ENHANCED DYNAMIC MULTI-PERIOD PORTFOLIO OPTIMIZATION: INTEGRATING ADVANCED TRANSACTION COST FUNCTIONS

A Thesis

by

Gülşah Bıçkı

Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in the
Department of Financial Engineering

Özyeğin University
September 2024

ENHANCED DYNAMIC MULTI-PERIOD PORTFOLIO OPTIMIZATION: INTEGRATING ADVANCED TRANSACTION COST FUNCTIONS

Approved by:

Asst. Prof. Özcan Ceylan, Advisor
Dept. of Hotel Management
Özyeğin University

Prof. Taylan Akdoğan
Dept. of Natural and Mathematical Sci.
Özyeğin University

Asst. Prof. Yasin Kütük
Dept. of Business Administration
Gebze Technical University

Date Approved: 13 August 2024

To my dear daughter, Gaye Merve, whose presence has been a constant source of joy and inspiration throughout this journey

ABSTRACT

This study investigates the impact of transaction costs, often overlooked in multi-period portfolio optimization. Traditional portfolio optimization models tend to underestimate the influence of transaction costs such as bid-ask spreads, market impact costs, and commission fees by assuming these costs to be constant. However, these costs, which can vary over time, were analyzed using daily data from the most liquid ETFs between January 2009 and May 2024 to identify the exogenous variables (e.g., VIX, trading volume) that best predict transaction cost function. These variables were integrated into advanced forecasting models, including ARIMA-X, SARIMA-X, and GARCH-X, to improve transaction cost forecasting accuracy. The performance of these models was compared using metrics such as RMSE and MAE, with the SARIMA-X model achieving the lowest RMSE and MAE values, indicating the most accurate predictions. The transaction costs predicted by the SARIMA-X model were then integrated into the multi-period portfolio optimization framework developed by Boyd [1]. The findings of this study indicate that accurately forecasting these time-varying factors plays an important role in modeling transaction costs and that incorporating these costs into the optimization process enhances multi-period portfolio performance.

Keywords: Multi-period portfolio optimization, transaction costs, bid-ask spread, exogenous variables, GARCH-X, ARIMA-X, SARIMA-X, RMSE, MAE, ETF, volatility, trading volume, portfolio performance.

ÖZETÇE

Bu çalışma, çok dönemli portföy optimizasyonunda genellikle göz ardı edilen işlem maliyetlerinin etkilerini araştırmaktadır. Geleneksel portföy optimizasyonu modelleri, işlem maliyetlerini, örneğin alım-satım farkları, piyasa etki maliyetleri ve komisyon ücretlerini sabit kabul ederek bu maliyetlerin etkisini küçümseme eğilimindedirBu maliyetler zamanla dalgalanabilmektedir. Bu çalışmada, Ocak 2009-Mayıs 2024 arası en likit ETF'lerin günlük verileri kullanılarak işlem maliyet fonksiyonunu en iyi tahmin eden dışsal değişkenler (örneğin VIX, işlem hacmi) belirlenmiştir. Bu değişkenler, işlem maliyeti tahmin doğruluğunu artırmak amacıyla ARIMA-X, SARIMA-X ve GARCH-X gibi ileri düzey tahmin modellerine entegre edilmiştir. Bu modellerin performansları RMSE ve MAE gibi metriklerle karşılaştırılmış ve SARIMA-X modeli en düşük RMSE ve MAE değerlerine ulaşarak en doğru tahminleri sağlamıştır. SARIMA-X modeliyle tahmin edilen işlem maliyeti ise Boyd tarafından geliştirilen çok dönemli portföy optimizasyonuna entegre edilmiştir [1]. Çalışmanın bulguları, zamanla değişen bu faktörlerin işlem maliyetlerinin doğru bir şekilde tahmin edilmesinde kritik bir rol oynadığını ve bu maliyetlerin optimizasyona dahil edilmesinin, çok dönemli portföy optimizasyonunun performansını anlamlı derecede artırdığını göstermektedir.

Anahtar Kelimeler: Çok dönemli portföy optimizasyonu, işlem maliyetleri, alım-satım farkları, dışsal değişkenler, GARCH-X, ARIMA-X, SARIMA-X, RMSE, MAE, ETF, volatilite, işlem hacmi, portföy performansı.

ACKNOWLEDGEMENTS

I am profoundly grateful for the guidance and mentorship provided by Assistant Professor Özcan Ceylan, whose involvement in this thesis has been nothing short of transformative. His ability to identify and refine the focus of this research, coupled with his steadfast support and invaluable feedback, has been crucial in navigating the intricacies of portfolio optimization and the detailed analysis of transaction costs. His commitment to fostering academic rigor and his encouragement have been a constant source of inspiration and have significantly shaped the direction and outcome of this study.

My deepest appreciation is also extended to Prof. Dr. Taylan Akdoğan, who has been an integral part of my academic journey, both at the university and throughout the development of this thesis. Reuniting in this context and receiving his support on my thesis topic has been a remarkable and serendipitous experience. Prof. Akdoğan's extensive knowledge in physics and programming has not only enriched my research but has also provided me with invaluable support in coding aspects of the thesis. His willingness to share his expertise and the joy of collaborating once again have deeply enriched my research experience and contributed significantly to the success of this study.

Additionally, I am thankful for the access to Bloomberg's extensive databases, which have been indispensable in providing the high-quality data and analytical tools necessary for my research. The availability of such resources has played a crucial role in ensuring the accuracy and reliability of my findings.

To all who have contributed to my journey, your support, guidance, and belief in the potential of my work have been invaluable. This thesis not only reflects my efforts but also embodies the collective wisdom and encouragement of many esteemed individuals.

Thank you for being part of this fulfilling journey and for your pivotal roles in the successful completion of my thesis. The mentorship and support I have received have profoundly impacted my growth as a researcher and will undoubtedly inspire my future endeavors in the field of finance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The constantly evolving nature of financial markets necessitates the continuous reassessment of portfolio management strategies and techniques. In this complex environment, portfolio optimization is a crucial tool for investors aiming to navigate market uncertainties while striving to maximize returns and minimize risks. Essentially, portfolio optimization is an analytical strategy used to select the best mix of assets for an investment portfolio. This process considers various factors such as the investor's risk tolerance, investment horizon, market conditions, transaction costs, economic indicators, tax implications, and overall investment objectives.

This thesis focuses on the importance of realistic transaction cost in multi-period portfolio optimization.

1.1 Motivation

Portfolio optimization has significantly evolved since Harry Markowitz's groundbreaking work. In 1952, Markowitz published his pioneering paper, "Portfolio Selection", in the Journal of Finance [2]. This paper laid the foundational principles of what would later be known as Modern Portfolio Theory (MPT). Markowitz introduced the concepts of diversification and the efficient frontier, demonstrating how investors could construct portfolios to maximize expected return for a given level of risk or minimize risk for a given level of expected return. In his 1959 book, "Portfolio Selection: Efficient Diversification of Investments", he provided a more comprehensive treatment of these principles and detailed their applications [3]. Markowitz's contributions laid the groundwork for subsequent advancements in portfolio optimization and risk management, influencing many financial theorists and practitioners. The 2014 study by

Kolm, Tütüncü, and Fabozzi provides a detailed account of the sixty-year historical development of this field and the practical challenges encountered [4].

Following Markowitz's pioneering contributions, numerous scholars and practitioners expanded the domain of portfolio optimization. Notable among these is James Tobin, who in 1958 introduced the concept of the risk-free asset into the portfolio selection model, further refining the efficient frontier's applicability. Tobin's separation theorem, which divides investment decision-making into two distinct phases, security selection and asset allocation enhanced the practical utility of Markowitz's MPT [5]. William Sharpe's development of the Capital Asset Pricing Model (CAPM) marked another milestone, providing insights into how market risk influences the expected return on assets and portfolios [6]. The CAPM, by establishing a linear relationship between expected return and beta, offered a mechanism to calculate the expected return on investment, considering both the risk of the security and the market.

Portfolio optimization has evolved significantly from the foundational contributions of Single Period Optimization by pioneers like Harry Markowitz, William Sharpe, and James Tobin [2], [3], [5], [6]. Their groundbreaking work laid the groundwork for modern portfolio theory, which primarily focused on optimizing portfolios over a single period. However, the inherent limitations of Single-Period Optimization became increasingly evident as the financial markets evolved. Optimizing portfolios for individual periods in isolation led to sub-optimal strategies that often incurred high transaction costs and failed to account for the dynamic nature of investment opportunities and market conditions [7]. Recognizing these limitations, researchers began to explore multi-period portfolio selection frameworks. Lobo, Fazel, and Boyd addressed these challenges by introducing a portfolio optimization model that incorporates both linear and fixed transaction costs [8]. Their model highlights the importance of considering transaction costs in portfolio optimization to avoid sub-optimal strategies and improve overall portfolio performance.

This dynamic and interconnected approach to portfolio management has been significantly influenced by the foundational research of Samuelson, Merton and Nystrup who were instrumental in developing the principles of multi-period optimization [9], [10], [11]. Building on Samuelson's discrete-time formulation, Constantinides extended these concepts to include proportional transaction costs [12]. Similarly, Davis and Norman, along with Dumas and Lucian, adapted these principles for continuous-time models, achieving comparable results [13], [14]. In the continuous-time model, investment decisions are updated continuously, allowing investors to constantly review and rebalance their portfolios. Multi-period settings are more adept at handling transaction costs, constraints, and time-varying forecasts, making them inherently suitable for more complex and realistic portfolio management strategies. Skaf and Boyd, in their work "Multi-period Portfolio Optimization with Constraints and transaction costs", examine how to optimize portfolios over multiple periods while considering these various factors [15]. On the other hand, Tütüncü highlights the practical challenges in implementing this method, including data and forecasting accuracy, computational difficulties, and market constraints [4].

Following the pioneering work of Bertsekas and Bellman, much of the literature on multi-period portfolio selection has relied heavily on dynamic programming techniques to optimize investment decisions over time [16], [17]. Research on multi-period portfolio selection largely relies on dynamic programming, which effectively incorporates the concept of recourse and utilizes updated information as trade sequences are determined [18]. However, applying dynamic programming for trade selection becomes highly impractical in real-world applications due to the "curse of dimensionality" [19], [20]. This challenge arises because the state space, the set of all possible combinations of asset holdings, grows exponentially with the number of assets and time periods. As a result, the computational resources and time required to solve these problems increase dramatically, making it infeasible for larger portfolios or longer time horizons. Consequently, most studies focus on a very limited number of assets and simplified objectives and constraints. For instance, Cesarone examined the Limited Asset Markowitz model, which includes quantity and cardinality constraints to address the practical issues of managing portfolios with a limited number of assets, emphasizing the computational complexities and the need for advanced optimization techniques in such scenarios [21].

A substantial body of literature examines multi-period portfolio selection without considering transaction costs [22]. In these special cases, dynamic programming remains feasible and can be effectively applied. However, to make dynamic programming problems more manageable, several approximation techniques are utilized. These include methods like approximate dynamic programming, which simplifies the computational complexity, and more straightforward formulations that extend single-period optimization frameworks into multi-period optimization models. Boyd used these approaches to help balance accuracy and computational feasibility in practical applications.

The method discussed in this article, introduced by Boyd, involves a relaxation of dynamic programming's comprehensive time horizon approach [1]. For practical purposes, various approximations of the dynamic programming approach are often employed, including methods like approximate dynamic programming and simpler models that extend single-period formulations into multi-period optimization problems [1], [23]. MPC enhances the optimization process by integrating new information at each time step and has been used in many industrial applications [24]. Specifically, at any given time T, a multi-period optimization problem is solved for H periods into the future using the current information. Despite determining optimal actions for several future periods, only the actions for the first period are executed. The optimization is then re-evaluated at time T + 1 with updated information. This receding horizon approach simplifies the full-horizon dynamic programming problem while allowing for rapid responses to evolving financial markets. In the realm of finance, applications of MPC include portfolio optimization, optimal trade execution and index tracking [25]. Nystrup et al. applied the same Multi-Period Optimization framework as Boyd to utilize multi-period forecasts. Their goal was to minimize the risk of the portfolio dropping below a certain level relative to its previous peak, thereby achieving a lower maximum drawdown [26].

To effectively implement MPC in financial applications such as portfolio optimization, optimal trade execution, and index tracking, it is crucial to predict key financial variables like bid-ask spreads, transaction costs, volatility, returns, and risk [27]. These predictions require robust time series models, which can analyze historical data to forecast future values [28]. Therefore, the integration of time series analysis techniques, such as ARIMA, becomes essential in providing the necessary predicted inputs for the multi-period optimization framework. A time series is a sequence of observations recorded at regular time intervals, such as hours, days, weeks, months, or years. Time series data is essential for understanding how events, incidents, symptoms, or changes evolve over time. Time series analysis can be divided into two categories based on the number of variables analyzed: univariate and multivariate. Univariate time series analysis involves a single time series data variable, while multivariate analysis incorporates multiple time series data variables [29], [30]. ARIMA (Autoregressive Integrated Moving Average) is primarily used for univariate time series analysis, focusing on modeling a single data variable by leveraging its past values to forecast future values. The ARIMA model predicts future values as a linear function of past values and random errors [31]. It uses three methods: autoregression, moving average, and differencing processes, collectively represented as ARIMA (p, d, q) [32]. This model is widely used for various types of time series data, aiming to uncover the underlying stochastic behavior of a series to forecast future values [33], [34]. On the other hand, SARIMA (Seasonal ARIMA) extends ARIMA by modeling seasonal patterns, improving its performance on seasonal series [35], [36], [37], [38]. GARCH (Generalized Autoregressive Conditional Heteroscedasticity), developed by Bollerslev from Engle's ARCH model, is invaluable for modeling time-varying volatility in financial time series. GARCH models effectively capture and predict financial market volatility [39].

In this paper, ARIMA-X, SARIMA-X, and GARCH-X models will be used. These advanced multivariate analysis methods build on existing time series models by incorporating multiple data variables and exogenous factors. Traditional ARIMA models typically ignore independent variables, limiting their accuracy when the time series is influenced by external factors. The ARIMA-X model addresses this limitation by including exogenous variables [40], [41]. SARIMA-X improves predictive capabilities by accounting for both seasonal patterns and exogenous variables, allowing for a more accurate modeling of complex time series data [42], [43]. Similarly, GARCH-X enhances the GARCH framework by adding exogenous, which helps better capture and predict financial market volatility.

The SARIMA-X model has proven to be highly effective in forecasting financial metrics by incorporating both seasonal patterns and exogenous variables. This model's ability to capture complex dynamics in time series data, such as transaction costs, has led to more accurate predictions, as demonstrated by our empirical analysis. The integration of SARIMA-X into the Multi-Period Optimization framework has significantly improved the model's forecasting accuracy, particularly in key metrics like RMSE and MAE, making it a robust choice for financial forecasting in dynamic environments.

This paper builds on the work of Boyd, who demonstrated that Multi-Period Optimization using convex programming is computationally feasible and can incorporate various costs and constraints [19]. To improve the accuracy of bid-ask spread predictions, thereby enhancing overall transaction cost predictions, advanced methods such as GARCH-X, ARIMA-X, and SARIMA-X are employed. These methods are relatively novel in this context compared to traditional approaches used by Oprisor, Boyd, and Uysal [1], [23], [44]. The best-performing model among these methods is integrated into the Multi-Period Optimization framework using Model Predictive Control [1]. This allows for dynamic adjustments based on new information at each time step, enhancing the optimization process. This approach shows that advanced forecasting methods can significantly improve the accuracy and reliability of transaction costs predictions, ultimately improving the effectiveness of multi-period portfolio optimization performance. This study analyzes the results of portfolio allocation performance over a five-period horizon, demonstrating how this approach improves portfolio optimization by dynamically adjusting to market conditions and enhancing transaction cost predictions.

By employing these advanced forecasting methods, this study aims to enhance the accuracy and reliability of transaction cost predictions, ultimately improving the performance of multi-period portfolio optimization. This approach addresses the research question: "Which advanced forecasting methodology (ARIMA-X, SARIMA-X, or GARCH-X) provides the most accurate transaction cost predictions for integration into a multi-period optimization framework?" Accurate transaction cost forecasts allow us to examine their impact on the optimization process and portfolio returns, using a more comprehensive analysis of market conditions with external influences.

1.2 Outline

Section 1 (Introduction and Literature Review) provides an overview of portfolio optimization, explaining Single Period Optimization and its limitations, and discusses the historical development of Multi-Period Optimization. It highlights the importance of accurately forecasting transaction costs and introduces Model Predictive Control in the context of dynamic portfolio management. This section also explains the selection of advanced forecasting methods, including ARIMA-X, SARIMA-X, and GARCH-X using exogenous variables, and concludes with the presentation of the research question and contributions of the study.

Section 2 (Methodology) explains the Multi-Period Optimization model and the integration of advanced forecasting methods such as ARIMA-X, SARIMA-X, and GARCH-X. It covers the selection and application of these forecasting models to predict key financial variables using heterogeneous variables.

Section 3 (Empirical Analysis) applies the Multi-Period Optimization framework using the best-performing forecasting method identified through a comparison of ARIMA-X, SARIMA-X, and GARCH-X models based on RMSE, and MAE metrics. It analyzes portfolio allocation performance over a five-period horizon and evaluates the accuracy of transaction cost predictions and their impact on portfolio optimization.

Section 4 (Conclusion) summarizes the findings and their implications for financial management. It addresses the research question and discusses the contributions of the study. Outlines the related possible future work.

Section 5 (VITA) gives a brief academic biography of Gülşah Bıçkı, including educational background, research focus, and relevant professional experience, will be provided.

1.3 Contribution

The main contributions of this study to the existing literature are two-fold. First, the accuracy of transaction cost predictions is enhanced by employing advanced forecasting methods such as ARIMA-X, SARIMA-X, and GARCH-X. By incorporating exogenous variables, the most accurate model is identified through rigorous evaluation using RMSE and MAE metrics. The findings indicate that the SARIMA-X model provides the best predictive accuracy among the models evaluated. Second, the best-performing forecasting method, SARIMA-X, is integrated into a Multi-Period Optimization framework using Model Predictive Control. This integration allows for dynamic adjustments to portfolio strategies based on new information at each time step, significantly improving portfolio optimization performance and transaction cost management.

CHAPTER II

METHODOLOGY

2.1 Multi Period Portfolio Optimization

Multi period optimization model forms the core of the advanced portfolio management strategy in this study. Multi period optimization offers a comprehensive framework for making optimal portfolio decisions over multiple periods, considering the dynamic nature of financial markets. Unlike single period optimization, which only optimizes portfolio allocation for a single period without considering future implications, Multi period optimization facilitates the development of long-term investment strategies and enables more informed and strategic decision-making by considering potential future market conditions and changes in asset return dynamics [15], [1], [44], [45].

Multi period optimization requires a more realistic modeling of transaction costs as it considers both the cost of entering and exiting positions over multiple periods. This realistic approach helps in minimizing total transaction costs, unlike single period optimization, which often underestimates these costs by only considering the immediate period's expenses. For example, an investor using single period optimization might buy stocks based on a short-term price increase prediction, only to sell them at a high cost shortly after due to a long-term market downturn. Multi period optimization, on the other hand, accounts for such costs upfront, allowing for cheaper position adjustments that align with both short and long-term expectations, leading to more suitable investment strategies.

By incorporating future return, risk estimates, and transaction cost estimates, Multi Period Optimization can better manage the tradeoff between short-term and long-term benefits. It takes into account the decay of return signals and adjusts the portfolio accordingly to maximize overall performance [15], [18], [1], [44]. A multi period optimization problem is solved for a given number of periods, 'H', ahead using the current information and it is resolved at each subsequent point in time with updated information. This approach simplifies the full horizon dynamic programming problem and allows for quick reactions to changes in financial markets.

The curse of dimensionality poses significant challenges in portfolio optimization due to the exponential increase in computational complexity as the number of assets and constraints grows. However, multi period optimization's ability to dynamically adjust to new information and optimize over multiple periods, combined with the advanced convex optimization techniques employed by CVX Portfolio package, effectively addresses these challenges.

CVX Portfolio, developed by Boyd, is a versatile open-source Python package designed to facilitate portfolio simulation and optimization. By utilizing two auxiliary Python libraries, Pandas for robust data management, and CVXPY for constructing and solving optimization problems, CVX Portfolio seamlessly integrates data handling with advanced optimization techniques while providing user defined functions and pre-cooked data for forecasting and transaction cost. Pandas provides structured data types and a comprehensive API, enabling efficient data manipulation and easy coupling with database backends [46], [47], [48].

CVXPY, based on disciplined convex programming principles, allows for the straightforward specification and modification of complex optimization problems. This flexibility makes it possible to incorporate non-standard trading and holding constraints, as well as intricate risk and cost functions. Through an object-oriented framework, CVX Portfolio offers classes representing various components of the optimization model, such as returns, risk measures, transaction costs, and constraints. These classes generate CVXPY expressions and constraints for any given period, facilitating the creation of both single period and multi period optimization models.

The practical advantages of CVX Portfolio are notable. For example, replacing 3/2-power transaction cost terms with square transaction cost terms can significantly speed up the optimization process. Additionally, using custom solvers based on operator-splitting methods instead of the default ECOS solver in CVXPY can further enhance computational efficiency [49]. These features make CVX Portfolio a powerful tool for handling the complexities of multi period optimization.

In summary, when $H = 1$, multi period optimization reduces to single period optimization, focusing only on a single time period. Conversely, when $H = T$, it considers all future periods up to T, representing the theoretical maximum trading activity. Multi period optimization's ability to dynamically adjust to new information and optimize over multiple periods leads to improved return performance and better management of transaction costs and risks. To address these issues, the CVX Portfolio library will be utilized, leveraging advanced convex optimization techniques and efficient data management practices to enhance portfolio optimization and transaction cost management [19], [1].

2.2 Multi Period Portfolio Optimization Model

A portfolio composed of several assets and a cash account, over a finite time horizon split into discrete time periods labeled $t = 1, \cdots, T$ is considered with a definition of a 'period' as single day. It is evaluated in this portfolio to decide the best allocation for a given horizon to maximize returns while minimizing risks and transaction costs at each day.

Suppose $h_t \in \mathbb{R}^{n+1}$ represents the value of each asset i in the portfolio at the start of period t. A negative value indicates a short position in asset i, $(h_t)_i < 0$ while a positive value indicates a long position $(h_t)_i > 0$. Since the asset $n+1$ represents the cash account, $(h_t)_i = 0$ implies that at time t , the portfolio is fully invested, with no cash held. The total value of the portfolio v_t, in dollars, at time t is expressed by

$v_t = 1^T h_t$.

Portfolio Weights are fractions of the total portfolio value that each asset represents. They are calculated by dividing the value of each asset by the total portfolio value.). $w_t \in \mathbb{R}^{n+1}$ are defined as $w_t = h_t/v_t$. The portfolio weights always sum to one, by definition, $1^T w_t = 1$, and are unitless (dollar holdings divided by portfolio total dollar value). Equivalent to the holding's scenario, the last weight $(w_t)_i = 0$ is the fraction of the total portfolio value being held in cash.

Trades and Normalized Trades (u_t and $z_t \in \mathbb{R}^n$)

Normalized trades represent the dollar value of trades made in period t, adjusted relative to the holdings of each asset in the portfolio. Let $u_t \in \mathbb{R}^n$, be the dollar value of the trades in period t. It is assumed that all trading occurs at the beginning of each time period. If $(u_t)_i > 0$, it indicates that $(u_t)_i$ dollars of asset i have been bought while $(u_t)_i < 0$ indicates the opposite, for $i = 1, \cdots, n$. The dollar value trades u_t are divided by the holdings of each asset h_t to obtain normalized trades z_t, $z_t = u_t/h_t$.

Single-Period Optimization focuses solely on the most recent trade decision z_t without taking future periods into account in the current optimization process. Essentially, Single-Period Optimization is a particular instance of Multi-Period Optimization where the planning horizon encompasses only one period. In contrast, Multi-Period Optimization determines the current trade vector z_t by solving an optimization problem over a planning horizon that spans H periods into the future, from $t, t+1, \cdots, t+H–1$. This approach is aligned with Model Predictive Control, where Multi-Period Optimization is implemented by continuously re-optimizing the portfolio at each time step with updated information, ensuring that the strategy remains adaptive to market changes.

The Multi-Period Optimization problem is formulated based on the core mean-variance objective: maximizing returns while minimizing risks and transaction costs.

Let $z_t, z_{t+1}, \cdots, z_{t+H-1}$ represent the sequence of planned trades over the investment horizon. Using estimated returns $\hat{r}$, risk $\hat{\psi}$, and transaction cost $\hat{\phi}^{\text{trade}}$ derived from the top-performing forecasting model, SARIMA-X, rather than Boyd and Oprisor's methods, a logical objective within the multi period framework is to maximize the cumulative risk-adjusted return over the entire investment horizon [1], [44].

$$\text{Maximize} \sum_{\tau=t}^{t+H-1} \left(\hat{r}_\tau^T (w_\tau + z_\tau) - \gamma^{\text{risk}} \hat{\psi}_\tau (w_\tau + z_\tau) - \gamma^{\text{trade}} \hat{\phi}_\tau^{\text{trade}} (z_\tau) \right) \tag{1}$$

subject to $1^T z_\tau = 0$, $z_\tau \in \mathcal{Z}_\tau$, $w_\tau + z_\tau \in \mathcal{W}_\tau$, $w_{\tau+1} = w_\tau + z_\tau$.

The first constraint $1^T z_\tau = 0$ (Trade Neutrality), ensures that trades do not change the total value of the portfolio; the sum of the trades must be zero. The second constraint, $z_\tau \in \mathcal{Z}_\tau$ Trade Bounds, limits the trades within feasible bounds specifying upper and lower limits on how much of each asset can be bought or sold. The third constraint, $w_\tau + z_\tau \in \mathcal{W}_\tau$ (Weight Bounds), ensures that the new portfolio weights after the trades remain within acceptable limits. The final constraint, $w_\tau + z_\tau \in \mathcal{W}$, updates the portfolio weights for the next period based on the current weights and the trades made by equation (1).

In the context of multi period optimization, the variables $z_t, z_{t+1}, \cdots, z_{t+H-1}$ and $w_{t+1}, \cdots, w_{t+H-1}$ are optimized over the planning horizon. The parameters γ_t^{risk} and γ_t^{trade} are positive scaling factors used to adjust the respective costs of risk and trading. The notation indicates that these values are estimates, rather than known or realized quantities. These parameters are often referred to as hyper-parameters, similar to those used in fitting statistical models. Hyperparameters can significantly influence the effectiveness of the Multi-Period Optimization method, and therefore, must be carefully selected through back testing.

As discussed earlier, the multi period optimization model leverages the SARIMA-X model to derive transaction costs. These estimates are crucial inputs for the optimization process, ensuring that the multi period optimization framework operates dynamically and adjusts to new information effectively.

2.3 Degrees of Freedom

The optimization model intentionally leaves several degrees of freedom to be defined [1]. Performance evaluation for the original model is conducted retrospectively using realized future data that could not have been anticipated during optimization, without providing future projections. To construct a comprehensive trading model, the following components are addressed:

Return and Risk Estimates:

Returns $\hat{r}$ and $\hat{\psi}$ are substituted by $r^{\mathrm{SARIMA-X}}$ and $\psi^{\mathrm{SARIMA-X}}$ generated from the SARIMA-X model. These estimates utilize ETF data spanning from January of 2009 to May of 2024, with testing data from 2022 to May 2024. Trading volume data averaged over a 10-day period is also incorporated.

In calculating risk, the quadratic risk formula derived from the SARIMA-X model is used. Assuming the returns $\hat{r} \in \mathbb{R}^{n+1}$, are randomly distributed, with covariance matrix $\Sigma_t \in R^{(n+1)\times(n+1)}$ is obtained from SARIMA-X model, the variance of the portfolio return R_t^p (quadratic risk) is given by equation (2).

$$\psi^{\text{SARIMA–X}} = \operatorname{var}(R_t^p) = (w_t + z_t)^T \Sigma_t (w_t + z_t) \tag{2}$$

It is noted that $\psi^{\text{SARIMA–X}}$ is an estimate of the return covariance based on sample data and model assumptions. The exact distribution of the process generating real asset price returns can never be known.

2.4 Transaction Costs

In financial markets, transactions come up with a transaction cost, denoted as $\phi_\tau^{\mathrm{trade}}(u_\tau)$, where $\phi_\tau^{\mathrm{trade}} : \mathbb{R}^{n+1} \to \mathbb{R}$ represents the dollar transaction cost function. The model assumes that this transaction cost function is separable, meaning the total transaction cost is the sum of individual transaction costs for each asset. This assumption simplifies the model by ignoring the high-frequency cointegration of asset prices, which is reasonable given that the analysis period is daily. The transaction cost function for asset i in period t is denoted as $\left(\phi_t^{\text{trade}}\right)_i$.

Following a similar methodology to Boyd, Grinold and Kahn, the transaction cost function $\left(\phi_t^{\text{trade}}\right)_i$ is typically modeled as [1], [50]:

$$x \to a|x| + b\sigma\frac{|x|^{3/2}}{V^{1/2}} + cx \tag{3}$$

However, unlike Boyd, Grinold and Kahn, the model developed incorporates a time-varying bid-ask spread (a_t), trading volume V_t and volatility σ_t. This leads to a modified transaction cost function. The updated transaction cost function, which accounts for the time-varying bid-ask spread a_t is given by:

$$x_t \to a_t|x_t| + b\sigma_t\frac{|x_t|^{3/2}}{V_t^{1/2}} + cx \tag{4}$$

where b, and c are numbers and x is a dollar trade amount [1]. a_t represents the asset's bid-ask spread at the beginning of the trading period and varies over time by equation (4). This term can also include broker fees if desired. The second term, with b being a positive constant with unit of 1/dollars represents the temporary impact cost of trading. V denotes the total dollar value of the asset traded in the market during the current period, and σ is the standard deviation of the asset price over recent periods, both of which vary over time. Boyd suggests that trading one day's volume typically moves the price by one day's volatility, implying a value of b around one. The term cx allows for expressing differences between buying and selling, with c being positive indicating lower costs for selling. If this term is ignored ($c = 0$), the cost is the same regardless of the trade direction. However, when $c < 0$, selling is more expensive than buying, which could reflect a market with difficulty borrowing stock to short sell or otherwise increased selling pressure (more sellers than buyers).

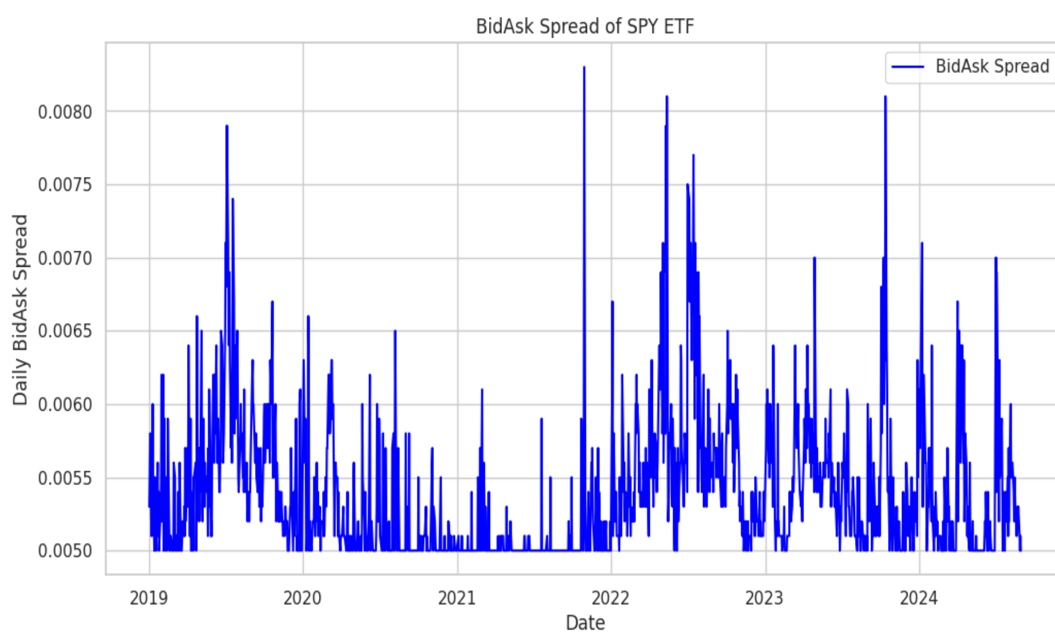

Figure 1: SPY ETF Bid-Ask Spread

Additionally, evidence from Bloomberg data demonstrates that the bid-ask spread varies over time, as shown in Figure 1. The transaction cost model is enhanced with new estimates generated via the SARIMA-X model, utilizing data from the seven most traded ETFs (QQQ, HYG, SPY, VEA, TLT, GLD, and EEM) spanning from January 2009 to May 2024, including test data from 2022 to May 2024.This approach ensures that the model utilizes the most current and relevant data, thereby improving its accuracy. The SARIMA-X model provides more precise transaction cost predictions by dynamically adjusting to market conditions and accounting for time-varying volatility. By incorporating exogenous variables such as trading volume, market volatility, and return, the model captures a comprehensive view of the factors affecting transaction costs.

Performance evaluation is conducted using realized future data that was unavailable during the initial optimization phase, thus validating the model's effectiveness with actual data. Among the evaluated models (ARIMA-X, SARIMA-X, and GARCH-X), the SARIMA-X model demonstrated the best performance. Its ability to manage volatility and include multiple exogenous variables makes it the most effective model for transaction cost prediction. Therefore, the SARIMA-X model is used for transaction cost estimates, enhanced with the findings of Boyd and Oprisor, ensuring accuracy and effectiveness in transaction cost prediction [1], [23].

2.5 Advanced Forecasting Models

Multi-Period Optimization requires several sets of forecasted data. This forecasted data includes asset returns, as well as the data needed to calculate transaction costs such as volatility, volume, and bid-ask spread. There are many different forecasting methods available in the literature. This section evaluates several advanced forecasting models that best fit the purpose: ARIMA-X, SARIMA-X, and GARCH-X. Following subsections gives these evaluations.

2.5.1 ARIMA-X Model

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful method used to forecast time series data. This model integrates autoregressive (AR), differencing (I), and moving average (MA) components to predict future values. Box and Jenkins emphasized the effectiveness of the ARIMA model in capturing the dynamics of financial data [29]. The ARIMA model is widely used to predict trends and patterns in dynamic and volatile environments such as financial markets.

Building on the foundational principles of the ARIMA model, the ARIMA-X (Autoregressive Integrated Moving Average with Exogenous Variables) model extends these concepts by incorporating exogenous variables which are any other related time series data than the data being forecasted. This allows the ARIMA-X model to consider not only past data points but also external factors that may influence the time series, resulting in more accurate predictions. Various studies have demonstrated the efficacy of the ARIMA and ARIMA-X models in financial time series analysis, with ARIMA being suitable for univariate time series forecasting, while ARIMA-X performs well in multivariate analysis [36], [37]. These models have been already applied to forecast commodity prices, such as electric city prices or natural gas [40], [51], [52].

The ARIMA-X model is represented by equation (5)

$$\begin{aligned} y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\ + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \\ + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + \epsilon_t \end{aligned} \tag{5}$$

where y_t represents the data value at time t, c is a constant term, ϕ_i are the autoregressive model parameters, ϵ_t is the error term, θ_i are the moving average model parameters, β_i are the parameters for the exogenous variables, and X_{t-i} represents the exogenous variables at time t.

For instance, ARIMA-X model was used in the context of the price of natural gas forecasting on monthly and quarterly periods. It incorporated temperature as an exogenous variable into the ARIMA framework, which highlighted the effectiveness of this approach in enhancing the accuracy of natural gas price forecast [51].

Sun and Koch analyzed water quality parameters in Apalachicola Bay using Box-Jenkins ARIMA models and cross-correlation techniques, identifying the impact of various control variables such as river discharge, wind stress, and water levels on salinity [53].

These studies illustrate the versatile application of the ARIMA model across different fields. The ARIMA-X model, by incorporating external factors, provides a powerful tool for making more precise predictions in financial data analysis. The inclusion of exogenous variables in financial analysis helps better capture market conditions and economic indicators, enhancing the understanding of the complex and dynamic nature of financial markets, leading to more accurate and reliable predictions.

The main goal of this study is to predict transaction costs with high precision by leveraging advanced forecasting methods. The ARIMA-X model's ability to include external influences makes it particularly well-suited for this task as a candidate. This holistic approach ensures that all relevant factors are considered, resulting in improved prediction accuracy and better-informed decision-making in portfolio optimization. By enhancing the forecasting of transaction cost dynamics through external influences, more effective investment strategies can be developed that adapt to changing market conditions.

Using Recursive Feature Elimination (RFE) and Time Series Split methods, the exogenous variables that best predict each target variable were selected.

All single exogenous variables, along with pairs of these variables, were compiled into a list of potential exogenous-variable-sets. This list includes Adjusted-Close, Bid, Ask, Volume, High, VIX-Adjusted-Close, VIX-Open, VIX-High, VIX-Low, Bid-Ask-Spread, and Return variables. With 11 single exogenous variables and 55 pairs of two-exogenous-sets, there are a total of 66 possible initial candidates for each target variable.

While forming these exogenous-variable-set candidates, any sets containing highly correlated variables were removed. For instance, sets such as {Bid-Ask-Spread} (a single exogenous variable candidate) and Bid, Ask (an exogenous-variable-set with two variables), {Bid, Bid-Ask-Spread}, and others were eliminated from the candidate list for the Bid-Ask-Spread target, because Bid-Ask-Spread is the target itself, and Bid, Ask is a highly correlated pair of variables related to the target.

Subsequently, Recursive Feature Elimination (RFE) was performed on the remaining 54 candidate exogenous variable sets. This process was applied to all target variables. The first 10 of the exogenous variable sets with lowest RMSE for Bid Ask Spread target for SPY ETF are shown, as an example, in Table 1. Table 2 shows the best exogenous sets for the Bid Ask Spread and return targets.

Table 1: First 10 of the exogenous variable sets with lowest RMSE values for the target of Bid Ask Spread forecasting and their corresponding RMSE values in ARIMA-X performance search.

Exogenous Variable	RMSE
VIX Open, VIX Low	0.006328095791
VIX High, VIX Low	0.006352514764
VIX Adj Close, VIX High	0.006497670905
VIX Adj Close, VIX Low	0.006903593221
VIX Low	0.006935191358
VIX Adj Close	0.006956465868
VIX Open	0.006977805469
VIX Low, Return	0.006985509293
VIX High, Return	0.006986638317
VIX Adj Close, VIX Open	0.006988341708

Table 2: Exogenous Variables Selected for Target Variables using the RFE Method for ARIMA-X.

Target Variables	Exogenous Variables
Bid-Ask Spread	VIX Open, VIX Low
Return	VIX Adjusted Close, VIX Open

In the model, when incorporating the best predictors, the highly correlated nature of certain variables led to adjustments in their use. For predicting **Bid Ask Spread**, only **VIX Open** and **VIX Low** were used as exogenous variables to avoid multicollinearity. For predicting **Return**, **VIX Adjusted Close** and **VIX Open** were utilized as the exogenous variable. These are all related to return, or differential of close prize. On the other hand, the volume was forecasted using simply a moving average due to its highly unpredictable nature. In addition to these, volatility was also forecasted with the same method. The transaction cost was calculated in Multi-Period Optimization using the forecasted data for the corresponding variables in the transaction cost model, which are the volume, volatility and bid-ask spread.

This approach aims to prevent unnecessary multicollinearity in the model and to achieve more accurate predictions. By focusing on these selected exogenous variables, the analysis provides deeper insights into the dynamics of ETF data, ultimately improving the investment decision-making process. While this model was tested on seven different ETFs, to maintain focus and clarity, only the results for the SPY ETF are presented here. Detailed analyses for the other ETFs can be provided upon request.

2.5.2 SARIMA-X Model

The SARIMA (Seasonal ARIMA) model is used for time series data with seasonal variations by incorporating seasonal components into the ARIMA model.

Hyndman and Athanasopoulos explained that SARIMA models improve accuracy in seasonal data [54]. Au, Saldana Jr., Spanswick, and Santerre utilized the SARIMA-X (SARIMA with exogenous variables) model to forecast electricity consumption in Pennsylvania during the COVID-19 pandemic. The study used electricity consumption data from PPL Electric Utilities, unemployment claims, and COVID-19 case/death data to make predictions with both SARIMA and SARIMA-X models. The results indicated that the SARIMA-X model, with the inclusion of exogenous variables, performed better than the SARIMA model, though the SARIMA model also provided significant accuracy [42]. Arunraj, Ahrens, and Fernandes used the SARIMA-X model for daily forecasts of food retail sales. The study included daily sales data, holiday effects, seasonal influences, and price changes as exogenous variables to build SARIMA and SARIMA-X models. The results showed that the SARIMA-X model achieved higher accuracy with the inclusion of exogenous variables and better captured the seasonal effects in sales data [43].

The SARIMA-X model is mathematically represented as follows equation (6)

$$
\begin{aligned}
y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \\
+ \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-s} \\
+ \Theta_1 E_{t-1} + \Theta_2 E_{t-2} + \cdots + \Theta_r E_{t-r} \\
+ \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + \epsilon_t
\end{aligned} \tag{6}
$$

where y_t represents the data value at time t, c is a constant term, ϕ_i are the autoregressive model parameters, ϵ_t is the error term, θ_i are the moving average model parameters, Φ_i are the **seasonal autoregressive parameters**, ϵ_t are the seasonal error terms, β_i are the parameters for the exogenous variables, and X_{t-i} represents the exogenous variables at time $t-i$. This model formulation allows the inclusion of various influencing factors in the time series analysis.

Traditional ARIMA models, while powerful, do not account for exogenous variables or seasonal effects, limiting their accuracy in complex financial environments. In the context of transaction cost estimation, external factors such as bid ask spread, trading volume, market volatility, and return, along with seasonal patterns, may play a crucial role in influencing transaction costs. By incorporating some of these exogenous variables and seasonal components, the SARIMA-X model provides a more comprehensive view of the factors affecting the time series data, leading to more accurate and reliable predictions. As before, SPY ETF was used to evaluate the efficacy of this model to compare with other models studied in this section.

A search of a suitable exogenous variable set was undertaken with an identical search approach that has been done for ARIMA-X among roughly 50 possible candidates. This process was applied to all target variables. The first 10 of the exogenous variable sets with lowest RMSE for Bid Ask Spread target of SPY ETF are shown, as an example, in Table 3. Table 4 shows the best exogenous sets for the Bid Ask Spread and return targets.

In the SARIMA-X model, the best predictors were carefully chosen to mitigate the

Table 3: First 10 of the exogenous variable sets with lowest RMSE values for the target of Bid Ask Spread forecasting and their corresponding RMSE values in SARIMA-X performance search.

Exogenous Variable	RMSE
Adj Close, Ask	0.005948896044
Bid, VIX Low	0.006341827484
Adj Close, VIX Low	0.006354175975
Ask, VIX Low	0.006366898956
High, VIX Low	0.006378544769
VIX Adj Close	0.006397666288
Adj Close, High	0.006512953804
Ask, High	0.006528826393
Bid, High	0.006534549486
High	0.006570495921

Table 4: Exogenous Variables Selected for Target Variables using the RFE Method for SARIMA-X.

Target Variables	Exogenous Variables
Bid Ask Spread	Adj Close, Ask
Return	VIX Adjusted Close, VIX Open

effects of multicollinearity. For predicting Bid Ask Spread, Bid and VIX Low were used as exogenous variables, as they proved to be the most significant predictors. For predicting Return, VIX Adjusted Close and VIX Open were employed as the exogenous variables due to its relatively better predictive power. The volume was forecasted using simply a moving average due to its highly unpredictable nature, as before. In addition to these, volatility was also forecasted with the same method. Similarly, although the model was tested on seven different ETFs, the results for the SPY ETF are presented to maintain focus and clarity. Detailed analyses for the other ETFs are available upon request.

2.5.3 GARCH-X Model

ARCH (Autoregressive Conditional Heteroskedasticity) Model developed by Robert F. Engle recognizes that financial market time series data are sensitive to volatility, which can change over time [38]. Building on the principles of the ARCH model, Tim Bollerslev introduced the GARCH (Generalized Autoregressive Conditional Heteroskedasticity). The GARCH model extends the ARCH model by incorporating both autoregressive and moving average components, allowing for a more comprehensive capture of long-term volatility dynamics in time series data [55].

The success of the GARCH model led to further enhancements by including exogenous variables, resulting in the GARCH-X model (Generalized Autoregressive Conditional Heteroskedasticity with Exogenous Variables). This model is particularly effective in complex systems like financial markets. Engle and Rangel demonstrated that adding exogenous variables improves the model's ability to explain financial volatility [40]. The formula for the GARCH-X model is equation (7):

$$\begin{aligned}\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \cdots + \alpha_q\epsilon_{t-q}^2 \\ +\beta_1\epsilon_{t-1}^2 + \beta_2\epsilon_{t-2}^2 + \cdots + \beta_p\epsilon_{t-p}^2\end{aligned} \tag{7}$$

The GARCH-X model was chosen for this study primarily to enhance the accuracy of transaction cost predictions, a critical component of the research. Traditional GARCH models, while effective in modeling time-varying volatility, do not account for exogenous variables. By incorporating these exogenous variables, the GARCH-X model provides a more comprehensive view of the factors affecting the time series data, leading to more accurate and reliable predictions. This model has also demonstrated strong performance in forecasting prices and interest rates, as noted by Han and Kristensen [56].

In this study, the GARCH-X model was applied to forecast various target variables such as Bid-Ask Spread and Return using SPY ETF data, with SPY being one of the seven ETFs analyzed.

A search of a suitable exogenous variable set was undertaken with an identical search approach that has been done for ARIMA-X among roughly 50 possible candidates. This process was applied to all target variables. The first 10 of the exogenous variable sets with lowest RMSE for Bid Ask Spread target of SPY ETF are shown, as an example, in Table 5. Table 6 shows the best exogenous sets for the Bid Ask Spread and return targets.

Table 5: First 10 of the exogenous variable sets with lowest RMSE values for the target of Bid Ask Spread forecasting and their corresponding RMSE values in GARCH-X performance search.

Exogenous Variable	RMSE
VIX Adj Close, VIX Low	0.006370735099
VIX Low	0.006376101106
VIX Open, Return	0.006459251950
Return	0.006475220131
VIX High	0.006480641736
VIX Adj Close	0.006526830882
VIX Adj Close, Return	0.006727377464
VIX Open, VIX High	0.006750071613
VIX Adj Close, VIX Open	0.009576818684
VIX Open	0.019390152086

Table 6: Exogenous Variables Selected for Target Variables using the RFE Method for GARCH-X.

Target Variables	Exogenous Variables
Bid-Ask Spread	VIX Adjusted Close, VIX Low
Return	VIX Adjusted Close, VIX Open

VIX Adjusted Close and VIX Low were used for predicting bid ask spread. VIX Adjusted Close and VIX Open were used as the exogenous variable for predicting the asset return. Although the model was tested on seven different ETFs, the results for the SPY ETF are presented to maintain focus and clarity. Detailed analyses for the other ETFs are available upon request.

As a result, the SARIMA-X was chosen to forecast data due its relatively better performance compared to the other two for the rest of this study.

CHAPTER III

EMPIRICAL ANALYSIS

This section outlines the empirical analysis conducted to evaluate the effectiveness of the multi period optimization framework using advanced forecasting methods. The main objectives are to compare ARIMA-X, SARIMA-X, and GARCH-X models based on AIC, BIC, RMSE, and MAE metrics, and to assess their impact on portfolio optimization. Then, the results of multi period optimization for several baseline cases and for this study is given.

3.1 Data Description

3.1.1 Data Sources

The data used in this analysis comprises daily financial data from the seven most traded ETFs: QQQ, HYG, SPY, VEA, TLT, GLD, and EEM. These ETFs were selected because they represent a diverse range of asset classes and market sectors, making them ideal for robust portfolio optimization. The data was sourced from the Bloomberg Terminal, covering the period from January 2009 to May 2024. Below are brief descriptions of each ETF:

3.1.2 Data Preprocessing

Data was preprocessed as follows:

Data Cleaning: Handled missing values and normalized the data. Outliers and gaps were identified and treated to ensure the data integrity and consistency.

Stationarity Check: The ADF test is used to check if the time series data is stationary. Dickey and Fuller developed the ADF test to determine if a time series contains a unit root, thereby assessing its stationarity [35]. Meerschaert and Scheffler

Table 7: Overview and Description of the Selected Seven ETFs.

ETF	Description
QQQ:	Tracks the Nasdaq-100 Index, focusing on large technology and growth companies in the U.S.
HYG:	Provides exposure to high-yield corporate bonds, offering higher income but with increased credit risk.
SPY:	Tracks the S&P 500 Index, representing the largest companies across various sectors of the U.S. stock market.
VEA:	Focuses on large and mid-sized companies in developed markets outside the U.S.
TLT:	Offers exposure to long-term U.S. Treasury bonds, often used to hedge against stock market volatility.
GLD:	Reflects the price of gold, commonly used as a safe-haven asset during economic uncertainty.
EEM:	Tracks large and mid-sized companies in emerging markets, providing higher growth potential with added risk.

noted that the ADF test is widely used in financial time series analysis [57]. As including the ADF test results for all seven ETFs would occupy extensive space, the results for the SPY ETF are presented as an example in Table 8.

Table 8: ADF Test Results for SPY ETF Target Variables.

Metric	Bid-Ask Spread	Return	Transaction Cost
ADF Statistics	-55.7666	-11.766	-3.7709
p-value	0.0	1.1e-21	0.0032

3.1.3 Descriptive Statistics

3.1.3.1 Interpretation of Descriptive Statistics for ETF Target Variables

The descriptive statistics of the seven ETFs "SPY, GLD, VEA, TLT, HYG, QQQ, and EEM" provide important insights into the behavior of these funds over the analysis period, given in Table 9. These ETFs represent a range of asset classes, from equities to bonds and commodities, making them critical in portfolio management and optimization.

Table 9: Descriptive statistics for ETF data.

	SPY	**GLD**	**VEA**	**TLT**	**HYG**	**QQQ**	**EEM**
Count	5619	5619	5619	5619	5619	5619	5619
Mean	0.00041	0.00021	0.0002	0.000066	0.000562	0.00022	0.00018
Standard Dev.	0.00939	0.00831	0.0104	0.008123	0.010900	0.01216	0.00505
Minimum	-0.1094	-0.0878	-0.1118	-0.06668	-0.11979	-0.12479	-0.0549
25th (Q1)	-0.0014	-0.0022	-0.0022	-0.00284	0.001658	-0.00341	-0.0007
Median (Q2)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
75th (Q3)	0.00322	0.00304	0.0036	0.003228	0.004091	0.00461	0.00142
Maximum	0.09060	0.04912	0.0889	0.075196	0.084705	0.09246	0.06546
Skewness	-0.4212	-0.3941	-0.6659	0.07266	-0.29165	-0.2266	0.01255
Kurtosis	15.0240	7.3128	13.224	6.05777	9.546663	9.526932	24.8509

Mean and Median: The mean daily returns for all ETFs are close to zero, with SPY having the highest mean return (0.00041) and EEM the lowest (0.00018). This indicates that, on average, these ETFs have generated modest daily gains over the period. The median (Q2) for all ETFs is also zero, suggesting that half of the observed returns are either zero or slightly positive/negative.

Standard Deviation: Standard deviation reflects the volatility of the ETFs. QQQ (0.01216) and SPY (0.00939) exhibit the highest volatility, indicating that their daily returns fluctuate more compared to TLT (0.00812), which is less volatile. This aligns with the risk profile of these assets, where equity ETFs like QQQ are expected to be more volatile than bond

Minimum and Maximum: The minimum and maximum values provide insights into the extreme return observations for each ETF. For instance, QQQ experienced the largest negative return (-0.12479) and also one of the highest positive returns (0.09246), highlighting its susceptibility to larger price swings. On the other hand, TLT and EEM show narrower ranges, with lower maximum and minimum returns, indicating more stable returns.

Skewness: Most ETFs exhibit negative skewness, particularly VEA (-0.6659), indicating a longer tail on the left side of the return distribution. This suggests that these ETFs may experience more extreme negative returns than positive ones. EEM, with a slight positive skewness (0.01255), stands out as having a more balanced return distribution, with no particular bias towards extreme losses or gains.

Kurtosis: The kurtosis values are significantly higher than a normal distribution for most ETFs, with EEM showing an extremely high kurtosis of 24.8509. This indicates that the returns of these ETFs exhibit "fat tails" or a higher probability of extreme events, either positive or negative. This is particularly important for investors, as it highlights the risk of outliers in the returns of these assets.

Conclusion: These descriptive statistics provide a comprehensive overview of the ETFs' volatility, and return, in general their behavior. The higher volatility and extreme kurtosis observed in QQQ and SPY make them attractive to risk-seeking investors, while the lower volatility and moderate skewness in TLT and HYG suggest more stability. These insights are valuable for constructing a diversified portfolio, balancing high-growth potential with low-risk investments.

3.2 Model Implementation

3.2.1 ARIMA-X Model

The ARIMA-X model extends the ARIMA model by incorporating external variables that influence the time series. This model helps in capturing the effects of exogenous factors on the target variables, leading to more accurate forecasts.

Parameter Selection: ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) graphs are used to determine the order of AR and MA components in the ARIMA model. Box and Jenkins highlighted the importance of ACF and PACF graphs in determining ARIMA model parameters [29].

Exogenous Variable Selection: RFE (Recursive Feature Elimination) is used to identify the most important features to enhance the model's predictive accuracy. Guyon and Elisseeff explained that RFE is an effective technique for improving model performance [58]. The method involves the following steps:

- Ranking the features by importance
- Iteratively removing the least important features
- Re-evaluating the model performance

Then, an exhaustive search over all possible cases and all passible pair of those cases were studied as a second method, which is computation intensive but yielded the most accurate selection by brute force.

3.2.2 SARIMA-X Model

The SARIMA-X model extends the ARIMA model by including seasonal components and exogenous variables. This model is particularly useful for time series data with strong seasonal patterns.

Parameter Selection: Similar to the ARIMA-X model, ACF and PACF graphs were used to select the parameters (p, d, q) for the non-seasonal part and (P, D, Q, s)

for the seasonal part. The seasonal differencing parameter (D) was determined based on the seasonality observed in the data. Hyndman and Athanasopoulos provided comprehensive guidelines on identifying and estimating these parameters for SARIMA models [54].

Exogenous Variable Selection: The RFE method was applied to select the exogenous variables for each target variable. Then, an exhaustive search over all possible cases and all passible pair of those cases were studied as a second method, which is computation intensive but yield the most accurate selection by brute force.

SARIMA-X Model Configurations:

- Bid-Ask Spread model, a SARIMA-X $(1, 0, 1)$ $x(1, 0, 1, 12)$ model was used.
- For the Return model, a SARIMA-X $(1, 0, 1)$ $x(1, 0, 1, 12)$ model was used.
- TransactionCost model, a SARIMA-X $(1, 0, 1)$ $x(1, 0, 1, 12)$ model was used.

3.2.3 GARCH-X Model

The GARCH-X model extends the GARCH model by including exogenous variables. This model is effective in modeling time-varying volatility in financial time series, incorporating the influence of external factors [56].

Parameter Selection: Parameters for the AR and GARCH components were selected based on AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) to ensure the best fit. Exogenous variables were selected using the RFE method, consistent with the approach suggested by Engle and Rangel [55].

Exogenous Variable Selection: The RFE method was applied to select the exogenous variables for each target variable. Then, an exhaustive search over all possible cases and all possible pair of those cases were studied as a second method, which is computation intensive but yield the most accurate selection by brute force.

Model Configuration: For each target variable, the GARCH-X model was configured as follows:

- BidAskSpread: AR-X + GARCH(1,1)
- Return: AR-X + GARCH(1,1)
- TransactionCost: AR-X + GARCH (RMSE, and MAE metrics to assess their forecasting accuracy. The results showed that the SARIMA-X model consistently outperformed the other models across the key metrics, particularly RMSE and MAE, which are crucial for achieving accurate forecasts in financial contexts.

Given SARIMA-X's superior performance based on these metrics, it was chosen for integration into the Multi-Period Optimization (Multi-Period Optimization) framework. By using SARIMA-X, the framework ensures robust and reliable forecasts, leading to more effective portfolio optimization and better management of transaction costs in dynamic financial environments.

3.3 Optimization Results

3.3.1 Integration into the Multi-Period Optimization Framework

The Multi-Period Optimization requires several types of forecasted data, including return forecasts and all other data related to transaction costs, such as volatility, volume, and bid-ask spread forecasts. These forecasted data were produced using various methods. Forecasting volume is challenging due to its unpredictable nature over shorter time frames, such as days and weeks. Therefore, a moving average is used to address this issue. All other forecasted data is produced using the SARIMA-X model, which was chosen for its performance in general compared to other methods. The optimization was executed across several cases by varying the forecast horizon of $H = 1, 2, 5, 10$ days as illustrated in Table 10. The longest horizon was chosen to be 10 days, because forecasting beyond about 10 days is expected to introduce both statistical and systematic errors which are large enough to be not reliable.

Outer search included all combinations of $\gamma^{\text{risk}} = 0.1, 0.3, 1, 3, 10, 30, 100, 300, 1000$ and $\gamma^{\text{trade}} = 1, 2, 5, 10, 20$.

Integration of advanced forecasted transaction cost to multi period optimization was examined by comparing it with several benchmark optimization setups which includes a setup for the expected lower limit for the performance and expected theoretical upper limit for the performance.

Boyd's Original Work: This benchmark uses Boyd's original setup, with asset forecasting done via moving averages and fixed transaction costs alone. It serves as a baseline for comparison as it lacks both a proper forecasting of returns values and transaction cost is not properly factored in.

Return Forecasting with Gaussian Noise: Here, the forecast values for the ETFs are determined by adding Gaussian noise to the real return values, as suggested by Boyd in his study, to assess the performance of the Multi-Period Optimization framework without focusing on the forecasting which is not the main focus of the study. This is another benchmarking scenario. The sigma of the Gaussian noise was determined to be comparable to the volatility of the asset, the ETF.

Real transaction costs: Instead of forecasting transaction costs, this case incorporates actual transaction costs for yet another benchmarking purpose. The goal is to evaluate how well the Multi-Period Optimization framework performs when the real transaction costs are taken into account in a precise manner. This represents an upper limit for the expected improvement when the transaction cost function devised by this study is incorporated into the Multi-Period Optimization. The case for the forecasted transaction cost is not expected to be better than this benchmark.

SARIMA-X transaction cost forecasting: This is the basis for this study. Here, the data required to calculate the transaction cost were forecasted using the moving average or SARIMA-X model for the variables needed by the transaction cost, which are bid ask spread, volume and volatility. An annualized growth rate

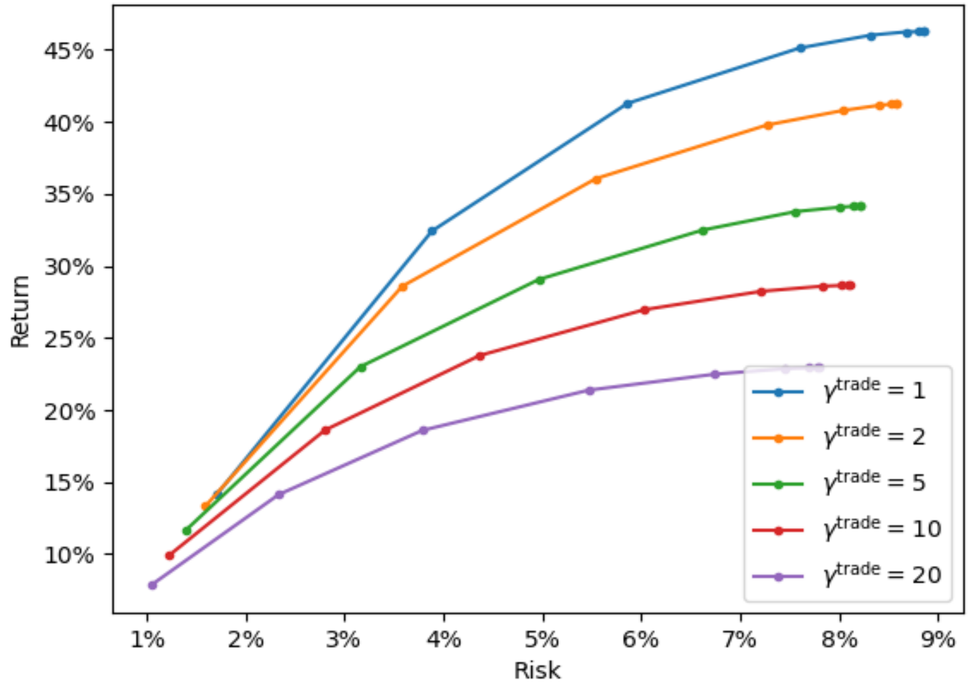

Figure 2: Return versus risk relation ($H = 2$) for $\gamma^{\text{trade}} = 1, 2, 5, 10, 20$ and $\gamma^{\text{risk}} = 0.1, 0.3, 1, 3, 10, 30, 100, 300, 1000$.

of 34% with a Sharpe ratio close to 3 was achieved as shown in Table 10, indicating strong performance. However, this result is contingent on accurate asset return/value forecasting, which is beyond the scope of this study. The analysis revealed that using the SARIMA-X model for transaction cost forecasting, in conjunction with the Multi-Period Optimization framework, led to superior results, particularly in maximizing the Sharpe ratio.

Figure 2 illustrates the testing of various scenarios across 45 different combinations of risk and trade factors, with the primary objective of maximizing the Sharpe ratio a critical metric that assesses the return of an investment relative to its risk.

Figure 3 illustrates the cumulative growth of the portfolio, demonstrating a consistent upward trend in portfolio value. This growth underscores the framework's ability to capture market opportunities and generate substantial returns over the analysis

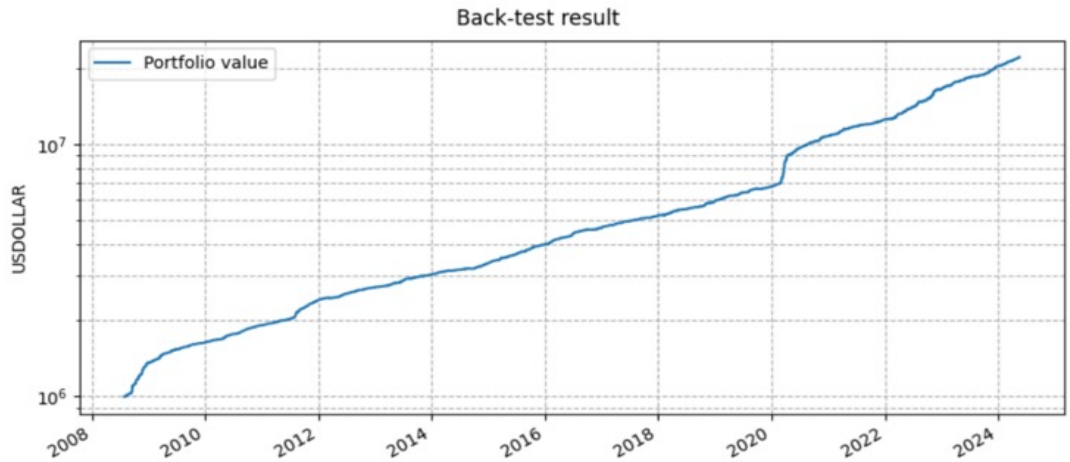

Figure 3: Cumulative growth of the portfolio.

period. It corresponds roughly annualized average yearly return of 34%.

Figure 4 provides a detailed view of the portfolio's asset allocation dynamics. The frequent adjustments in portfolio weights reflect the framework's responsiveness to market changes, ensuring an optimal balance between risk and return at all times.

Figure 5 shows the portfolio's leverage and turnover rates, highlighting the strategic use of leverage to enhance returns and the efficiency of the rebalancing process. Despite the higher turnover, the use of SARIMA-X helped manage transaction costs, contributing to the framework's overall performance.

Figure 4: Portfolio allocation over time.

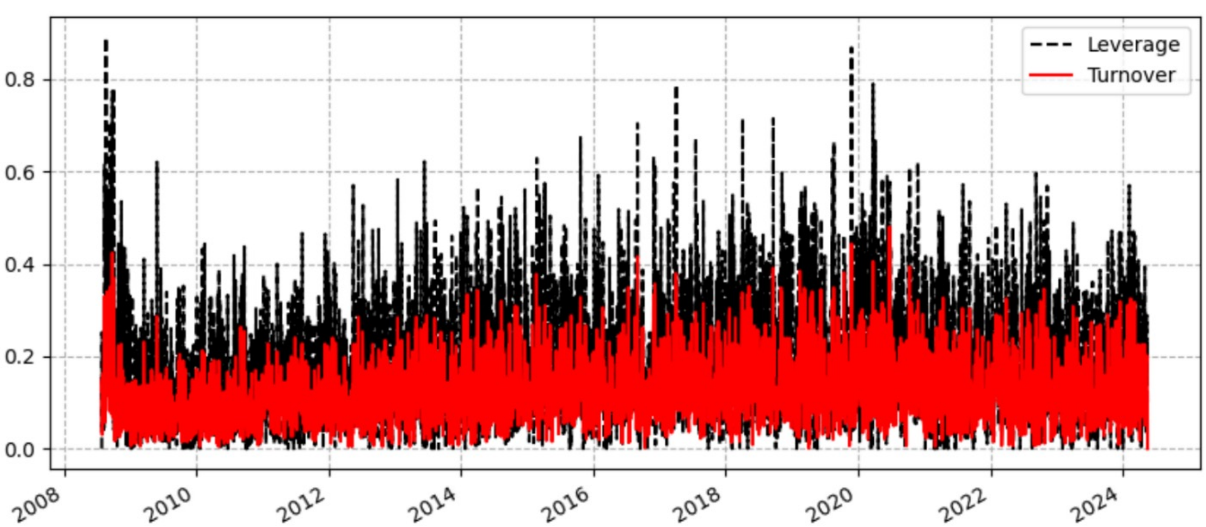

Figure 5: Portfolio's leverage and turnover rates.

Table 10: Largest Sharpe Ratio. Gammas are indicating risk-aversion coefficient as defined in MPO.

H=1	γ^{trade}	γ^{risk}	Growth (%)	Sharpe
Boyd Benchmark (MA-FC) Flat T_{cost}	1	1	2.3	0.15
Return FC=r+Gauss Flat T_{cost}	1	300	30.3	2.72
Return FC=r+Gauss Real T_{cost}	1	300	29.8	2.93
Return FC=r+Gauss T_{cost} FC = SARIMA-X	2	1000	12.4	2.98

H=2	γ^{trade}	γ^{risk}	Growth (%)	Sharpe
Boyd Benchmark (MA-FC) Flat T_{cost}	1	0.1	3.3	0.17
Return FC=r+Gauss Flat T_{cost}	1	300	33.2	2.79
Return FC=r+Gauss Real T_{cost}	2	1000	13.4	2.97
Return FC=r+Gauss T_{cost} FC = SARIMA-X	1	300	32.8	2.97

H=5	γ^{trade}	γ^{risk}	Growth (%)	Sharpe
Boyd Benchmark (MA-FC) Flat T_{cost}	2	0.1	3.4	0.17
Return FC=r+Gauss Flat T_{cost}	1	300	34.8	2.90
Return FC=r+Gauss Real T_{cost}	1	300	34.1	2.96
Return FC=r+Gauss T_{cost} FC = SARIMA-X	1	300	33.9	2.96

H=10	γ^{trade}	γ^{risk}	Growth (%)	Sharpe
Boyd Benchmark (MA-FC) Flat T_{cost}	5	0.1	3.3	0.17
Return FC=r+Gauss Flat T_{cost}	1	300	35.1	2.81
Return FC=r+Gauss Real T_{cost}	1	300	34.3	3.08
Return FC=r+Gauss T_{cost} FC = SARIMA-X	1	300	34.0	2.98

CHAPTER IV

CONCLUSION

Integration of advanced forecasting models, specifically GARCH-X, ARIMA-X, and SARIMA-X, into a multi-period portfolio optimization framework, with a particular focus on transaction cost modeling was explored in this study. The primary objective was to study the often overlooked but significant impact of transaction costs on portfolio performance. By incorporating exogenous variables such as bid-ask spreads, and trading volume into time series models, this study has demonstrated that advanced forecasting techniques on transaction cost can enhance both accuracy and portfolio optimization outcomes.

Throughout the empirical analysis, multiple benchmark cases were employed to evaluate the performance of the enhanced optimization framework. The baseline case used Boyd's original setup, which assumed fixed transaction costs and relied on simple moving averages for asset return forecasting. In comparison to this traditional approach, the inclusion of more sophisticated forecasting techniques in this study provided a notable improvement in portfolio returns and risk-adjusted performance metrics.

In particular, the SARIMA-X model for transaction cost forecasting demonstrated a noticeable improvement over simpler approaches, yielding an annualized average growth rate of 34% with a Sharpe ratio approaching to 3. This was compared to the baseline case where fixed transaction costs were used, which resulted in much lower growth rates and Sharpe ratios. The results highlighted the importance of accurate transaction cost forecasting in improving the overall performance of the optimization model.

Additional benchmarks were employed to assess the role of different transaction cost models. For example, the "real transaction costs" benchmark, which incorporated actual historical transaction costs into the optimization process, was used as an upper bound for the expected performance. While the SARIMA-X model did not surpass this benchmark, it came remarkably close, illustrating that the predictive power of SARIMA-X is reliable even in highly volatile and dynamic market conditions. The "Gaussian noise" forecast, simulating returns with added noise, served as a lower-bound benchmark, with significantly poorer performance metrics compared to SARIMA-X, further reinforcing the latter's effectiveness.

The study also explored different forecast horizons (H = 1, 2, 5, 10 days) to understand the sensitivity of portfolio performance to the length of the forecast period. The SARIMA-X model consistently outperformed in all scenarios, demonstrating its robustness across different time horizons and market conditions.

The results of this thesis underscore the critical role of dynamic transaction cost modeling in portfolio optimization. Traditional single-period optimization approaches, which either ignore or simplify transaction costs, often lead to suboptimal investment strategies that fail to account for the fluctuating nature of costs in real-world trading environments. By integrating dynamic transaction cost forecasts into a multi-period optimization framework, this study demonstrated how investors can more accurately balance risks and returns while minimizing unnecessary costs. This approach not only improves portfolio returns but also ensures more efficient asset allocation and trading decisions.

While this study has made contributions in improving multi-period portfolio optimization through advanced transaction cost modeling, there are several areas for future research.

1. Exploring More Complex Models: The models employed in this thesis, while effective, can be extended to include even more sophisticated machine learning

techniques which have been shown to capture complex time dependencies and non-linearities in financial data.

2. Incorporating Liquidity Constraints: Future work could also investigate the incorporation of liquidity constraints into the portfolio optimization framework. Transaction costs are often influenced by market liquidity, and integrating liquidity forecasting models could lead to even more accurate predictions and optimized trade execution strategies.

3. Macro-economic and Behavioral Factors: Another avenue for exploration is the inclusion of macroeconomic indicators and behavioral factors that influence market sentiment. Factors such as interest rates, GDP growth, and investor sentiment could be modeled as additional exogenous variables to further enhance the predictive capabilities of the transaction cost model.

4. Real-time Adaptive Optimization: Developing a real-time adaptive optimization model could allow the portfolio to react more dynamically to market conditions as new data becomes available. This could involve integrating reinforcement learning techniques into the existing multi-period framework, enabling the model to learn and adapt its trading strategies continuously.

5. Application to Different Asset Classes: Although this study focused on ETFs, future research could apply the framework to other asset classes such as commodities, cryptocurrencies, or foreign exchange markets. Each asset class has unique cost structures and volatility dynamics, and testing the model across these different markets could provide valuable insights into its generalizability.

Ultimately, this study not only advances the field of portfolio optimization by demonstrating the critical value of dynamic transaction cost modeling but also lays the groundwork for future innovations that can further refine investment strategies in increasingly complex and unpredictable financial markets.

REFERENCES

[1] S. Boyd, E. Busseti, S. Diamond, R. N. Khan, K. Koh, P. Nystrup, and J. Speth, "Multi-period trading via convex optimization," *Foundations and Trends in Optimization*, vol. 3, no. 1, pp. 1–76, 2017.

[2] H. M. Markowitz, "Portfolio selection," *Journal of Finance*, vol. 7, pp. 77–91, 1952.

[3] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. Wiley, 1959.

[4] P. N. Kolm, R. Tütüncü, and F. J. Fabozzi, "60 years of portfolio optimization: Practical challenges and current trends," *European Journal of Operational Research*, vol. 234, no. 2, pp. 356–371, 2014.

[5] J. Tobin, "Liquidity preference as behavior towards risk," *The Review of Economic Studies*, vol. 25, no. 2, pp. 65–86, 1958.

[6] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The Journal of Finance*, vol. 19, no. 3, pp. 425–442, 1964.

[7] J. Mossin, "Optimal multiperiod portfolio policies," *The Journal of Business*, vol. 41, no. 2, pp. 215–229, 1968.

[8] M. Lobo, M. Fazel, and S. Boyd, "Portfolio optimization with linear and fixed transaction costs," *Annals of Operations Research*, vol. 152, no. 1, pp. 341–365, 2007.

[9] P. A. Samuelson, "Lifetime portfolio selection by dynamic stochastic programming," in *Stochastic Optimization Models in Finance*, pp. 517–524, 1975.

[10] R. C. Merton, "Optimum consumption and portfolio rules in a continuous-time model," in *Stochastic Optimization Models in Finance*, pp. 621–661, Academic Press, 1975.

[11] P. Nystrup, H. Madsen, and E. Lindström, "Dynamic portfolio optimization across hidden market regimes," *Quantitative Finance*, vol. 18, no. 1, pp. 83–95, 2018.

[12] G. M. Constantinides, "Multiperiod consumption and investment behavior with convex transactions costs," *Management Science*, vol. 25, no. 11, pp. 1127–1137, 1979.

[13] B. Dumas and E. Luciano, "An exact solution to a dynamic portfolio choice problem under transaction costs," *The Journal of Finance*, vol. 46, no. 2, pp. 577–595, 1991.

[14] M. H. A. Davis and A. R. Norman, "Portfolio selection with transaction costs," *Mathematics of Operations Research*, vol. 15, no. 4, pp. 676–713, 1990.

[15] J. Skaf and S. Boyd, "Multi-period portfolio optimization with constraints and transaction costs." Working Manuscript, 2009.

[16] D. P. Bertsekas, *Dynamic Programming and Optimal Control.* Athena Scientific, 1995.

[17] R. Bellman, "Dynamic programming and lagrange multipliers," *Proceedings of the National Academy of Sciences*, vol. 42, no. 10, pp. 767–769, 1956.

[18] N. Gârleanu and L. H. Pedersen, "Dynamic trading with predictable returns and transaction costs," *The Journal of Finance*, vol. 68, no. 6, pp. 2309–2340, 2013.

[19] S. Boyd, M. T. Mueller, B. O'Donoghue, and Y. Wang, "Performance bounds and suboptimal policies for multi–period investment," *Foundations and Trends in Optimization*, vol. 1, no. 1, pp. 1–72, 2013.

[20] H. Anis and R. Kwon, "A shrinking horizon optimal liquidation framework with lower partial moments criteria," *Journal of Computational Finance*, 2019. Forthcoming.

[21] F. Cesarone, A. Scozzari, and F. Tardella, "A new method for mean-variance portfolio optimization with cardinality constraints," *Annals of Operations Research*, vol. 205, pp. 213–234, 2013.

[22] J. Y. Campbell and L. M. Viceira, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors.* Clarendon Lectures in Economics, Oxford University Press, 2002.

[23] X. Li, A. S. Uysal, and J. M. Mulvey, "Multi-period portfolio optimization using model predictive control with mean-variance and risk parity frameworks," *European Journal of Operational Research*, vol. 299, no. 3, pp. 1158–1176, 2022.

[24] R. Cho, K. Thoney, and T. Hodgson, "Rolling horizon scheduling of multi-factory supply chains," in *Proceedings of the 2003 Winter Simulation Conference*, vol. 2, pp. 1409–1416, IEEE, 2003.

[25] J. A. Primbs and C.-H. Sung, "A stochastic receding horizon control approach to constrained index tracking," *Asia-Pacific Financial Markets*, vol. 15, pp. 3–24, 2008.

[26] P. Nystrup, S. Boyd, E. Lindström, and H. Madsen, "Multi-period portfolio selection with drawdown control," *Annals of Operations Research*, vol. 282, no. 1, pp. 245–271, 2019.

[27] F. Herzog, G. Dondi, and H. P. Geering, "Stochastic model predictive control and portfolio optimization," *International Journal of Theoretical and Applied Finance*, vol. 10, no. 2, pp. 203–233, 2007.

[28] G. E. P. Box, G. M. Jenkins, and J. F. MacGregor, "Some recent advances in forecasting and control," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 23, no. 2, pp. 158–179, 1974.

[29] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[30] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[31] V. c. Ediger and S. Akar, "Arima forecasting of primary energy demand by fuel in turkey," *Energy Policy*, vol. 35, no. 3, pp. 1701–1708, 2007.

[32] H. J. Bierens, "Armax model specification testing, with an application to unemployment in the netherlands," *Journal of Econometrics*, vol. 35, no. 1, pp. 161–190, 1987.

[33] M. As'ad, "Finding the best arima model to forecast daily peak electricity demand," Master's thesis, 2012.

[34] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the arima model on the covid-2019 epidemic dataset," *Data in Brief*, vol. 29, p. 105340, 2020.

[35] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427–431, 1979.

[36] Y. Li, Y. Su, and L. Shu, "An armax model for forecasting the power output of a grid connected photovoltaic system," *Renewable Energy*, vol. 66, pp. 78–89, 2014.

[37] P. Ďurčák and S. Považanová, "Arima vs. arima-x–which approach is better to analyze and forecast macroeconomic time series," in *Proceedings of 30th International Conference Mathematical Methods in Economics*, vol. 2, pp. 136–140, 2012.

[38] M. Y. Anwar, J. A. Lewnard, S. Parikh, and V. E. Pitzer, "Time series analysis of malaria in afghanistan: using arima models to predict future trends in incidence," *Malaria Journal*, vol. 15, pp. 1–10, 2016.

[39] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[40] W. Anggraeni, K. B. Andri, and F. Mahananto, "The performance of arima-x model and vector autoregressive (var) model in forecasting strategic commodity prices in indonesia," *Procedia Computer Science*, vol. 124, pp. 189–196, 2017.

[41] G. P. Zhang, “Time series forecasting using a hybrid arima and neural network model,” *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[42] J. Au, J. Saldaña Jr, B. Spanswick, and J. Santerre, “Forecasting power consumption in pennsylvania during the covid-19 pandemic: A sarima-x model with external covid-19 and unemployment variables,” *SMU Data Science Review*, vol. 3, no. 2, p. 6, 2020.

[43] N. S. Arunraj, D. Ahrens, and M. Fernandes, “Application of sarima-x model to forecast daily sales in the food retail industry,” *International Journal of Operations Research and Information Systems*, vol. 7, no. 2, pp. 1–21, 2016.

[44] R. Oprisor and R. Kwon, “Multi-period portfolio optimization with investor views under regime switching,” *Journal of Risk and Financial Management*, vol. 14, no. 1, p. 3, 2020.

[45] H. Wang, X. Li, W. Hong, and K. Tang, “Multi-objective approaches to portfolio optimization with market impact costs,” *Memetic Computing*, vol. 14, no. 4, pp. 411–421, 2022.

[46] S. Diamond and S. Boyd, “Cvxpy: A python-embedded modeling language for convex optimization,” *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[47] W. McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, 2012.

[48] E. Busseti, S. Diamond, S. Boyd, and BlackRock, “Cvxportfolio.” urlhttps://github.com/cvxgrp/cvxportfolio, 2017.

[49] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[50] Grinold and Kahn, *Active portfolio management*. McGraw-Hill, 2000.

[51] L. M. Liu and M. W. Lin, “Forecasting residential consumption of natural gas using monthly and quarterly time series,” *International Journal of Forecasting*, vol. 7, no. 1, pp. 3–16, 1991.

[52] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, “Forecasting dynamic day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark,” *Applied Energy*, vol. 293, p. 116983, 2021.

[53] H. Sun and M. Koch, “Time series analysis of water quality parameters in an estuary using box-jenkins arima models and cross correlation techniques,” in *Computational Methods in Water Resources XI*, pp. 230–239, 1996.

[54] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice.* OTexts, 2018.

[55] R. F. Engle and J.-M. Rangel, "The spline-garch model for low-frequency volatility and its global macroeconomic causes," *The Review of Financial Studies*, vol. 21, no. 3, pp. 1187–1222, 2008.

[56] H. Han and D. Kristensen, "Asymptotic theory for the qmle in garch-x models with stationary and nonstationary covariates," *Journal of Business & Economic Statistics*, vol. 32, no. 3, pp. 416–429, 2014.

[57] M. M. Meerschaert and H.-P. Scheffler, *Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice*, vol. 321. John Wiley & Sons, 2001.

[58] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

VITA

She completed her Bachelor's degree in Physics at Boğaziçi University in 2009. Subsequently, she completed her MS degree in Biomedical Engineering, specializing in Medical Systems and Informatics, in 2012. After her academic studies, she worked as a Sales Support Director at a foreign biomedical devices company for approximately 10 years, where she specialized in providing strategic sales support and managing customer relations within the biomedical sector.

In the past three years, she has been working as a Financial Specialist at Center of Financial Engineering of Özyeğin University, where she focuses on financial analysis, portfolio management, and overseeing research projects. Her research is in the area of portfolio management strategies.