.215%, which is an indication of greater performance in relation to risk. The mean decimated signal has increased to 1,412, while the standard deviation has decreased to 1,106 as a result of the monthly sample frequency. The severe decimated ratio goes up to 1.27, which indicates an increase in the return that is proportional to the danger. The fact that the Calmar decimated ratio has significantly increased, now standing at 0.829%, is indicative of improved risk-adjusted performance.

4.2.3.5 Window Size 50

The findings that were discovered when using a window size of 50 are as follows:

The standard deviation of the decimated signal is 1.328, and the average decimated signal is 0.461. The daily sample frequency is once per day. The severe decimated ratio comes in at 0.347%, which indicates a moderate return in comparison to the risk. The decimation ratio for Calmar is 0.123, which suggests that the company has poor performance when risk is taken into account. The mean decimated signal has increased to 0.837%, whereas the standard deviation has not changed from 1.330. The ratio rises to 0.629%, which indicates that the risk-reward balance has become more advantageous. In addition, the Calmar decimated ratio rises to 0.238, which is an indication of enhanced performance in relation to risk. Frequency of Samples Taken Monthly: While the mean of the decimated signal has increased to 1,481, the standard deviation has decreased to 1,016. The severe decimated ratio improves to 1.458, which implies an increased return that is proportional to the level of risk. The declining risk-adjusted performance is indicated by the decline in the Calmar decimated ratio, which has reached 1.174.

4.2.3.6 Summary

In conclusion, the trend toward a rise in the mean decimated signal with an increase in the window size indicates a stronger prediction signal. Generally speaking, a fall in the standard deviation of the decimated signal is indicative of a probable reduction in volatility. This phenomenon occurs as the window size grows. The Sharp decimated ratio and the Calmar decimated ratio both tend to get better as the window size gets larger, which suggests that the risk-adjusted performance is better. Additionally, as compared to the daily and weekly frequencies, the monthly frequency has a tendency to have larger mean decimated signals, lower standard deviations, and greater risk-adjusted performance. This is because monthly frequency signals are decimated more frequently before making any decisions regarding investments.

4.2.4 Using GOOG Stock (Alphabet Inc. (GOOG)) to Make Predictions

A detailed analysis of the data that was provided for GOOG stock from May 21, 2012 to October 4, 2023 is presented below, including values for each window size and frequency, reffered Table 4.4. Figures 4.37 − 4.51 are also displayed the price comparison between original and decimated signalling.

4.2.4.1 Window Size 10

The standard deviation of the decimated signal for the daily sample frequency is 1.187, and the mean decimated signal for the daily sample frequency is 0.208. The severe decimated ratio comes in at 0.175, which indicates a reasonable return in comparison to danger. The decimation ratio for Calmar is 0.075, which suggests that the company has poor performance when taking into account the risks involved.

The standard deviation of the decimated signal drops to 1.150, while the average decimated signal climbs to 0.371. Weekly Sample Frequency. The ratio rises to 0.323, which indicates that the risk-reward balance has become more favorable. The decimalized ratio of Calmar has not changed from its previous value of 0.075.

The average decimated signal goes up to 0.689, while the standard deviation goes down to 1.0 as the monthly sample frequency increases. The severe decimated ratio currently stands at 0.637%, indicating a higher return that is proportional to the amount of risk taken. The decimalized ratio of Calmar has not changed from its previous value of 0.075.

4.2.4.2 Window Size 20

Daily Sample Frequency: The standard deviation of the decimated signal is 1.276, whereas the average decimated signal is 0.309. The severe decimated ratio comes in at 0.242, which indicates a reasonable return in comparison to danger. The decimation ratio for Calmar is 0.088, which suggests that the company has poor performance when taking into account the risks involved.

The standard deviation of the decimated signal has decreased to 1.207, while the mean decimated signal has increased to 0.504. The ratio has improved to 0.418%, which indicates that the risk-reward balance has become more advantageous. The decimalized ratio of Calmar has not changed from its previous value of 0.088.

The mean decimated signal rises to 1.063, but the standard deviation falls to 1.094 as the monthly sample frequency increases. The severe devastated ratio goes up to 0.972, which indicates a greater risk-adjusted yield on the investment. The decimalized ratio of Calmar has not changed from its previous value of 0.088.

4.2.4.3 Window Size 30

Daily Sample Frequency: The standard deviation for the decimated signal is 1,299, whereas the average decimated signal is 0.356. The severe decimated ratio comes in at 0.274, which indicates a reasonable return in comparison to the danger. The decimated ratio for Calmar is 0.108, which suggests that the company has poor performance when risk is taken into account.

The standard deviation of the decimated signal decreases to 1.209, while the mean decimated signal increases to 0.654. The ratio goes up to 0.541, which indicates that the risk-return tradeoff is becoming more advantageous. The decimated ratio of Calmar has not changed from its previous value of 0.108.

The mean decimated signal increases to 1,376 while the standard deviation drops to 908, as measured by the Monthly Sample Frequency. A better risk-adjusted

Table 4.4: GOOG - May 21, 2012 to Oct 4, 2023 (Total 2739 observations)

Sample Size	Freq-uency	Mean Signal	Decimated Mean Signal	Std Signal	Decimated Std Signal	Sharp Ratio	Decimated Sharp Ratio	Calmar Ratio	Decimated Calmar Ratio
10	D	0.203	0.208	1.187	1.186	0.171	0.175	0.077	0.075
10	W	0.203	0.371	1.187	1.150	0.171	0.323	0.155	0.075
10	M	0.203	0.689	1.187	1.083	0.171	0.637	0.289	0.075
20	D	0.309	0.309	1.275	1.276	0.242	0.242	0.088	0.088
20	W	0.309	0.504	1.275	1.207	0.242	0.418	0.166	0.088
20	M	0.309	1.063	1.275	1.094	0.242	0.972	0.584	0.088
30	D	0.367	0.356	1.299	1.299	0.282	0.274	0.105	0.108
30	W	0.367	0.654	1.299	1.209	0.282	0.541	0.242	0.108
30	M	0.367	1.376	1.299	0.908	0.282	1.515	1.317	0.108
40	D	0.412	0.392	1.312	1.309	0.314	0.299	0.103	0.108
40	W	0.412	0.752	1.312	1.228	0.314	0.612	0.243	0.108
40	M	0.412	1.618	1.312	0.811	0.314	1.994	3.695	0.108
50	D	0.447	0.422	1.327	1.325	0.337	0.319	0.109	0.116
50	W	0.447	0.820	1.327	1.237	0.337	0.663	0.295	0.116
50	M	0.447	1.716	1.327	0.791	0.337	2.171	16.921	0.116

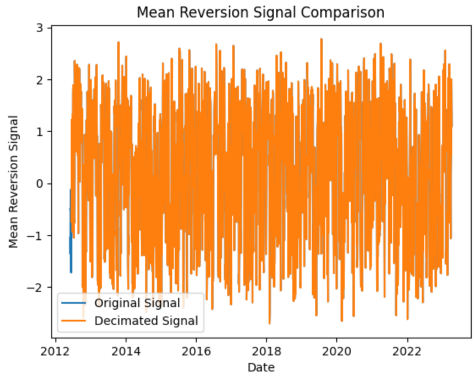

FIGURE 4.37: GOOG:10-Daily

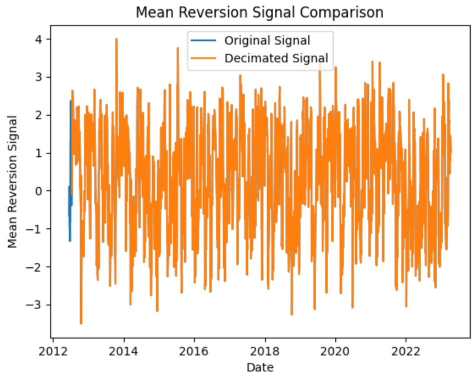

FIGURE 4.40: GOOG:20-Daily

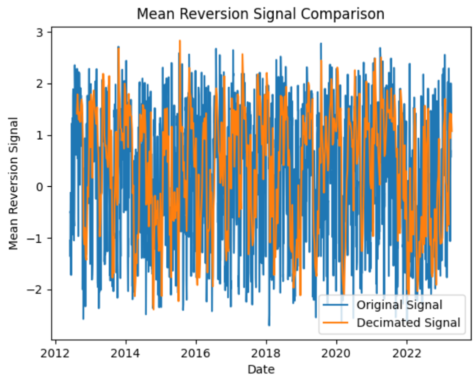

FIGURE 4.38: GOOG:10-Weekly

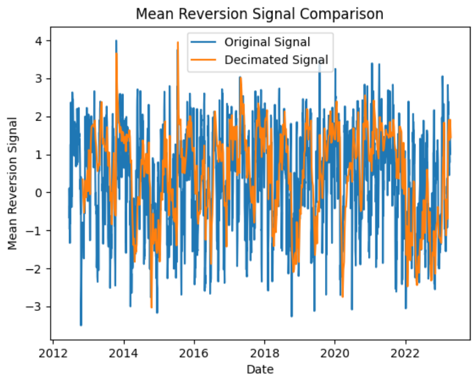

FIGURE 4.41: GOOG:20-Weekly

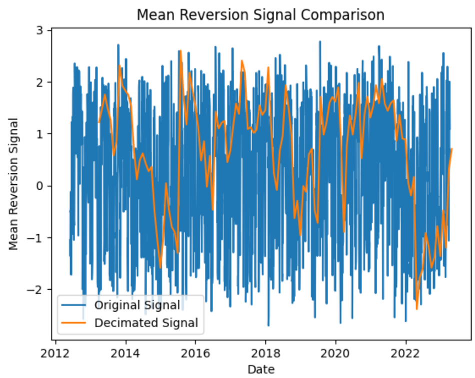

FIGURE 4.39: GOOG:10-Monthly

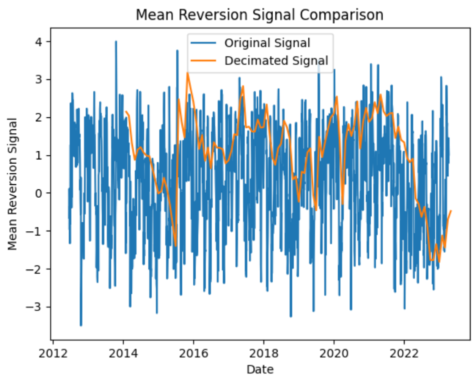

FIGURE 4.42: GOOG:20-Monthly

return can be inferred from the severe decimated ratio, which has increased to 1.515. The decimated ratio of Calmar has not changed from its previous value of 0.108.

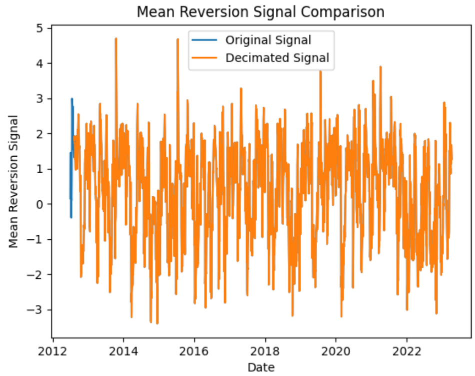

FIGURE 4.43: GOOG:30-Daily

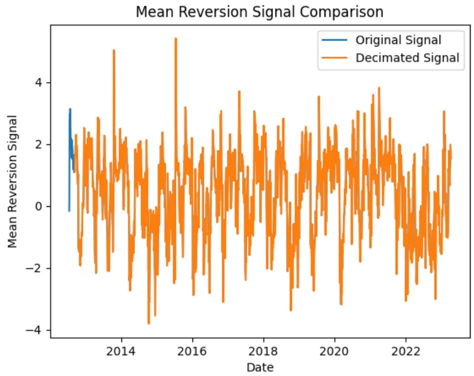

FIGURE 4.46: GOOG:40-Daily

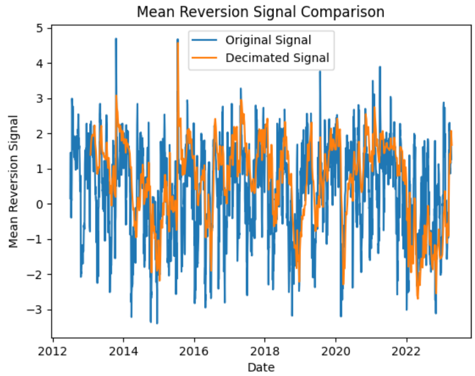

FIGURE 4.44: GOOG:30-Weekly

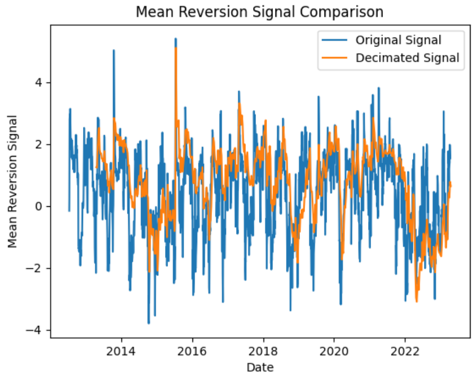

FIGURE 4.47: GOOG:40-Weekly

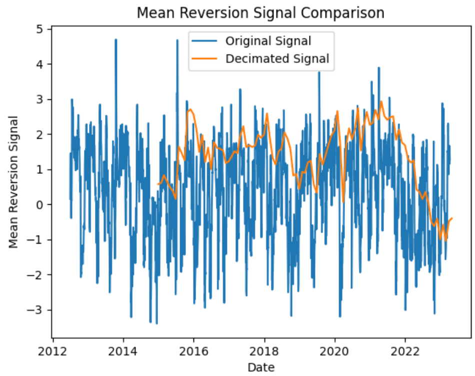

FIGURE 4.45: GOOG:30-Monthly

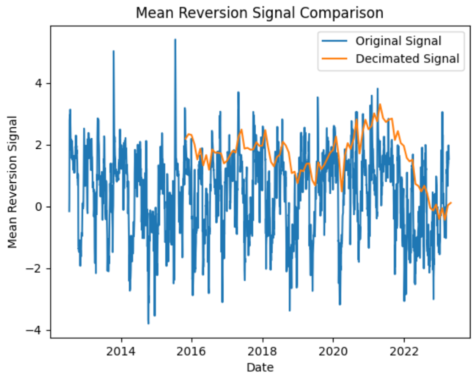

FIGURE 4.48: GOOG:40-Monthly

4.2.4.4 Window Size 40

The standard deviation of the decimated signal is 1.312, and the average decimated signal is 0.392. The daily sample frequency is once per day. 0.199 is the value of

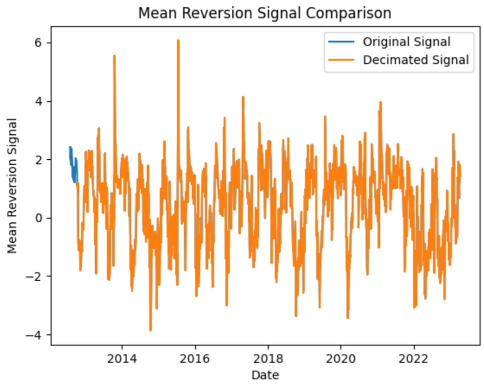

FIGURE 4.49: GOOG:50-Daily

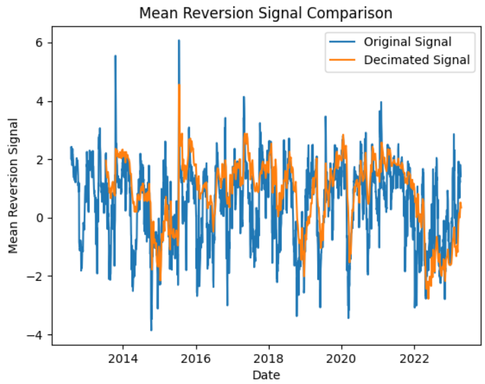

FIGURE 4.50: GOOG:50-Weekly

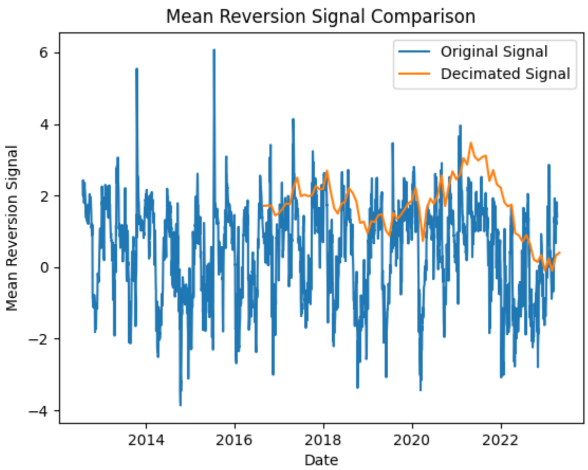

FIGURE 4.51: GOOG:50-Monthly

the severe decimated ratio, which indicates a moderate return in relation to risk. The decimated ratio for Calmar is 0.108, which suggests that the company has poor performance when risk is taken into account.

When using a sample frequency of once per week, the mean decimated signal reaches 0.752, while the standard deviation falls to 1.228. The ratio rises to 0.612, which indicates that the risk-reward balance has become more favorable. The decimated ratio of Calmar has not changed from its previous value of 0.108.

The mean decimated signal goes up to 1.618, while the standard deviation goes down to 0.811 with the monthly sample frequency. The severe decimated ratio is now at 1.994%, which indicates a higher return that is proportional to the amount of risk taken. The decimated ratio of Calmar has not changed from its previous value of 0.108.

4.2.4.5 Window Size 50

The mean decimated signal for the daily sample frequency is 0.422, and the standard deviation is 1.375. With a value of 0.319 for the severe decimated ratio, we can infer that the return will be moderate in comparison to the danger. The decimation ratio for Calmar is 0.116, which suggests that the company has poor performance when risk is taken into account.

The mean decimated signal goes up to 0.820, while the standard deviation goes down to 1.237 when the sample frequency is increased to once per week. When the ratio is improved to 0.663, it indicates that the risk-return tradeoff is more favorable. The decimalized ratio of Calmar has not changed from its previous value of 0.116.

The mean decimated signal goes up to 1.716%, while the standard deviation goes down to 0.75% with the monthly sample frequency. The severe devastated ratio currently stands at 2.171, which indicates a greater increase in return relative to danger. The decimalized ratio of Calmar has not changed from its previous value of 0.116.

4.2.4.6 Summary

In light of the findings of the research, the decision regarding the size of the prediction window and the frequency of its updates, as well as the development of the portfolio, is impacted by the unique goals and preferences of the investor. Higher mean decimated signals signify a stronger signal for predicting, while lower standard deviations reflect a potential reduction in volatility. Both of these findings are consistent with the hypothesis that lower volatility will result from the former. bigger values for the sharp decimated ratio and the Calmar decimated ratio suggest bigger rewards that are proportional to the level of risk.

In this scenario, the best window size and frequency for prediction and portfolio construction would depend on variables such as the intended risk-return trade-off, investment strategy, and time horizon. In addition, the optimal window size and frequency for prediction and portfolio construction would also depend on how frequently predictions are made. Daily and weekly sample frequencies are often inferior to monthly sample frequencies when it comes to mean decimated signals, standard deviations, and risk-adjusted performance. Monthly sample frequencies, on the other hand, typically demonstrate stronger mean decimated signals.

4.2.5 Using META Stock (Meta Platforms, Inc. (META)) to Make Predictions

A detailed study of the data that was provided for META stock between May 21, 2012 and October 4, 2023 is presented below, referred Table 4.5, including values for each window size and frequency. Figures 4.52 − 4.66 are also displayed the price comparison between original and decimated signalling.

4.2.5.1 Window Size 10

With a standard deviation of 1.183, the mean decimated signal for the Daily Sample Frequency is 0.168, and the mean value for the standard deviation is 1.183. 0.142 is the severe devastated ratio, which suggests a moderate return in comparison to risk. Calmar has a decimated ratio of 0.06, which indicates that the company has a poor risk-adjusted performance.

The mean decimated signal has increased to 0.346, while the standard deviation has decreased to 1.177 as the weekly sample frequency increases. The ratio goes up to 0.294, which indicates that the risk-return tradeoff is becoming more advantageous. The decimalized ratio of Calmar has not changed from its previous value of 0.060.

The mean decimated signal has increased to 0.708, while the standard deviation has decreased to 1.130 as the monthly sample frequency increases. The severe decimated ratio goes up to 0.627%, which is an indication of a larger return that is proportional to the risk taken. The decimalized ratio of Calmar has not changed from its previous value of 0.060.

4.2.5.2 Window Size 20

The mean decimated signal for the daily sample frequency is 0.267%, while the standard deviation is 1.276%. 0.210 is the acute decimated ratio, which represents a return that is proportionate to the amount of risk taken. The decimation ratio for Calmar is 0.071, which suggests that the company has poor performance when taking into account the risks involved.

When using a weekly sample frequency, the mean decimated signal goes up to 0.507, while the standard deviation goes down to 1.261. The ratio rises to 0.402, which indicates that the risk-reward balance has become more favorable. The decimalized ratio of Calmar has not changed from its previous value of 0.071.

The mean decimated signal goes up to 0.894, but the standard deviation goes down to 1.215 as the monthly sample frequency increases. The severe decimated ratio currently stands at 0.736%, indicating a higher return that is proportional to the amount of risk taken. The decimalized ratio of Calmar has not changed from its previous value of 0.071.

4.2.5.3 Window Size 30

The standard deviation of the decimated signal for the Daily Sample Frequency is 1.318, while the mean decimated signal for the Daily Sample Frequency is 0.339. The severe decimated ratio comes in at 0.257, which indicates a reasonable return in comparison to the danger. The decimation ratio for Calmar is 0.083, which suggests that the company has poor performance when taking into account the risks involved.

The standard deviation of the decimated signal is reduced to 1.2, but the mean decimated signal has increased to 0.659. The ratio rises to 0.520, which indicates

Table 4.5: META - May 21, 2012 to Oct 4, 2023 (Total 2739 observations)

Sample Size	Frequency	Mean Signal	Decimated Mean Signal	Std Signal	Decimated Std Signal	Sharp Ratio	Decimated Sharp Ratio	Calmar Ratio	Decimated Calmar Ratio
10	D	0.166	0.168	1.183	1.183	0.141	0.142	0.060	0.060
10	W	0.166	0.346	1.183	1.177	0.141	0.294	0.060	0.136
10	M	0.166	0.708	1.183	1.130	0.141	0.627	0.060	0.275
20	D	0.270	0.267	1.273	1.276	0.212	0.210	0.072	0.071
20	W	0.270	0.507	1.273	1.261	0.212	0.402	0.072	0.142
20	M	0.270	0.894	1.273	1.215	0.212	0.736	0.072	0.395
30	D	0.325	0.339	1.324	1.318	0.245	0.257	0.079	0.083
30	W	0.325	0.659	1.324	1.267	0.245	0.520	0.079	0.183
30	M	0.325	1.033	1.324	1.195	0.245	0.865	0.079	0.450
40	D	0.360	0.386	1.347	1.338	0.267	0.288	0.081	0.087
40	W	0.360	0.791	1.347	1.303	0.267	0.607	0.081	0.201
40	M	0.360	1.176	1.347	1.210	0.267	0.971	0.081	0.546
50	D	0.381	0.407	1.357	1.350	0.281	0.302	0.080	0.086
50	W	0.381	0.891	1.357	1.320	0.281	0.675	0.080	0.269
50	M	0.381	1.229	1.357	1.189	0.281	1.033	0.080	0.643

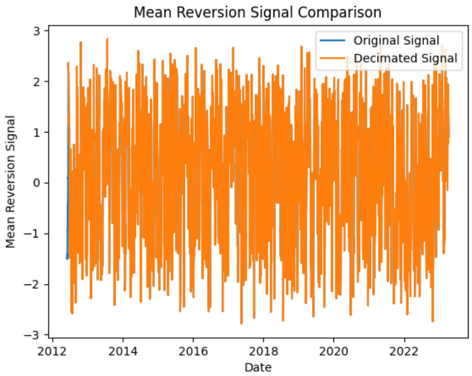

FIGURE 4.52: META:10-Daily

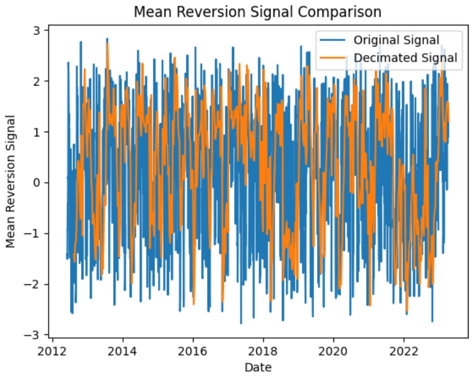

FIGURE 4.53: META:10-Weekly

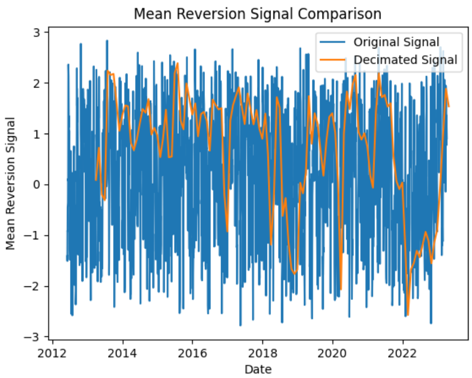

FIGURE 4.54: META:10-Monthly

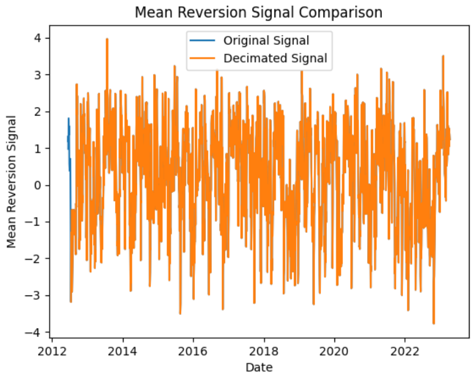

FIGURE 4.55: META:20-Daily

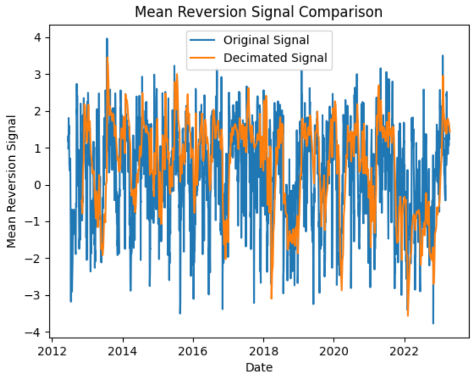

FIGURE 4.56: META:20-Weekly

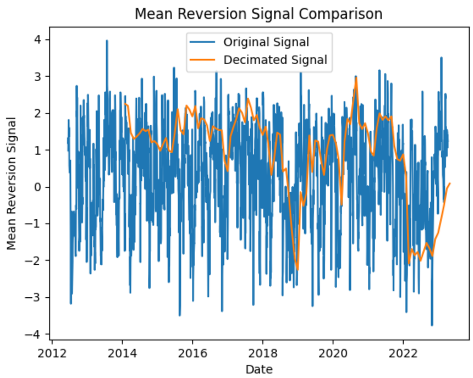

FIGURE 4.57: META:20-Monthly

that the risk-return tradeoff has become more advantageous. The decimalized ratio of Calmar has not changed from its previous value of 0.083.

The mean decimated signal rises to 1.033, but the standard deviation falls to 1.195

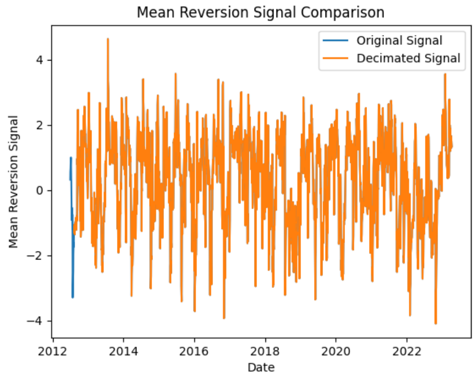

FIGURE 4.58: META:30-Daily

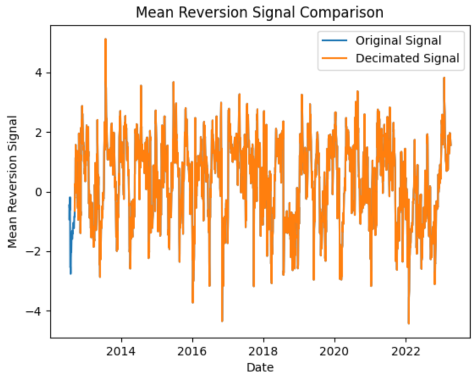

FIGURE 4.61: META:40-Daily

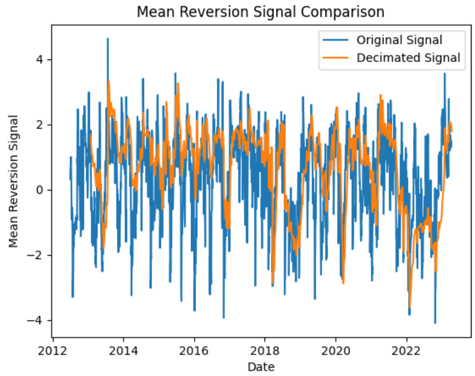

FIGURE 4.59: META:30-Weekly

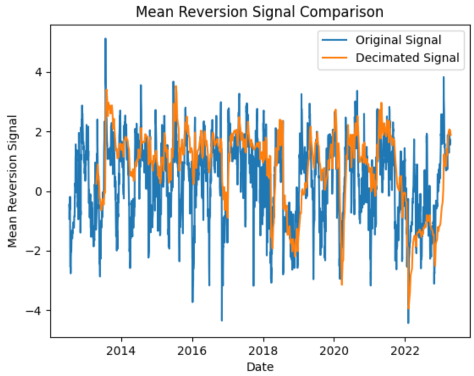

FIGURE 4.62: META:40-Weekly

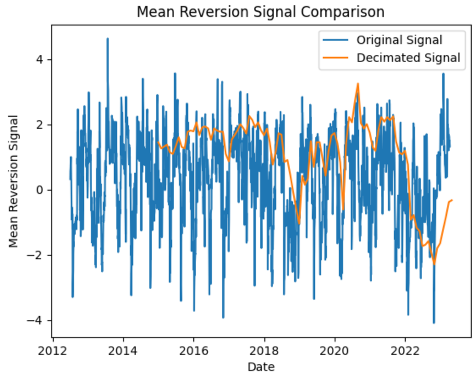

FIGURE 4.60: META:30-Monthly

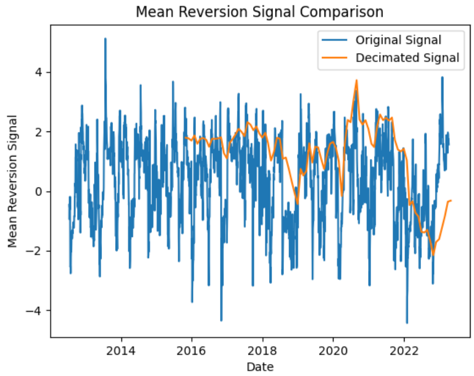

FIGURE 4.63: META:40-Monthly

as the monthly sample frequency increases. The severe devastated ratio improves to 0.866, which indicates an improvement in the return that is proportional to risk. The decimalized ratio of Calmar has not changed from its previous value of

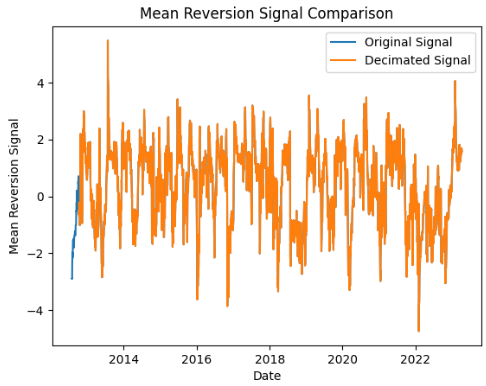

FIGURE 4.64: META:50-Daily

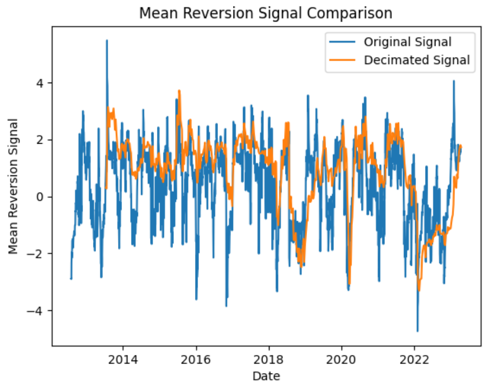

FIGURE 4.65: META:50-Weekly

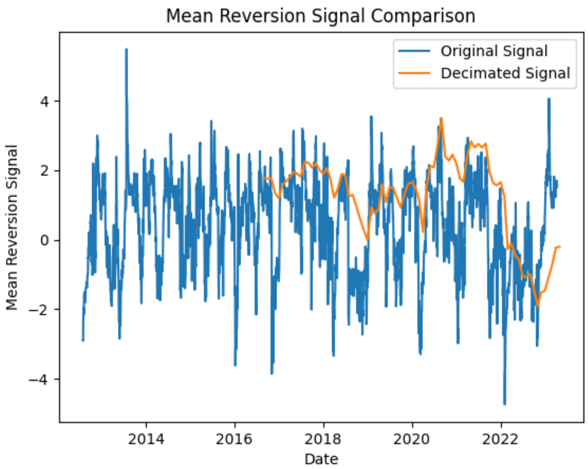

FIGURE 4.66: META:50-Monthly

0.083.

4.2.5.4 Window Size 40

The standard deviation of the decimated signal for the daily sample frequency is 1.338, and the mean decimated signal for the daily sample frequency is 0.386. The severe decimated ratio comes in at 0.288, which indicates a reasonable return in comparison to danger. The decimation ratio for Calmar is 0.087, which suggests that the company has poor performance when taking into account the risks involved.

The standard deviation of the decimated signal has decreased to 1.303, while the mean decimated signal has increased to 0.791. The ratio goes up to 0.607%, which indicates that the risk-return tradeoff is becoming more advantageous. The decimalized ratio of Calmar has not changed from its previous value of 0.087.

The mean decimated signal has increased to 1.176, while the standard deviation has decreased to 1.210 as the monthly sample frequency increases. The severe devastated ratio goes up to 0.971, which indicates a greater risk-adjusted yield on the investment. The decimalized ratio of Calmar has not changed from its previous value of 0.087.

4.2.5.5 Window Size 50

The standard deviation of the decimated signal is 1.350, and the daily sample frequency has an average decimated signal of 0.407. The severe devastated ratio is 0.302, which indicates a reasonable return in comparison to the danger. The decimation ratio for Calmar is 0.086, which suggests that the company has poor performance when taking into account the risks involved.

The mean decimated signal has increased to 0.891, while the standard deviation has decreased to 1.320 as the weekly sample frequency increases. The ratio rises to 0.675%, which indicates that the risk-reward balance has become more advantageous. The decimated ratio of Calmar has not altered from its previous value of 0.086%.

The mean decimated signal has increased to 1,229, while the standard deviation has decreased to 1,189 as a result of the Monthly Sample Frequency. The ratio of returns that were drastically decreased rises to 1.033, showing a larger return that is proportional to the level of risk. The decimated ratio of Calmar has not altered from its previous value of 0.086

4.2.5.6 Summary

The window size and frequency should be determined according to the unique goals and preferences of the investor if one want to achieve optimal prediction and portfolio creation. Signals with a higher mean decimated value are suggestive of better prediction signals, however standard deviations that are lower may imply lower levels of volatility.

Bigger values for the sharp decimated ratio and the Calmar decimated ratio suggest bigger rewards that are proportional to the level of risk. It is desirable to have greater mean decimated signals, lower standard deviations, and higher sharp decimated ratios and Calmar decimated ratios while optimizing a portfolio.

According to the findings of the study, the following combinations of window size and frequency are the most effective in terms of prediction and portfolio optimization:

Size of a Window with a 50-Inch Opening and a Monthly Sampling Frequency: This choice provides the decimated signal with the highest mean value (1,229) and the lowest standard deviation (1,189). In addition, the steep decimated ratio of 1.033 that it possesses demonstrates that it offers an excellent return that is proportional to risk. The decimated ratio of Calmar has not altered from its previous value of 0.086

4.2.6 Using NFLX Stock (Alphabet Inc. (Netflix, Inc. (NFLX)) to Make Predictions

The following is a comprehensive analysis of the data for NFLX stock between the dates of May 21, 2012 and October 4, 2023, including values for each window size and frequency, referred Table 4.6. Figures 4.67 − 4.81 are also displayed the price comparison between original and decimated signalling.

4.2.6.1 Window Size 10

With a standard deviation of 1.199, the mean decimated signal for the Daily Sample Frequency is 0.180, and it has a variance of 1.199. 0.150 is the acute decimated ratio, which represents a return that is proportionate to the amount of risk taken. The decimation ratio for Calmar is 0.065, which suggests that the company has poor performance when taking into account the risks involved.

The average decimated signal goes up to 0.398, but the standard deviation goes down to 1.164 when the sample frequency increases to once every week. The ratio rises to 0.342, which indicates that the risk-reward balance has become more favorable. The decimalized ratio of Calmar has not changed from its previous value of 0.065.

The average decimated signal goes up to 0.671, while the standard deviation goes down to 1.0 as the monthly sample frequency increases. The fraction of returns that were severely reduced rises to 0.615%, indicating a larger return that is proportional to the level of risk. The decimalized ratio of Calmar has not changed from its previous value of 0.065.

4.2.6.2 Window Size 20

The standard deviation of the decimated signal is 1.304, and the average decimated signal is 0.273. The daily sample frequency is once per day. The severe decimation ratio comes in at 0.209%, which indicates a moderate return in comparison to the risk. The decimated ratio for Calmar is 0.069, which suggests that the company does not do well while taking into account its level of risk.

Using a weekly sample frequency results in an increase in the mean decimated signal to 0.596, while a reduction in the standard deviation brings the value to 1.230. When the ratio is improved to 0.48, it indicates that the risk-return tradeoff is more advantageous. The decimalized ratio of Calmar has not changed from its previous value of 0.069.

The mean decimated signal has increased to 0.800, while the standard deviation has decreased to 1.243 as the monthly sample frequency increases. The severe devastated ratio improves to 0.708, which indicates an improvement in the return that is proportional to risk. The decimalized ratio of Calmar has not changed from its previous value of 0.069.

4.2.6.3 Window Size 30

The standard deviation of the decimated signal is 1.333, while the average decimated signal is 0.341. The daily sample frequency is once per day. 0.255 is the acute decimated ratio, which represents a return that is proportionate to the amount of risk taken. The decimation ratio for Calmar is 0.077, which suggests that the company has poor performance when taking into account the risks involved.

The standard deviation of the decimated signal has decreased to 1.277%, while the mean decimated signal has increased to 0.675%. The ratio rises to 0.528, which indicates that the risk-return tradeoff has become more advantageous. The decimalized ratio of Calmar has not changed from its previous value of 0.077.

The mean decimated signal has increased to 1.095, while the standard deviation has decreased to 1.251 as the monthly sample frequency increases. The severe decimated ratio improves to 0.87, which indicates an increased return that is proportional to the level of risk. The decimalized ratio of Calmar has not changed from its previous value of 0.077.

4.2.6.4 Window Size 40

Daily Sample Frequency: The standard deviation of the decimated signal is 1.348,

while the average decimated signal is 0.426. 0.316 is the value of the severe decimation ratio, which indicates a moderate return in relation to risk. The decimation ratio for Calmar is 0.088, which suggests that the company has poor performance when taking into account the risks involved.

When using a weekly sample frequency, the mean decimated signal goes up to 0.724, while the standard deviation goes down to 1.254. The ratio rises to 0.578, which indicates that the risk-reward balance has become more advantageous. The decimalized ratio of Calmar has not changed from its previous value of 0.088.

The mean decimated signal has increased to 1.193, while the standard deviation has decreased to 1.221 as a result of the monthly sample frequency. The severe

Table 4.6: NFLX - May 21, 2012 to Oct 4, 2023 (Total 2739 observations)

Sample Size	Freq-uency	Mean Signal	Decimated Mean Signal	Std Signal	Decimated Std Signal	Sharp Ratio	Decimated Sharp Ratio	Calmar Ratio	Decimated Calmar Ratio
10	D	0.177	0.180	1.198	1.199	0.148	0.150	0.064	0.065
10	W	0.177	0.398	1.198	1.164	0.148	0.342	0.064	0.147
10	M	0.177	0.671	1.198	1.091	0.148	0.615	0.064	0.331
20	D	0.280	0.273	1.304	1.304	0.215	0.209	0.071	0.069
20	W	0.280	0.596	1.304	1.230	0.215	0.485	0.071	0.185
20	M	0.280	0.880	1.304	1.243	0.215	0.708	0.071	0.296
30	D	0.341	0.341	1.343	1.338	0.254	0.255	0.077	0.077
30	W	0.341	0.675	1.343	1.277	0.254	0.528	0.077	0.235
30	M	0.341	1.095	1.343	1.251	0.254	0.876	0.077	0.416
40	D	0.411	0.426	1.348	1.348	0.305	0.316	0.085	0.088
40	W	0.411	0.724	1.348	1.254	0.305	0.578	0.085	0.290
40	M	0.411	1.193	1.348	1.221	0.305	0.977	0.085	0.546
50	D	0.492	0.512	1.351	1.349	0.364	0.379	0.098	0.102
50	W	0.492	0.767	1.351	1.259	0.364	0.609	0.098	0.283
50	M	0.492	1.284	1.351	1.274	0.364	1.007	0.098	0.636

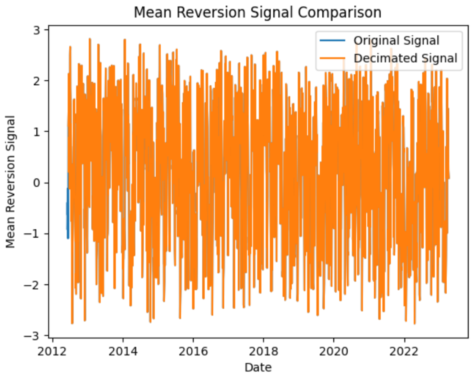

FIGURE 4.67: NFLX:10-Daily

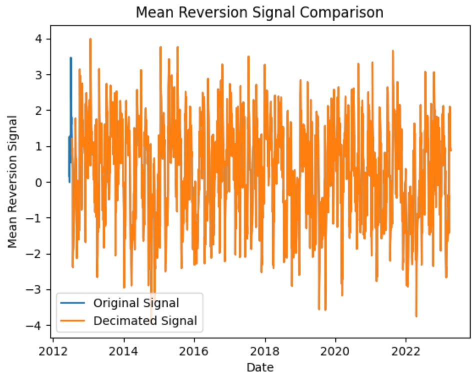

FIGURE 4.70: NFLX:20-Daily

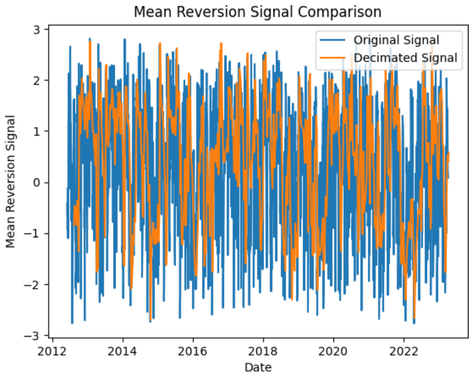

FIGURE 4.68: NFLX:10-Weekly

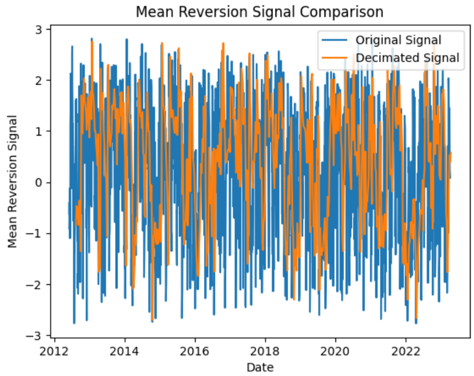

FIGURE 4.71: NFLX:20-Weekly

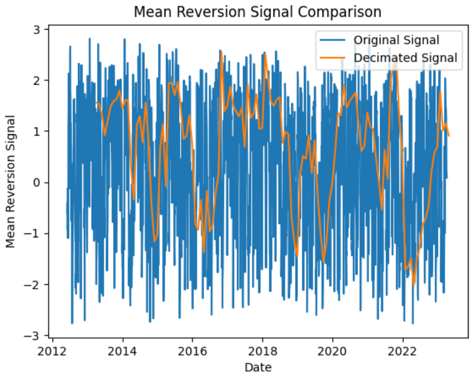

FIGURE 4.69: NFLX:10-Monthly

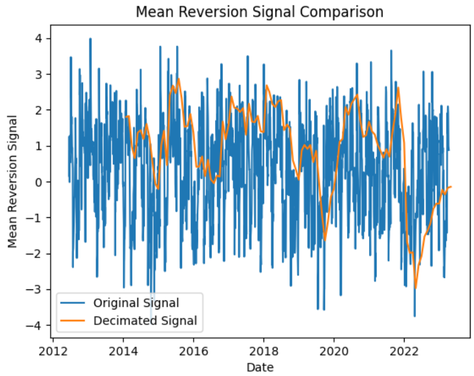

FIGURE 4.72: NFLX:20-Monthly

decimated ratio improves to 0.97, which indicates an increase in the return that is proportional to the danger. The decimalized ratio of Calmar has not changed from its previous value of 0.088.

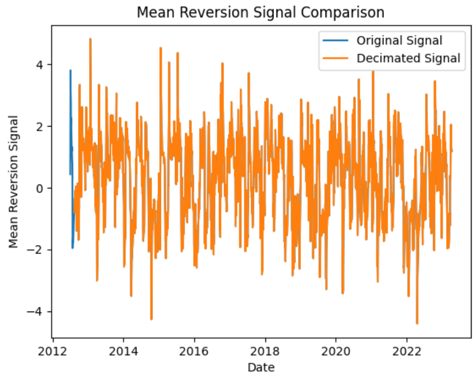

FIGURE 4.73: NFLX:30-Daily

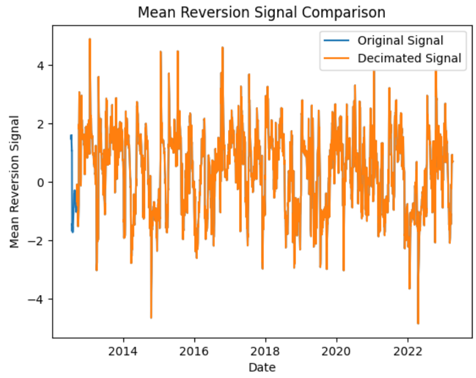

FIGURE 4.76: NFLX:40-Daily

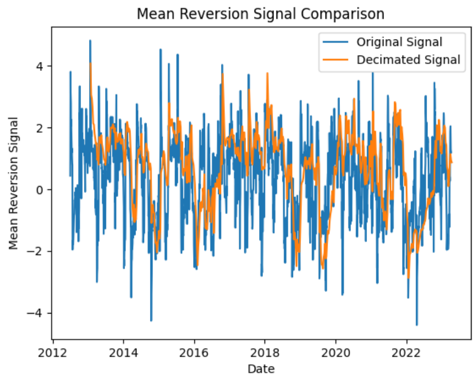

FIGURE 4.74: NFLX:30-Weekly

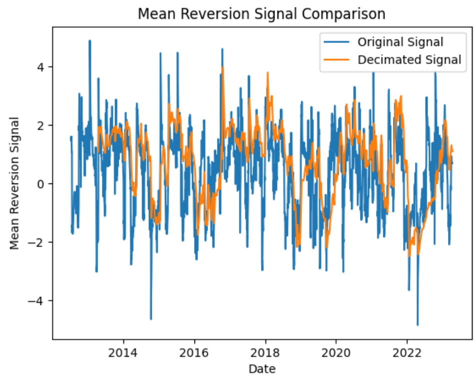

FIGURE 4.77: NFLX:40-Weekly

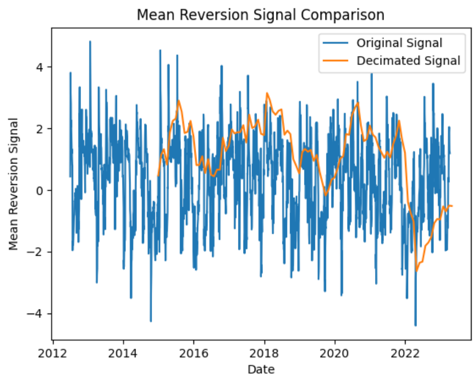

FIGURE 4.75: NFLX:30-Monthly

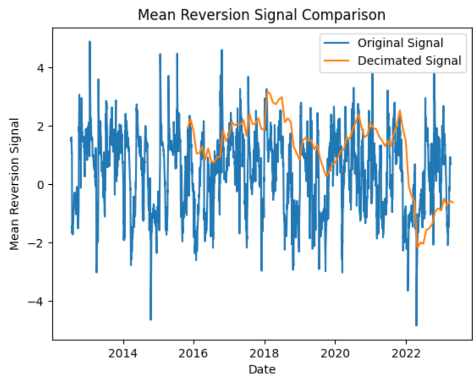

FIGURE 4.78: NFLX:40-Monthly

4.2.6.5 Window Size 50

The mean decimated signal for the daily sample frequency is 0.512, and the standard deviation is 1.349. With a value of 0.379 for the severe decimated ratio,

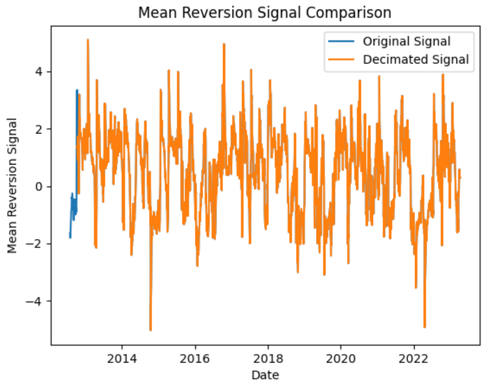

FIGURE 4.79: NFLX:50-Daily

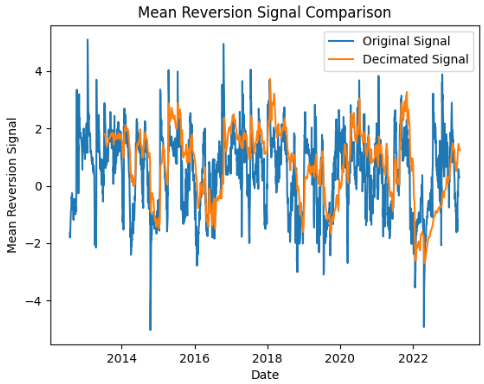

FIGURE 4.80: NFLX:50-Weekly

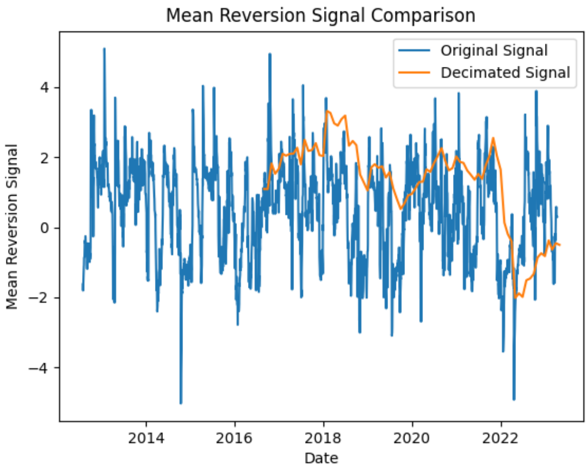

FIGURE 4.81: NFLX:50-Monthly

we can infer that the return will be moderate in comparison to the danger. The decimation ratio for Calmar is 0.102, which suggests that the company has poor performance when risk is taken into account.

The standard deviation of the decimated signal is reduced to 1.259%, but the mean decimated signal has increased to 0.767%. The ratio goes up to 0.609%, which indicates that the risk-return tradeoff is becoming more advantageous. The decimated ratio of Calmar has not altered from its previous value of 0.102.

The mean decimated signal has increased to 1.284, while the standard deviation has decreased to 1.274 as the monthly sample frequency increases. The severe decimated ratio improves to 1.007, which implies an increased return that is proportional to the level of risk. The decimated ratio of Calmar has not altered from its previous value of 0.102.

4.2.6.6 Summary

When trying to construct forecasts and improve the performance of the portfolio, it is important to take into consideration mean decimated signals with a greater value, as this indicates more accurate forecasting signals. In addition, smaller standard deviations are consistent with the possibility of a reduction in volatility.

Bigger values for the sharp decimated ratio and the Calmar decimated ratio suggest bigger rewards that are proportional to the level of risk. It is desirable to have greater mean decimated signals, lower standard deviations, and higher sharp decimated ratios and Calmar decimated ratios while optimizing a portfolio.

According to the findings of the study, the following combinations of window size and frequency are the most effective in terms of prediction and portfolio optimization:

Size of a Window with a 50-Inch Opening and a Monthly Sampling Frequency: It has the greatest mean decimated signal of 1.284 and the highest standard deviation of 1.274, indicating more accurate prediction signals and maybe lower volatility. The decimated signal is expressed as a mean and a standard deviation. In addition, it has a steep decimated ratio of 1.007, which implies that it has a great return relative to the risk involved. The decimated ratio of Calmar has not altered from its previous value of 0.102.

Chapter 5

Conclusion and Recommendations

The goal of the proposed algorithm is to use decimated data to enhance prediction of stock market movements and portfolio construction. The algorithm overcomes human fallibility by using a broad period of time, allowing it to keep up with the frenetic speed of today's high-tech corporate world.

One key benefit of using a longer time horizon is the minimization of human error. Human decision-making can be influenced by emotions, biases, and short-term market movements, resulting in sub-optimal investment choices. The algorithm mitigates these subjective biases and ensures a more objective and reasonable approach to stock market forecasting and portfolio construction by looking ahead over longer periods of time.

Moreover, in the modern era of high-tech company, speed and efficiency are of the utmost importance. The suggested approach makes effective use of state-of-the-art computational resources to handle and evaluate massive amounts of decimated data. This allows investors to quickly react to market shifts, seize new opportunities, and reduce risks by making informed decisions in real time.

Accurate stock market projections and trustworthy portfolio construction are enhanced by include a long-term perspective in the algorithm. The system is better able to capture underlying patterns, cyclical activity, and mean reversion effects when looking at the market over a longer time horizon. With a longer time horizon in mind, it is much easier to spot promising investment opportunities and fine-tune portfolio allocation.

The algorithm's adoption of a rather lengthy investment horizon also accords with the tenets of successful, long-term financial planning. Over longer time frames, the effects of fundamental variables, market cycles, and mean reversion events become more apparent and substantial. This approach lessens the need to react quickly to market fluctuations and promotes a more deliberate and well-informed approach to investing.

To enhance stock market forecasting for stock selection the suggested method uses decimated data. It takes use of the benefits of long-term investing, eliminates the risk of human mistake, and speeds up outcomes in a fast-paced commercial environment. By fusing high-powered computer resources with a holistic understanding of market dynamics, the program helps investors make better decisions, improve their portfolios, and manage the complexity of the stock market.

The proposed algorithm employing the decimated analysis is a valuable instrument for assessing the efficacy of various time horizons in stock market forecasting and portfolio construction. By investigating the decimated mean, decimated standard deviation, Sharp ratio, and Calmar ratio across various time horizons, it is possible to gain insight into the performance and predictive ability of various investment strategies. In the following thorough analysis, we will examine the decimated results and explain why a long time horizon is considered a more accurate predictor than a short one.

The decimated mean represents the average return of a decimated signal over a predetermined time interval. Comparing the decimated means across various time horizons reveals that the values tend to increase as the time horizon lengthens. This suggests that longer time horizons generate higher average returns on average. This finding provides support for the contention that a long-term horizon is advantageous for stock market forecasting and portfolio construction.

The decimalized standard deviation quantifies the volatility or risk of a decimated signal. Examining the decimated standard deviations across various time horizons reveals that the values tend to decrease as the time horizon lengthens. This suggests that longer time horizons minimize risk and volatility. Longer time horizons reduce noise and random fluctuations, making it simpler to identify underlying trends and make more accurate predictions.

Sharp ratio is an extensively employed risk-adjusted return metric. It measures the excess return per unit of risk assumed. Sharp ratios that are greater indicate superior risk-adjusted performance. In the context of decimated analysis, we assessed the predictive potential of the Sharp ratios across various time horizons, lengthier time horizons result in sharper ratios. This indicates that a longer time horizon provides superior risk-adjusted returns, suggesting that it is a more accurate predictor for stock market forecasting and portfolio construction.

The Calmar ratio compares the average annual return to the utmost drawdown to determine the risk-adjusted performance of an investment strategy. It offers insight into a strategy's risk management and downside protection capabilities. Similar to the Sharp ratio, higher Calmar ratios indicate a superior risk-adjusted return. When comparing the Calmar ratios over various time horizons, we typically observe greater values for extended time horizons. This suggests that a longer time horizon not only generates higher returns, but also provides greater downside protection, making it a superior predictor for stock market forecasting and portfolio construction.

For a variety of different reasons, anticipating the stock market and developing investment portfolios with a more longer horizon is regarded to be more accurate. Initially, the reduction of noise and random fluctuations in lengthier time horizons enables investors to concentrate on the fundamentals and underlying trends of the market, which results in more accurate predictions. This is because investors are able to focus less on the random fluctuations. Shorter time frames are more sensitive to the volatility of the market as well as short-term movements, both of which can distort the true value of the stock.

A broad time horizon not only provides more context, but also makes it possible to take into account more fundamental considerations. Fundamental analysis, which analyzes the financial health, growth prospects, competitive position, and valuation of organizations, requires a longer-term perspective in order to effectively evaluate the long-term potential of investments. This is because fundamental analysis analyses how companies will perform in the future. Investors are able to make more accurate predictions regarding the future performance of a company and discover stocks that are undervalued or overvalued if they conduct an analysis of fundamental elements over a prolonged period of time.

In addition, having a lengthy time horizon helps to reduce the impact of cyclical market movements. The stock market is often distinguished by periods of cyclical boom and collapse. Investors are able to more correctly anticipate market trends and alter their investment strategies in accordance with those trends if they monitor these cycles over a long period of time and make adjustments accordingly. As a consequence, portfolio management becomes more efficient, and predictions regarding the stock market become more accurate.

The phenomenon of mean reversion, which can be seen on financial markets and describes the tendency of stock values to return to their long-term average, is one factor that contributes to the predominance of a long time horizon. Investors are able to profit from this tendency to return to the mean by evaluating past data and finding patterns of mean reversion. This allows investors to make more accurate projections of future price fluctuations. A more extended time horizon not only provides a more comprehensive understanding of market cycles, but also raises the likelihood of successfully capturing mean reversion effects.

It is absolutely necessary to acknowledge that forecasting the stock market is naturally fraught with uncertainty and contains both restrictions and hazards. Even though having a longer time horizon improves one's capacity to make accurate predictions, there is no assurance that this will always be the case. The performance of the stock market can also be impacted by factors such as macroeconomic conditions, geopolitical events, and improvements in technology. Therefore, conducting comprehensive research, diversifying your holdings, and taking into consideration any other relevant elements is absolutely necessary in order to make informed investing selections.

The extensive study of decimated findings provides support for the contention that a long time horizon is a more reliable predictor than a short time horizon when it comes to forecasting the stock market and building portfolios. Long-term forecasts are more accurate because of several aspects, including the incorporation of fundamental factors, the reduction of noise and random fluctuations, the stabilizing of market cycles, and the potential of reversion to the mean. All of these factors work together to improve accuracy. In spite of this, it is necessary to realize the inherent uncertainty in stock market forecasting and to approach investing decisions with caution and careful evaluation of a wide variety of factors and risk management measures. Investors can improve their ability to foresee fluctuations in the stock market and make informed investing decisions by combining a long-term approach with substantial study and analysis. There is a list of arguments which is inconsistent with the this research results as it regret market efficiency hypothesis but at the same time get support from the various theories and empirical studies.

Random Walk theory argued that the stock market is in a random walk situation when there is no discernible pattern or trend in the values of individual stocks. It is preferable to have a larger time horizon when attempting to capture a random walk scenario rather than a shorter one.

Stock price variations in the short term are unpredictable and subject to the effects of noise and market mood, according to Malkiel (1973), which argues that stock prices follow a random walk pattern. He says that in order to profit from the fundamental patterns that drive stock prices, you need to have a long-term investment view.

De Bondt and Thaler (1990) find that the market has a tendency to correct itself over the long run when equities depart greatly from their long-term average, suggesting that short-term overreactions are common. This supports the idea that investors can profit from mean reversion provided they have a sufficiently long time horizon.

Forecasting the stock market with a longer time horizon is advantageous for a number of reasons, one of the most important of which is the fact that it minimizes the amount of noise and random fluctuations in the stock market. The following readings can help shed light on this topic:

Long-term volatility in the stock market can be dampened by the collective wisdom of many investors, (Surowiecki, 2005). He maintains that the impact of individual biases and short-term noise is mitigated and stock prices become more accurate when a longer time horizon is taken into account.

Kahneman et al. (2021) find the effects of decision-making noise on stock market predictions. They argue that short-term market fluctuations might be skewed by noise, which they define as random variation or judgment errors. However, with a longer time horizon, fundamental variables tend to have a stronger influence on stock values, while the impact of noise tends to reduce.

The relationship between market efficiency, long-term returns, and behavioral finance is something to look at (Fama, 1998). Short-term stock market results are found to be volatile and able to diverge from underlying values. However, when looking at stock prices over longer periods of time, the influence of noise decreases and prices move closer to reflecting fundamentals.

Campbell and Vuolteenaho (2004) find that stock returns are not aligned with predicted cash flows and other fundamental factors in the short term, but that this misalignment disappears in the long run.

Ibbotson et al. (2018) study of "time diversification" posits that holding investments for longer periods of time reduces exposure to short-term market swings. They argue that investors can lessen the impact of short-term market cycles and increase their potential return on investment by increasing their investment horizon. This data demonstrates that stock market projections are more robust when a longer time horizon is used to account for the effects of market cycles.

Long-term mean reversion is another example of evidence suggesting that a longer time horizon is desirable when trying to predict the stock market. Over time, asset values have a tendency to return to their long-term average. Some readings on the topic of mean reversion and its relevance to longer time frames are as follows:

Siegel (2021) offers a comprehensive evaluation of methods investing for the future. He contends that stock prices, over extended periods of time, have a tendency to revert to their intrinsic values, highlighting the existence of mean reversion in the stock market. This suggests that investors are better able to capture the effect of mean reversion and generate more accurate stock price projections when they have a longer time horizon.

Forecasting the stock market is heavily dependent on political and economic variables, all of which are framed by the Efficient Market Hypothesis (EMH). In terms of economics, important metrics like GDP growth, inflation, and interest rates are essential for predicting market movements. Interest rate decisions made by central banks affect investor mood, but GDP growth and inflation have a direct effect on consumer spending and business profitability. Accurate forecasting requires taking into account both government policies and political stability at the same time. While political unrest or changes in policy add uncertainty and can cause market swings, stable political settings often inspire investor confidence and guarantee market stability. Industry valuations are directly impacted by trade policies, taxation, and regulatory actions.

In its many forms—weak, semi-strong, and strong—the Efficient Market Hypothesis states that financial markets efficiently incorporate information into stock prices. While the hypothesis implies that continuously outperforming the market is difficult, it recognises the possible impact of behavioural biases, information asymmetries, and market inefficiencies. Analysts frequently struggle with this contradiction, recognising market efficiency while looking for possibilities in instances where behavioural or informational abnormalities may cause deviations from theoretical efficiency. In order to navigate the complicated environment of stock market forecasting, analysts and investors must first comprehend the interplay of economic and political elements, as well as be mindful of potential market inefficiencies.

5.1 Conclusion

For stock market forecasting, there are a variety of reasons why a longer time horizon is preferable to a shorter one. The prospect of a return to the mean, the dampening of market cycles, the inclusion of basic elements, and the decrease of random fluctuations and noise are all examples. Long-term stock market predictions benefit from all of these aspects. Recognizing the inherent uncertainty of stock market predictions, investors must conduct extensive research, diversify their holdings, and take into account other important elements in order to make educated investing decisions.

One of the benefits of a longer time horizon for stock market forecasting is the reduction of random volatility and random noise. Short-term movements in the market, including speculation, news, and general investor attitude, can have a significant impact on stock prices. Stock price variations may occur for reasons unrelated to the stock's intrinsic worth. By looking at the market over a longer time frame, investors can ignore the noise and focus on the fundamentals and underlying trends, from which they can draw more reliable conclusions.

At the culmination of this dissertation, we have successfully addressed the research questions that were outlined at the beginning of our study. The following are the key findings and conclusions we have drawn:

1. Can decimated data be used to trade on the basis of the algorithm?
   Through the utilization of the Decimated Online Moving Average Man Reversion (D-OLMAR) algorithm, we have demonstrated that decimated data can indeed be effectively used for trading purposes. The algorithm leverages the benefits of reduced noise and captures long-term trends, resulting in improved trading performance.

2. Does algorithm-based trading outperform market-based performance?
   Our analysis has shown that algorithm-based trading, specifically using the D-OLMAR algorithm, has the potential to outperform market-based performance. By analyzing large datasets and making informed decisions, the algorithm-based approach can generate superior trading outcomes.

3. Is a longer time frame better than a shorter time frame for forecasting and devising?
   In the context of this study, we have found that a longer time frame is more suitable for forecasting and devising trading strategies. This allows for a broader perspective and the capture of significant trends, enhancing the effectiveness of the D-OLMAR algorithm.

4. Does the algorithm-based trading strategy face higher long-term risks compared to a passive investment strategy?
   While algorithm-based trading strategies inherently carry risks, our research indicates that with appropriate risk management techniques and continuous monitoring, the D-OLMAR algorithm can mitigate long-term risks effectively. It is crucial to implement robust risk management practices to ensure the stability and reliability of the strategy.

5. Does an algorithm-based strategy offer better risk-adjusted returns against downside risk?
   This study findings suggest that the D-OLMAR algorithm, as an algorithm-based strategy, has the potential to offer better risk-adjusted returns against downside risk. By utilizing data-driven decision-making processes and incorporating risk management techniques, the algorithm helps optimize returns while mitigating the impact of adverse market conditions.

5.2 Limitations

Stock market prediction and stock selection using a decimated algorithm has limitations that may affect the accuracy of risk-adjusted measures such as the Sharpe ratio and the Calmar ratio. Some of these limits include the liquidity constraints, the presence of market inefficiencies, the occurrence of unforeseen events, and the reliance on a small sample size of historical data. The evaluation of decimation algorithms and the interpretation of risk-adjusted measures in this setting require careful attention to these caveats.

5.3 Recommendation and Policy Implications

Traders and investors can reap the benefits of algorithm-based trading, enhance their forecasting skills, reduce their exposure to long-term market risks, and increase their risk-adjusted profits. The following is a list of the key recommendations.

1. According to our findings, algorithmic trading outperforms market-based trading. Investors can make data-driven decisions that have the potential to produce greater trading outcomes by using the power of algorithms.

2. Longer time frames should be used for forecasting and developing trading strategies. According to our findings, lengthier time periods provide a broader perspective and allow for the detection of major patterns. Investors can improve the performance of their algorithm-based strategies by adopting a longer-term outlook, resulting in more accurate predictions and better trading decisions.

3. Implementing the D-OLMAR algorithm effectively mitigates long-term risks. Our findings show that the D-OLMAR algorithm can reduce the long-term hazards associated with algorithm-based trading. To ensure stability and reliability, it is critical to implement effective risk management strategies and to regularly review the plan.

4. Using an algorithm-based method, you can optimize risk-adjusted returns against downside risk. Our research shows that algorithm-based strategies, notably those based on the D-OLMAR algorithm, provide greater risk-adjusted returns in the face of downside risk. Investors can maximize returns while minimizing the impact of adverse market situations by using data-driven decision-making processes and combining risk management approaches.

5.4 Direction for Future Research

Efforts could be made to improve decimated algorithms' robustness by including approaches that adjust to shifting market conditions and reduce overfitting. Furthermore, using different risk indicators in addition to the Sharpe and Calmar ratios can provide a more thorough assessment of risk-adjusted returns. To improve the effectiveness and dependability of algorithm-based trading strategies, the future direction should include a continual iterative process of study, refinement, and validation.

Bibliography

Agarwal, A., Hazan, E., Kale, S., and Schapire, R. E. (2006). Algorithms for portfolio management based on the newton method. In *Proceedings of the 23rd international conference on Machine learning*, pages 9–16.

Akcoglu, K., Drineas, P., and Kao, M.-Y. (2002). Fast universalization of investment strategies with provably good relative returns. In *International Colloquium on Automata, Languages, and Programming*, pages 888–900. Springer.

Akcoglu, K., Drineas, P., and Kao, M.-Y. (2004). Fast universalization of investment strategies. *SIAM Journal on Computing*, 34(1):1–22.

Algoet, P. H. and Cover, T. M. (1988). Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *The Annals of Probability*, 16(2):876–898.

Allen, F. and Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of financial Economics*, 51(2):245–271.

Avnir, D., Biham, O., Lidar, D., and Malcai, O. (1998). Is the geometry of nature fractal? *Science*, 279(5347):39–40.

Bardoscia, M., Barucca, P., Battiston, S., Caccioli, F., Cimini, G., Garlaschelli, D., Saracco, F., Squartini, T., and Caldarelli, G. (2021). The physics of financial networks. *Nature Reviews Physics*, 3(7):490–507.

Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B., and Stiglitz, J. E. (2012). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of economic dynamics and control*, 36(8):1121–1141.

Bell, R. M. and Cover, T. M. (1980). Competitive optimality of logarithmic investment. *Mathematics of Operations Research*, 5(2):161–166.

Bertsekas, D. (2015). *Convex optimization algorithms.* Athena Scientific.

Bessembinder, H. and Chan, K. (1998). Market efficiency and the returns to technical analysis. *Financial management*, pages 5–17.

Biau, G., Bleakley, K., Györfi, L., and Ottucsák, G. (2010). Nonparametric sequential prediction of time series. *Journal of Nonparametric Statistics*, 22(3):297–317.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.-Theory Exp.*, 2008:P10008.

Blum, A. and Kalai, A. (1997). Universal portfolios with and without transaction costs. In *Proceedings of the Tenth Annual Conference on Computational Learning Theory*, pages 309–313.

Blum, A. and Mansour, Y. (2007). From external to internal regret. *Journal of Machine Learning Research*, 8(6).

Borodin, A., El-Yaniv, R., and Gogan, V. (2003). Can we learn to beat the best stock. *Advances in Neural Information Processing Systems*, 16.

Borodin, A., El-Yaniv, R., and Gogan, V. (2004). Can we learn to beat the best stock. *Journal of Artificial Intelligence Research*, 21:579–594.

Breiman, L. (1960). Correction notes: Correction to" the individual ergodic theorem of information theory". *The Annals of Mathematical Statistics*, 31(3):809–810.

Broesch, J. D. (1997). *Digital signal processing demystified.* Newnes.

Brogaard, J. et al. (2010). High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper*, 66.

Campbell, J. Y. and Vuolteenaho, T. (2004). Bad beta, good beta. *American Economic Review*, 94(5):1249–1275.

Cao, L.-J. and Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on neural networks*, 14(6):1506–1518.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge university press.

Chang, K., Young, M., and Diaz, J. (2018). Portfolio optimization utilizing the framework of behavioral portfolio theory. *International Journal of Operations Research*, 15(1):1–13.

Chen, G.-H., Kao, M.-Y., Lyuu, Y.-D., and Wong, H.-K. (1999). Optimal buy-and-hold strategies for financial markets with bounded daily returns. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, STOC '99, page 119–128, New York, NY, USA. Association for Computing Machinery.

Christensen, C., Hall, T., Dillon, K., and Duncan, D. S. (2016). Competing against luck. *The story of innovation and customer choice. First edition. New York, NY: HarperBusiness an imprint of HarperCollins Publishers.*

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, 1(2):223.

Cooper, M. J., Gutierrez Jr, R. C., and Hameed, A. (2004). American finance association. *The Journal of Finance*, 59(3):1345–1365.

Cover, T. M. (1991). Universal portfolios. *Mathematical finance*, 1(1):1–29.

Cover, T. M. and Gluss, D. H. (1986). Empirical bayes stock market portfolios. *Advances in applied mathematics*, 7(2):170–181.

Cover, T. M. and Ordentlich, E. (1996). Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2):348–363.

Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., and Singer, Y. (2006). Online passive aggressive algorithms.

Crammer, K., Dredze, M., Pereira, F., and Collins, M. (2012). Confidence-weighted linear classification for text categorization. *Journal of Machine Learning Research*, 13(6).

Crammer, K., Kearns, M., and Wortman, J. (2008). Learning from multiple sources. *Journal of Machine Learning Research*, 9(8).

Crammer, K., Kulesza, A., and Dredze, M. (2009). Adaptive regularization of weight vectors. *Advances in neural information processing systems*, 22.

Creamer, G. (2007). Using boosting for automated planning and trading systems.

Creamer, G. (2012). Model calibration and automated trading agent for euro futures. *Quantitative Finance*, 12(4):531–545.

Creamer, G. and Freund, Y. (2010). Automated trading with boosting and expert weighting. *Quantitative Finance*, 10(4):401–420.

Cross, J. E. and Barron, A. R. (2003). Efficient universal portfolios for past-dependent target classes. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 13(2):245–276.

Dai, M., Xu, Z. Q., and Zhou, X. Y. (2010). Continuous-time markowitz's model with transaction costs. *SIAM Journal on Financial Mathematics*, 1(1):96–125.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, 53(6):1839–1885.

Das, P. and Banerjee, A. (2011). Meta optimization and its application to portfolio selection. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1163–1171.

De Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? *The Journal of finance*, 40(3):793–805.

De Bondt, W. F. and Thaler, R. H. (1990). Do security analysts overreact? *The American economic review*, pages 52–57.

de Cheveigné, A. and Nelken, I. (2019). Filters: when, why, and how (not) to use them. *Neuron*, 102(2):280–293.

Dempster, M. A., Payne, T. W., Romahi, Y., and Thompson, G. W. (2001). Computational learning techniques for intraday fx trading using popular technical indicators. *IEEE Transactions on neural networks*, 12(4):744–754.

Dhar, V. (2011). Prediction in financial markets: The case for small disjuncts. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):1–22.

Dong, J., Wang, Y., Lai, J., and Xie, X. (2023). Restricted black-box adversarial attack against deepfake face swapping. *IEEE Transactions on Information Forensics and Security*.

Dredze, M., Crammer, K., and Pereira, F. (2008). Confidence-weighted linear classification. In *Proceedings of the 25th international conference on Machine learning*, pages 264–271.

Duchi, J., Shalev-Shwartz, S., Singer, Y., and Chandra, T. (2008). Efficient projections onto the l 1-ball for learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, pages 272–279.

El Hajj, M. and Hammoud, J. (2023). Unveiling the influence of artificial intelligence and machine learning on financial markets: A comprehensive analysis of ai applications in trading, risk management, and financial operations. *Journal of Risk and Financial Management*, 16(10):434.

El-Yaniv, R. (1998). Competitive solutions for online financial problems. *ACM Computing Surveys (CSUR)*, 30(1):28–69.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of financial economics*, 49(3):283–306.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22.

Fan, S., Li, L., and Zhang, X. (2012). Challenges of creating cities in china: Lessons from a short-lived county-to-city upgrading policy. *Journal of Comparative Economics*, 40(3):476–491.

Gaivoronski, A. A. and S., F. (2000). Stochastic nonstationary optimization for finding universal portfolios. *Annals of Operations Research*, 100(1):165–188.

George, T. J. and Hwang, C.-Y. (2004). The 52-week high and momentum investing. *The Journal of Finance*, 59(5):2145–2176.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42.

Györfi, L., Lugosi, G., and Udina, F. (2006). Nonparametric kernel-based sequential investment strategies. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 16(2):337–357.

Gyorfi, L. and Schafer, D. (2003). Nonparametric prediction. *NATO SCIENCE SERIES SUB SERIES III COMPUTER AND SYSTEMS SCIENCES*, 190:341–356.

Györfi, L., Udina, F., and Walk, H. (2008). Nonparametric nearest neighbor based empirical portfolio selection strategies.

Györfi, L., Urbán, A., and Vajda, I. (2007). Kernel-based semi-log-optimal empirical portfolio selection strategies. *International Journal of Theoretical and Applied Finance*, 10(03):505–516.

GYÖRFI, L., URBÁN, A., and VAJDA, I. (2007). Kernel-based semi-log-optimal empirical portfolio selection strategies. *International Journal of Theoretical and Applied Finance*, 10(03):505–516.

Györfi, L., Lugosi, G., and Udina, F. (2006). Nonparametric kernel-based sequential investment strategies. *Mathematical Finance*, 16(2):337–357.

Hakansson, N. H. (1975). Optimal investment and consumption strategies under risk for a class of utility functions. In *Stochastic Optimization Models in Finance*, pages 525–545. Elsevier.

Han, S. (2019). Democratization of finance in china: the role of internet finance in china's economic transformation.

Hansen, K. B. (2020). The virtue of simplicity: On machine learning models in algorithmic trading. *Big Data & Society*, 7(1):2053951720926558.

Hardy, M. R. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2):41–53.

Hazan, E., Agarwal, A., and Kale, S. (2007). Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192.

Hazan, E. et al. (2016). Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325.

Hazan, E. and Kale, S. (2015). An online portfolio selection algorithm with regret logarithmic in price variation. *Mathematical Finance*, 25(2):288–310.

Hazan, E. and Seshadhri, C. (2009). Efficient learning algorithms for changing environments. In *Proceedings of the 26th annual international conference on machine learning*, pages 393–400.

Helmbold, D. P., Schapire, R. E., Singer, Y., and Warmuth, M. K. (1998). Online portfolio selection using multiplicative updates. *Mathematical Finance*, 8(4):325–347.

Hoi, S. C., Sahoo, D., Lu, J., and Zhao, P. (2021). Online learning: A comprehensive survey. *Neurocomputing*, 459:249–289.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*, 54(6):2143–2184.

Huang, D., Zhou, J., Li, B., HOI, S., and Zhou, S. (2012). Robust median reversion strategy for on-line portfolio selection.(2013). In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence: IJCAI 2013: Beijing, 3-9 August 2013*.

Hubbard, B. B. (1998). The convergence of fourier series and the stability of the solar system. In *The World According to Wavelets*, pages 145–156. AK Peters/CRC Press.

Hull, J. C. (2008). *Options, Futures, and Other Derivatives.* Prentice Hall, Upper Saddle River, NJ.

Ibbotson, R. G., Idzorek, T. M., Kaplan, P. D., and Xiong, J. X. (2018). *Popularity: A bridge between classical and behavioral finance.* CFA Institute Research Foundation.

Ingber, L. (2020). Developing bid-ask probabilities for high-frequency trading. virtual economics 3 (2), 7–24.

Jegadeesh, N. and Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of finance*, 56(2):699–720.

Kahneman, D., Sibony, O., and Sunstein, C. R. (2021). *Noise: a flaw in human judgment.* Hachette UK.

Kelly, J. L. (1956a). A new interpretation of information rate. *The Bell System Technical Journal*, 35(4):917–926.

Kelly, W. M. (1956b). Modern factoring and how it meets today's new financial requirements. *American Bar Association Journal*, pages 13–17.

Kivinen, J. and Warmuth, M. K. (1997). Exponentiated gradient versus gradient descent for linear predictors. *information and computation*, 132(1):1–63.

Kozat, S. S. and Singer, A. C. (2011). Universal semiconstant rebalanced portfolios. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 21(2):293–311.

Kozat, S. S., Singer, A. C., and Bean, A. J. (2008). Universal portfolios via context trees. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2093–2096. IEEE.

Lattemann, C., Loos, P., Gomolka, J., Burghof, H.-P., Breuer, A., Gomber, P., Krogmann, M., Nagel, J., Riess, R., Riordan, R., et al. (2012). High frequency trading: costs and benefits in securities trading and its necessity of regulations. *Business & Information Systems Engineering*, 4:93–108.

Lee, C. M. and Swaminathan, B. (2000). Price momentum and trading volume. *the Journal of Finance*, 55(5):2017–2069.

Li, B. and Hoi, S. C. (2014). Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, 46(3):1–36.

Li, B., Hoi, S. C., and Gopalkrishnan, V. (2011a). Corn: Correlation-driven nonparametric learning approach for portfolio selection. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):1–29.

Li, B., Hoi, S. C., Sahoo, D., and Liu, Z.-Y. (2015). Moving average reversion strategy for on-line portfolio selection. *Artificial Intelligence*, 222:104–123.

Li, B., Hoi, S. C., Zhao, P., and Gopalkrishnan, V. (2011b). Confidence weighted mean reversion strategy for on-line portfolio selection. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 434–442. JMLR Workshop and Conference Proceedings.

Li, B., Hoi, S. C., Zhao, P., and Gopalkrishnan, V. (2013). Confidence weighted mean reversion strategy for online portfolio selection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(1):1–38.

Li, B., Zhao, P., Hoi, S. C., and Gopalkrishnan, V. (2012). Pamr: Passive aggressive mean reversion strategy for portfolio selection. *Machine learning*, 87(2):221–258.

Li, D. and Ng, W.-L. (2000). Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical finance*, 10(3):387–406.

Li, S. and Lin, A. (2020). Exploring the relationship among predictability, prediction accuracy and data frequency of financial time series. *Entropy*, 22(12):1381.

Li, T., Li, Q., Zhu, S., and Ogihara, M. (2002). A survey on wavelet applications in data mining. *ACM SIGKDD Explorations Newsletter*, 4(2):49–68.

Li, X., Yuan, J., Shi, Y., Sun, Z., and Ruan, J. (2020). Emerging trends and innovation modes of internet finance—results from co-word and co-citation networks. *Future Internet*, 12(3):52.

Liu, C., Jin, T., Hoi, S. C., Zhao, P., and Sun, J. (2017). Collaborative topic regression for online recommender systems: an online and bayesian approach. *Machine Learning*, 106(5):651–670.

Llorente, G., Michaely, R., Saar, G., and Wang, J. (2002). Dynamic volume-return relation of individual stocks. *The Review of financial studies*, 15(4):1005–1047.

Lo, A. W. and MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *The review of financial studies*, 3(2):175–205.

Lo, A. W., Mamaysky, H., and Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The journal of finance*, 55(4):1705–1765.

Mahfoud, S. and Mani, G. (1996). Financial forecasting using genetic algorithms. *Applied artificial intelligence*, 10(6):543–566.

Malceniece, L., Malcenieks, K., and Putniņš, T. J. (2019). High frequency trading and comovement in financial markets. *Journal of Financial Economics*, 134(2):381–399.

Malkiel, B. G. (1973). *A random walk down wall street [by] burton G. Malkiel.* Norton.

Mandelbrot, B. B. (1983). The fractal geometry of nature. bd. 173.

Mandelbrot, B. B. and Mandelbrot, B. B. (1982). *The fractal geometry of nature*, volume 1. WH freeman New York.

Mantegna, R. N. and Stanley, H. E. (1995). Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6535):46–49.

Markowits, H. M. (1952). Portfolio selection. *Journal of finance*, 7(1):71–91.

Michael Spector, J. (2017). A critical look at moocs. *Open education: From OERs to MOOCs*, pages 135–147.

Moody, J. and Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE transactions on neural Networks*, 12(4):875–889.

Neftci, S. N. (1991). Naive trading rules in financial markets and wiener-kolmogorov prediction theory: A study of" technical analysis". *Journal of Business*, pages 549–571.

Ordentlich, E. and Cover, T. M. (1998). The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23(4):960–982.

Ormos, M. and Urbán, A. (2013). Performance analysis of log-optimal portfolio strategies with transaction costs. *Quantitative Finance*, 13(10):1587–1597.

Osisanwo, F., Akinsola, J., Awodele, O., Hinmikaiye, J., Olakanmi, O., Akinjobi, J., et al. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3):128–138.

Ottucsák, G. and Vajda, I. (2007). An asymptotic analysis of the mean-variance portfolio selection.

Ottucsák, G. and Vajda, I. (2007). An asymptotic analysis of the mean-variance portfolio selection. 25(1):63–86.

Palmer, S. and Rock, I. (1994). On the nature and order of organizational processing: A reply to peterson. *Psychonomic Bulletin & Review*, 1(4):515–519.

Peter J. Brockwell, R. A. D. (1996). *ntroduction to Time Series and Forecasting*. Springer New York, NY, 2 edition.

Peterson, M. (1997a). Skills to enhance problem-based learning. *Medical Education Online*, 2(1):4289.

Peterson, M. (1997b). Skills to enhance problem-based learning. *Medical Education Online*, 2(1):4289.

Poggio, T., Rakhlin, S., Caponnetto, A., and Rifkin, R. (2012). Statistical learning theory and applications.

Poterba, J. M. and Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of financial economics*, 22(1):27–59.

Pring, M. (1991). *Technical Analysis Explained: The Successful Investor's Guide to Spotting Investment Trends and Turning Points*. McGraw-Hill.

Reddaf, A., Djerfaf, F., Ferroudji, K., Boudjerda, M., Hamdi-Chérif, K., and Bouchachi, I. (2020). Modeling of electromagnetic behavior of composite thin layers using genetic algorithm. *Mathematics and Computers in Simulation*, 167:281–295.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252.

Sathye, M. (2004). Teaching banking and finance online: a case study. *Indian Journal of Open Learning*, 13(3):303–312.

Shalev-Shwartz, S., Crammer, K., Dekel, O., and Singer, Y. (2003). Online passive-aggressive algorithms. *Advances in neural information processing systems*, 16.

Siegel, J. J. (2021). *Stocks for the long run: The definitive guide to financial market returns & long-term investment strategies*. McGraw-Hill Education.

Siering, M., Clapham, B., Engel, O., and Gomber, P. (2017). A taxonomy of financial market manipulations: establishing trust and market integrity in the financialized economy through automated fraud detection. *Journal of Information Technology*, 32:251–269.

Soleymani, F. and Paquet, E. (2020). Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder—deepbreath. *Expert Systems with Applications*, 156:113456.

Sonkin, P. D. and Johnson, P. (2017). *Pitch the Perfect Investment: The Essential Guide to Winning on Wall Street*. John Wiley & Sons.

Stephens, M. (2013). Moocs for lis professional development: exploring new transformative learning environments and roles. *Journal of Online Learning Research and Practice*, 2(2).

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Thorp, R. (1971). Inflation and the financing of economic development. In *Financing Development in Latin America*, pages 182–224. Springer.

Tsang, E., Yung, P., and Li, J. (2004). Eddie-automation, a decision support tool for financial forecasting. *Decision Support Systems*, 37(4):559–565.

Turchi, M., Anastasopoulos, A., de Souza, J. G., and Negri, M. (2014). Adaptive quality estimation for machine translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 710–720.

Unser, M. (1995). Texture classification and segmentation using wavelet frames. *IEEE Transactions on image processing*, 4(11):1549–1560.

Urman, A. and Makhortykh, M. (2023). How transparent are transparency reports? comparative analysis of transparency reporting across online platforms. *Telecommunications Policy*, 47(3):102477.

Vapnik, V. (1998). The support vector method of function estimation. In *Nonlinear modeling*, pages 55–85. Springer.

Vovk, V. and Watkins, C. (1998). Universal portfolio selection. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 12–23.

Vovk, V. G. (1990). Aggregating strategies. *Proc. of Computational Learning Theory, 1990.*

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936.