TESIS DE LA UNIVERSIDAD
DE ZARAGOZA

2025

54

Yifu Jiang

Portfolio Selection Using Advanced Optimization Methods

Director/es
Atwi Saab, Majed
Olmo Badenas, José

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

PORTFOLIO SELECTION USING ADVANCED OPTIMIZATION METHODS

Autor

Yifu Jiang

Director/es

Atwi Saab, Majed
Olmo Badenas, José

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

2024

Universidad Zaragoza

1542

Tesis Doctoral

Portfolio Selection Using Advanced Optimization Methods

Autor

Yifu Jiang

Director/es

Olmo Badenas, José
Atwi Saab, Majed

Departamento de Análisis Económico
Universidad De Zaragoza

2024

Universidad Zaragoza
1542

Doctoral Thesis

Portfolio Selection Using Advanced Optimization Methods

Author

Yifu Jiang

Supervisors

Olmo Badenas, José
Atwi Saab,Majed

Department of Economic Analysis
University of Zaragoza

2024

Acknowledgments

During my doctoral study, I am deeply grateful for the multitude of individuals who have supported and guided me throughout this academic journey. I extend my thanks with the utmost sincerity to all my supervisors, professors, peers, and family who have been a part of this remarkable chapter in my life.

I want to express my gratitude to my supervisors, Professors José Olmo and Majed Atwi, who supported me during my doctoral study. At the forefront of my acknowledgments is Professor José Olmo, whose unwavering support and guidance have been instrumental in my academic growth. From my initial steps in the world of research to now, where I can confidently navigate the complexities of my field, Professor José Olmo's profound knowledge, professional ethos, and rigorous approach to scholarship have been invaluable. The open-minded mentorship and the freedom to explore my research interests have shaped my academic pursuits. The approachability and patience with which Professor José Olmo has guided me through numerous challenges and provided me with the motivation and courage to persevere. Studying under Professor José Olmo has indeed been an enriching and treasured experience in my life.

I extend my gratitude to Professor Majed Atwi for his meticulous instruction and wise advice. The stringent requirements and the emphasis on rigorous research methodology while writing my dissertation have taught me discipline. I will carry forward in my future endeavors. Professors José Olmo and Majed Atwi have been actively involved in the evolution of my research, from topic selection to the finalization of my dissertation. Their extensive knowledge and dedication to academic excellence have profoundly influenced my academic development, and I am eternally grateful for the lessons I have learned.

During my time in Zaragoza, I have been fortunate to engage in various academic activities that have broadened my horizons. I am immensely thankful to my coordinator, Professor Lola Gadea, whose insights and expertise have been a source of inspiration. I would also like to thank the departmental secretary, Teresa Ortas, for her assistance and guidance. Additionally, I am grateful to the friends I have met in Spain, for their support and encouragement.

Lastly, I give special thanks to my family: my father, mother, and husband. Their unwavering support and understanding have been my strength. Their love and encouragement have been the foundation upon which I have built my academic success.

In conclusion, this dissertation is not just evidence of my academic achievements but also a reflection of the collective efforts and contributions of the many individuals who have been a part of this journey. I am deeply grateful to each of you, and I look forward to carrying the lessons learned and the relationships forged into the next chapter of my life.

Agradecimientos

Durante mis estudios de doctorado, estoy profundamente agradecida a la multitud de personas que me han apoyado y guiado a lo largo de este viaje académico. Extiendo mi agradecimiento con la mayor sinceridad a todos mis supervisores, profesores, compañeros y familiares que han formado parte de este notable capítulo de mi vida.

Quiero expresar mi gratitud a mis supervisores, los profesores José Olmo y Majed Atwi, que me han apoyado durante mis estudios de doctorado. En el primer lugar de mis agradecimientos se encuentra el profesor José Olmo, cuyo inquebrantable apoyo y orientación han sido fundamentales en mi crecimiento académico. Desde mis primeros pasos en el mundo de la investigación hasta ahora, cuando puedo navegar con confianza por las complejidades de mi campo, los profundos conocimientos del profesor José Olmo, su ética profesional y su riguroso enfoque de la erudición han sido de un valor incalculable. Su tutoría abierta y la libertad para explorar mis intereses de investigación han dado forma a mis actividades académicas. La accesibilidad y paciencia con la que el profesor José Olmo me ha guiado a través de numerosos retos y me ha proporcionado la motivación y el coraje para perseverar. Estudiar con el profesor José Olmo ha sido una experiencia enriquecedora y valiosa en mi vida.

Extiendo mi gratitud al profesor Majed Atwi por su meticulosa instrucción y sabios consejos. Los estrictos requisitos y el énfasis en una metodología de investigación rigurosa durante la redacción de mi tesis me han enseñado disciplina. La llevaré adelante en mis futuros empeños. Los profesores José Olmo y Majed Atwi han participado activamente en la evolución de mi investigación, desde la selección del tema hasta la finalización de mi tesis. Sus amplios conocimientos y su dedicación a la excelencia académica han influido profundamente en mi desarrollo académico, y les estoy eternamente agradecida por las lecciones que he aprendido.

Durante mi estancia en Zaragoza, he tenido la suerte de participar en diversas actividades académicas que han ampliado mis horizontes. Estoy inmensamente agradecida a mi coordinadora, la profesora Lola Gadea, cuyos conocimientos y experiencia han sido una fuente de inspiración. También me gustaría dar las gracias a la secretaria del departamento, Teresa Ortas, por su ayuda y orientación. Además, agradezco a los amigos que he conocido en España, por su apoyo y aliento.

Por último, doy las gracias especialmente a mi familia: mi padre, mi madre y mi marido. Su apoyo inquebrantable y su comprensión han sido mi fuerza. Su amor y su aliento han sido los cimientos sobre los que he construido mi éxito académico.

En conclusión, esta tesis no es sólo una prueba de mis logros académicos, sino también un reflejo de los esfuerzos colectivos y las contribuciones de las muchas personas que han formado parte de este viaje. Estoy profundamente agradecida a cada una de ellas y espero poder trasladar las lecciones aprendidas y las relaciones forjadas al próximo capítulo de mi vida.

Abstract

Portfolio selection is a critical area in financial economics and investments. However, optimal portfolio selection still faces many challenges, like the dynamic nature of the market, the uncertainty of extreme events, and the complexity of high-dimensional data. Therefore, robust portfolio selection models are crucial in financial investment to improve portfolios' risk management capability and preserve investors' wealth, especially in extreme events, such as financial crises and COVID-19 pandemics.

It is noticeable that recent literature on financial and portfolio theory is rapidly incorporating machine learning (ML) techniques and deep reinforcement learning (DRL) for better decision-making. Applying ML techniques not only enhances the ability to process time-series data but also improves insights to make optimal decisions for investors. Moreover, DRL has gained attention for its ability to solve complex financial decision-making problems, especially showing efficiency in large-scale financial markets. This thesis is devoted to optimal portfolio selections from the following three perspectives.

The first study proposes a dynamic robust portfolio selection model using the worst-case conditional value (WCVaR) at an objective function. The proposed robust model for the dynamics of portfolio constituents has three main features: i) accommodates tail dependence between assets employing a mixture of copula functions; ii) conditional heteroscedasticity and leverage effects are considered through the implementation of a GJR-GARCH model; and iii) extreme events are taken into account by considering parametric and semiparametric hybrid models for the marginal distribution of asset returns. Empirical results verify the portfolio performance superiority (i.e., Sharpe ratio, cumulative returns, and volatility) of the proposed WCVaR portfolio method before and during the COVID-19 pandemic against benchmark portfolios commonly used by practitioners.

The second study designs an advanced model-free DRL framework to construct optimal portfolio strategies in dynamic, complex, and large-dimensional financial markets. Investors' risk aversion and transaction cost constraints are embedded in an extended Markowitz's mean-variance reward function. To do this, this study implements a twin-delayed deep deterministic policy gradient (TD3) algorithm. The proposed DRL-TD3-based risk and transaction cost-sensitive portfolio method combines advanced exploration strategies and dynamic policy updates, which effectively addresses the challenges of the high-dimensional portfolio optimization problem. An empirical application illustrates this methodology to obtain two optimal portfolios by flexibly controlling both transaction cost and portfolio risk with (i) the constituents of the Dow Jones Industrial Average and (ii) the constituents of the S&P100 index. The results show better portfolio performances of the proposed DRL portfolio method compared to several competitors from the traditional DRL methods under different scenarios.

The third study proposes a novel investment strategy based on DRL for long-term portfolio allocation in the presence of transaction costs and risk aversion. We design an advanced portfolio policy framework to model the price dynamic patterns using convolutional neural networks (CNN), capture group-wise asset dependence using WaveNet, and solve the optimal asset allocation problem using DRL. These methods are embedded within a multi-period Bellman equation framework. An additional appealing feature of our investment strategy is its ability to optimize dynamically over a large set of potentially correlated risky assets. The performance of this portfolio is tested empirically over different holding periods, risk aversion levels, transaction cost rates, and financial indices. The results demonstrate the effectiveness and superiority of the proposed long-term portfolio allocation strategy compared to several competitors based on machine learning methods and traditional optimization techniques.

All three studies are closely related to modern portfolio theory, but each focuses on different market conditions and decision-making challenges. These investigations can provide investors with more comprehensive and flexible investment strategies under various market conditions.

Resumen

La selección de carteras es un área crítica de la economía financiera y las inversiones. Sin embargo, la selección óptima de carteras sigue enfrentándose a muchos retos, como la naturaleza dinámica del mercado, la incertidumbre de los acontecimientos extremos y la complejidad de los datos de alta dimensión. Por lo tanto, los modelos sólidos de selección de carteras son cruciales en la inversión financiera para mejorar la capacidad de gestión del riesgo de las carteras y preservar la riqueza de los inversores, especialmente en acontecimientos extremos, como las crisis financieras y las pandemias.

Es notable que la literatura reciente sobre teoría financiera y de carteras esté incorporando rápidamente técnicas de aprendizaje automático (ML) y aprendizaje por refuerzo profundo (DRL) para una mejor toma de decisiones. La aplicación de técnicas de aprendizaje automático no solo mejora la capacidad de procesar datos de series temporales, sino que también mejora la información para tomar decisiones óptimas para los inversores. Además, DRL ha llamado la atención por su capacidad para resolver problemas complejos de toma de decisiones financieras, especialmente mostrando eficiencia en mercados financieros a gran escala. Esta tesis está dedicada a la selección óptima de cartera desde las siguientes tres perspectivas.

El primer estudio propone un modelo dinámico de selección de cartera robusto utilizando el valor condicional del peor de los casos (WCVaR) en una función objetivo. El modelo robusto propuesto para la dinámica de los componentes de la cartera tiene tres características principales: i) se adapta a la dependencia de cola entre activos que emplean una combinación de funciones de cópula; ii) se consideran los efectos de heterocedasticidad condicional y apalancamiento mediante la implementación de un modelo GJR-GARCH; y iii) los eventos extremos se tienen en cuenta considerando modelos híbridos paramétricos y semiparamétricos para la distribución marginal de los rendimientos

de los activos. Los resultados empíricos verifican la superioridad del desempeño de la cartera (es decir, el índice de Sharpe, los rendimientos acumulados y la volatilidad) del método de cartera WCVaR propuesto antes y durante la pandemia de COVID-19 frente a las carteras de referencia comúnmente utilizadas por los profesionales.

El segundo estudio diseña un marco DRL avanzado sin modelos para construir estrategias de cartera óptimas en mercados financieros dinámicos, complejos y de grandes dimensiones. La aversión al riesgo de los inversores y las limitaciones de los costos de transacción están integradas en una función de recompensa de varianza media ampliada de Markowitz. Para hacer esto, este estudio implementa un algoritmo de gradiente de política determinista profundo (TD3) doble retardado. El método de cartera sensible a los costos de transacción y riesgo basado en DRL-TD3 propuesto combina estrategias de exploración avanzadas y actualizaciones dinámicas de políticas, lo que aborda de manera efectiva los desafíos del problema de optimización de cartera de alta dimensión. Una aplicación empírica ilustra esta metodología para obtener dos carteras óptimas controlando de manera flexible tanto el costo de transacción como el riesgo de la cartera con (i) los componentes del Dow Jones Industrial Average y (ii) los componentes del índice S&P100. Los resultados muestran mejores desempeños de cartera del método de cartera DRL propuesto en comparación con varios competidores de los métodos DRL tradicionales en diferentes escenarios.

El tercer estudio propone una nueva estrategia de inversión basada en DRL para la asignación de carteras a largo plazo en presencia de costos de transacción y aversión al riesgo. Diseñamos un marco de política de cartera avanzado para modelar los patrones dinámicos de precios utilizando redes neuronales convolucionales (CNN), capturar la dependencia de activos grupales utilizando WaveNet y resolver el problema de asignación óptima de activos utilizando DRL. Estos métodos están integrados dentro de un marco de ecuaciones de Bellman de períodos múltiples. Una característica atractiva adicional de nuestra estrategia de inversión es su capacidad de optimizar dinámicamente un gran conjunto de activos de riesgo potencialmente correlacionados. El desempeño de esta cartera se prueba empíricamente en diferentes períodos de tenencia, niveles de aversión al riesgo, tasas de costos de transacción e índices financieros. Los resultados demuestran la efectividad y superioridad de la estrategia de asignación de cartera a largo plazo propuesta en comparación con varios competidores basados en métodos de aprendizaje automático y técnicas de optimización tradicionales.

Los tres estudios están estrechamente relacionados con la teoría moderna de carteras, pero cada uno se centra en diferentes condiciones del mercado y desafíos en la toma de decisiones. Estas investigaciones pueden proporcionar a los inversores estrategias de inversión más completas y flexibles en diversas condiciones del mercado.

List of Publications

The following works were published/submitted for publication during the process of this dissertation.

1. Jiang Yifu, Olmo Jose, & Atwi Majed. (2024). Dynamic robust portfolio selection under market distress. *North American Journal of Economics and Finance, 69*, 102037. https://doi.org/10.1016/j.najef.2023.102037.
2. Jiang Yifu, Olmo Jose, & Atwi Majed. (2024). Deep reinforcement learning for portfolio selection. Accepted by *Global Finance Journal*.
3. Jiang Yifu, Olmo Jose, & Atwi Majed. (2024). High-dimensional multi-period portfolio allocation using deep reinforcement learning. Submitted to *International Review of Financial Analysis.*

Index of Contents

List of Figures

List of Tables

CHAPTER 1

GENERAL INTRODUCTION

1.1 Background and importance

Due to the financial crisis, financial markets have experienced dramatic volatility and increased uncertain events, like the European debt crisis, the COVID-19 pandemic, and the collapse of Silicon Valley Bank (SVB). While the markets have generally maintained a stable and upward trend over the long term due to a series of measures taken by governments and central banks, it cannot be denied that the short-term influence caused by unforeseen occurrences is indeed substantial. These events can trigger unexpected fluctuations in the market, weaken investor confidence, and have a wider effect on economic activity. Therefore, portfolio selection is a crucial subject in both financial economics and investment decision-making. The evolution of the mean-variance optimization method by Markowitz (1952) provides a theoretical framework for constructing portfolios.

Investors themselves face increasingly complex market dynamics and diverse constraints in the rapidly changing financial markets (Winkelmann, 2004). From fluctuations in economic cycles to constraints on individual risk preferences, these factors have a profound impact on decision-making (Vieira & Filomena, 2020). Consequently, studying optimal portfolio selection under market dynamics and constraints is of great theoretical and practical significance. It not only helps investors to make more analytical and rational investment decisions in the changing market environment but also is the key to increasing investment returns and controlling risks for investors. Moreover, this thesis explores the strategy of adaptive adjustment through portfolio optimization under diversified market conditions. At the algorithmic level, this thesis introduces a comparative analysis of machine learning (ML) with traditional portfolio methods, which provides innovative perspectives on financial economics, as well as practical decision-making tools for financial decision-makers. The next subsections explain the importance of these problems in portfolio theory.

1.1.1 Portfolio allocation and robust optimization

Value-at-Risk (VaR) has been widely explored in the investment literature, which aims to construct optimal portfolios that consider quantile risk measures. The literature on portfolio allocation using this measure to capture risk has grown steadily over the last twenty years. Related literature includes the mean-risk model introduced by Fishburn (1977), which can be considered an early extension of standard mean-variance formulations. Other important contributions considering the VaR quantiles as constraints in the asset allocation optimization exercise are Basak and Shapiro (2001), Krokhmal et al. (2001), Campbell et al. (2001), Wu and Xiao (2002), Bassett et al. (2004), Engle and Manganelli (2004) and Ibragimov and Walden (2007), among others. This literature illustrates the properties of VaR-optimal portfolios while acknowledging considerable computational difficulties (Gaivoronski & Pflug, 2005; Rachev et al., 2007).

The central idea of robust portfolio optimization is to use uncertainty sets for the unknown parameters characterizing the distribution functions of the asset returns instead of only point estimates and to compute portfolios whose worst-case performance is optimal. In this scenario, worst-case performance is considered portfolio performance under the least favorable combination of the model parameters within the uncertainty set. Examples of robust portfolio optimization using the CVaR as an objective function include Ghaoui et al. (2003), Zhu and Fukushima (2009), and Hellmich and Kassberger (2011). More recently, Deng and Liang (2021) explored time series copula models, and Su et al. (2021) solved the portfolio selection problem using regime-switching models.

A major factor influencing the optimal portfolio decision is the behavior of the portfolio constituents in the tails and, more specifically, the likelihood of extreme events. A flexible specification of the multivariate distribution of the portfolio constituents is the use of copula models, see Sklar (1959). This modeling strategy can capture different forms of dependence in the tails of the multivariate return distribution. These techniques have been used for portfolio allocation problems in Low (2018) and Garcia and Tsafack (2011), among many others. A particular case of copula model in high dimensions is vine copulas, see Weiß and Scheffer (2015). These copulas model the multivariate dependence in a vector of random variables pairwise. Another example of utilizing the copula model in portfolio allocation problems is given in Kakouris and Rustem (2014). These authors construct a mixture copula model embedded in a portfolio selection problem using CVaR and WCVaR measures as objective functions. However, the heavy-tailed nature of the distribution of asset returns is not considered in these models. Belhajjam et al. (2017) proposed a multivariate extreme VaR method to obtain the optimal portfolio weights. This method applies extreme value theory (EVT) techniques to model the behavior of asset returns in the tails. It considers a copula model for the joint tail dependence between the assets in the portfolio. An important drawback of these studies is the reliance on bivariate copula models, which limits the scope of methods' application in practice.

The presence of dynamics in the conditional multivariate distribution of asset returns is a stylized fact. Under the assumption of linearity and Gaussianity of asset returns, it is sufficient to model the dynamics of the multivariate distribution through the second moments of pairs of asset returns. In particular, generalized autoregressive conditional heteroskedasticity (GARCH) type models, see Bollerslev (1986), and versions of it considering different features of the data, have been the main workhorse for modeling the presence of conditional heteroscedasticity in asset returns. The Conditional Constant Correlation model of Bollerslev (1990) and the Dynamic Conditional Correlation (DCC) model of Engle (2002), see also Aielli (2013), are among the most popular. It is well documented, though, that the dynamics of asset returns are nonlinear and exhibit stylized facts such as asymmetries and heavy tails. In this scenario, modeling the conditional variance and correlations may not be sufficient to capture the effect of tail dependencies. To correct for this, a more suitable approach is to consider dynamic copula models, see Patton (2006) as seminal contribution. Weiß (2013), Karmakar and Paul (2019) or Sun et al. (2020) combine time varying multivariate GARCH models with dynamic copulas to describe the dependence and risk of asset returns. Han, Li, and Xia (2017) and Han et al. (2020) presented several dynamic robust portfolio optimization models obtained from DCC-Copula-GARCH model specifications for describing the dynamic dependence between bivariate assets. These authors capture high-dimensional dependencies using R-vine and C-vine copulas but do not consider extreme portfolio losses.

1.1.2 Application of deep reinforcement learning algorithm in financial asset allocation

In recent years, some literature has incorporated methods from the ML literature to improve the predictions of stock returns and, with it, the dynamic allocation of assets to investment portfolios. In particular, Rubesam (2022) investigated ML investment portfolio construction techniques and demonstrated the advantages of portfolio learning over traditional approaches. Mavruk (2022) considered ML methods to preselect stocks prior to the portfolio formation stage, and Fereydooni and Mahootchi (2023) implemented ML methods to improve investment decisions in financial markets. RL models are a type of ML that applies a different set of mathematical tools used for perception and representation learning. RL models are useful for sequential decision-making (Ngo et al., 2023). Goodell et al. (2021) provided a comprehensive review of ML finance tools, including how their hybridization can effectively address large and complex portfolio problems across various domains. They covered innovative techniques like the Markov decision process (MDP; Almahdi & Yang, 2017; Peck & Yang, 2011), Q-learning (QL), deep QL (DQL), proximal policy optimization (PPO), and deep deterministic strategy gradients (DDPG; Lillicrap et al., 2015). Additional techniques are commonly used for portfolio management, such as actor-critic (AC) networks (Aboussalah & Lee, 2020). Moody et al. (1998, 2001) designed a recurrent RL asset allocation method that used market data from the S&P500 index and US Treasury bill data.

QL has been applied to search for the optimal weights used to optimize portfolio asset allocations (Halperin, 2019; Zeng & Klabjan, 2018). The long short-term memory (LSTM) network improves the forecasting ability of DL and RL models, and Li et al. (2021a) used one to predict financial returns in global commodity markets. Similarly, Ta et al. (2020) investigated their application to quantitative trading and optimization-based stock prediction. Furthermore, Almahdi and Yang (2017, 2019) combined recurrent RL and a particle-swarm technique for portfolio weight allocation and market constraint, demonstrating that this method achieves better cumulative returns than those obtained from maximizing the contemporary Sharpe ratio (SR) and mean-variance approaches. Zhang and Maringer (2016) used a genetic algorithm to enhance portfolio trading performance with RL models.

Pigorsch and Schafer (2022) were among the first to develop a DQL portfolio trading strategy in a cross-sectional setting. Park et al. (2020) and Shavandia and

Khedmati (2022) employed a multi-agent DQL algorithm to construct high-return portfolios. Huang and Tanaka (2022) designed a modularized and scalable multi-agent deep Q-network (DQN) to handle large-scale portfolios with heterogeneous data. Notably, DQL provides the algorithmic framework for DQN, which was initially tailored for discrete action spaces. Aboussalah and Lee (2020) addressed this limitation by suggesting stacked deep dynamic recurrent RL models. These models provide real-time portfolio management based on continuous actions in multidimensional state spaces. This approach maximizes expected returns while guaranteeing portfolio risk-tolerance constraints. Lin et al. (2020), Wang and Ku (2022) applied DDPG portfolio strategies to maximize total returns while maintaining a risk diversification objective.

An important challenge for portfolio trading management is accommodating the underlying market risk while accounting for transaction costs (Kircher & Rsch, 2021; Moallemi & Saglam, 2015). Notably, transaction costs were ignored in earlier developments (Almgren & Chriss, 2001; Choi et al., 2019; Gaivoronski & Pflug, 2005; Li et al., 2018; Park et al., 2019). However, ignoring transaction costs can lead models to recommend over-aggressive portfolio trading. Some studies that include these costs as objective functions have partially addressed this issue (Ma et al., 2019; Qureshi et al., 2017; Roni & Jean-Luc, 1996; Zhang et al., 2011). Model-free RL algorithms were developed to control the impact of transaction costs adaptively. Building upon the work of Betancourt and Chen (2021), Zhao et al. (2023) applied deep reinforcement learning (DRL) to develop portfolio policies and automatically execute transactions in such settings. Notably, a DRL-based portfolio management strategy on markets with transaction costs is proposed by Betancourt and Chen (2021). Similarly, Xu and Dai (2022) used RL to derive hedging strategies that maximize profits while considering the effects of various transaction costs and underlying assets.

Jang and Seong (2023) combined DRL with traditional portfolio theory, showing that DRL portfolios outperform other ML strategies. However, they did not tackle risk aversion and transaction costs in optimizing the investment portfolio. Recent work by Cui et al. (2023) provided a DRL algorithm that explicitly embeds the presence of risk alongside risk aversion objectives for portfolio decision-making. Garcia-Galicia et al. (2019) and Wang and Zhou (2020) proposed models that consider transaction costs and mean-variance linear constraints in the portfolio objective functions. Furthermore, Moody and Saffell (2001) applied RL to optimize risk-adjusted returns while handling the impact of different transaction costs. Bühler et al. (2018) investigated how to use standard DRL algorithms to build nonlinear reward structures under transaction costs, liquidity, and risk-sensitive constraints, thus verifying the effectiveness of the portfolio trading method. Zhang et al. (2022) introduced transaction costs and risk-sensitive portfolio management methods based on DRL to maximize total returns. Li et al. (2018) and Yang et al. (2020) developed ensemble trading models using AC, PPO, and DDPG methods and embedded transaction costs and risk aversion factors into a novel cost-sensitive reward function. Moreover, Sebastian et al. (2021) proposed a DRL-based repeated portfolio method that accounts for asset variability and mutual asset correlations. The empirical results confirmed the ability of the method to achieve superior performance over previous RL and traditional portfolio selection methods.

High-dimensional portfolios benefit greatly from the advantages of diversification when containing a large number of assets. However, their management strategies require data-rich environments that present an obstacle to normal portfolio construction (Fernandez-Arjona & Filipovic, 2022). Pigorsch and Schafer (2022) is one of the first studies to propose a DQL method that constructs high-dimensional portfolios in a cross-sectional setting. Meanwhile, Bühler et al. (2018) developed an advanced DRL method for hedging a portfolio of derivatives under market constraints, including transaction costs, liquidity limits, and risk-tolerance levels. The empirical results confirm the superiority of the proposed DRL methods against standard approaches to portfolio optimization in large dimensions. Notably, the performance advantage increases with the number of assets in the portfolio.

1.1.3 Long-term portfolio allocation in the dynamic market

Long-term portfolio allocation has been a critical area of research in finance, focusing on how investors can optimally allocate their assets over extended periods to maximize returns while managing risks. The dynamic market portfolio allocation problem involves allocating assets among multi-periods under changing market conditions. This problem includes fluctuations in asset prices, changes in the economic environment, and other factors that affect investment decisions. Unlike the single-period portfolio problem, the multi-period dynamic market portfolio allocation involves considering multiple points in time. At each point in time, investors need to re-value the market conditions and adjust their portfolios accordingly. To extend the single-period portfolio management to multi-period portfolio management, Zhang et al. (2013) developed the possibilistic expected value and variance to solve the multi-period fuzzy portfolio selection problems, and this method is proven effective to maximize the portfolio value after multi-period. Liu et al. (2019) identified the multi-period fuzzy market portfolio selection problem and presented an approach-based differential evolution method that takes realistic constraints into account and performs beneficially for complex portfolio selection models. To determine the effects of the optimal solution in multi-period portfolio allocation, Hibiki (2006) compared the hybrid model to the tree stochastic model, and the result indicates that the hybrid model is more suitable for identifying the optimal asset. However, traditional algorithms may struggle to capture and handle complex nonlinear relationships, especially when multiple assets and multi-periods are involved. This may lead to a degradation of the model's performance in solving complex investment problems. Reinforcement learning (RL), on the other hand, is an emerging field of research that can be used to make investment methods adaptable to dynamic markets. Moreover, a novel DRL method was suggested by Cui et al. (2023) and applied to multi-period portfolio selection. The approach proved beneficial to various investors in terms of making risk-adjusted return-based decisions.

1.2 Motivation and objectives

Given the dynamic nature of financial markets, investors prioritize portfolio diversification and risk management nowadays. Therefore, it is important to provide investors with a more robust and efficient investment strategy to maximize portfolio returns at a given level of risk during periods of market fluctuations. Considering the prevailing conditions, the motivations of the thesis can be outlined in the following points.

- Various methods (e.g., mean-variance, min-variance) have been used to address portfolio selection problems, which involve allocating investment funds among different financial assets to maximize investment returns or minimize risk. However, traditional methods typically assume static market conditions and may fail to consider the dynamic nature of financial markets. When constructing a portfolio, it is crucial to consider the impact of extreme market conditions on investment decisions and managing portfolio tail risks, especially for low-probability events that can lead to large losses. In this case, developing robust modeling can be a powerful tool for identifying and quantifying tail risks under market uncertainty.

- With the development of ML and deep learning (DL) technologies, investors can leverage these innovative instruments to maximize the process of making investment decisions. Despite the success of traditional ML methods in financial decision-making, there are still challenges in dealing with complex and highly dimensional markets. How to effectively extract the market dynamics and perform model-free portfolio optimization still needs to be investigated. This thesis is also motivated to provide a new perspective on financial decision-making by combining DL with RL, especially when dealing with high-dimensional and nonlinear optimization problems.
- Traditional portfolio approaches usually struggle to adapt to the complexity of multi-period investing, especially in capturing the dynamics of asset dependencies and considering the impact of investment constraints, which are critical in long-term investing. Due to the limitations of traditional models, this thesis is inspired to develop an advanced DRL framework combined with DL to address these challenges and construct optimal portfolio strategies in a multi-period investment problem.

The objectives of this thesis can be summarized as follows:

- Firstly, Chapter 2 aims to develop a dynamic, robust portfolio selection model based on worst-case scenarios. Moreover, the objective is to capture the tail dependence of asset returns and extreme events by combining the GJR-GARCH model and EVT, ensuring that the model can perform well during the market stress period. By incorporating a mixed copula model, this chapter aims to capture the dependency relationships of asset returns and adjust portfolio weights under uncertainty and extreme market conditions to improve the robustness and adaptability of the portfolio.

- The application of DRL has become a major breakthrough in finance. The nonlinear and non-stationary nature of financial markets makes it difficult for traditional portfolio selection models to predict asset return accurately. DRL models, particularly those based on neural networks, can capture these complex nonlinear relationships. Chapter 3 intends to address the limitations of modern portfolio theory by incorporating advanced approaches. By doing so, this chapter designs efficient DRL models for investors to construct portfolios that are better suited to dynamic financial environments. Through comparative analysis with traditional algorithms, the goal is to demonstrate the potential benefits of DRL methods in portfolio management for providing practical decision support for investors under different transaction cost rates and risk aversion levels.

- To address the problems faced by portfolios in multi-period scenarios, the objective of Chapter 4 is to explore and implement an advanced portfolio optimization strategy that utilizes a novel DRL-TD3 framework to adapt to dynamic market changes while considering market constraints. Moreover, the developed portfolio selection model is capable of handling multi-period investment decisions and high-dimensional asset portfolios for long-term investment requirements. Meanwhile, the innovative instrument provided for financial decision-makers should allow them to make rational decisions not only in the short term but also in the long term.

1.3 Contributions

Given portfolio management's vital role in the financial markets, this thesis's findings are valuable for further academic research and industry implications. The primary contributions of this thesis are summarized below.

- A dynamic and robust portfolio optimization method is developed to minimize the consequences of the tail events (see Chapter 2). The proposed dynamic, robust portfolio selection model is based on conditional value-at-risk and optimizes portfolios during periods of market stress, reducing extreme risks. Moreover, the dynamic volatility of asset returns and tail dependencies is captured by combining the GJR-GARCH-EVT model with mixed copula models, thereby enhancing the risk management capabilities of the portfolio. At the same time, the impact of transaction costs on portfolio returns is considered. Most importantly, the empirical analysis demonstrates the superiority of the proposed model compared to traditional portfolio strategies, such as minimum variance and equally weighted portfolios, particularly during the COVID-19 pandemic.

- To efficiently process and analyze large-scale datasets, an advanced model-free DRL method is proposed. The approach combines DL and RL approach to construct optimal portfolios (see Chapter 3). The developed model performs particularly well in high-dimensional settings by implementing the Twin-Delayed deep deterministic policy gradient (TD3) algorithm. Additionally, to make the strategies become more practical and better meeting investor needs, the risk aversion and transaction cost constraints are embedded in the reward function. Significantly, the empirical result shows the outperformance of the proposed model compared to traditional portfolio strategies, including the recent models proposed in the DRL literature.

- A multi-period risk-averse and transaction cost-awareness portfolio selection model is presented to maximize the returns for investors (see Chapter 4). Meanwhile, an advanced portfolio policy framework is designed to extract asset price dynamic patterns and capture the cross-asset dependencies to enhance decision-making capabilities under the DRL framework. Investors can leverage their strategies to process multi-period investment decisions for high-dimensional assets effectively. Moreover, the findings suggest that the model not only achieves a higher investment portfolio return and Sharpe ratio but also is notably effective and superior in resolving constraints within multi-period portfolio models using real-world data.

1.4 Thesis outline

The structure of this thesis includes a total of five chapters as follows.

Chapter 1 presents a general introduction to portfolio selection and management background, as well as the related literature review of the thesis.

In Chapter 2, a dynamic and robust portfolio selection model is proposed, which is based on minimizing the portfolio's worst-case scenarios using the Conditional Value at Risk as a relevant risk measure. I illustrate the performance of this portfolio before and during the COVID-19 pandemic using statistical measures. The results show the outperformance of the WCVaR portfolio during the turmoil period against benchmark portfolios commonly used by practitioners.

In Chapter 3, an advanced model-free DRL framework is investigated for constructing optimal portfolio strategies in dynamic, complex, and large-dimensional financial markets. Additionally, a TD3 algorithm is implemented to address the challenges posed by high-dimensional state and action spaces in the financial markets.

Chapter 4 describes an advanced portfolio policy framework considering the portfolio selection problem. This framework optimizes the reallocation of high-dimensional assets over multiple periods. Moreover, dependencies between assets are taken into account to more accurately assess the portfolio's risk and build a robust portfolio.

Lastly, Chapter 5 concludes a comprehensive thesis summary and describes recommendations for future research directions based on current research.

CHAPTER 2

DYNAMIC ROBUST PORTFOLIO SELECTION UNDER MARKET DISTRESS

2.1 Introduction

Portfolio selection is an important area in financial economics and investments. Markowitz (1952), in his pioneering work, introduced mean-variance optimal portfolios. The underlying risk of an investment position in these portfolios is determined by the portfolio variance that is, in turn, constructed from the covariance matrix of asset returns. In many settings, the fluctuations of the portfolio return around its mean may not be an appropriate measure of portfolio risk. This is, for example, the case when the interest is in minimizing the portfolio's downside.

An alternative that has been widely explored in the investment literature is to construct optimal portfolios that consider quantile risk measures, a typical example being the Value-at-Risk (VaR). The literature on portfolio allocation using this measure to capture risk has grown steadily over the last twenty years.

Another strand of the literature has focused on the Expected Shortfall, also denominated as Conditional VaR (CVaR), which measures the expected portfolio loss once the return on the portfolio is below the corresponding VaR. This measure has become popular because of its simplicity, but more importantly, because it satisfies a set of properties that characterize the risk measure as coherent. Unfortunately, quantile risk measures such as the VaR do not satisfy these properties, see Artzner et al. (1999). The development of investment portfolios considering the CVaR as the risk measure of interest include Topaloglou et al. (2002), Rockafellar and Uryasev (2002), and Guo et al. (2019), among many others. A related literature (Zhu & Fukushima, 2009) applies robust optimization techniques to construct investment portfolios. This chapter differentiates between two types of events that affect the risk of a portfolio when considering VaR and CVAR measures. The first type corresponds to the probability of extreme losses due to the existence of heavy tails in the marginal distribution of asset returns. The second type concerns the probability of joint dependence in the tails of the multivariate distribution of asset returns.

In this study, I propose a dynamic robust portfolio optimization problem that focuses on minimizing tail events. This strategy is particularly interesting during market distress episodes. The portfolio is robustified by optimizing the CVaR risk measure over a confidence set, that is, estimates of the model parameters are replaced by confidence regions for such parameters. In this way, the proposed portfolio is optimized over the combination of model parameters that results in the worst-case scenario. Moreover, the portfolio strategy can be interpreted as a maximin optimization rule in which investors optimize the conditional VaR measure under the worst-case scenario. Portfolio weights are dynamically adjusted by optimization of the WCVAR measure over rolling windows.

The proposed specification for the dynamics of asset returns is a GJR-GARCH-EVT model, with GJR standing for the asymmetric conditional volatility model introduced in Glosten et al. (1993) and EVT for modeling the tails of the marginal distribution of asset returns. This model is combined with a dynamic specification of a multivariate mixture copula model capable of describing the various potential dependence structures of multiple asset returns. My interest is in capturing joint dynamics in the tails of the multivariate distribution of asset returns. To do this, I combine the GJR-GARCH-EVT with a dynamic mixture copula model given by a mixture of a Gaussian copula, a Student-t copula, and a Clayton copula. The weights defining the mixture copula are obtained by maximum likelihood estimation. This modeling strategy allows us to capture the presence of heavy tails, conditional heteroscedasticity, and nonlinearities in the dependence structure between asset returns. The mixture of copula functions allows us to capture important stylized facts of the joint distribution of asset returns, such as the presence of asymmetric dependence between the lower and upper tail of the joint distribution of asset returns. These stylized facts are particularly important during episodes of market distress. Alternative multivariate parametric models can also be used to capture the joint dynamics of asset returns, such as the multivariate Student-t distribution or the generalized hyperbolic distribution function. Copula functions are, however, superior modeling devices for several reasons. Copulas consider the dependency between the marginal distributions of the random variables instead of focusing directly on the dependency between the random variables themselves. This makes them more flexible than standard distributions because it is possible to separate the selection of the multivariate dependency from the selection of the univariate distributions. These functions are also capable of capturing different forms of extreme tail dependence between the left and right tails of the multivariate distribution. The parametric specification of copula functions allows natural extensions to the time-varying case by introducing dynamics in the parameters characterizing the copula function. Dynamic specifications of copula functions are a simple way of obtaining h-period ahead forecasts of the vector of asset returns that can be applied in scenario-based portfolio selection. In particular, I apply Monte-Carlo methods to simulate the predictive distribution function of multivariate asset returns necessary to obtain the robust optimal portfolio WCVaR model.

The performance of this novel portfolio allocation model is assessed during turmoil periods. To do this, I consider two differentiated periods given by the episode before the outbreak of the COVID-19 pandemic and the episode right afterward. These periods are characterized by very different macroeconomic and financial conditions. This work conducts an empirical study to compare the in-sample and out-of-sample performance of two versions of the proposed GJR-GARCH-EVT model with a mixture copula function against popular benchmark portfolios widely used by practitioners, namely, the minimum variance portfolio, the shrinkage minimum variance portfolio, and the equally-weighted portfolio that are computed over these two periods. The empirical results show strong performance of the proposed approach in both evaluation periods using different performance measures such as the gross return and the Sharpe ratio and highlight the ability of flexible specification to adapt to distress episodes of the market.

This study is closely related to Zhu and Fukushima (2009) and Han et al. (2017). These authors also consider optimal portfolios obtained from maximizing the WCVaR risk measure. However, in contrast to these studies, the introduced model adopts the mixture copula model for the dynamics of multivariate returns and the GJR-GARCH-EVT model to capture the occurrence of extreme events and the presence of conditional heteroscedasticity. These additional features of the proposed model prove the strong performance of the method during stress periods.

2.2 Econometric model

This section is divided into four blocks. The first subsection discusses the GJR-GARCH-EVT model to describe the dynamics of asset returns. The second block introduces a mixture copula model as a modeling device that improves the estimation of the dependence structure for multiple asset returns. The third subsection presents a dynamic robust portfolio selection method that uses the WCVaR as objective function. The last section details an algorithm that implements the proposed methodology.

2.2.1 The GJR-GARCH-EVT model

An interesting feature of financial markets is that asset price volatility is more sensitive to bad news than good news. This is reflected in a strong negative correlation between current returns and future volatility. Volatility decreases when returns increase and increases when returns decrease. This trend is generally referred to as the leverage effect. Traditional GARCH models are not able to capture this stylized fact of the data. This phenomenon is captured, though, by Glosten et al. (1993) in the GJR-GARCH model and Zakoian (1994) in the threshold-GARCH model. These models can reflect the asymmetry inherent in the volatility process through the leverage coefficient. Asset returns have other particular characteristics, such as conditional heteroscedasticity, heavy tails, and negative skewness of the distribution (Low, 2018). In this case, the AR(1)-GJR-GARCH(1,1) model with student-t distribution is an interesting time series process to model the dynamics of asset returns.

In this chapter, I consider the optimal portfolio allocation for a universe of n financial assets, and let $r_{i,t}(i = 1, \dots n;\ t = 1, \dots T)$ denote the continuously compounded daily returns of the i-th asset, defined as:

$$r_{i,t} = 100 \times \left(ln(pr_{i,t}) - ln(pr_{i,t-1})\right) \tag{2.1}$$

where$r_{i,t}$and $pr_{i,t}$ denote the percentage daily returns and the closing price of the i-th asset at the t-th day, respectively. Mathematically, the AR(1)-GJR-GARCH(1, 1) model is:

$$\begin{cases} r_{it} = \varphi_0 + \varphi_1 r_{it-1} + \varepsilon_{it} \\ \varepsilon_{it} = h_{it}^{\frac{1}{2}} z_{it}, z_{it} \sim F_i(\cdot) \\ h_{it} = \alpha_{i0} + \alpha_i \varepsilon_{i,t-1}^2 + \gamma_i \varepsilon_{i,t-1}^2 I_{\varepsilon_{i,t-1}} + \beta_i h_{i,t-1}, \end{cases} \tag{2.2}$$

where ε_{it} and z_{it} denote the errors and innovations, respectively, of the time series process $r_{it}(i = 1, \dots n;\ t = 1, \dots T)$; $F_i(\cdot)$ denotes the cumulative distribution function of the innovations, andφ_0 and $\varphi_1$are the parameters characterizing the dynamics of the autoregressive process. The conditional volatility process is denoted as$h_{it}$and $I_{\varepsilon_{i,t-1}}$is an indicator function which equals 1 if $\varepsilon_{i,t-1}$ is positive and zero, otherwise. Here $\alpha_{i0} \geq 0$, $\gamma_i \geq 0$, $\alpha_i \geq 0$,and $\beta_i \geq 0$ are the intercept, ARCH, GARCH and leverage coefficients characterizing the conditional volatility process. The coefficient γ_i determines the asymmetric effect on conditional volatility. If $\gamma_i > 0$, negative shocks (bad news) have larger impact $(\alpha_i + \gamma_i)\varepsilon_{i,t-1}^2$ on the conditional variance h_{it} than positive shocks (good news) given by $\alpha_i \varepsilon_{i,t-1}^2$. If $\gamma_i < 0$, the stock market is more responsive to good news. An additional condition to ensure the weak stationarity of the above conditional volatility process is $\alpha_i + 0.5\gamma_i + \beta_i < 1$.

To complete the dynamics of asset returns, this section proposes a flexible parametric model for the marginal distribution $F_i(\cdot)$ of the innovations in Eq. (2.2). The distribution in Eq. (2.3) is obtained by applying the conditional probability theorem and exploiting EVT results for the tails of the distribution. For the middle domain defined by the interval $[\eta^L, N_{\eta^R}]$, with η^L and η^R the lower and upper thresholds characterizing the tails of the distribution, this work considers two modeling strategies: (a) the empirical distribution function[1] and (b) a standard Normal distribution. The distribution in Eq. (2.3) focuses on the semiparametric model (a). More importantly, the conditional distribution in both tails is modeled as a Generalized Pareto distribution (GPD). This parametric choice is motivated by EVT theory and in particular, by a theoretical result in Pickands (1975) that shows that the conditional distribution of a random variable in the tail can be approximated by a GPD distribution for sufficiently large values of the thresholds η^L and N_{η^R}. This is done to capture the presence of extreme events beyond the range provided by standard parametric models such as the Normal or Student-t distributions, see also Alexander and Rüdiger (2000) and Embrechts et al (2001). Mathematically,

[1] See McNeil and Frey (2000) and Engle and Gonzalez-Rivera (2001) for semiparametric formulations of the distribution of asset returns.

$$F_i(z) = \begin{cases} \frac{N_{\eta^L}}{N}\left\{1 + \varsigma^L \frac{\eta^L - z}{\chi^L}\right\}^{-\frac{1}{\varsigma^L}}, z < \eta^L, \\ \frac{N_z}{N}, \eta^L \leq z \leq \eta^R, \\ \frac{N_{\eta^R}}{N}\left\{1 + \varsigma^R \frac{z - \eta^R}{\chi^R}\right\}^{-\frac{1}{\varsigma^R}}, z > \eta^R, \end{cases} \tag{2.3}$$

where ς^L and ς^R denote the shape parameters (tail index coefficients) corresponding to the left and right tails of the distribution of the sequence of innovations z; χ^L andχ^R are, respectively, the left and right tail scale parameters. Similarly, N denotes the total number of observations and N_z the number of observations below or equal to z, thus, N_z/N is the empirical distribution function.

Estimation of the model parameters is carried out using maximum likelihood methods. In the first step, I fit the GJR-GARCH model to the return prices and obtain estimates of the conditional mean and variance equations. In the second step, the tail indices and scale parameters of the GPD tail distributions are fitted by maximum likelihood, as in McNeil and Frey (2000). In particular, the tail indices characterizing the decay of the distribution function in each tail are fitted using the Hill estimator, see Hill (1975), after suitable selection of the threshold estimates $\eta^L$and η^R.

2.2.2 Dynamic mixture copula model

Copulas have been widely adopted to model the dependence structures between financial markets with the purpose of monitoring risk and modeling returns in portfolios of assets. Popular examples of copula functions in multivariate settings are the Gaussian, the multivariate Student-t, and the multivariate Archimedean copula. These copulas are simple to represent and capture different stylized facts (e.g., asymmetric tail dependence across tails, asymptotic dependence in the tails) of multivariate returns.

The theory on copula functions was developed in Sklar (1959). This author shows that the multivariate distribution function G(z) of a vector of random variables $z = (z_1, \cdots, z_n)$ with marginal distributions $F_1, F_2, \cdots, F_n$ can be expressed in terms of a copula function C such that

$$G(z) = C(F_1(z_1), F_2(z_2), \ldots, F_n(z_n)) = C(u_1, \ldots, u_n), \tag{2.4}$$

where $u_i = F(z_i) \in [0,1]$ for i=1,…,n.

In this paper, I adopt a mixture of three copula functions: Gaussian, Student-t, and Clayton copula functions. This flexible approach allows us to capture complex dependence structures among asset returns. The Gaussian copula aims to capture dependence in the middle regimes. The Student-t copula is sensitive to extreme events in the tails of the distributions, but it does not exhibit extreme tail dependence for very large values of asset returns. The Clayton copula function describes asymptotic tail dependence in the lower tail of the joint distribution of asset returns. The distribution function of the multivariate Gaussian copula is expressed as:

$$C^{\text{Gau}}(u_1, \ldots, u_n; \rho^g) = \Psi_{\rho^g}\left(\Psi^{-1}(u_1), \ldots, \Psi^{-1}(u_n)\right), \tag{2.5}$$

where $\Psi_{\rho^g}(\cdot)$ is the standard multivariate Normal distribution function with correlation matrix ρ^g; $\Psi(\cdot)$ is the univariate Normal distribution and $\Psi^{-1}(\cdot)$ its inverse function. The distribution function of the Student-t copula is expressed as:

$$C^{\text{Stu}}(u_1, \ldots, u_n; \nu, \rho^s) = \Phi_{\nu,\rho^s}\left(\Phi_{\nu_1}^{-1}(u_1), \ldots, \Phi_{\nu_n}^{-1}(u_n)\right), \tag{2.6}$$

where $\Phi_{\nu,\rho^s}(\cdot)$ is the standard multivariate Student-t distribution function with correlation matrix ρ^s and degrees of freedom parameter ν. The model allows for different degrees of freedom across marginal distributions. Thus, $\Phi_{\nu_i}^{-1}(\cdot)$ denotes the inverse of the distribution function $\Phi_{\nu_i}(\cdot)$. Finally, the distribution function of the Clayton copula is expressed as:

$$C^{\text{Cla}}(u_1,\ldots,u_n;\theta)=\left(\sum_{i=1}^{n}u_i^{-\theta}-n+1\right)^{-\frac{1}{\theta}},\theta\in(0,\infty). \tag{2.7}$$

Thus, the multivariate mixture copula model can be expressed as:

$$\begin{aligned}MC(u_1,\ldots,u_n;\rho^g,\rho^s,\gamma,\nu,\theta)\\ =\lambda^G C^{\text{Gau}}(u_1,\ldots,u_n;\rho^g)+\lambda^S C^{\text{Stu}}(u_1,\ldots,u_n;\nu,\rho^s)\\ +\lambda^C C^{\text{Cla}}(u_1,\ldots,u_n;\theta),\end{aligned} \tag{2.8}$$

where λ^G, λ^S, and λ^C are respectively the weights of Gaussian, Student-t, and Clayton copula functions, satisfying $\lambda^G,\lambda^S,\lambda^C\geq 0$and $\lambda^G+\lambda^S+\lambda^C=1$.

The estimation of the mixture copula parameters ($\rho^g,\rho^s,\nu,\theta,\lambda^G,\lambda^S,\lambda^C$) is done by applying maximum likelihood estimation (MLE) methods. In practice, it is difficult to obtain an analytical solution using MLE. To overcome this, I resort to simulation methods by combining the MLE approach with the expectation maximization (EM) algorithm. The EM algorithm for missing data transforms the likelihood function maximization problem for incomplete data into the maximization of the likelihood function for complete data. This is done by assuming the presence of latent variables in the maximization exercise. Through the iteration of the expectation and maximization processes up to convergence, I obtain an optimal solution of the likelihood function. More details of the estimation of the mixture copula parameters can be found in the recent studies by Sahamkhadam (2018) and Ben Nasr and Chebana (2022).

Considering the fact that the dependence of multiple assets is dynamic and time varying, this study combines the GJR-GARCH-EVT model with a mixture copula to dynamically describe the dependence and risk of asset returns, where the parameters λ^{Gaus}, $\lambda^{\text{stu-t}}$, and λ^{Cla} also changes over time. The implementation of the dynamic mixture copula model is detailed in Section 2.2.4.

2.2.3 Dynamic robust portfolio optimization

To set up the framework for the robust optimization problem, I define the loss function $L(w,r)$ associated to a decision $w \in \mathbb{R}^n$ and a random return $r \in \mathbb{R}^n$. For the optimal portfolio allocation problem, the loss function is $L(w,r) = -w'r$ but other configurations of the loss function are also possible. For the sake of generality, the following expressions define the relevant tail risk measures as a function of the general loss function $L(w,r)$. By doing so, I stay close to the formulations of the CVaR in Rockafellar and Uryasev (2002) and Kakouris and Rustem (2014).

The portfolio is characterized by the allocation of different assets. This allocation is determined by the vector $w = (w_1, w_2, \cdots, w_n)$ and a random return vector $r = (r_1, r_2, \cdots, r_n)$. For each w, I denote by $\Psi(w,a)$ the distribution function for the loss function $L(w,r)$ such that $\Psi(w,\xi) = \int_{L(w,r)\leq\xi} g(r)\,dr$, where $g(\cdot)$ is the joint density function of the vector of random returns r and ξ denotes a tolerance level for the loss function. Furthermore, let $\alpha \in (0,1)$ denote a confidence level associated to the maximum loss of the portfolio. In applications, this value is in the range [0.95,0.99]. The $VaR_\alpha(w)$ associated to the loss function is defined as:

$$VaR_\alpha(w) = min\{\xi \in \mathbb{R}: \Psi(w,\xi) \geq \alpha\}. \qquad (2.9)$$

Similarly, the $CVaR_\alpha(w)$ is defined as the expected loss once the risk measure $VaR_\alpha(w)$ is exceeded. More formally,

$$CVaR_\alpha(w) = \frac{1}{1-\alpha}\int_{L(w,r)\geq VaR_\alpha(w)} L(w,r)\,g(r)dr. \qquad (2.10)$$

As mentioned above, the $CVaR_\alpha(w)$ measure has some appealing properties not shared by the VaR such as the sub-additivity and coherence, see Arztner et al. (1999), Topaloglou et al. (2014) and Guo et al. (2019). Rockafellar and Uryasev (2002), in their seminal contribution, show that the $VaR_\alpha(w)$ and $CVaR_\alpha(w)$ of the loss function associated to the vector $w = (w_1, w_2, \cdots, w_n)$ can be calculated simultaneously by solving the following convex optimization problem:

$$G_\alpha(w,\xi) = \xi + \frac{1}{1-\alpha}\int_{r\in\mathbb{R}^n}[L(w,r)-\xi]^+ g(r)dr, \qquad (2.11)$$

such that $CVaR_\alpha(w) = \underset{\xi\in\mathbb{R}}{min} G_\alpha(w,\xi)$. The associated $VaR_\alpha(w)$ is defined as the value $\xi \in \mathbb{R}$ that solves the minimization problem. More formally, $\xi_\alpha(w) = \underset{\xi\in\mathbb{R}}{argmin} G_\alpha(w,\xi)$ such that $CVaR_\alpha(w) = G_\alpha(w, VaR_\alpha(w))$. Note that throughout this chapter uses $VaR_\alpha(w)$ and $\xi_\alpha(w)$ indistinctively.

Kakouris and Rustem (2014) extend this approach to obtain the VaR_α and $CVaR_\alpha$ of a portfolio of assets with multivariate dependence modeled using copula models. To do this, these authors transform the loss function such that $L(w,r) = \tilde{L}(w,u)$, with $u = (u_1,\ldots,u_n) = (F_1(r_1),\ldots,F_n(r_n))$. The loss function $\tilde{L}(w,u)$ maps the domain of the loss function from $\mathbb{R}^n$ to $[0,1]^n$. The relevant tail risk measures with the copula representation are defined as:

$$\begin{cases} VaR_\alpha(w) = min\{\xi \in \mathbb{R}: C(u|\tilde{L}(w,u) \le \xi) \ge \alpha\}, \\ CVaR_\alpha(w) = \frac{1}{1-\alpha}\int_{\tilde{L}(w,u)\ge VaR_\alpha(w)} \tilde{L}(w,u)\, c(u)du. \end{cases} \qquad (2.12)$$

Note that $C(u|\tilde{L}(w,u) \le \xi)$ denotes the multivariate distribution function associated to the copula function for the vector $u = (u_1,\ldots,u_n)$ and conditional on the event $\tilde{L}(w,u) \le \xi$. Similarly, $c(u)$ denotes the corresponding copula density. The $CVaR_\alpha(w)$ is obtained, as before, from minimizing $G_\alpha(w,\xi)$ that, under the transformation of the loss function, it can be expressed as $G_\alpha(w,\xi) = \xi + \frac{1}{1-\alpha}\int_{u\in[0,1]^n}[\tilde{L}(w,u)-\xi]^+ c(u)du$.

These formulations of the tail risk measures involve knowledge of the copula function modeling the multivariate dependence structure between the asset returns. A robust approach to compute the CVaR that incorporates uncertainty about the specific choice of the copula function is the worst-case CVaR. This risk measure considers the worst outcome from a set of possible scenarios. These scenarios are modelled by different choices of the copula density function. Following Kakouris and Rustem (2014), I define the worst-case CVaR as $WCVaR_\alpha(w) = \sup_{c(\cdot)\in\mathbb{C}} CVaR_\alpha(w)$, where $\mathbb{C}$ denotes a set of candidate copulas within the list introduced above. This approach can be extended to consider the mixture copula $MC(u_1,\dots,u_n;\rho^g,\rho^s,\gamma,\nu,\theta)$ introduced in Eq. (2.8). In this case the $WCVaR_\alpha$ is obtained from minimizing the function

$$G_\alpha(w,\xi,\lambda^g,\lambda^s,\lambda^c) = \lambda^g G_\alpha^g(w,\xi) + \lambda^s G_\alpha^s(w,\xi) + \lambda^c G_\alpha^c(w,\xi). \tag{2.13}$$

More formally, $WCVaR_\alpha(w) = \min_{\xi\in\Re}\max_{\lambda\in\Delta} G_\alpha(w,\xi,\lambda^g,\lambda^s,\lambda^c)$, with $\Delta = \{\lambda^g,\lambda^s,\lambda^c : \lambda^g+\lambda^s+\lambda^c = 1, \lambda^g,\lambda^s,\lambda^c \geq 0\}$.

In this setup, the optimal portfolio allocation is obtained from $\min_{w\in W} WCVaR_\alpha(w)$, with W the set of feasible portfolio weights. More specifically, the results of Zhu and Fukushima (2009) allow us to obtain the optimal portfolio allocation as the solution of the following robust optimization problem:

$$WCVaR_\zeta(w) = \min_{w\in\mathbb{W}}\min_{\xi\in\mathbb{R}}\max_{\lambda\in\Delta} G_\alpha(w,\xi,\lambda) = \min_{(w,\xi,\phi)\in\mathbb{W}\times\mathbb{R}\times\mathbb{R}}\Big\{\phi : \lambda^g G_\alpha^g(w,\xi) + \lambda^s G_\alpha^s(w,\xi) + \lambda^c G_\alpha^c(w,\xi) \leq \phi, \forall\lambda\in\Delta\Big\}, \tag{2.14}$$

with ϕ satisfying that $G_\alpha^i(w,\xi) \leq \phi$, for i=g,s,c. The objective function can be reduced to $\min_{(w,\xi,\phi)\in\mathbb{W}\times\mathbb{R}\times\mathbb{R}}\{\phi : G_\alpha^i(w,\xi) \leq \phi, i = g,s,c\}$.The solution to this problem can be obtained using Monte Carlo simulation methods. Rockafellar and Uryasev (2000) provides an approximation of the functions $G_\alpha^i(w,\xi)$ based on simulation of S^i random samples. Then, the optimal portfolio allocation is obtained from the minimization of the simulated function

$$\min_{(w,\xi,\phi)\in\mathbb{W}\times\mathbb{R}\times\mathbb{R}}\left\{\phi : \xi + \frac{1}{S_i(1-\alpha)}\sum_{k=1}^{S^i}[\tilde{L}(w,u_i^{[k]}) - \xi]^+ \leq \phi, i = g,s,c\right\}, \tag{2.15}$$

where $u_i^{[k]}$ is the k^th sample vector associated to copula $C^i(\cdot)$.Following Zhu and Fukushima (2009), the minimization problem can be expressed as

$$\begin{cases} \min \phi \\ s.t.\ w \in \mathbb{W}, v \in \mathbb{R}^m, \xi \in \mathbb{R}, \phi \in \mathbb{R}, \\ \xi + \frac{1}{S^i(1-\alpha)} (\mathbf{1}^i)^T v_i \leq \phi, i = g, s, c \\ v_i^k \geq \tilde{L}(w, u_i^{[k]}) - \xi, k = 1,2,\cdots, S_i, i = g, s, c, \\ v_i^k \geq 0, k = 1,2,\cdots, S_i, k = 1,2,\cdots, S_i, i = g, s, c, \end{cases} \tag{2.16}$$

where $v = (v^1, v^2, v^3) \in \mathbb{R}^m$ with $m = S_1 + S_2 + S_3$ and $\mathbf{1}^i = (1,1,\cdots,1)^T \in \mathbb{R}^{S^i}$. The set of feasible portfolio weights $\mathbb{W}$ is defined as $\mathbb{W} = \left\{ \sum_{j=1}^{n} \boldsymbol{w}_j = 1,\ \boldsymbol{w}_j \geq \mathbf{0},\ \text{j=1,...,n} \right\}$.

2.2.4 Dynamic algorithm

The above methods are implemented using the algorithm developed in Weiß (2013), Nikoloulopoulos et al. (2012), Han, Li, and Xia (2017) and Sahamkhadam (2018). The parameters of the GJR-GARCH-EVT model with mixture copula $MC(u_1, \ldots, u_n; \rho^g, \rho^s, \gamma, \nu, \theta)$ introduced in Eq. (2.3) and Eq. (2.8) are estimated over rolling windows of T observations from the vector of daily log returns $r_t = (r_{1t}, \cdots, r_{nt})$. This study considers a Monte Carlo procedure to simulate one-period ahead returns of the assets in the portfolio. The algorithm is as follows;

Step 1: The vector of parameters characterizing the conditional volatility processes $h_{i,t}$ and the marginal distribution function in Eq. (2.3) are estimated by maximum likelihood to obtain estimates of the standardized residuals z_{it} in Eq. (2.2) and marginal distribution functions $u_{it} = \hat{F}_i(z_{it})$ for i=1,…,n.

Step 2: The vector of dependence parameters $(\rho^g, \rho^s, \gamma, \nu, \theta)$ characterizing the mixture copula is estimated by maximizing the log-likelihood of the mixture copula function for the vectors $(u_{1t}, \ldots, u_{nt}) = (\hat{F}_1(z_{1t}), \ldots, \hat{F}_n(z_{nt}))$ for t=1,...,T. This procedure also includes the estimation of the vector $(\lambda^g, \lambda^s, \lambda^c)$ defining the mixture copula.

Step 3: The parameter estimates $(\widehat{\rho^g}, \widehat{\rho^s}, \hat{\gamma}, \hat{v}, \hat{\theta})$ and $(\hat{\lambda}^g, \hat{\lambda}^s, \hat{\lambda}^c)$ obtained from Step 2 are used to simulate random samples of vectors of residuals $(z_{1t}^{[k]}, \ldots, z_{nt}^{[k]})$ indexed by *k=1,…,S* from the estimated mixture copula$MC(u_1, \ldots, u_n; \widehat{\rho^g}, \widehat{\rho^s}, \hat{\gamma}, \hat{v}, \hat{\theta})$; *S* denotes the number of simulated samples obtained from the Monte Carlo exercise.

Step 4: Plug in the simulated vector of observations $(z_{1t}^{[k]}, \ldots, z_{nt}^{[k]})$ obtained from Step 3 into the location-scale process introduced in Eq. (2.2):

$$r_{it}^{[k]} = \hat{\varphi}_0 + \hat{h}_{it}^{1/2} z_{it}^{[k]},$$

for *i*=1,…,n and *t*=1,…,*T*, where $\hat{\varphi}_0$ denotes the intercept of the location-scale process estimated in Step 1 and $\hat{h}_{it}$are the forecasts of the conditional volatility process in Eq. (2.3) (i.e, GJR-GARCH) constructed from the corresponding parameter estimates.

Step 5: Input the forecasted returns $(r_{1t}^{[k]}, r_{2t}^{[k]}, \ldots, r_{nt}^{[k]}]$ into the WCVaR optimization problem in Eq. (2.16), and compute the optimal portfolio weights w^* by solving the minimization problem.

Step 6: The in-sample rolling window is extended by one day such that the in-sample vector of returns covers observations from 2 to *T*. This sample is used to estimate the model parameters, compute the portfolio returns, and construct the optimal portfolio allocation obtained from minimizing the worst-case CVaR at a confidence level α.

2.3. Data

Daily closing prices of four representative assets given by the S&P500, gold, Bitcoin, and United States 5-Year Treasury Bond are used in this study. Daily data from 2 January 2015 to 13 December 2021 are obtained from https://www.investing.com/ and Bloomberg. The sample period is divided into three sub-periods: the Pre-COVID-19 period (from January 2, 2015 to January 29, 2020); the COVID-19 period (from January 29, 2020 to July 31, 2020), and the After-COVID-19 period (from August 1, 2020 to December 13, 2021). Similar periods are used in Ali et al. (2021) and Raj et al. (2022), among many others. In this work, I mainly evaluate portfolio allocation during and after the epidemic outbreak periods but the whole sample period is used for estimation of the model parameters. This study estimates the model from the daily logarithmic returns using data from 01/02/2015 to 01/29/2020 and uses data from 01/30/2020 to 12/13/2021 for the out-of-sample evaluation exercise. The latter period covers the outbreak of the COVID-19 pandemic and a period that denominates After-COVID-19.

Figure 2.1 plots the daily prices (black color) and returns (blue color) for S&P500, gold, Bitcoin, and 5-year bond for the constituents of the portfolio. The S&P500 index and gold price have a similar positive trend in the Pre-COVID-19 period, while bond prices fluctuate considerably. It is also noticeable that large drops were experienced by the S&P500 index and Bitcoin at the start of the COVID-19 pandemic. In contrast, the After-COVID-19 period witnesses large increases in the price of these assets that exhibit a strong positive trend whereas gold and bond prices remain stable or even show a slight decline during this period. All the assets exhibit strong volatility clustering that motivates the application of conditional volatility models such as the GJR-GARCH to capture these dynamics and remove serial dependence from the squared returns.

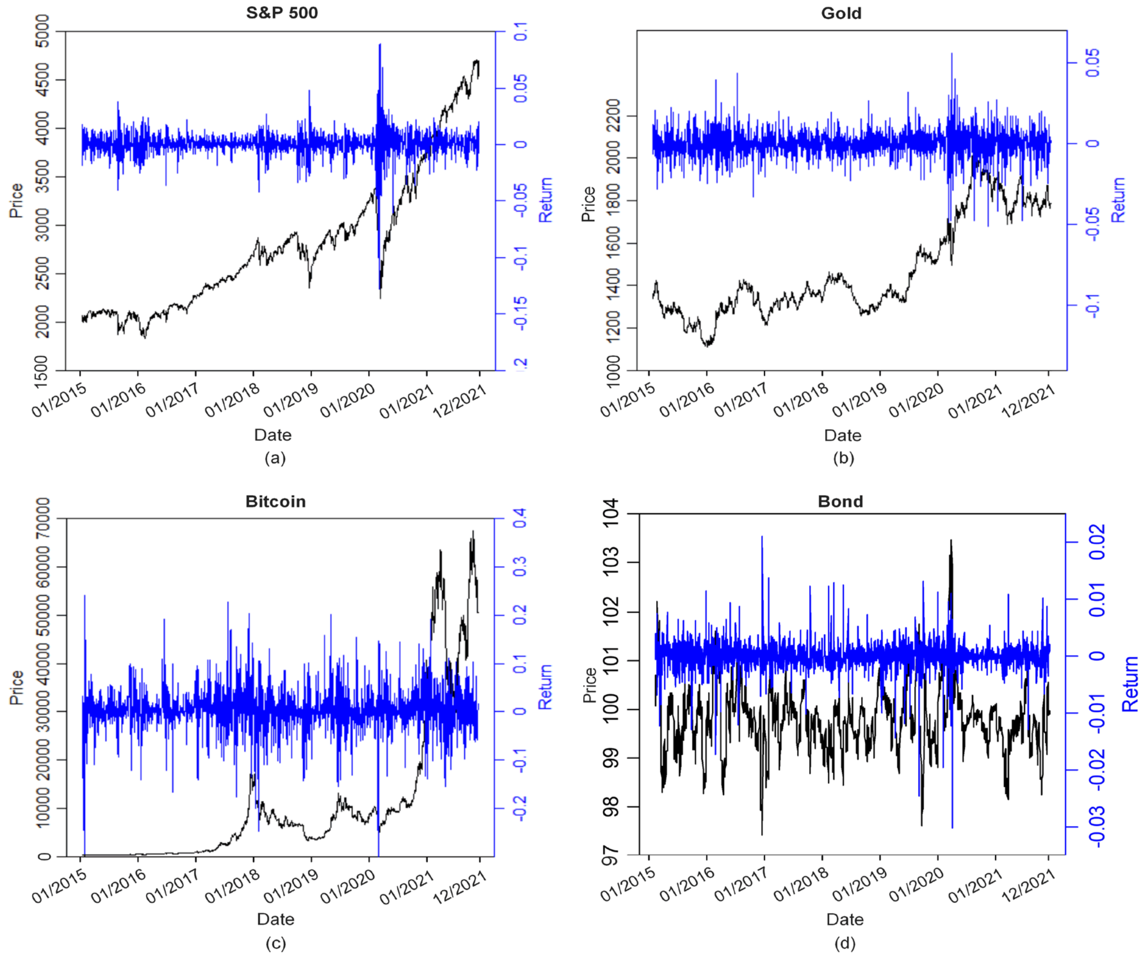


Figure 2. 1. .Daily prices and returns for the four portfolio constituents.

Table 2.1 reports summary statistics for the logarithmic returns for the four assets, including the Jarque–Bera test of normality. With the only exception of the bond market, the other three assets have positive mean returns in both the in-sample as well as out-of-sample periods. Importantly, the volatility in asset returns, captured by the standard deviation, is significantly higher during/after the COVID-19 pandemic outbreak than in the pre-COVID-19 period for S&P 500, gold, and the US bond. Table 1 also shows the high excess kurtosis for all four assets in both periods, especially for Bitcoin, but the skewness differs considerably across periods. Bitcoin achieves the highest average log returns, which might have attracted investment, but it has the highest volatility leading to high portfolio risk. The Jarque–Bera test statistics applied to all assets in both periods reject the null hypothesis of normality at 1% significance level.

Table 2.1. Descriptive statistics of daily logarithmic returns for the portfolio constituents.

Index	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis	Jarque–Bera
In-sample (pre-COVID-19 period): 01/02/2015-01/29/2020							
S&P 500	0.00036	0.00846	-0.0418	0.0484	-0.5283	3.8517	97.8501 (*)
Gold	0.00015	0.00763	-0.0337	0.0439	0.16057	2.7220	9.57671 (*)
Bitcoin	0.00266	0.04579	-0.2942	0.2408	-0.2024	5.5266	347.847 (*)
Bond	$-4.6*10^{-6}$	0.00281	-0.0245	0.0212	-0.4671	13.2992	5681.55(*)
Out-of sample (during/post COVID-19 period): 01/30/2020-12/13/2021							
S&P 500	0.00076	0.01693	-0.1277	0.0897	-1.0424	14.316	2609.51 (*)
Gold	0.00021	0.01183	-0.0508	0.0563	-0.4365	4.2782	47.2222 (*)
Bitcoin	0.00347	0.05090	-0.4973	0.1918	-1.9932	20.215	6154.12 (*)
Bond	$-1.9*10^{-5}$	0.00286	-0.0301	0.0110	-0.2.914	32.426	17847.5(*)
Overall period: 01/02/2015-12/13/2021							
S&P 500	0.00047	0.01138	-0.1277	0.0897	-1.0468	21.1361	24289.4 (*)
Gold	0.00017	0.00896	-0.0508	0.0563	-0.1942	5.12853	341.170 (*)
Bitcoin	0.00287	0.04720	-0.4973	0.2408	-0.8048	10.8708	4703.36 (*)
Bond	$-7.6*10^{-7}$	0.00278	-0.0301	0.0212	-1.020	18.5688	17967.61(*)
Note: This table applies the natural logarithmic returns for the indexes. (*) denotes rejection of the null hypothesis of normality at a 1% significance level.							

2.4. Empirical results

This section presents an empirical exercise that compares portfolio performance (cumulative returns, volatility, Sharpe ratio, maximum drawdown) between the following five investment strategies: 1) The dynamic robust WCVaR portfolio, which uses the GJR-GARCH-EVT model and multivariate dynamic mixture copula model. This portfolio is denoted throughout as G-E-D-M-C-WCVaR; 2) The dynamic robust WCVaR portfolio, which uses the dynamic multivariate mixture copula model but fits a Normal distribution for the marginal distribution of asset returns. This portfolio is denoted as N-D-M-C-WCVaR; 3) The minimum-variance portfolio (MV). This procedure is a popular investment strategy that minimizes the global portfolio variance without setting a target portfolio return. The choice of this portfolio instead of the mean-variance portfolio introduced in Markowitz (1952) is because the expected return is usually measured with error and can lead to noisy portfolios with significant estimation error, see the seminal contribution of Jagannathan and Ma (2003). The variance of the portfolio is given by:

$$Var(R) = w^T \sum w, \tag{2.17}$$

where $\sum$ is the covariance matrix of asset returns. The global MV portfolio allocation is obtained by solving the standard Markowitz asset allocation problem defined as:

$$\begin{cases} \min\{w^T \sum w\} \\ s.t.\ \sum_{i=1}^{n} w_i = 1, \\ 0 \le w_i \le 1, i \in \{1, \cdots, n\}, \end{cases} \tag{2.18}$$

where $w = (w_1, w_2, \cdots, w_n)$ and n is the number of assets in the portfolio. 4) The fourth proposed investment strategy is the MV portfolio with shrinkage estimation. In practice, the population covariance matrix $\sum$ is not known and is replaced by its sample counterpart, denoted by S. The sample covariance matrix has appealing properties, such as being maximum likelihood under normality. However, when the number of assets in the portfolio is larger than the sample size the estimation of the covariance matrix is very noisy. Ledoit and Wolf (2003, 2004, 2012, 2017, 2022) in a sequel of seminal contributions propose shrinkage methods to correct the effect of estimation error. These methods based on the ideas of Stein (1956) and James and Stein (1961) impose some structure on the estimator of the covariance matrix. A simple approach is to use a multiple of the identity matrix as shrinkage target, see Ledoit and Wolf (2004), such that $\sum^{*} = \alpha \, \mathrm{I}_n + (1 - \alpha)\hat{S}$, where I_n is the $n \ x \ n$ identity matrix and α is the shrinkage parameter obtained from minimizing the quadratic loss function $E[\|\Sigma^{*} - \Sigma\|^{2}]$. In this context, estimation error is not an issue given that the sample size is much larger than the number of candidate assets in the portfolio. Nevertheless, for completeness and to assess the role of shrinkage, I also consider this investment strategy. 5) The last investment strategy is the equally-weighted portfolio (EWP), characterized by the same allocation (1/n) to all assets in the portfolio. The size of the rolling window T is set to 475 throughout the empirical section.

2.4.1 Performance evaluation during/after COVID-19

To show the superiority and robustness of the proposed method, this subsection uses the data during/after the COVID-19 period (January 29, 2020 to December 13, 2021) as evaluation period. This period is characterized by the presence of large uncertainty in the economy and financial markets.

Figure 2.2 reports the composition of the portfolios for the different investment strategies. The allocation to the S&P500 index is in red, the allocation to Gold is in brown, the allocation to Bitcoin is in yellow, and the allocation to the US 5-year Treasury bond is in green. A rapid inspection of the different panels reveals the importance of the S&P500 index and gold for the four investment strategies. Portfolios in Panels (a) to (c) are quite diversified in the sense that the share of investment to the S&P500, gold, and Bitcoin in each asset is similar. The smallest allocation is to the US Treasury bond but this choice is consistent across strategies. This is due to the poor performance of the bond during this period. The strategies based on the G-E-D-M-C-WCVaR and N-D-M-C-WCVaR portfolios in Panels 2.2(a) and 2.2(b) are constructed using a confidence level of $\alpha = 0.95$. These strategies show more variation in the portfolio allocation than MV portfolio.

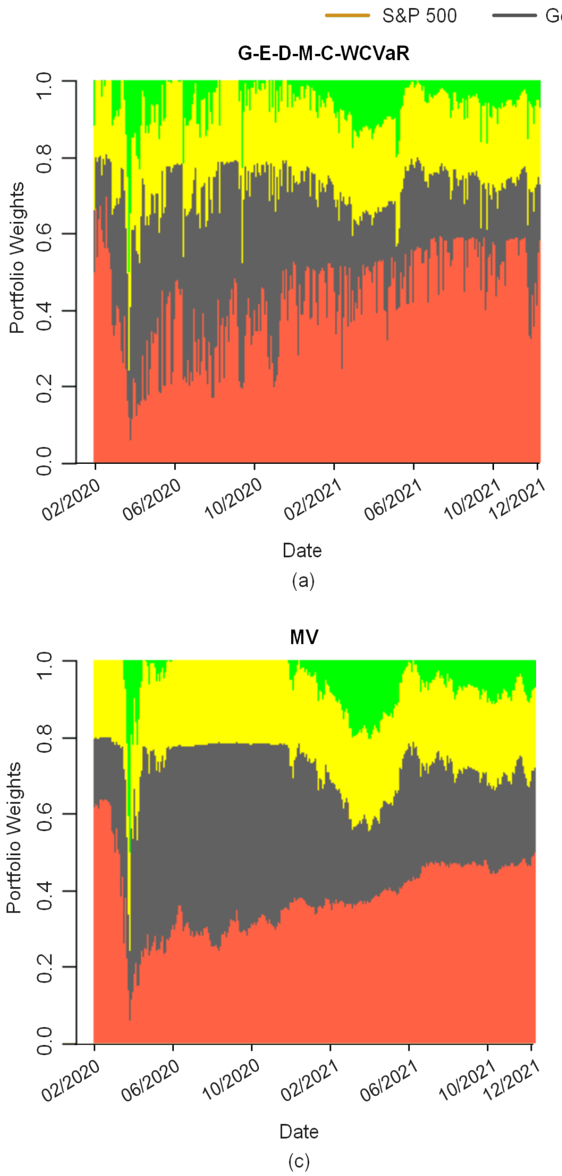


Figure 2.2. Panels (a) to (d) present the time-varying portfolio allocation to each of the four assets in the portfolio. The evaluation period is January 29, 2020 to December 13, 2021.

The comparison of Panels 2.2(c) and 2.2(d) also reveals the effect of shrinkage in the optimal allocation of assets to the investment portfolios. The contribution of Gold gains importance at the expense of the S&P500 index in the MV-Shrinkage portfolio throughout the COVID-19 period. This may be due to the improved ability of this portfolio to get exposure to Gold, which acts as a safe haven in market distress episodes. On the downside, this study notes that the portfolio weights of the MV-Shrinkage method fluctuate more than the MV strategy, which results in higher portfolio turnover and transaction costs.

Figure 2.3 reports the cumulative returns of the five investment portfolios over the evaluation period January 29, 2020 to December 13, 2021 at a confidence level $\alpha = 0.95$. Similarly, Table 2.2 presents performance statistics given by the average daily return (AR), volatility (Vol), Sharpe ratio (SR), CVaR, maximum drawdown (MDD), and total return (TR). The performance of the G-E-D-M-C-WCVaR portfolio selection method is superior to the performance of the remaining competitors across all metrics. The profitability is higher as shown by the AR and TR measures. Risk is lower as shown by Vol, MD and CVaR. Correspondingly, the tradeoff between risk and return is also superior as shown by the SP out-of-sample statistic. Interestingly, the dynamics in Figure 2.3 and the results in Table 2.2 also reveal the outperformance of the MV-Shrinkage method compared to the standard MV methodology and confirm the suitability of shrinkage methods for the MV strategy even in small dimensions. As shown in Panel 2.2(d), the outperformance during the crisis period may be due to a larger exposure to Gold compared to the MV portfolio, which has similar exposure to the S&P500 index and Gold.

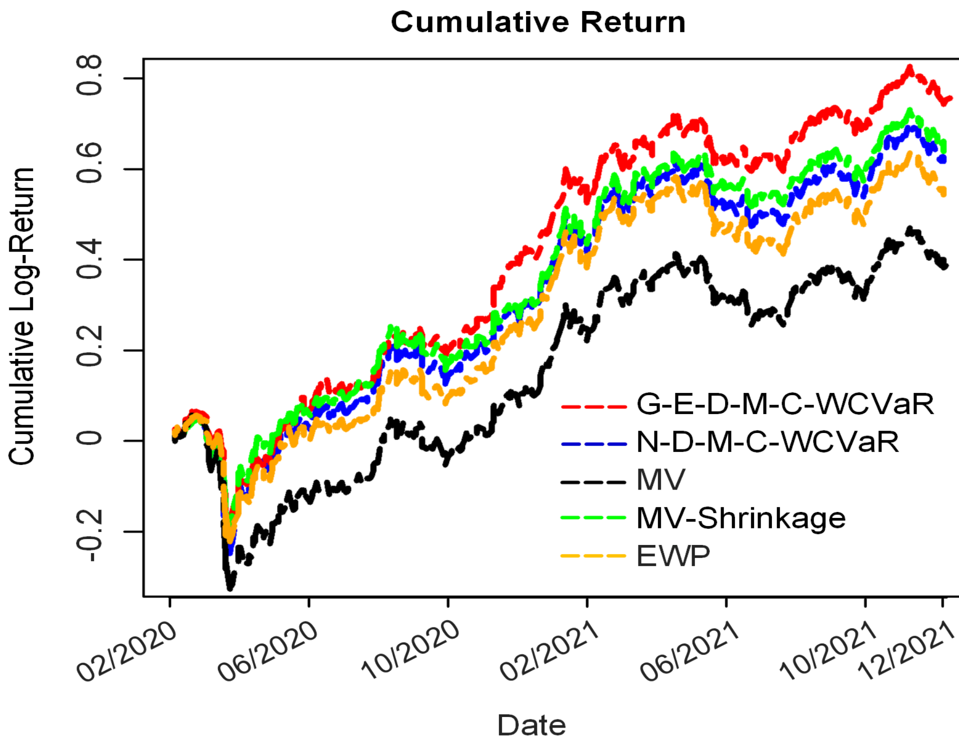


Figure 2.3. The cumulative return process for the five investment portfolios at $\alpha = 0.95$ during the evaluation period January 29, 2020 to December 13, 2021.

Table 2.2. Out-of-sample performance comparisons for the five competing portfolios at $\alpha = 0.95$ during the evaluation period January 29, 2020 to December 13, 2021.

Method	AR (10^{-2})	Vol (10^{-2})	SR	CVaR	MDD	TR
G-E-D-M-C-WCVaR	0.1593	0.0229	0.0239	0.0666	0.1851	0.7581
D-M-C-WCVaR	0.1294	0.0243	0.0149	0.0866	0.1878	0.6158
MV	0.0815	0.0257	0.0089	0.0921	0.1988	0.3881
MV-Shrinkage	0.1357	0.0198	0.0217	0.0624	0.1486	0.6459
EWP	0.1146	0.0238	0.0133	0.0860	0.1769	0.5457

To assess the robustness of the results to different definitions of the tail of the distribution of the portfolio return, this work considers different values of the confidence level α. Figures 2.4 and 2.5 report the same performance analysis for higher confidence levels ($\alpha = 0.975$ and $\alpha = 0.99$).

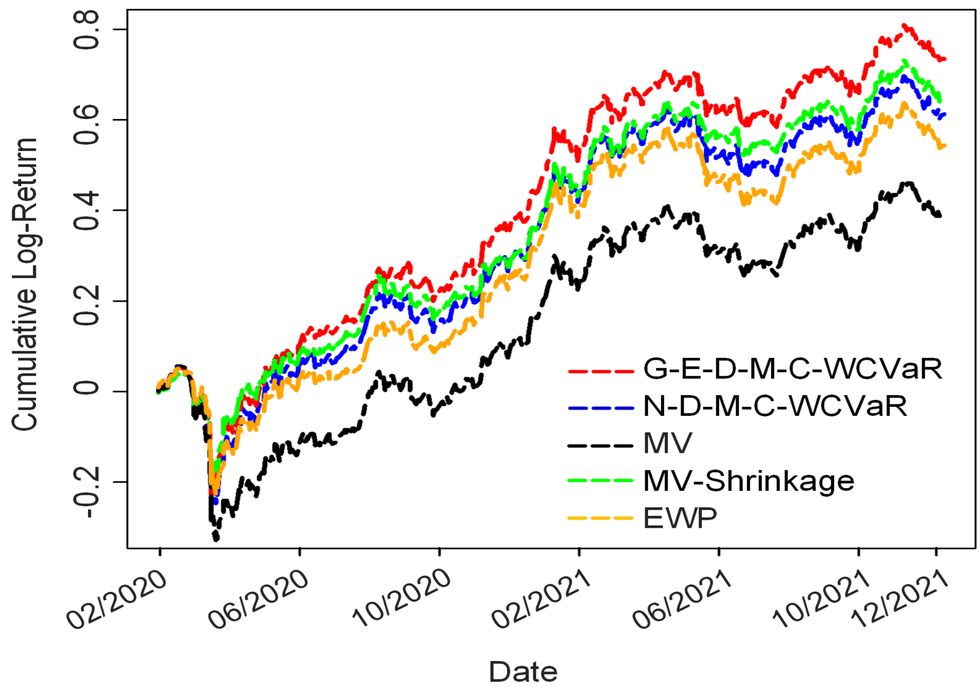


Figure 2.4. The cumulative return process for the five investment portfolios at $\alpha = 0.975$ during the evaluation period January 29, 2020 to December 13, 2021.

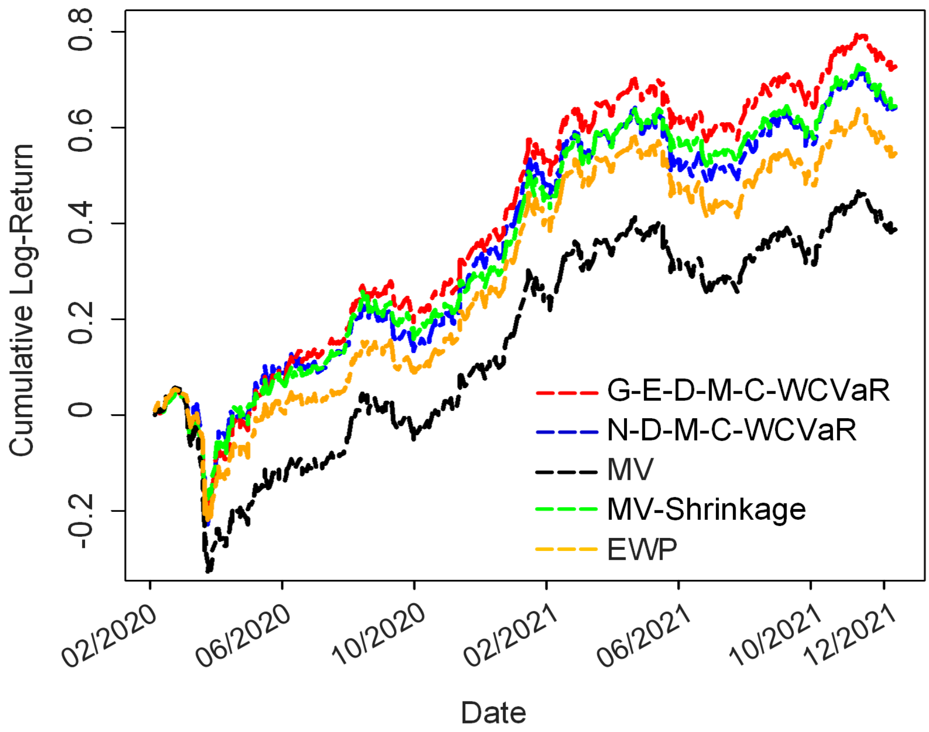


Figure 2.5. The cumulative return process for the four investment portfolios at $\alpha = 0.99$ during the evaluation period January 29, 2020 to December 13, 2021.

These values correspond to investment portfolios more concerned with downside risk events. Note that the MV and EWP strategies are identical to Figure 2.2. The cumulative returns for the other two strategies show slight variations compared to $\alpha = 0.95$. Tables 2.3 and 2.4 present the corresponding statistics for the performance measures for $\alpha = 0.975$ and $\alpha = 0.99$. The results confirm the outperformance of the G-E-D-M-C-WCVaR portfolio selection method for all metrics (return, risk and return/risk tradeoff) during this period under different tail risk tolerance levels. In contrast, the MV method obtains the lowest Sharpe ratio, as well as the highest volatility, downside risk, and maximum drawdown in the out-of-sample evaluation period. Importantly, this result highlights the poor performance of mainstream investment strategies such as the MV and EWP widely used by practitioners during distress episodes of the market. Interestingly, the MV-Shrinkage portfolio manages to report superior performance measures than the simple MV approach.

Table 2.3. The out-of-sample performance statistics for the five investment strategies at $\alpha = 0.975$ during the evaluation period January 29, 2020 to December 13, 2021.

Method	AR (10^{-2})	Vol (10^{-2})	SR	CVaR	MDD	TR
G-E-D-C-WCVaR	0.1548	0.0239	0.0195	0.0795	0.1866	0.7368
N-D-C-WCVaR	0.1287	0.0243	0.0150	0.0856	0.1885	0.6125
MV	0.0815	0.0257	0.0089	0.0921	0.1988	0.3881
MV-Shrinkage	0.1357	0.0198	0.0217	0.0624	0.1486	0.6459
EWP	0.1146	0.0238	0.0133	0.0860	0.1769	0.5457

Table 2.4. The out-of-sample performance statistics for the five investment strategies at $\alpha = 0.99$ during the evaluation period January 29, 2020 to December 13, 2021.

Method	AR (10^{-2})	Vol (10^{-2})	SR	CVaR	MDD	TR
G-E-D-M-C-WCVaR	0.1517	0.0240	0.0187	0.0810	0.1875	0.7222
N-D-C-WCVaR	0.1347	0.0246	0.0166	0.0812	0.1791	0.6410
MV	0.0815	0.0257	0.0089	0.0921	0.1988	0.3881
MV-Shrinkage	0.1357	0.0198	0.0217	0.0624	0.1486	0.6459
EWP	0.1146	0.0238	0.0133	0.0860	0.1769	0.5457

2.4.2 Performance evaluation in the Pre-COVID-19 period

This section studies portfolio performance over the period preceding the pandemic that I denominate as Pre-COVID 19 (from January 2, 2019 to January 28, 2020). The cumulative portfolio return of the five investment strategies is shown in Figure 2.6 and the performance statistics are reported in Table 2.5.

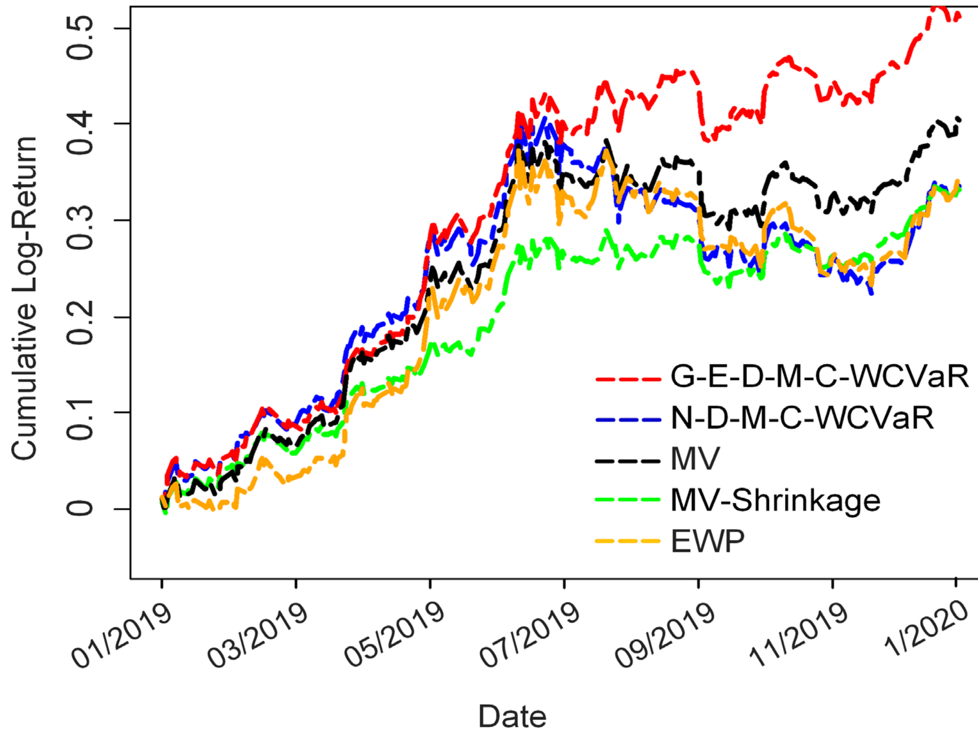


Figure 2.6. The cumulative return process for the four investment portfolios at $\alpha = 0.95$ during the evaluation period January 2, 2019 to January 28, 2020.

Figure 2.6 shows very similar performance for the five investment strategies during the Pre-COVID-19 period. The EWP slightly outperforms, but in all cases the cumulative return has a clear positive trend. Interestingly, though, the proposed G-E-D-M-C-WCVaR strategy maintains its performance showing a robust positive trend during the COVID-19 turmoil period whereas other three methods exhibit a negative trend during the second semester of 2019. Table 2.5 confirms these findings for different performance measures. The dynamic G-E-D-M-C-WCVaR portfolio significantly outperforms the different competitors in terms of Sharpe ratio, cumulative return, and CVaR. These results highlight the importance of considering the occurrence of extreme events and tail dependence along with the dynamics in the conditional volatility and correlation processes present in most modern portfolio allocation methods.

Table 2.5. The out-of-sample performance statistics for the five investment strategies at $\alpha = 0.95$ during the evaluation period January 2, 2019 to January 28, 2020.

Method	AR (10^{-2})	Vol (10^{-2})	SR	CVaR	MDD	TR
G-E-D-M-C-WCVaR	0.1892	0.00093	0.1152	0.0164	0.0783	0.5126
N-D-C-WCVaR	0.1237	0.01117	0.0550	0.0225	0.0905	0.3352
MV	0.1491	0.00095	0.0665	0.0224	0.0641	0.4042
MV-Shrinkage	0.1255	0.0124	0.0545	0.0231	0.0888	0.3400
EWP	0.1255	0.01241	0.0545	0.0230	0.0888	0.3400

2.4.3 Portfolio turnover and transaction costs

The above analysis highlights the strong performance of the portfolios proposed in this study. On the downside, this study observes in Figure 2.2 higher turnover in the optimal allocation of assets comprising the proposed optimal portfolios. The turnover in the EWP is zero, by construction. This rapid reaction to unexpected market movements may be the reason for the strong performance of the proposed portfolios. However, introducing additional flexibility to the portfolio comes at a price. Dynamic portfolios incur significant transaction costs due to the need of rebalancing the portfolios more often. Following the investment literature and, in particular, Kirby and Ostdiek (2012), Shen et al. (2014), Li et al. (2018) and Zhang et al. (2022), these costs are introduced as percentage rates of the total investment.

Let $\widehat{w}_{i,t}$ denote the portfolio weight before rebalancing the portfolio. At time t, the weight in asset i before the portfolio is rebalanced is:

$$\widehat{w}_{i,t} = \frac{w_{i,t-1}(1+r_{i,t})}{1+\sum_{j=1}^{n} w_{j,t-1}\, r_{j,t}}, \tag{2.19}$$

with $w_{i,t-1}(1 + r_{i,t})$ the portfolio wealth invested on asset i and $w_{i,t}$ is the portfolio weight at time t after rebalancing. Portfolio turnover (TO) at time t is defined as:

$$TO_t = \sum_{i=1}^{n} \left| w_{i,t} - \widehat{w}_{i,t} \right|, \tag{2.20}$$

and the associated transaction costs are defined as $TC_t = \pi TO_t$ where π is the assumed level of proportional costs per transaction. Kirby and Ostdiek (2012) show that the portfolio return, $\bar{r_{pt}}$, net of rebalancing transaction costs for period t is given by

$$\bar{r_{pt}} = (1 + w_{t-1} r_t)(1 - TC_t) - 1, \tag{2.21}$$

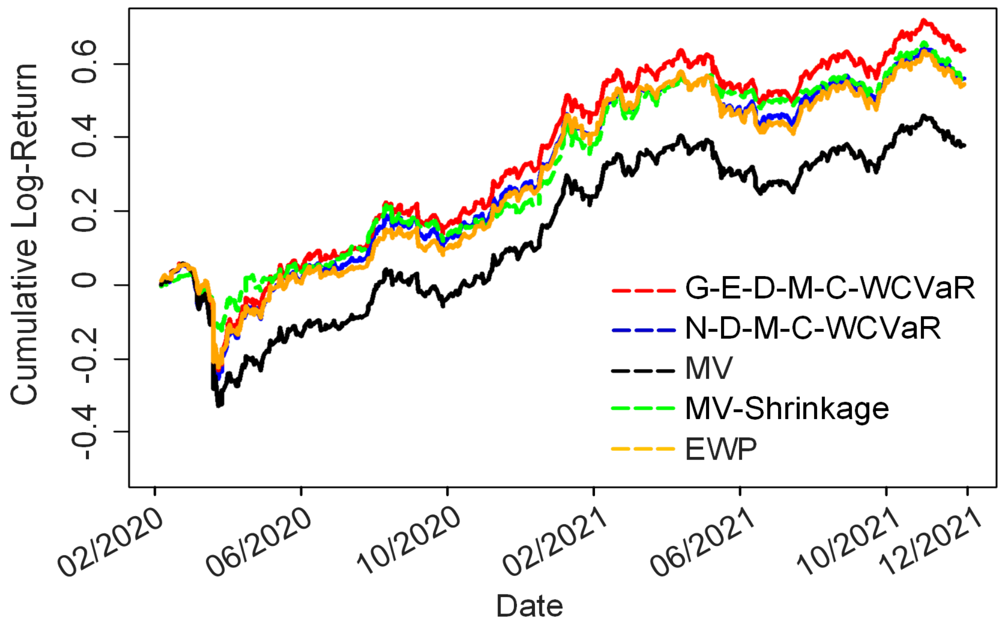
where w_{t-1} denotes the vector of portfolio weights after rebalancing and r_t is the vector of asset returns.

In order to control the amount of portfolio turnover during investment, I set a constraint TR, with $TO_t < TR$, for all t, that establishes the maximum portfolio turnover in each period. Thus, dynamics in the portfolio weights over time are not allowed if TR=0 whereas a large amount of flexibility is allowed if TR increases. The empirical application below considers TR=1 and penalizes such flexibility with different cost rates $\pi = 10{,}25$ and 50 basis points (bp). The constraint on portfolio turnover is used to smooth the portfolio weights over time and avoid wide fluctuations across consecutive periods.

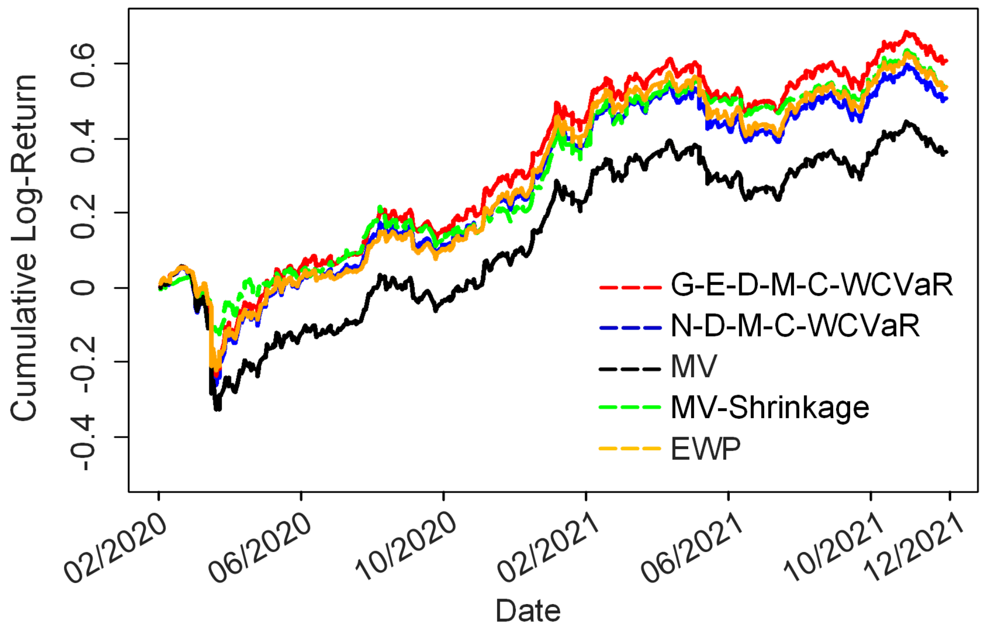
Figure 2.7 reports the cumulative portfolio returns after considering transaction costs $TC_t = \pi TO_t$, with $\pi = 10, 25$ and 50 bps, respectively.[2] As expected, compared with the results in Figure 2.3, the cumulative return process considering transaction costs, $\bar{r_{pt}}$, declines considerably for the three dynamic portfolios but most notably for the G-E-D-M-C-WCVaR and D-M-C-WCVaR investment portfolios. For example, the cumulative returns of the proposed G-E-

[2] Transaction costs of 10, 25 and 50 basis points are standard choices in the literature on portfolio investments, see DeMiguel et al (2009) and Kirby and Ostdiek (2012).

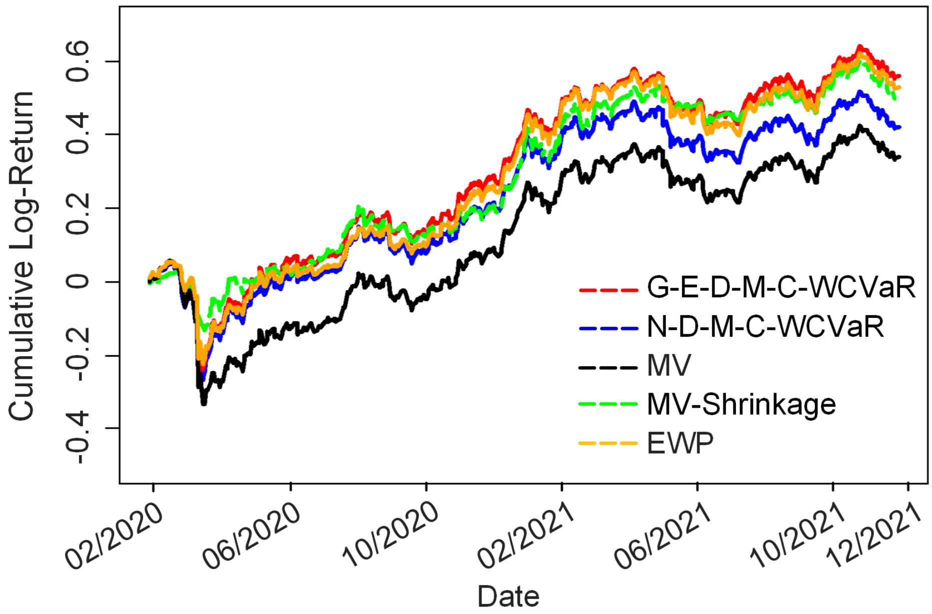
D-M-C-WCVaR method are 0.6379, 0.6075 and 0.5567, respectively, for $\pi = 10, 25$ and 50 bps, showing a monotonic decrease in the profitability of the portfolio as transaction costs rise. Nevertheless, the proposed G-E-D-M-C-WCVaR method still outperforms the other three methods. Interestingly, the bottom panel of Figure 2.7 shows the level of transaction costs that evens the profitability of the proposed strategy and the EWP.



(a) Transaction cost rate 10 basis point



(b) Transaction cost rate 25 basis point



(c) Transaction cost rate 50 basis point

Figure 2.7. The cumulative return process for the five investment portfolios during the evaluation period January 29, 2020 to December 13, 2021 at $\alpha = 0.95$. Transaction cost rates are 10, 25, and 50 basis points, respectively.

Table 2.6 reports the portfolio turnover statistic for the four investment portfolios and confidence levels $\alpha = 0.95, 0.975, 0.99$. As shown in Figure 2.2, both the G-E-D-M-C-WCVaR and D-M-C-WCVaR strategies have much higher portfolio turnover rates than the other two methods. These figures indicate the flexibility of the dynamic robust portfolios to adapt to market conditions. Unsurprisingly, this flexibility comes at a cost that is incorporated in the portfolio comparison through the presence of transaction costs. In addition, as the MV and EWP approaches do not take the confidence level constraint α into account the portfolio optimization problem, and thus their corresponding portfolio weights do not change under different confidence levels, leading to the same TO value with $\alpha = 0.95, 0.975, 0.99$ Shen et al. (2014), Li et al. (2018) and Zhang et al. (2022).

Table 2.6. Average portfolio turnover for the four investment portfolios for $\alpha = 0.95, 0.975, 0.99$ when the level of proportional costs per transaction is $\pi = 10$bps.

Method	G-E-D-M-C-WCVaR	D-M-C-WCVaR	MV	MV-Shrinkage	EWP
α = 0.95	0.1192	0.1461	0.0410	0.0808	0.0143
α = 0.975	0.1490	0.1446	0.0410	0.0808	0.0143
α = 0.99	0.1311	0.2071	0.0410	0.0808	0.0143

2.5. Conclusion

This article proposes an optimal portfolio allocation based on the minimization of a tail risk measure in the worst-case scenario. The optimization under the worst case adds robustness to the portfolio that is shielded against the presence of parameter uncertainty. Conceptually, the portfolio is optimized over a confidence set rather than over maximum likelihood point estimates. This investment strategy is particularly fruitful over distress episodes of the market that are characterized by abrupt changes in asset returns. The asset allocation obtained from our robust optimal portfolio strategy is dynamic and obtained by considering rolling windows over which to minimize the objective risk measure WCVAR. The optimization is achieved in several steps. First, the existence of conditional volatility clustering is filtered out using the GJR-GARCH-EVT model for the dynamics of returns. In the second stage, I model the joint dependence between the asset returns using a multivariate mixture copula model with particular emphasis on capturing tail events and negative extreme dependence.

Using this model, this work has examined the portfolio performance of different portfolios constructed from a set of four representative assets (S&P 500, Gold, Bitcoin, and US 5-year Treasury bond) during two periods characterized by market turmoil and the occurrence of extreme events caused by the COVID-19 pandemic crisis. The empirical investigation shows that the introduced dynamic WCVaR model allocates more weight to the assets with lower tail risk and lower tail dependencies. During the COVID-19 distress episode, this strategy achieves superior performance in terms of higher cumulative returns, Sharpe ratio, lower maximum drawdown, and less risk compared to several benchmark portfolios widely used in the investment literature, such as the minimum-variance and the equally-weighted portfolio. The strategy is also competitive before the outbreak of the pandemic, suggesting that portfolios focused on minimizing tail risk are also useful in calm periods in which extreme returns are not a cause of concern.

CHAPTER 3

DEEP REINFORCEMENT LEARNING FOR PORTFOLIO SELECTION

3.1 Introduction

With Markowitz (1952) introduced the mean-variance framework to construct optimal portfolios, this method has evolved over the years in various directions. It includes the construction of portfolios based on the maximization of investor preferences or the development of sophisticated tools to optimize investors' short-term and long-term objective functions. However, in recent years, machine learning techniques, in particular, have been widely applied to risk management and optimal portfolio management tasks in dynamic financial markets (Henriques & Sadorsky, 2023). RL (Sutton & Barto, 2018) and DL (Goodell et al., 2021; Mavruk, 2022) methods have been widely used to solve portfolio selection optimization problems. Furthermore, RL is a promising method that does not require a specific portfolio baseline to make investment decisions and is solely based on financial market information obtained at a given time period. The RL-based learning process is optimized by maximizing portfolio return or minimizing portfolio risk (Chaouki et al., 2020; Sutton & Barto, 2018) and enhancing portfolio forecasting accuracy by learning from errors.

The recent success of RL in portfolio allocation problems is related to the model's ability to improve expected return reductions and portfolio risk. However, several important issues remain that must be resolved prior to applying these methods for portfolio investment to capture the complex and volatile characteristics of financial markets within the optimal portfolio decision rubric (Aboussalah & Lee, 2020; Bühler et al., 2018; Li et al., 2018; Moody et al., 1998; Moody & Saffell, 2001; Yang et al., 2020; Zhao et al., 2023). In addition to maximizing investment returns, other factors, such as transaction costs and investor risk tolerance, must be considered in portfolio decision-making (Bühler et al., 2018; Li et al., 2018). Moreover, the selection of historical data for RL model training may be problematic (Yang et al., 2020). The characterization of important model features and state variables depends on the training phase, which is based on time-series data. Thus, the selection of specific datasets affects these characterizations and, in turn, the effectiveness of the corresponding portfolio selection strategy (Zhao et al., 2023). Furthermore, RL performance depends on model stability, convergence speed, and the ability of the algorithm to learn the underlying dynamic relationships between variables. A key advantage of these methods is their ability to manage high-dimensional portfolios very efficiently (Li et al., 2018; Yang et al., 2020).

This chapter proposes an advanced model-free DRL framework that constructs optimal portfolios in high-dimensional settings. Specifically, this model combines DL and RL methods to create a DRL framework and optimize portfolio strategies in dynamic, complex, and large-dimensional financial markets. Investor's risk aversion and transaction cost constraints are embedded in the model. Notably, DRL algorithms adopt neural networks as function approximators to automatically learn complex representations from high-dimensional data, thereby allowing them to capture key features and relationships among various assets. The proposed model applies an adaptive portfolio trading method based on a twin-delayed deep deterministic policy gradient (TD3) algorithm. Thus, this study contributes to the DRL portfolio allocation literature by (i) building an investment strategy that naturally accommodates high-dimensional portfolios; (ii) embedding market risk, individual risk aversion, and transaction costs into the objective function; and (iii) solving the complex Bellman equation by combining twin networks, target networks, exploration strategies, and delayed policy updates that address the challenges posed by high-dimensional state and action spaces with nonlinear relationships.

The empirical performance of this strategy-building tool is assessed in an extensive out-of-sample exercise for high-dimensional portfolios with stocks from the Dow Jones Industrial Average and the S&P100 index using cumulative return, maximum drawdown, Sharpe ratio, and additional performance metrics. The introduced model outcomes are compared with those of minimum variance (MV), maximum Sharpe ratio, and other popular DRL methods designed for similar tasks. The empirical results confirm that the proposed model is superior in terms of profitability and risk trade-offs in the presence of transaction costs and different degrees of risk aversion.

3.2 Model and methods

The first block of this section introduces the portfolio choice problem under the presence of transaction costs and investor risk aversion constraints. The second introduces MDP for portfolio trading decisions, and the third describes the proposed DRL-based portfolio and portfolio selection methodology.

3.2.1 Asset allocation problem under portfolio constraints

Consider a financial market with N risky assets that has been trading for T periods, and let $\boldsymbol{p}_t = [p_{1,t}, \cdots, p_{N,t}]^{\mathrm{T}} \in \mathbb{R}_+^N$ denote the vector of asset prices such that $p_{n,t}$ indicates the closing price of the n-th asset at time t. Similarly, let $m_{n,t} \in \mathbb{Z}_+$ and $P_t \in \mathbb{R}$ denote the n-th asset shares and portfolio value at time period t, respectively. Investors perform portfolio trading strategies over the N assets during each period, including buying, selling, and holding, which results in increasing, reducing, and no changes to asset shares, respectively. The portfolio selection decision variable is denominated in this literature as the trading action and is characterized by $k_{n,t}$ for asset n at time t. For each asset, this variable takes three possible values $\{-1, 0, 1\}$, denote selling, holding, and buying, respectively. This strategy can be extended to consider multiple shares of each asset such that the domain of $k_{n,t}$is $\{-k, \cdots, -1, 0, 1, \cdots, k\}$, where -$k$ and k denote the net number of shares that a trader can sell or buy, respectively, in a given period. Traders can set a maximum number of shares,$k \leq m_{max}$, so that at each period t, the number of shares on each asset n is given by $m_{n,t+1} = m_{n,t} + k_{n,t}$. The value of the portfolio at time t is given by the following equation:

$$P_t = P_{t-1} + \boldsymbol{p}_t^{\mathrm{T}} \boldsymbol{k}_t, \tag{3.1}$$

where $\boldsymbol{p}_t^{\mathrm{T}}$ is the transpose of price vector $\boldsymbol{p_t}$, and $\boldsymbol{k_t} = [k_{1,t}, \cdots, k_{N,t}]^{\mathrm{T}}$.

Each trading process (buying, holding, or selling) generally involves transaction costs and it is necessary to consider this factor in the portfolio selection problem. Let ξ denotes a positive constant that characterizes the transaction cost rate of issuing a new trade. Thus, the transaction cost is expressed as (Yang et al., 2020):

$$c_t^{\mathrm{tran}} = \xi \times \boldsymbol{p}_t^{\mathrm{T}} |\boldsymbol{k}_t|. \tag{3.2}$$

For simplicity, this study assumes that the constant is fixed across assets, although this can be modified to accommodate the presence of different transaction costs and different liquidities.

Underlying risk is an important aspect to consider in an objective function. I differentiate two sources of risk that influence asset allocation, namely, true risk of the portfolio position, captured by portfolio variance σ_t^2, and investor risk aversion coefficient β. Here, the variance of the portfolio position is proxied by the sample variance of the portfolio return over the last t days, defined as $\sigma_t^2 = \frac{1}{t}\sum_{i=1}^{t}((P_i - P_{i-1})/P_{i-1} - \mu_t)^2$, where μ_t is the average return of assets over the last t periods. The literature takes different approaches to introduce risk aversion coefficient in portfolio optimization schemes. The most relevant tactic in financial economics is to leverage a utility function to model investor's preferences. Popular utility functions include constant absolute risk aversion and constant relative risk aversion (Chambers & Quiggin, 2007). Herein, this work follows Markowitz's approach and introduce coefficient β that reflects the investor's degree of risk aversion. To model the investor's attitude toward the underlying portfolio risk, I incorporate a cost function in the objective function as follows:

$$c_t^{\text{risk}} = \beta \sigma_t^2. \tag{3.3}$$

3.2.2. Markov decision process for portfolio trading

The portfolio trading decision is modeled as an MDP, characterized by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, r)$ in which $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}(s_{t+1}|s_t, a_t)$ is the transition probability of state $s_{t+1} \in \mathcal{S}$, given the realization of action $a_t \in \mathcal{A}$ based on the observed state $s_t \in \mathcal{S}$ at the t-th learning time step. Similarly, $r_t(s_t, a_t, s_{t+1})$ is the reward function obtained from taking action a_t at state s_t under the realization of state s_{t+1} at time t + 1.

The state space is defined by a vector of state variables such that $s_t = [\boldsymbol{p}_t, \boldsymbol{m}_{t-1}] \in \mathcal{S}$. Similarly, action space $\mathcal{A}$ contains the set of possible actions, $a_t \in \mathcal{A}$. In the portfolio allocation problem, each action is defined as the quantity of an asset a trader wishes to buy, sell, or hold in a given period such that $a_t \in \{-k, \cdots, -1, 0, 1, \cdots, k\}$. Here, this work uses a_t and $k_{n,t}$ interchangeably. The action space can be normalized to [−1, 1], which is a continuous action space that can be used for continuous-action RL algorithms.

For each state of nature and possible action, I define a portfolio policy $\pi(a_t, s_t)$ that describes the trading strategy in a given scenario. Similarly, the reward function $r_t(s_t, a_t, s_{t+1})$ is obtained by implementing the portfolio policy after state s_{t+1} is realized. This study proposes the following reward function that includes the variation in the value of the portfolio as well as the presence of transaction costs. The function accounts for risk by penalizing the volatility of the portfolio by the degree of the investor's risk aversion. The objective reward function is expressed as:

$$r_t = \boldsymbol{p}_t^{\mathrm{T}} \boldsymbol{m}_t - \underbrace{\beta \sigma_t^2}_{\text{Risk cost}} - \underbrace{\xi \boldsymbol{p}_t^{\mathrm{T}} |\boldsymbol{k}_t|}_{\text{Transaction cost}} . \tag{3.4}$$

In DRL, the agent is the entity that executes decision-making, interacts with the environment, and learns how to choose actions to maximize cumulative returns and improve strategy to obtain better results. Specifically, the objective of the agent is to learn an optimal trading policy $\pi(s_t, a_t)$ that selects an action to maximize the reward function. Policies are evaluated using an action-value function, $Q_\pi(s_t, a_t)$ (i.e., the Q-function). Wiering and Otterlo (2012) expressed this function as a Bellman equation:

$$Q_\pi(s_t, a_t) = \mathbb{E}_{s_{t+1}} \left[r_t(s_t, a_t, s_{t+1}) + \gamma \mathbb{E}_{s_{t+1}, a_{t+1}} [Q_\pi(s_{t+1}, a_{t+1})] \right], \tag{3.5}$$

where $0 < \gamma < 1$ is the discount factor, and $\mathbb{E}_{s_{t+1}}[\cdot]$is the conditional expectation on the realization of state s_{t+1}.

QL is one of the most common algorithm in RL. During the learning process, the action-value $Q_\pi(s_t, a_t)$ function is continuously updated recursively to obtain the maximum cumulative reward, thereby obtaining the optimal trading strategy. The temporal difference error quantifies the difference between the expected and actual Q-value as the agent interacts with the environment and moves from one state to another by taking an action. The updated Q-value is determined by multiplying the temporal difference error and learning rate and adding it to the current Q-value (Wiering & Otterlo, 2012). Thus, we have:

$$Q_\pi^{New}(s_t, a_t) \leftarrow Q_\pi^{Old}(s_t, a_t) + \alpha \Bigg(r_t(s_t, a_t, s_{t+1}) +$$

$$\gamma \max_{a_{t+1} \in \mathcal{A}} Q_\pi(s_{t+1}, a_{t+1}) - Q_\pi^{Old}(s_t, a_t) \Bigg), \tag{3.6}$$

where α is the learning rate, $0 < \alpha < 1$, γ is the discount factor, $Q_\pi^{New}(s_t, a_t)$ is the updated Q-value of $Q_\pi^{Old}(s_t, a_t)$, $r_t(s_t, a_t, s_{t+1})$ is the reward when the agent executes action a_t at state s_t and receives the new state, s_{t+1}, from the market environment. $\max_{a_{t+1} \in \mathcal{A}} Q_\pi(s_{t+1}, a_{t+1})$ indicates the maximum Q-value after selecting action a_{t+1} at state s_{t+1}, $Q_\pi^{Old}(s_t, a_t)$ on the right side of Eq. (3.6) denotes the estimated Q-value under action a_t at state s_t, and $r_t(s_t, a_t, s_{t+1}) + \gamma \max_{a_{t+1} \in \mathcal{A}} Q_\pi(s_{t+1}, a_{t+1}) - Q_\pi^{Old}(s_t, a_t)$ is the temporal difference error. This function represents the signal used to adjust estimates over time (Wiering & Otterlo, 2012).

Iterating Eq. (3.6) allows the Q-value function to gradually converge to the optimal Q-function as the agent iteratively interacts with the environment and learns how to make better decisions that maximize the cumulative reward. Moreover, learning rate α plays a key role as it controls the frequency of updates for each Q-value. Specifically, it determines the degree of balance between the new and old values. The investor observes the arrival of new information in each state (e.g., number of assets and asset prices), followed by selecting an available action with optimized policy π that maximizes the Q-value. Subsequently, the investor obtains an instantaneous reward from the market environment, which is used to evaluate the quality of the selected action, thereby resulting in an adjusted portfolio strategy based on the evaluation. The detailed implementation of the learning processes by the investor (agent) is described at the end of the following section.

3.2.3. TD3-based portfolio trading algorithm

This study adopted the TD3 algorithm to search the optimal portfolio trading strategy in dynamic financial markets. TD3 extends and improves the DDPG algorithm by introducing new features to make it increasingly stable during training and to improve convergence speeds. Furthermore, DDPG is an AC algorithm that extends the deterministic policy gradient algorithm to continuous action spaces. It employs a neural network as the actor to approximate the optimal policy and another neural network as the critic to approximate the state-action value function. The actor network directly outputs the deterministic action, given the current state. The critic network learns the Q-value function by taking the state and action as inputs. TD3 algorithm includes a convolutional neural network (CNN) or LSTM actor network with weight $\boldsymbol{\psi}$ and two critic networks with weights $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. The critic networks more effectively avoid overestimating the Q-value function and are, therefore, able to achieve better learning performance. After receiving the portfolio trading policy, $\pi(s_t, a_t)$, from the actor network, the two critic networks update their loss functions as follows:

$$L(\boldsymbol{\theta}_i) = \frac{1}{I}\sum_t^{t+i}\left(y_t - Q_\pi(s_t, a_t|\boldsymbol{\theta}_l)\right)^2, l = 1,2, \tag{3.7}$$

where I is the size of the replay buffer, which is the amount of information used for model training, i is the i-th sample of the replay buffer, and the information set in the replay buffer is denoted as $\mathcal{D}$. As the TD3 has two critic networks, l denotes the l-th critic network, and $l = 1,2$. The difference between the Q-value function$Q_\pi(s_t, a_t|\boldsymbol{\theta}_l)$in Eq. (3.7) and $Q_\pi(s_t, a_t)$ in Eq. (3.6) is that $Q_\pi(s_t, a_t|\boldsymbol{\theta}_l)$ adopts the CNN backbone and optimizes neural network weights, $\boldsymbol{\theta}_l$, using information from the available dataset. y_t is the target Q-value function, expressed as:

$$y_t = r_t(s_t, a_t, s_{t+1}) + \gamma Q_\pi(s_{t+1}, a_{t+1}|\boldsymbol{\theta}_l'), \tag{3.8}$$

where $\boldsymbol{\theta}_l'$ is the updated neural network weights of the l-th critic network.

In TD3, the target Q-value function, y_t, plays a key role in training the portfolio learning model. A target Q-value is adopted in various DRL algorithms to update the Q-values during the learning process, which represents the estimated maximum one-period ahead reward that the agent can achieve by selecting the best action in the next state and discounting it by γ. This target Q-value y_t is used to update the current Q-value function $Q_{\pi}(s_t, a_t|\boldsymbol{\theta}_l)$. Then, the actor network uses the policy gradient method is updated:

$$\nabla_{\psi} J(\boldsymbol{\psi}) \approx \mathbb{E}_{s_t}[\nabla_a Q(s_t, a_t|\boldsymbol{\theta}_l) \mid_{a_t=\pi(s_t;\psi)} \nabla_{\psi}\pi(s_t|\boldsymbol{\psi})]. \tag{3.9}$$

Because a traditional DDPG can sometimes converge to the local optimal portfolio policy, TD3 introduces the following three techniques to avoid this. The first involves target policy smoothing, a technique that reduces the estimation bias of the Q-value and improves model generalization by adding noise on target policy output. The equation for target strategy smoothing is expressed as follows:

$$\tilde{a} = \pi(s_{t+1}|\boldsymbol{\psi}) + \varsigma, \varsigma \sim clip(\mathcal{N}(0, \tilde{\delta}), -\kappa, \kappa), \tag{3.10}$$

where $\pi(s_{t+1}|\boldsymbol{\psi})$ is the action output from the main network, $\tilde{a}$ is the target action after adding noise, ς is Gaussian noise after clipping with variance $\tilde{\delta}^2$ for policy smoothing, and κ is the clipping amplitude.

The second technique is double-QL, which reduces the overestimation of Q-values by maintaining two critic network Q-value functions. The formula for double-QL is expressed as:

$$y = r_t + \gamma \cdot min(Q_1'(s_{t+1}, a'|\boldsymbol{\theta}_1'), Q_2'(s_{t+1}, a'|\boldsymbol{\theta}_2')), \tag{3.11}$$

where y is the target Q-value, r_t is the actual reward, γ is the discount factor, Q_1' and Q_2' are the Q-value functions of the two target networks used to calculate the minimum Q-value.

Next, according to Eq. (3.8) and Eq. (3.9), the portfolio trading agent updates the weights of the three online networks using the following equations:

$$\boldsymbol{\theta}_{l,t} = \boldsymbol{\theta}_{l,t-1} + \alpha_c L(\boldsymbol{\theta}_{l,t-1})\boldsymbol{\theta}_{l,t-1}, \tag{3.12a}$$

$$\boldsymbol{\psi}_t = \boldsymbol{\psi}_{t-1} + \alpha_a \nabla_{\psi_{t-1}} J(\boldsymbol{\psi}_{t-1})\boldsymbol{\psi}_{t-1}, \tag{3.12b}$$

where α_a and α_c are the learning rates of the actor and critic networks, respectively.

The target networks are updated periodically by tracking the main networks using the following update rule:

$$\boldsymbol{\theta}'_{l,t} = \tau\boldsymbol{\theta}_{l,t-1} + (1-\tau)\boldsymbol{\theta}'_{l,t-1}, \tag{3.13a}$$

$$\boldsymbol{\psi}'_{t} = \tau\boldsymbol{\psi}_{t-1} + (1-\tau)\boldsymbol{\psi}'_{t-1}, \tag{3.13b}$$

where τ is the hyperparameter that controls the rate of target network updates.

The third technique is the delayed policy update, which reduces the update frequency of the actor network and forces the critic network to update more frequently than the actor, thus improving stability and convergence speed.

Figure 3.1 presents the TD3-based portfolio trading framework, which aims to devise a portfolio trading strategy that maximizes the portfolio reward $r_t(s_t, a_t, s_{t+1})$ in a dynamic environment. The learning agent (i.e., investor) collects the market state s_t to train the learning model, where the risk awareness behavior of the investor and transaction cost are input to the model through the reward function and included in the training step.

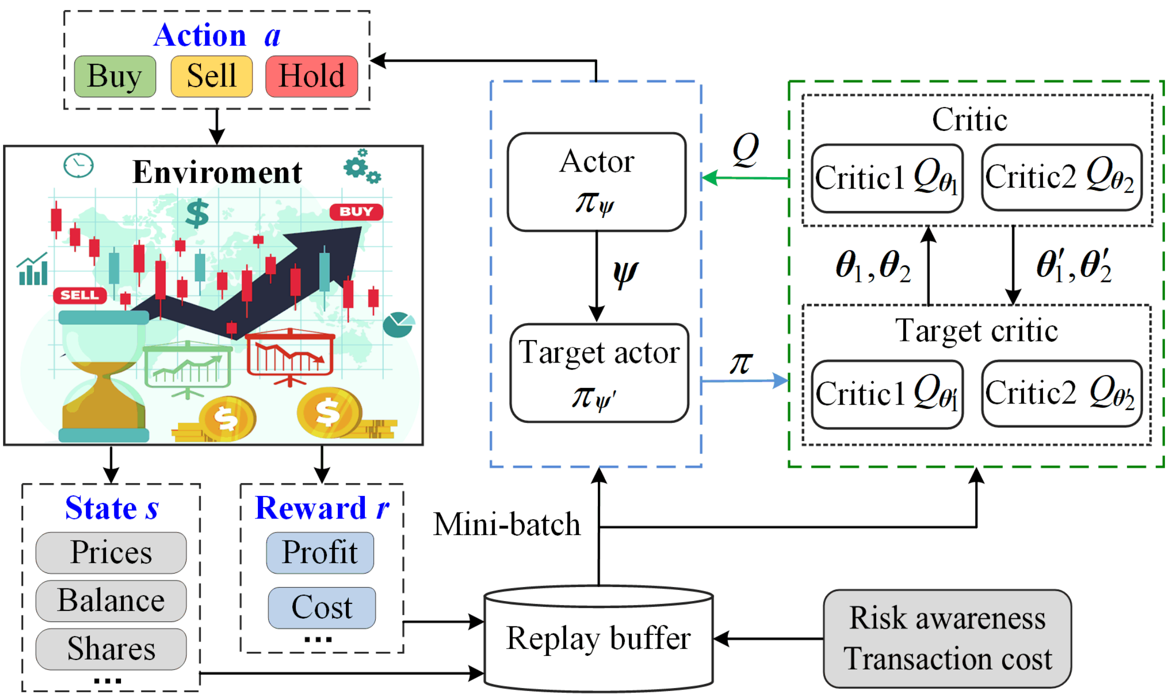


Figure 3.1. TD3-based portfolio trading framework.

Subsequently, a portfolio actiona_t (i.e., buy, sell, or hold), is selected as a combination to maximize the Q-value function. After executing the selected action, the investor receives reward r_t from the financial market and moves one period ahead. Thus, a new set of states, s_{t+1}, and possible actions are encountered at time $t + 1$. The replay buffer stores all training samples from which the agent randomly selects a mini-batch sample set at each training epoch to train the portfolio learning model. The TD3 method for optimal portfolio allocation is listed in Algorithm 3.1. To ease the interpretation of the method, the main steps of the algorithm are summarized as follows:

Step 1: Initialize the TD3-based learning framework, including actor $\pi(s|\boldsymbol{\psi})$ and critics $Q_1(s,a|\boldsymbol{\theta_1})$ and $Q_2(s,a|\boldsymbol{\theta_2})$ with random weights $\boldsymbol{\psi}$, $\boldsymbol{\theta_1}$, and $\boldsymbol{\theta_2}$, respectively, and initialize the replay buffer set, $\mathcal{D}$.

Step 2: Observe state s from the trading market, which contains asset prices and market shares.

Step 3: Input the financial market information into the learning framework, including the asset prices, number of assets, etc.

Step 4: Select one available action, a (buying, holding, or selling), to maximize the reward function, r_t, and execute the action before observing state s_{t+1}, which is realized in the next period.

Step 5: Store transition (s_t, a_t, r_t, s_{t+1}) into the replay buffer, $\mathcal{D}$, where the learning agent will sample a mini-batch of transitions (s_t, a_t, r_t, s_{t+1}).

Step 6: Update the actor and critic networks by applying the gradient based method of Eq. (3.7), (3.8), (3.12), and (3.13).

Step 7: The TD3 model is fully trained after a certain number of learning episodes, when the training loss becomes less than a given tolerance level or all sampled data are completely trained.

Algorithm 3.1. Portfolio risk and transaction cost-aware TD3-based portfolio trading.
Input the maximum number of episodes, I, the maximum number of steps per episode, T, exploration noise standard deviation, δ, policy noise standard deviation, $\tilde{\delta}$, clip factor, κ, mini-batch size, M, discount factor, γ, target policy smoothing coefficient, τ, and delay parameter, d. The financial market information is also inputted in the learning system.
Initialize actor network $\pi(s\|\boldsymbol{\psi})$ and critic networks $Q_1(s,a\|\boldsymbol{\theta_1})$ and $Q_2(s,a\|\boldsymbol{\theta_2})$ with random weights $\boldsymbol{\psi}$, $\boldsymbol{\theta_1}$, and $\boldsymbol{\theta_2}$, respectively.
Initialize target actor network $\pi(s\|\boldsymbol{\psi}')$ and target critic networks $Q_1(s,a\|\boldsymbol{\theta}'_1)$ and $Q_2(s,a\|\boldsymbol{\theta}'_2)$ with weights $\boldsymbol{\psi}' \leftarrow \boldsymbol{\psi}$, $\boldsymbol{\theta}'_1 \leftarrow \boldsymbol{\theta_1}$, and $\boldsymbol{\theta}'_2 \leftarrow \boldsymbol{\theta_2}$, respectively.
Initialize replay buffer $\mathcal{D}$.
For episode i in $1,2,\ldots,I$ do the following:
For step t in $1,2,\ldots,T$do the following:
Observe state s_t in the financial market.
Select the action with exploration noise $a = \pi(s_t\|\boldsymbol{\psi}) + \varsigma, \varsigma \sim clip(\mathcal{N}(0,\delta), -\kappa, \kappa)$.
Execute action a_t and observe the next state, s_{t+1}, and reward r_t.
Store transition (s_t, a_t, r_t, s_{t+1}) in replay buffer $\mathcal{D}$.
Sample the mini-batch of transitions (s_t, a_t, r_t, s_{t+1}) from replay buffer $\mathcal{D}$.
Compute target actions $\tilde{a} = \pi(s_{t+1}\|\boldsymbol{\psi}) + \varsigma, \varsigma \sim clip(\mathcal{N}(0,\delta), -\kappa, \kappa)$.
Compute target values $y = r_t + \gamma \cdot min(Q'_1(s_{t+1}, a'\|\boldsymbol{\theta}'_1), Q'_2(s_{t+1}, a'\|\boldsymbol{\theta}'_2))$.
Update the critic networks by minimizing the mean-squared error loss function:
$L(\boldsymbol{\theta_i}) = \frac{1}{I}\sum_{t=1}^{t+i}\left(y_t - Q_\pi(s_t, a_t\|\boldsymbol{\theta_l})\right)^2, l = 1,2.$
If $t \ mod \ d = 0$,
Update the policy by applying the gradient:
$\nabla_\psi J(\boldsymbol{\psi}) \approx \mathbb{E}_{s_t}[\nabla_a Q(s_t, a_t\|\boldsymbol{\theta}_l) \|_{a_t=\pi(s_t;\psi)} \nabla_\psi \pi(s_t\|\boldsymbol{\psi})]$.
Update the target networks:
$\boldsymbol{\theta}'_{l,t} = \tau\boldsymbol{\theta}_{l,t-1} + (1-\tau)\boldsymbol{\theta}'_{l,t-1}$; $\boldsymbol{\psi}'_t = \tau\boldsymbol{\psi}_{t-1} + (1-\tau)\boldsymbol{\psi}'_{t-1}$.
End for
End for

3.3 Empirical application

This section provides an empirical illustration of the aforementioned methodology and algorithms to construct investment portfolios. This study uses the Yahoo Finance database and selects the 30 constituent stocks of the Dow Jones Industrial Average (DJIA) and the 100 constituents of the S&P100 index as the trading stock pool. Furthermore, I exploit historical daily closing price data from April 1, 2010 to March 9, 2023 for the performance evaluation. The empirical application has three stages, including training, validation, and trading testing, where the entire dataset is divided into three parts. [3] The daily closing price data from April 1, 2010 to January 2, 2020 and January 3, 2020 to April 29, 2021 are utilized for the learning agent model training and validation, respectively. Notably, the validation stage assesses the performance of the method using in-sample information to evaluate the quality of the portfolio training model. Subsequently, the trained model was used to test the trading performance based on the testing dataset in an out-of-sample period from April 30, 2021 to March 9, 2023.

Portfolio performance is usually assessed by comparing the metrics of portfolio competitors. The first metric that I consider is the cumulative portfolio value, which measures the increase in portfolio value at the end of the investment period. Here, the cumulative portfolio value P_T is obtained as the net increment of the portfolio over time. Combining Eq. (3.1) and (3.2), I obtain the net value of the portfolio P_T^{net}, defined as follows:

$$P_T^{\text{net}} = P_0 + \sum_{t=1}^{T} \boldsymbol{p}_t^{\mathrm{T}} (\boldsymbol{k}_t - \xi |\boldsymbol{k}_t|), \tag{3.14}$$

where P_0 is the initial portfolio trading wealth, which is set to $P_0 = 1$ and T denotes the length of the investment time period. The transaction cost rate (ξ) is

[3] This work utilizes Python as the primary software and employ packages including but not limited to torch, multiprocessing, and pyfolio. The computer used for running the empirical results is Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz and RAM is 20 GB.

taken into account in the cumulative portfolio. The final cumulative return, S_T, after T, is defined as follows:

$$S_T = \frac{P_T^{\text{net}} - P_0}{P_0}. \tag{3.15}$$

The cumulative value neglects the presence of risk in the portfolio as it reports only the cumulative gain. A complementary performance measure widely used in the literature is the Sharpe ratio, which measures the portfolio return per risk unit. The Sharpe ratio is defined as:

$$SR = \frac{\mathbb{E}[\rho_t] - \rho_f}{\sigma_{\rho_t}}, \tag{3.16}$$

where ρ_t denotes the portfolio return, $\rho_t = (P_t - P_{t-1})/P_{t-1}$, ρ_f is the risk-free portfolio return, and σ_{ρ_t} is the unconditional volatility of ρ_t, defined as $\sigma_{\rho_t} = \sqrt{\text{Var}(\rho_t)}$.

A related metric that captures the portfolio's maximum potential loss faced by investors is the MDD measure. This is defined as the largest loss from peak to trough:

$$MDD = \max_{t:l>t} \frac{P_t^{\text{net}} - P_l^{\text{net}}}{P_t^{\text{net}}}. \tag{3.17}$$

A mixture of the MDD risk measure and the Sharpe ratio is the Calmar ratio (CR), which provides an alternative characterization of the portfolio risk-adjusted value to evaluate performance. Investors can use this metric to identify portfolios that align with their risk appetite and investment objectives, where CR is defined as:

$$CR = \frac{P_T^{\text{net}}}{MDD}. \tag{3.18}$$

Trading performance comparisons are presented using the following models:

(1). The proposed risk and transaction cost-sensitive (RTC) portfolio based on the TD3 algorithm combined with CNN, denoted RTC-CNN-TD3.

(2). The proposed RTC portfolio based on the TD3 algorithm combined with LSTM, denoted RTC-LSTM-TD3.

(3). The risk and transaction cost-sensitive portfolio based on the RTC-DDPG algorithm combined with CNN, denoted RTC-CNN-DDPG.

(4). The risk and transaction cost-sensitive portfolio based on the RTC-PPO algorithm combined with CNN, denoted RTC-CNN-PPO, where PPO is a common RL portfolio trading method (Aboussalah et al., 2022; Li et al., 2021b).

(5). The minimum variance portfolio trading method, which aims to minimize portfolio risk, denoted min-variance (MV).

(6). The maximum Sharpe ratio portfolio, which maximizes the Sharpe ratio, denoted Max-Sharpe.

The hyperparameter values for portfolio optimization are provided in Table 3.1. To improve the learning efficiency of the proposed DRL algorithm, a suitable parameter must be selected. For the learning rate, if this study sets a learning rate that is too small (e.g., a = 0.0001), it will take longer to achieve convergence. On the contrary, if the learning rate is set too large (e.g., a = 0.01), the method leads to significant fluctuations in the model parameters and destabilizes the training processes. Hence, I select a suitable learning rate (i.e., a = 0.001) in the training model.

Similarly, the number of hidden layers and their sizes should be moderated. If this study sets numerous layers, it may render an overly complex model that demands high computational complexity. However, the limited number of layers may result in poor performance because it would lack the capacity to learn effectively. The remaining hyperparameters are general values for DRL frameworks (Sutton & Barto, 2018).

Table 3.1. Hyperparameter values for portfolio optimization.

Hyperparameter	Value	Hyperparameter	Value
Actor learning rate	10^{-4}	Second hidden layer size	512
Critic learning rate	10^{-4}	Target policy coefficient	10^{-4}
Optimizer	Adam	Max. episode number	1000
Discount factor	0.98	Replay buffer Size	106
Mini-batch size	64	Policy noise	0.02
First hidden layer size	512	Policy noise clip	0.05
Number of hidden layers	2	Exploration noise standard deviation	0.15

3.3.1 Empirical results for DJIA stocks

Figure 3.2 and Table 3.2 present the back-testing portfolio trading performance on 30 DJIA stocks for a risk aversion coefficient of $\beta = 0.005$ and a transaction cost rate of $\xi = 0.05\%$. The first scenario hardly penalizes the presence of risk and transaction costs, and all results were performed during the out-of-sample evaluation (testing) period. Figure 3.2 illustrates the strong performance of the three DRL-based portfolio methods (i.e., TD3, PPO, and DDPG). These portfolios achieve a higher cumulative return than those managed by the other two benchmark methods (i.e., Max-Sharpe and MV). Table 3.2 reports the performance metrics. The RTC-CNN-TD3 portfolio has an annualized return of 26.91%, which is significantly higher than those of Max-Sharpe (2.62%) and MV (3.25%). Moreover, the proposed DRL portfolio method effectively captures the dynamic patterns of each asset price, and the trading performance can be optimized accordingly. From Figure 3.2 and Table 3.2, it can be observed that both the RTC-CNN-TD3 and RTC-LSTM-TD3 portfolios have similar performance, whereas the CNN and LSTM methods have the most structurally comparable structure for portfolio feature extraction.

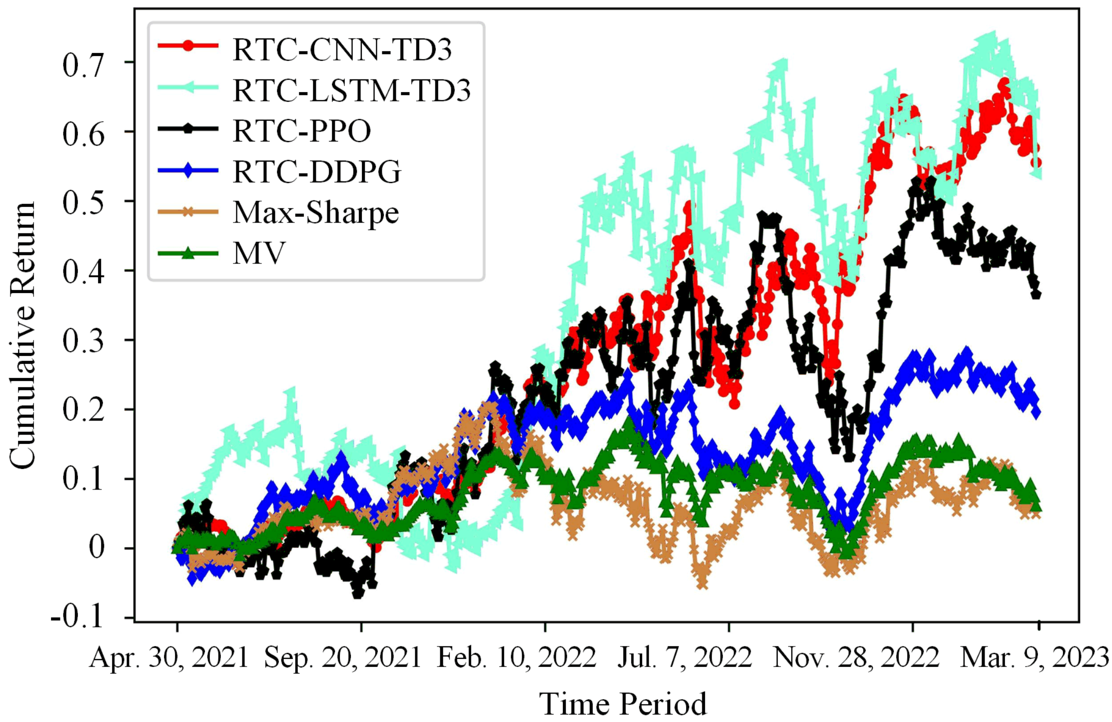


Figure 3.2. Cumulative return performance comparisons using different investment strategies for a risk aversion coefficient of $\beta = 0.005$ and a transaction cost rate of $\xi = 0.05\%$.

Table 3.2. Performance measures of different portfolio methods when $\beta = 0.005$ and $\xi = 0.05\%$.

Method	RTC-CNN-TD3	RTC-LSTM-TD3	RTC-CNN-DDPG	RTC-CNN-PPO	Max-Sharpe	MV
Annual Return (%)	26.91	26.20	10.18	18.28	2.62	3.25
Cum. Return (%)	55.53	53.92	19.62	36.50	4.91	6.10
Annual Volatility (%)	22.01	29.02	17.28	28.08	19.97	13.42
Sharpe Ratio	1.19	0.95	0.65	0.74	0.23	0.31
Max Drawdown (%)	19.11	20.61	18.58	23.52	21.38	16.01
Calmar Ratio	1.41	1.27	0.55	0.78	0.12	0.20

Table 3.2 shows that the Sharpe ratio of the RTC-CNN-TD3 method is the highest, indicating that this strategy outperformed the rest in terms of risk/return tradeoff. Furthermore, the RTC-CNN-TD3 method achieves the highest annual return and CR during the testing period. On the downside, this strategy exhibits higher annual volatility than the two benchmark methods. Interestingly, it is observed that the RTC-CNN-PPO portfolio may not be a suitable investment strategy because of its poor performance in terms of annual volatility (28.08%) and MDD (23.52%). These results demonstrate that DRL-driven portfolios can adequately handle dynamic financial data. In particular, the RTC-CNN-TD3 method generally outperforms the others.

This study further evaluates the portfolio performance of the six competing methods by increasing the values of the two cost-sensitive parameters (i.e., risk aversion coefficient β and transaction cost rate ξ). Figure 3.3 and Table 3.3 present the performance measures of the six methods when $\beta = 0.01$ and $\xi = 0.1\%$. Compared with the results shown in Figure 3.2 and Table 3.2, the cumulative returns of all portfolios decrease to different extents. This is because the increased risk aversion and transaction cost levels discourage risk-taking positions and high portfolio turnover. Moreover, the three DRL-based portfolio methods have similar cumulative returns in the period from April 30, 2021 to September 20, 2021. Interestingly, the proposed RTC-CNN-TD3-based portfolio achieves the highest value (approximately 33.69%) at the end of the out-of-sample period on March 9, 2023. In contrast, the Max-Sharpe and MV portfolios achieve cumulative returns of only 3.97% and 5.11%, respectively. As expected, as the two cost-sensitive parameters (i.e., β and ξ) increase, the annual volatility of the methods drops. The Sharpe ratio also takes lower values than the ones shown in Table 3.2; however, the proposed method still significantly outperforms the competitors, indicating that the RTC-CNN-TD3-based portfolio strategy balances risk and return. From Table 3.2 and Table 3.3, this study finds that although both RTC-CNN-TD3 and RTC-

LSTM-TD3 achieve comparable performances, RTC-LSTM-TD3 exhibits a higher volatility, suggesting that risk-averse investors may prefer RTC-CNN-TD3.

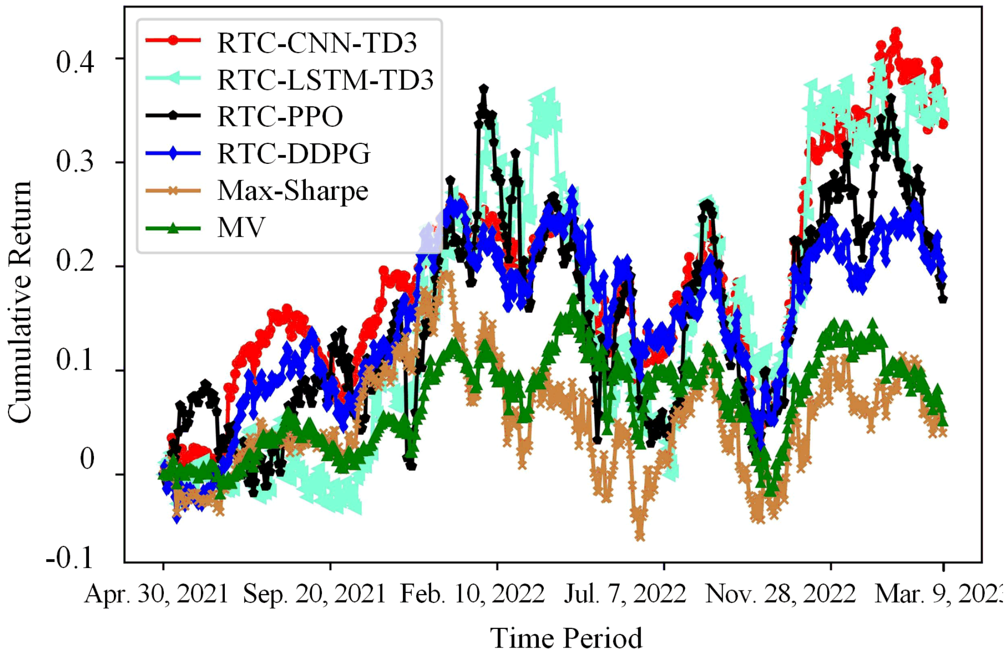


Figure 3.3. Cumulative return performance comparisons using different portfolio trading strategies for a risk aversion coefficient of $\beta = 0.01$ and a transaction cost rate of $\xi = 0.1\%$.

Table 3.3 Performance measures of different portfolio methods when $\beta = 0.01$ and $\xi = 0.1\%$.

Method	RTC-CNN-TD3	RTC-LSTM-TD3	RTC-CNN-DDPG	RTC-CNN-PPO	Max-Sharpe	MV
Annual Return (%)	16.96	17.40	9.87	8.77	2.12	2.73
Cum. Return (%)	33.69	34.62	19.05	16.86	3.97	5.11
Annual Volatility (%)	18.21	26.26	16.47	26.20	19.78	13.28
Sharpe Ratio	0.95	0.74	0.65	0.45	0.21	0.27
Max Drawdown (%)	19.82	26.84	19.19	25.64	21.20	15.87
Calmar Ratio	0.86	0.65	0.51	0.34	0.10	0.17

This work further gauges the effect of the transaction cost rate, ξ, on portfolio performance. Unsurprisingly, Figure 3.4 and Table 3.4 reveal that the cumulative return declines as the transaction cost rate increases. Particularly, the RTC-CNN-TD3-based portfolio strategy achieves high cumulative returns when the transaction cost rate is comparatively low (e.g., 0.01% and 0.05%) but reduces by approximately 50% when the transaction cost rate rises to 0.1% and 0.5%. The reason lies in the fact that the transaction cost level has a sizable effect on the profitability of the strategy and investor behavior. In addition, as shown in Table 3.4, increases of ξ are accompanied by decreases in Sharpe ratio, MDD, and CR, especially when ξ is relatively large (e.g., ξ = 0.5%). These empirical findings suggest that the investment strategy is very sensitive to frequent changes in portfolio composition.

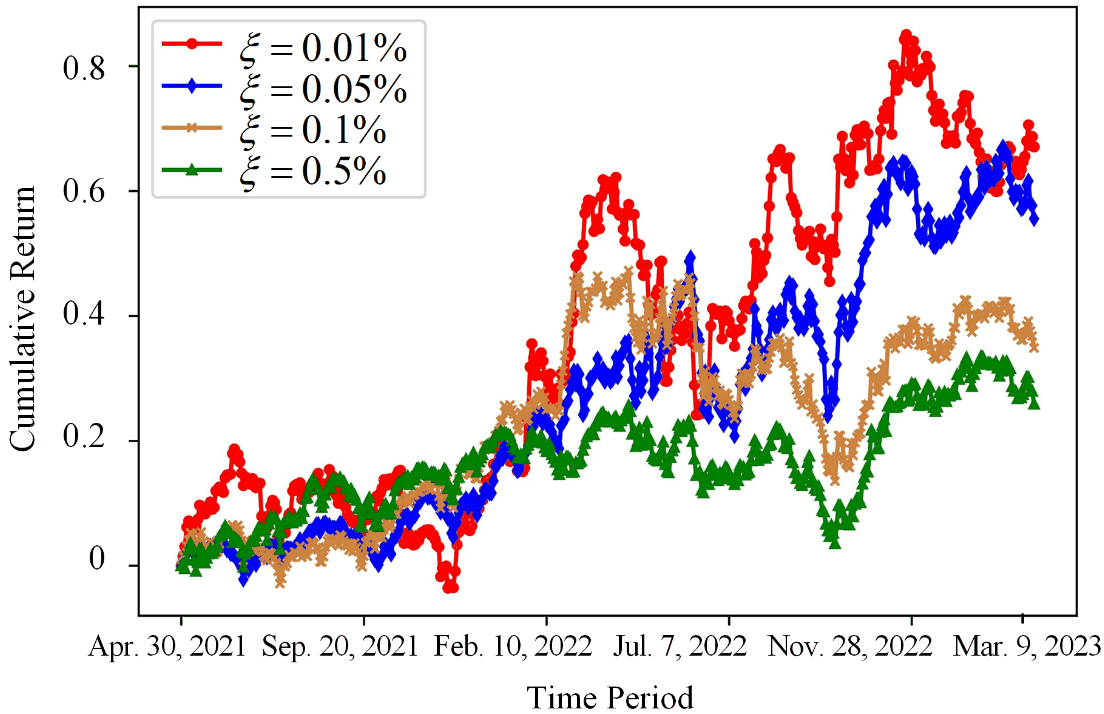


Figure 3.4. Cumulative return performance of the RTC-CNN-TD3 portfolio under different transaction cost rates, ξ, for a risk aversion coefficient of β = 0.005.

Table 3.4 Performance measures of the RTC-CNN-TD3 portfolio under different transaction cost rates, ξ, when β = 0.005.

ξ	0.01%	0.05%	0.1%	0.5%
Annual Return (%)	31.90	26.91	17.51	13.23
Cum. Return (%)	67.06	55.52	34.86	25.89
Annual Volatility (%)	28.42	22.01	20.12	17.38
Sharpe Ratio	1.12	1.19	0.90	0.80
Max Drawdown (%)	23.45	19.11	22.88	17.59
Calmar Ratio	1.36	1.41	0.77	0.75

Finally, I evaluate the effect of the risk aversion coefficient β on the performance of RTC-CNN-TD3 (see Figure 3.5 and Table 3.5).

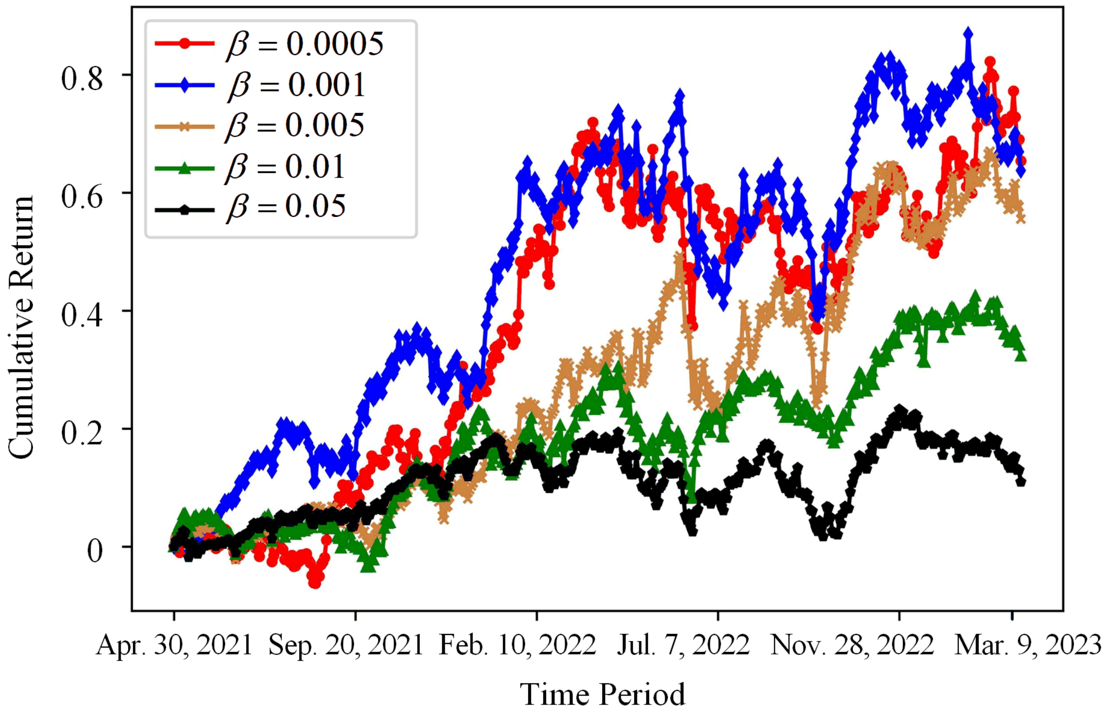


Figure 3.5. Cumulative return performance of the RTC-CNN-TD3 portfolio under different risk aversion coefficients, β, for a transaction cost rate of ξ = 0.05%.

As expected, increases in risk aversion are accompanied by decreases in the annual volatility of the proposed portfolio over the testing dataset, indicating the effectiveness of the proposed cost-sensitive method in incorporating investor's attitudes toward risk. Furthermore, the MDD value declines for larger values of β. As this value is determined by asset price volatility, the results confirm that constraining the volatility of portfolio returns can assist investors in managing downside risks.

Table 3.5 Performance measures of the RTC-CNN-TD3 portfolio under different risk aversion coefficients, β, when $\xi = 0.0005$.

β	0.0005	0.001	0.005	0.01	0.05
Annual Return (%)	31.19	30.50	26.91	16.32	5.79
Cum. Return (%)	65.38	63.75	55.53	32.32	11.00
Annual Volatility (%)	27.06	24.52	22.01	20.26	15.91
Sharpe Ratio	1.14	1.21	1.19	0.85	0.43
Max Drawdown (%)	20.39	21.07	19.11	17.05	14.80
Calmar Ratio	1.53	1.45	1.41	0.96	0.39

3.3.2 Empirical results for S&P100 stocks

This section compares the portfolio performance of the methods in a high-dimensional setting given by the constituents of the S&P100 index. Figure 3.6 illustrates the cumulative return performance of the out-of-sample dataset. As expected, the three portfolio methods based on DRL (i.e., TD3, PPO, and DDPG) achieve better performance than the traditional Max-Sharpe and MV strategies. This is because the proposed learning methods are capable of effectively searching for the optimal portfolio decision strategy in complex, uncertain, dynamic, and large-scale financial trading markets.

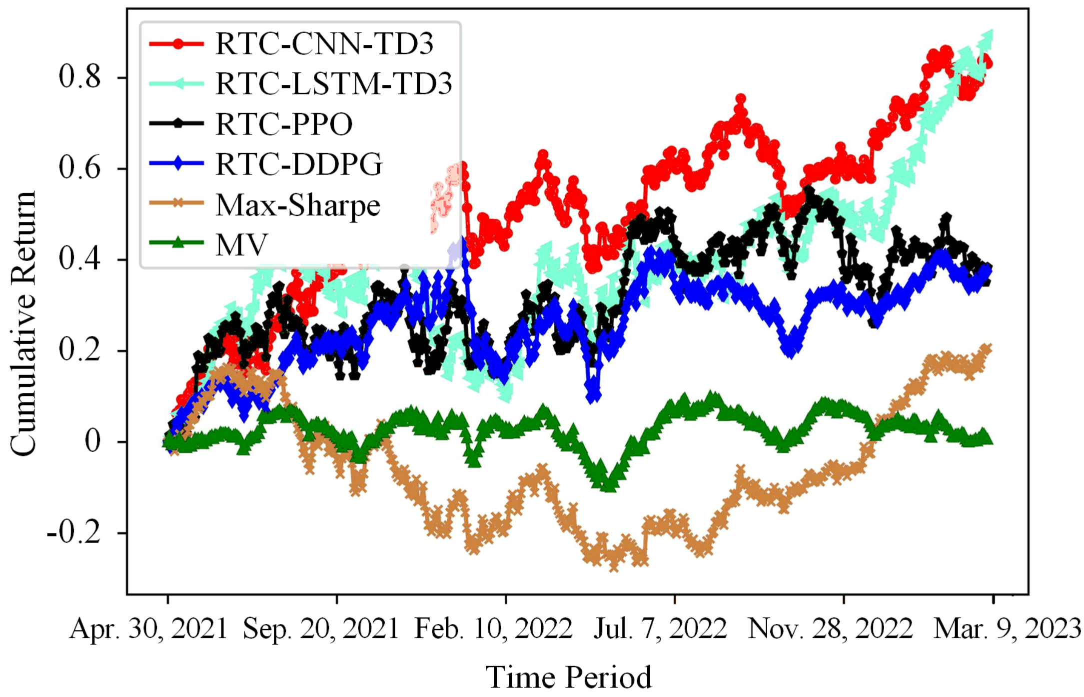


Figure 3.6. Cumulative return performance comparisons using different portfolio trading strategies for a risk aversion coefficient of β = 0.005 and a transaction cost rate of ξ = 0.05%.

Furthermore, as shown in Table 3.6, the proposed two TD3-based portfolio methods achieve the highest annual return and Sharpe ratio in this high-dimensional setting, confirming their ability to balance risk and return. Similar to the previous results, the MV method shows the worst portfolio performance in almost all indicators, apart from the MDD measure. From Figure 3.6 and Table 3.6, I observe that the RTC-LSTM-TD3 method obtains a slightly higher cumulative return than the RTC-CNN-TD3 portfolio at the expense of higher volatility and MDD, posing increased risk for investors.

Figure 3.7 and Table 3.7 illustrate the effect of considering different transaction cost rates, ξ, on the performance of the proposed RTC-CNN-TD3 portfolio for the constituents of the S&P100 index. The portfolio return and market performance measures decreased with increasing ξ. Because a large transaction cost rate induces investors to reduce trading activities, the portfolio return declines. Moreover, when ξ = 0.5%, the proposed learning-based portfolio had a significant loss compared with the scenario given by ξ = 0.01%.

Table 3.6 Performance measures of different portfolio methods when β = 0.005 and ξ = 0.05%.

Method	RTC-CNN-TD3	RTC-LSTM-TD3	RTC-CNN-DDPG	RTC-CNN-PPO	Max-Sharpe	MV
Annual Return (%)	36.92	39.34	17.99	18.37	10.25	3.08
Cum. Return (%)	83.09	89.35	37.49	38.34	20.67	5.93
Annual Volatility (%)	24.79	26.10	22.92	31.72	28.72	14.73
Sharpe Ratio	1.39	1.40	0.84	0.69	0.48	0.09
Max Drawdown (%)	15.90	27.45	23.52	18.95	38.85	15.94
Calmar Ratio	2.32	1.43	0.76	0.97	0.27	0.02

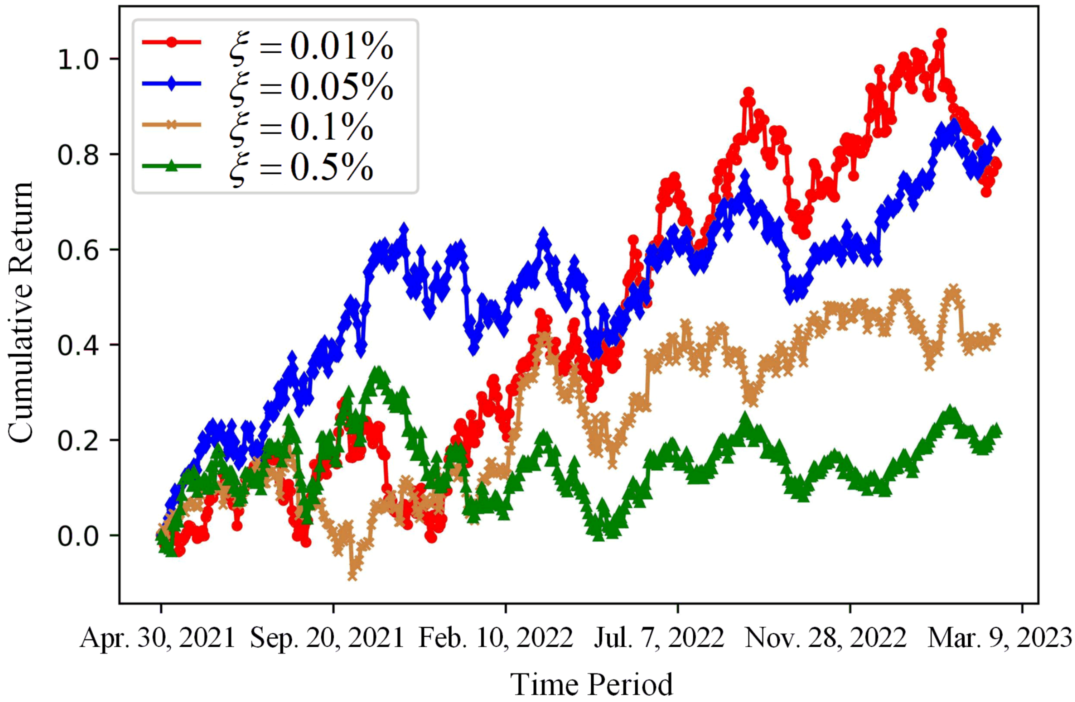


Figure 3.7. Cumulative return performance of the RTC-CNN-TD3 portfolio under different transaction cost rates, ξ, for a risk aversion coefficient of β = 0.005

Table 3.7 Performance measures of the RTC-CNN-TD3 portfolio under different transaction cost rates, ξ, when β = 0.005.

ξ	0.01%	0.05%	0.1%	0.5%
Annual Return (%)	34.85	36.92	20.23	11.03
Cum. Return (%)	77.78	83.09	42.57	22.30
Annual Volatility (%)	30.49	24.79	26.99	25.20
Sharpe Ratio	1.13	1.39	0.82	0.54
Max Drawdown (%)	22.42	15.90	23.90	25.62
Calmar Ratio	1.55	2.32	0.85	0.43

For completeness, Figure 3.8 and Table 3.8 report the cumulative return and performance measures, respectively, for different risk aversion levels. As before, the portfolio return and volatility decrease as β increases. There is an exception, with β = 0.005. For this level of investor risk aversion, it is observed that good cumulative return performance and high Sharpe and CR values, suggesting that this level of intermediate risk aversion is optimal from an investment perspective.

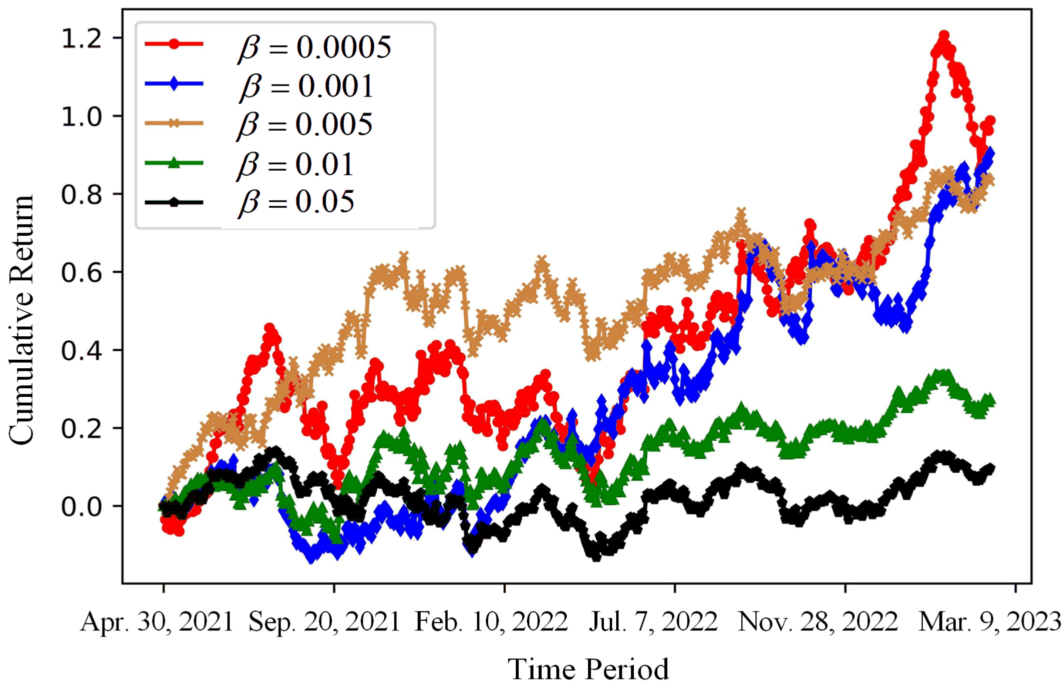


Figure 3.8. Cumulative return performance of the RTC-CNN-TD3 portfolio under different values of the risk aversion coefficient, β, for ξ = 0.05%.

Table 3.8 Performance measures of the RTC-CNN-TD3 portfolio under different risk aversion coefficients, β, when ξ = 0.05%.

β	0.0005	0.001	0.005	0.01	0.05
Annual Return (%)	42.91	39.71	36.92	13.43	4.9
Cum. Return (%)	98.80	90.34	83.09	27.444	9.73
Annual Volatility (%)	33.62	29.97	24.79	21.83	20.26
Sharpe Ratio	1.22	1.27	1.39	0.69	0.34
Max Drawdown (%)	28.27	22.14	15.90	16.69	23.93
Calmar Ratio	1.51	1.793	2.32	0.80	0.21

3.4 Conclusion

This study proposes a DRL method to construct optimal portfolios that perform particularly well, even in high-dimensional settings. The proposal model combines DL and RL methods into a new DRL model that optimizes portfolio allocation. Investor risk aversion and transaction cost constraints are embedded using an extended Markowitz's mean-variance reward function, implemented using a TD3 algorithm.

This work applied these strategies to the constituents of the DJIA and S&P100 and found extremely encouraging results. In particular, the proposed DRL method outperforms traditional investment strategies widely used by practitioners and recent models proposed in the deep reinforcement literature.

CHAPTER 4

HIGH-DIMENSIONAL MULTI-PERIOD PORTFOLIO ALLOCATION USING DEEP REINFORCEMENT LEARNING

4.1 Introduction

Portfolio management supports investors in making decisions on how to allocate resources and funds across a set of assets and over time. Traditional portfolio selection methods typically consider single-period returns. Markowitz's mean-variance optimization model has limitations, particularly in addressing long-term investment horizons and dynamic market conditions. Building on work, the Capital Asset Pricing Model (CAPM) developed by Sharpe (1964) and Lintner (1965) introduces the concept of a risk-free rate and a market portfolio, providing a single-period framework to evaluate asset returns based on systematic risk. While influential, CAPM's assumptions of market efficiency and investor homogeneity limit its practical applicability for long-term portfolio allocation.

Long-term portfolio allocation focuses on how investors can optimally allocate investment assets over extended periods of time to maximize returns while managing risks (Lucey & Muckley, 2011). One of the first contributions in this area was Merton (1969, 1971), who extended the portfolio selection framework to a continuous-time setting, incorporating intertemporal choice and dynamic strategies for long-term investors. His work introduced the concept of dynamic asset allocation, emphasizing the importance of adjusting portfolio weights over time in response to changes in market conditions and investor preferences. It is widely understood, at least since the work of this author, see also Samuelson (1969), that the solution to a multi-period portfolio choice problem can be very different from the solution to a static portfolio choice problem. Unfortunately, intertemporal asset allocation models are hard to solve in closed form unless strong assumptions on the investor's objective function or the statistical distribution of asset returns are imposed.

Traditionally, the extension from single-period to multi-period portfolio optimization has been addressed using stochastic dynamic programming.

Samuelson (1969) and Bellman (1957) developed methods for solving dynamic optimization problems, allowing for the consideration of future states and decisions in portfolio management. Nevertheless, the lack of closed-form solutions for optimal portfolios in multi-period settings has limited the applicability of Merton's model and has not displaced Markowitz's paradigm. This situation began to change due to several developments in numerical methods and continuous time finance models. More specifically, some authors such as Barberis (2000) and Brennan et al. (1997, 1999), among a few others, provide discrete-state numerical algorithms to approximate the solution of the portfolio problem over infinite horizons. Other articles obtain closed-form solutions to the Merton model in a continuous time framework with a constant risk-free interest rate and a single risky asset if long-lived investors have power utility defined over terminal wealth (Kim & Omberg, 1996) or if investors have power utility defined over consumption (Watcher, 2002), or if the investor has Epstein and Zin (1989, 1991) utility with intertemporal elasticity of substitution equal to one (Campbell & Viceira, 1999; Schroder & Skiadas, 1999). Approximate analytical solutions to the Merton model have been developed by Campbell et al. (2003) and Campbell and Viceira (1999, 2001, 2002) for models exhibiting an intertemporal elasticity of substitution not too far from one. An alternative to solving the investor's optimal portfolio choice problem is proposed by Ait-Sahalia and Brandt (2001), Brandt (1999), and Brandt and Clara (2006). These authors show how to select and combine variables to best predict the optimal portfolio weights, both in single-period and multi-period contexts. Laborda and Olmo (2017) focus directly on the dependence of the portfolio weights on the predictor variables through a linear parametric portfolio policy rule. This characterization allows them to apply the generalized method of moments estimation and testing methods to sample analogues of the multi-period Euler equations that characterize the optimal portfolio choice.

Standard methods based on solving Bellman equations struggle with large datasets. Hambly et al. (2023) noted that portfolio optimization often involves high dimensionality, complex non-linear relationships, and constraints, making it difficult for traditional algorithms to adapt to changing market environments and large-scale data. These limitations lead to suboptimal portfolio strategies in dynamic situations. Recent advances in machine learning and artificial intelligence have significantly impacted portfolio management. DRL, as explored by Jiang et al. (2017) and Wang and Zhou (2020), offers robust frameworks for developing adaptive and dynamic portfolio strategies. These methods leverage vast amounts of data and sophisticated algorithms to optimize asset allocation in real time. Compared with standard portfolio allocation methods such as Markowitz's paradigm, DRL aims to search for optimal sequences of actions and then obtain a multi-step task whose objective is to achieve the maximum cumulative reward (Sutton & Barto, 2018). This allows DRL to adapt to complex, high-dimensional, and dynamic environments, making it an attractive method for improving traditional portfolio allocation.

The literature applying these techniques for multi-period portfolio allocation is rapidly growing. Aboussalah et al. (2022) applied the DRL approach to construct optimal portfolios in a multi-period investment scenario. These authors introduce a novel approach based on convolutional neural networks. The method is based on a dynamic multi-period mean-variance optimization model that proves effective in portfolio management. Corsaro et al. (2022) investigated the application of L1-regularization with ML and neural networks-based automatic selection to perform multi-period portfolio selection. Wei et al. (2021) show the benefits of stochastic neural network algorithms to incorporate asymmetric investor sentiment and construct an investment portfolio that balances the return and risk over a multiple holding period. In a recent study, Cui et al. (2024) applied DRL to multi-period portfolio selection and considered different risk aversion levels.

Other challenges to construct optimal investment strategies over multi-period investment horizons are the presence of transaction costs and other market frictions. Research by Constantinides (1986) and Liu and Loewenstein (2002) incorporated these factors into dynamic models to obtain more realistic strategies that account for the costs associated with rebalancing portfolios. Eom and Park (2017) investigated the impact of common factors by implementing a comparative correlation matrix on stocks and emphasized that portfolios constructed by considering market factors significantly outperform other approaches in terms of diversification. However, with the increasing availability of high-dimensional and high-frequency financial data, more sophisticated models need to be developed to handle the complexity and volume of information. Work by Bernardi and Catania (2018) and Zhao et al. (2023) utilized copula models and Monte Carlo methods to capture dependencies and optimize portfolios in high-dimensional settings. Recent studies use the DRL framework to extract cross-asset dependence features in financial investments (Zhang et al., 2022; Xu et al., 2020). Marzban et al. (2023) introduced a WaveNet structure in the DRL framework, capturing cross-asset dependence with a new WaveCorr layer and improving portfolio optimization.

In this chapter, I propose a multi-period portfolio selection model that incorporates investors' risk aversion and different portfolio constraints such as box, turnover, and budget constraints. Given the importance of market features in investments, this work designs an advanced portfolio policy framework to model the dynamics of asset prices, capture asset dependence information, and make optimal asset allocation decisions in high-dimensional settings. Specifically, this chapter presents a multi-period Bellman equation combined with DRL to optimize long-term portfolio selection strategies. A risk-averse and constraint-aware reward function is proposed to maximize portfolio return while adhering to constraints.

Empirical results demonstrate the effectiveness and superiority of the proposed portfolio method in various real-world settings.

The main contributions of this chapter are summarized as follows. First, I formulate a multi-period portfolio selection model featuring strict market constraints, aiming to maximize portfolio wealth by considering risk aversion and transaction costs. To address multi-period investment challenges, I integrate the multi-period Bellman equation with DRL into a MDP framework. Second, by developing a multi-period investment portfolio model, I conduct an in-depth comparative analysis under various investment holding periods, revealing the impact of the investment horizon on portfolio management and comparing single- and multi-period investment approaches. Third, this work develops a new portfolio policy framework that effectively extracts time-series price dynamic patterns using convolutional neural networks. The proposed approach also captures group-wise asset dependence via WaveNet before performing portfolio decision-making using DRL in high-dimensional settings. An adaptive objective function is designed to maximize investors' expected utility over multi-period investment horizons while guaranteeing portfolio constraints, controlling both portfolio risk and transaction costs. Fourth, extensive empirical results under different real-world settings (e.g., various datasets, transaction cost rates, risk-aversion levels, and a large number of assets) verify the effectiveness and superiority of the proposed multi-period portfolio method in terms of cumulative return, Sharpe ratio, and representation abilities compared to existing methods.

4.2 Multi-period portfolio optimization

In this section, I introduce the theoretical framework for constructing multi-period optimal investment portfolios. The section is divided into three blocks. First, this work introduces the mathematical formalism of a multi-period investment portfolio model, specifically model portfolio returns over multiple horizons. Then, practical constraints are considered when establishing the investment portfolios. Finally, an optimized objective function for a multi-period investment portfolio is developed and aligned with real-world investment scenarios.

4.2.1 Mathematical formalism of multi-period portfolio model

In the real world of investments, it is important to figure out how to effectively manage a portfolio by achieving long-term financial goals. According to Barberis (2000), one widely adopted strategy is the buy-and-hold approach, where investors choose a set of assets and hold onto them for an extended period before adjusting portfolio allocation. In contrast, another strategy, known as the re-balancing strategy, emphasizes periodically adjusting the portfolio to adapt to market fluctuations.

In this work, I consider a dynamic multi-period portfolio optimization problem with a rebalancing strategy, which optimizes the reallocation of capital among a number of financial assets. The goal is to optimize the weighting between different assets in the portfolio for the best balance, taking into account profitability targets as well as market conditions. This study rebalances the allocation on a daily basis using updated information in each period.[4]

Here, I reallocate the fund at the beginning of each period over a planning horizon that extends h periods: t, t+1, t+2, ..., t+h. This work assumes that a financial market has N risky assets, and the closing prices of these assets comprise a price vector $\boldsymbol{x}_t \in \mathbb{R}_+^N$ at time period t, where $\boldsymbol{x}_t = (x_{1,t}, \cdots, x_{N,t})$ with $x_{i,t}$ being the price of asset i. A portfolio is managed with a vector of these asset weights $\boldsymbol{\omega}_t = (\omega_{1,t}, \cdots, \omega_{N,t})^T \in \mathbb{R}^N$, where $\omega_{i,t}$ denotes the proportion of capital

[4] It is acknowledged that the literature on multi-period asset allocation usually considers lower frequencies, implying that rebalancing takes place monthly or even quarterly. This chapter uses daily trading data such that the multiperiod asset allocation may involve one month, but the rebalancing is done daily.

reallocation invested in the i-th asset. At the end of each period, investors could actively adjust the value of their portfolios in accordance with the realized returns and the most recent data available from the financial markets. The vector of asset returns is expressed as:

$$\boldsymbol{r}_t = \left(r_{1,t}, r_{2,t}, \cdots, r_{N,t}\right)^{\mathrm{T}} = \left(\frac{x_{1,t}-x_{1,t-1}}{x_{1,t-1}}, \frac{x_{2,t}-x_{2,t-1}}{x_{2,t-1}}, \cdots, \frac{x_{N,t}-x_{N,t-1}}{x_{N,t-1}}\right)^{\mathrm{T}}, \tag{4.1}$$

where $\boldsymbol{r}_{i,t}$represents the i-th asset return on time period t. The portfolio value in period t is denoted by p_t, and it is given by:

$$p_t = p_{t-1}(1 + \boldsymbol{\omega}_{t-1}^{\mathrm{T}} \boldsymbol{r}_t) = p_{t-1}\left(1 + \sum_{i=1}^{N} \omega_{i,t-1} r_{i,t}\right). \tag{4.2}$$

The logarithmic rate of portfolio return on N risky assets in time period t is defined as:

$$\hat{r}_t = ln\left(\frac{p_t}{p_{t-1}}\right) = ln(1 + \boldsymbol{\omega}_{t-1}^{\mathrm{T}} \boldsymbol{r}_t) = ln\left(1 + \sum_{i=1}^{N} \omega_{i,t-1} r_{i,t}\right). \tag{4.3}$$

Sales and purchases of assets typically incur transaction costs, such as exchange fees and execution fees. Thus, it is necessary to consider the transaction cost in the portfolio trading selection. Here, let ξ denote the proportional cost level of each trading, and I set transaction cost rates for purchases and sales equal to ξ_t at time t. Furthermore, let $\psi_t \in [0,1]$ denote the total transaction cost, which is defined as:

$$\psi_t = \xi_t \sum_{i=1}^{N} \left|\omega_{i,t} - \omega_{i,t-1}\right|. \tag{4.4}$$

After considering transaction costs, the terminal portfolio value at time t+h is expressed as: (Zhang et al., 2022; Moody et al., 1998)

$$p'_{t+h} = p_t((1-\psi_{t+1})(1+\boldsymbol{\omega}_t^{\mathrm{T}} \boldsymbol{r}_{t+1}))((1-\psi_{t+2})(1+\boldsymbol{\omega}_{t+1}^{\mathrm{T}} \boldsymbol{r}_{t+2})) \cdots \Big((1-\psi_{t+h})\left(1+\boldsymbol{\omega}_{t+h-1}^{\mathrm{T}} \boldsymbol{r}_{t+h}\right)\Big) = p_t \prod_{k=1}^{h} \Big((1-\psi_{t+k})(1+\boldsymbol{\omega}_{t+k-1}^{\mathrm{T}} \boldsymbol{r}_{t+k})\Big). \tag{4.5}$$

Therefore, the updated logarithmic rate of return at the end of period t in Eq. (4.3) is given by:

$$\hat{r}_t = ln\left(\frac{p_t^{'}}{p_{t-1}^{'}}\right) = ln\left(\frac{(1-\psi_t)p_t}{p_{t-1}^{'}}\right) = ln\left(\frac{(1-\psi_t)p_{t-1}^{'}(1+\boldsymbol{\omega}_{t-1}^{\mathrm{T}}\boldsymbol{r}_t)}{p_{t-1}^{'}}\right)$$

$$= ln((1-\psi_t)(1+\boldsymbol{\omega}_{t-1}^{\mathrm{T}}\boldsymbol{r}_t)). \tag{4.6}$$

According to Eq. (4.6), the total portfolio return over a planning horizon of h holding periods is expressed as:

$$R_h = \sum_{k=1}^{h} \hat{r}_{t+k} = \sum_{k=1}^{h} ln\big((1-\psi_{t+k})(1+\boldsymbol{\omega}_{t-1}^{\mathrm{T}}\boldsymbol{r}_t)\big)$$

$$= \sum_{k=1}^{h} ln(1-\psi_{t+k}) + \sum_{k=1}^{h} ln(1+\boldsymbol{\omega}_{t-1}^{\mathrm{T}}\boldsymbol{r}_t). \tag{4.7}$$

4.2.2 Multi-period portfolio model formulation

I consider the mean-variance function as the one-period utility function modeling investor´s short-term preferences:

$$u_t = \hat{r}_t - \lambda\sigma_t^2, \tag{4.8}$$

where λ is a risk-aversion parameter that balances the quantity placed on the maximization of portfolio return rate $\hat{r}_t$ and the minimization of portfolio risk σ_t^2. The expression for the portfolio risk (variance) of returns over the h planning horizon is formulated as follows:

$$\sigma_h^2 = \frac{1}{h}\sum_{k=1}^{h}\sigma_{t+k}^2 = \frac{1}{h}\sum_{k=1}^{h}(\hat{r}_{t+k} - \bar{r}_h)^2, \tag{4.9}$$

where $\bar{r}_h$ denotes the average portfolio return over the h planning horizon holding periods.

The long-term portfolio allocation problem is characterized by the maximization of the investor's multi-period utility computed over h periods and denoted by $U_{t,h}$. To prioritize the immediate rewards and account for the uncertainty and potential risks associated with future returns, the discount factor γ is introduced (Jaisson, 2022; Olschewski et al., 2021). Moreover, it prevents the issue of infinite accumulation of returns in ongoing tasks and encourages the investor to achieve returns sooner. Therefore, the multi-period objective utility over h periods can be respectively re-defined as:

$$U_{t,h} = u_{t+1} + \gamma^1 u_{t+2} + \cdots + \gamma^{h-1} u_{t+h} = \sum_{k=1}^{h} \gamma^{k-1} u_{t+k}\,, \gamma \in [0,1]. \qquad (4.10)$$

When $\gamma = 0$, the investor's focus is solely on immediate return. For $\gamma < 1$, the reward sequence converges given that the individual reward is finite.

I assume the investor intends to look for a strategy that maximizes its expected long-term utility $U_{t,h}$ over the h-period investment horizon. The multi-period portfolio weight matrix strategy $\boldsymbol{\omega} = (\boldsymbol{\omega}_{t+1}, \boldsymbol{\omega}_{t+2}, \cdots, \boldsymbol{\omega}_{t+h})$ over h horizon periods can be achieved by addressing the following optimization problem that incorporates a set of portfolio constraints, namely, a budget constraint, a turnover constraint, and a box constraint. The budget constraint is given by the following condition:

$$\sum_{i=1}^{N} \omega_{i,t} = 1,\ \forall t. \qquad (4.11)$$

Similarly, the turnover constraint reduces the effect of the transaction costs on portfolio returns. Most studies use the average turnover when evaluating the influence of transaction costs, as it estimates the portfolio weight updates. The portfolio turnover (*TO*) constraint at time period *t* can be expressed as:

$$TO_t = \sum_{i=1}^{N} \left|\omega_{i,t} - \omega_{i,t-1}\right| \le TO_t^{max} \ \ \forall t. \qquad (4.12)$$

where TO_t^{max} is the maximum turnover rate at time period *t*, $0 \le TO_t \le 1$. Finally, this chapter also includes a box constraint that avoids extreme investment positions and fosters the presence of diversification. To do this, I set an upper and lower bound for the maximum and minimum weights in the portfolio. Thus, the box constraint is defined as:

$$0 \le \omega_i^{min} \le \omega_{i,t} \le \omega_i^{max}, \forall i, \forall t. \qquad (4.13)$$

Therefore, under transaction cost, portfolio risk, and market constraints, we develop a multi-period portfolio selection model as the following optimization objective model to optimize the portfolio weight $\boldsymbol{\omega}$, which is formulated as:

$$
\begin{cases}
\max_{\boldsymbol{\omega}} E[U_h] = \max_{\boldsymbol{\omega}} \sum_{k=1}^{h} \gamma^{k-1} E[u_{t+k}] \\
= \max_{\boldsymbol{\omega}} (E[\hat{r}_{t+1} - \lambda\sigma_{t+1}^2] + \cdots + \gamma^{h-1} E[\hat{r}_{t+h} - \lambda\sigma_{t+h}^2]) \\
s.t.\ \ \sum_{i=1}^{N} |\omega_{i,t} - \omega_{i,t-1}| \leq TO_t^{max}\ ,\ \forall t \\
\quad \sum_{i=1}^{N} \omega_{i,t} = 1, \omega_{i,t} \geq 0, \forall i,\ \forall t \\
\ 0 \leq \omega_i^{min} \leq \omega_{i,t} \leq \omega_i^{max}, \forall i, \forall t.
\end{cases}
\tag{4.14}
$$

4.3 Investor's long-term optimization problem

The multi-period optimization problem (Eq.4.14) defined in Section 4.2.2 is complex non-linear programming. However, traditional optimization techniques (Kamali et al., 2019; Bertsimas & Sim, 2004) face difficulties in achieving the optimal investment strategy. Moreover, financial markets are generally dynamic, complex, and large-scale, which increases the difficulty of solving the multi-period portfolio problem. Fortunately, RL is a model-free dynamic programming strategy that can be adopted to tackle the decision-making problem by learning the optimal policy in dynamic markets (Jaisson, 2022; Olschewski et al., 2021).

This section proposes a DRL-based portfolio management framework using a deterministic policy gradient algorithm, where a risk-averse approach for continuous action (i.e., asset allocation) maximizes the expected reward. This work designs an advanced portfolio policy framework to extract both price series dynamics and dependence between assets. In detail, I propose a portfolio framework based on RL with CNN and WaveNet structures for portfolio selection, which is shown in Fig. 4.1. The sequential information based on CNN is adopted to capture the dynamic patterns in each asset price, the dependence information based on WaveNet is exploited to model especially under high-dimensional environment, and the decision-making module is used to perform portfolio weight vector selection.

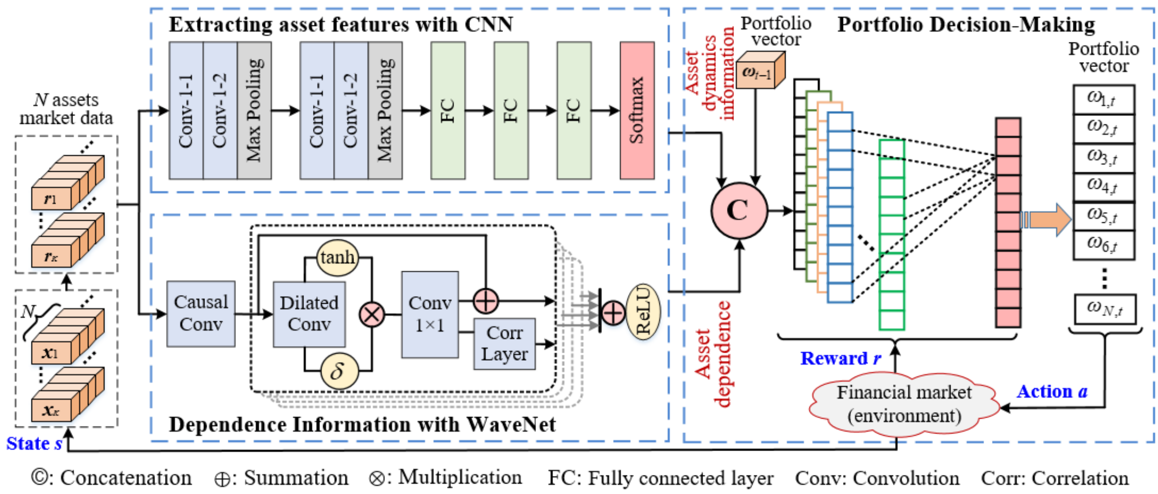


Figure 4.1. The proposed portfolio framework based on DRL with CNN and WaveNet.

4.3.1 Extraction of dynamic price sequence information based on CNN

Gu et al. (2018) stated that CNN is capable of combining features and achieving a higher accuracy on large-scale datasets compared with other DL methods. Therefore, I developed a sequential information module based on CNN to extract the changes in each asset price trend. This module aims to capture the temporal dynamic characteristics of high dimensional asset prices and effectively model time series data.

As illustrated in Fig. 4.1, CNN extracts the dynamic nonlinear features of each asset separately on its multidimensional input data, thereby greatly improving the accuracy of the price mapping features. The structure of CNN for price dynamic sequential information extraction includes not only input and output layers but also convolutional layers, pooling layers, and fully connected layers.

Input layer: This layer pre-processes the original asset returns, including normalizing the amplitude into the same range [0, 1], which reduces the interference caused by differences in the value range of data in various dimensions.

Convolutional layer: The convolutional kernels are used as effective methods for price feature extraction, and the result obtained from price data after convolution is called a Feature Map. The specific process can be expressed as:

$$y_{j,n} = \delta\left(b + \sum_{l=0}^{Z} \sum_{m=0}^{M} \varpi_{l,m} x'_{j+l,n+m}\right), \tag{4.15}$$

where δ represents the activation function, b is the shared bias parameter, Z and M are the length and width of the local receptive field, respectively, and $\varpi_{l,m}$ denotes the shared weight parameters between neurons, $x'_{j+l,n+m}$ is the data corresponding to the input matrix received by the convolutional layer.

The widely used activation function includes sigmoid, tanh, ReLU, etc. The first two are usually observed in fully connected layers, while the latter ReLU is more common in convolutional layers. The fully connected layer is prone to overfitting due to its large number of parameters and the relationship between all elements of the output and input. Therefore, ReLU functions are added between each layer in the model as non-linear activation units to prevent overfitting and increase non-linear expression ability. The specific expression form is given as follows:

$$\delta(x') = ReLU(x') = \begin{cases} x', & x' \geq 0 \\ 0, & x' < 0. \end{cases} \tag{4.16}$$

Pooling layer: The main objective of this layer is to remove unimportant samples from the Feature Map and thus reduce the number of parameters. Max pooling preserves the maximum value within each small block, which is equivalent to preserving the best matching result for that block.

Fully connected layer: Each node of the fully connected layer is connected to all the nodes of the previous layer and is used to synthesize the features extracted from the previous side. All neurons between the two layers are connected with weights, and the fully connected layer is usually at the tail of the convolutional neural network.

Softmax layer: This layer provides non-linear modeling capability by mapping the output results of the convolutional layer into nonlinear maps, which can effectively capture the asset price dynamics.

4.3.2 Cross-asset dependence information extraction based on WaveNet

WaveNet can effectively capture the cross-asset dependence information in portfolio management (Marzban et al., 2023). In this subsection, this study applies the WaveNet framework to estimate the time-varying cross-asset dependence $\boldsymbol{Q}_t$, where the specification of dependence $q_{i,j,t}$ is adjusted over time based on a neural functionφ. This work use the WaveCorr layer in Marzban et al. (2023) as the convolution layer for capturing asset dependence in the WaveNet, which is associated with the following correlation layer (Corr-layer) function set.

$$Q = \{q_{i,j,t} \in \mathbb{R}, a \in \mathbb{R}\}, \tag{4.17}$$

where the vector of asset dependence $q_{i,t}(r_t) = [q_{i,1,t}, q_{i,2,t}, \cdots, q_{i,N,t}]$ between asset i and other assets can be achieved depending on neural function φ which is expressed as:

$$\boldsymbol{q}_{i,t}(\boldsymbol{r}_t) = \left(\varphi(r_{i,t}) \odot (\mathbf{1}\boldsymbol{\omega}_0^{\mathrm{T}}) + \sum_{j=1}^{N} \varphi(r_{j,t}) \odot \left(\mathbf{1}\boldsymbol{\omega}_j^{\mathrm{T}}\right)\right)\mathbf{1} + a, \tag{4.18}$$

where a is the bias for accelerating neural network fitting. A general model operating directly on the assets' price returns is provided, where the joint probability of an input stream $r_t = [r_{1,t}, \cdots, r_{i,t} \cdots, r_{N,t}]$ is modeled as a conditional probability (Van Den Oord et al., 2016; Marzban et al., 2023), i.e.,

$$\eta(r_t) = \prod_{i=1}^{N} \eta(r_{i,t} | r_{1,t-1}, \cdots, r_{N,t-1}). \tag{4.19}$$

Each sample $r_{i,t}$ of the i-th asset is conditioned on the samples at all previous time steps. The causal convolution operation extracts the price dynamics, but it may need large kernel sizes and layers. Thus, apart from adopting causal convolution, the dilated operation is also applied to meet the exponentially large receptive fields with only a few layers while maintaining the network and computational efficiency. In the WaveNet structure, a softmax distribution is adopted to model the conditional distribution $\eta(\boldsymbol{r}_t)$ better, even if the asset returns are implicitly continuous. In addition, residual and parameterized skip connections are adopted to enhance the training convergence. With the help of WaveNet, the dependence information between assets can be constructed in a multi-block framework for obtaining more information and profits of the dynamic market, as illustrated in Fig. 4.1.

4.3.3 Multi-period portfolio decision-making based on DRL

The dynamic asset price features and dependence information obtained from CNN and Wavenet will be combined in 'C,' as shown in Fig.4.1, as the input of the deterministic policy gradient (DPG) model for Portfolio decision-making.

4.3.3.1 Markov Decision Processes (MDP) with multi-period Bellman equation

DRL typically combines the MDP framework to address the challenges posed by the multi-period Bellman equation. MDP provides a mathematical foundation for describing the interaction between an agent and an environment, incorporating elements such as states, actions, rewards, and state transitions.

The portfolio management problem (Eq.4.14) is defined as a Markov Decision Process with a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, u)$. Specifically, the learning agent (i.e., an investor) observes one state $s_t \in \mathcal{S}$ (i.e., the assets' daily prices, the latest asset returns, and portfolio weight) from the market and then chooses an action $a_t \in \mathcal{A}$ (i.e., portfolio weight vector $\boldsymbol{\omega}_t$). Afterwards, the agent will achieve an instantaneous reward u_t and observe the next state s_{t+1}. Here, let $\pi(a_t, s_t)$ denote a portfolio policy, mapping from observed states with a Markovian transition probability $\mathcal{P}(s_{t+1}|s_t = s, a_t = a)$ over available actions that the agent selects. Note that the objective function at the t-th period $u_t$in Eq. (4.12) is the immediate reward function, and the multi-period utility U_h over h period in Eq. (4.13) is the long-term reward function.

According to the discussions of the multi-period optimization problem (Eq.4.14) in Section 4.2.3, the investor aims to get trade-offs between the risk and return of the portfolio under several constraints. In this context, the risk-averse and constraint-awareness reward function at the t-th single period is designed as follows:

$$u_t = \hat{r}_t - \underbrace{\lambda\sigma_t^2}_{\text{Risk-averse}} - \underbrace{c_1\, max(0, TO_t - TO_t^{max})}_{\text{Turnover constraint}} - \underbrace{c_2 \sum_{i=1}^{N} max(0, \omega_{i,t} - \omega_i^{max})}_{\text{Box constraint}} \tag{4.20}$$

where c_1 and c_2 are the punishment parameter for the unsatisfied maximum turnover constraint (Eq.4.9) and maximum weight constraint (Eq.4.10). This reward function will be penalized by any condition that turnover TO_t or portfolio weight $\omega_{i,t}$ goes beyond $TO_t^{max}$and ω_i^{max}, respectively. Notably, as I apply DRL to maximize the objective function u_t which is also the reward function in DRL in a learning framework, the portfolio weight values at the beginning of the training stage may sometimes fail to meet both the turnover constraint and box constraint. Thus, I set these to cost terms to encourage the learning agent to guarantee these two constraints. In addition, the parameter values of c_1 and c_2 are set by balancing the trade-off between the portfolio return $\hat{r}_t$ and the two punishment values. Note that the values of $c_1$and $c_2$should be carefully selected. If parameters are too large, the reward function will sacrifice most of the portfolio return and portfolio risk performance to meet the turnover and box constraints. By contrast, the impact of the parameters on the reward function is limited if $c_1$and $c_2$values are too small Thus, after a number of empirical results of the selections of $c_1$andc_2, it is efficient to set $c_1 = c_2 = 0.5$.

The state-value function for policy in MDP over a planning horizon of h holding periods can be described as follows:

$$V_\pi(s) = E_\pi(U_h|s_t = s) = E_\pi\left(\sum_{k=1}^{h} \gamma^{k-1} u_{t+k}\,|s_t = s\right), \quad (4.21)$$

where $V_\pi(s)$ is the expected reward under the policy π and the state s. The expectation is computed based on the agent's policy mapping π. Similarly, I define the action-value function for the policy π by using Q_π as follows:

$$Q_\pi(s, a) = E_\pi(U_h|s_t = s, a_t = a) = E_\pi\left(\sum_{k=1}^{h} \gamma^{k-1} u_{t+k}\,|s_t = s, a_t = a\right), \quad (4.22)$$

where $Q_\pi(s, a)$ indicates an expected return commencing at state s with performing action a and following policy π. For simplicity, the transition probability is denoted by $\mathcal{P}_{ss'}^{a} = \mathcal{P}(s_{t+1}|s_t = s, a_t = a)$. Additionally, the expected return for transitioning from the current state s to the next state s' (s_{t+1}) by taking action a is denoted by $u_{ss'}^{a} = E(u_{t+1}|s_t = s, a_t = a, s_{t+1} = s')$.

The self-consistency of the value function indicates that certain recursive relationships are required to be met. The multi-period Bellman equation $V_\pi(s)$ can be interpreted as below:

$$\begin{aligned}
V_\pi(s) &= E_\pi(U_h|s_t = s) \\
&= E_\pi(u_{t+1} + \gamma u_{t+2} + \gamma^2 u_{t+3} + \cdots \gamma^{h-1} u_{t+h}|s_t = s) \\
&= E_\pi\left(u_{t+1} + \sum_{k=2}^{h} \gamma^{k-1} u_{t+k}\,|s_t = s\right) \\
&= \sum_a \pi(s, a) \sum_{s'} \mathcal{P}_{ss'}^{a}\left(u_{ss'}^{a} + \gamma E_\pi\left(\sum_{k=2}^{h} \gamma^{k-1} u_{t+k}|s_{t+1} = s'\right)\right) \\
&= \sum_a \pi(s, a) \sum_{s'} \mathcal{P}_{ss'}^{a}\left(u_{ss'}^{a} + \gamma V_\pi(s')\right) \\
&= \sum_a \pi(s, a)\, Q_\pi(s, a).
\end{aligned} \quad (4.23)$$

The Bellman equation for V_π is assigned to this crucial recursive relation. The solution for the Bellman equation is the value function.

4.3.3.2 Multi-period portfolio based on DRL

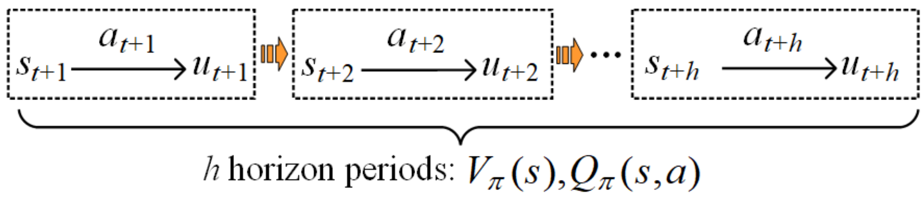


Figure 4.2. The multi-period portfolio trajectory based on DRL.

The learning agent (investor) aims to achieve the multi-period portfolio reward U_h over h horizon periods. Here, this work adopts a DRL-based deterministic policy gradient to obtain the optimal portfolio policy. As shown in Fig. 4.2, the multi-period investment has h state-action pairs i.e., (s_{t+1}, a_{t+1}),(s_{t+2}, a_{t+2}),...,(s_{t+h}, a_{t+h}), and the agent aims to maximize both the state-value function $V_\pi(s)$ and the action-value $Q_\pi(s, a)$ by selecting the optimal action vector $\boldsymbol{a} = (a_{t+1}, a_{t+2}, \cdots, a_{t+h})$ over h horizon periods based on the observed state vector $\boldsymbol{s} = (s_{t+1}, s_{t+2}, \cdots, s_{t+h})$.

To achieve this, DL with a set of neural network parameters $\boldsymbol{\theta}$ is used to specify the policy in the DRL framework, i.e.,$\pi_{\boldsymbol{\theta}}(s, a)$. The objective of DRL is to maximize U_h over the time period interval $[t+1, t+h]$ generated by $\pi_{\boldsymbol{\theta}}(s, a)$, i.e.,

$$\max_{\boldsymbol{\theta}} J(\pi_\theta) = \mathbb{E}_{\pi_{\boldsymbol{\theta}}(s,a)}\big(U(u_{t+1}(\boldsymbol{\omega}_{t+1}), u_{t+2}(\boldsymbol{\omega}_{t+2}), \cdots, u_{t+h}(\boldsymbol{\omega}_{t+h}))\big). \quad (4.24)$$

DPG learning methods enable the agents (i.e., investors) to learn portfolio strategies through real-time interaction with the financial markets. The agent continuously observes market information and learns adaptive strategies during their interaction. This method is usually applicable to real-time decision-making for financial markets based on the current observed state of the environment. It also requires addressing the trade-off between temporal issues and exploration of different strategies in order to find the best portfolio strategy in a constantly changing market.

This work uses the state distribution $\rho^{\pi}(s)$, and the objective function in Eq. (4.24) is given by:

$$J(\pi_{\boldsymbol{\theta}}) = \int_{\mathcal{S}} \rho^{\pi}(s) \int_{\mathcal{A}} \pi_{\boldsymbol{\theta}}(s,a)\, Q_{\pi}(s,a) dads$$

$$= \mathbb{E}_{s\sim\rho^{\pi}, a\sim\pi_{\boldsymbol{\theta}}}[Q_{\pi}(s,a)], \tag{4.25}$$

and the gradient of the objective function is expressed as:

$$\nabla_{\boldsymbol{\theta}} J(\pi_{\boldsymbol{\theta}}) = \int_{\mathcal{S}} \rho^{\pi}(s) \int_{\mathcal{A}} \nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}}(s,a)\, Q_{\pi}(s,a) dads$$

$$= \mathbb{E}_{s\sim\rho^{\pi}, a\sim\pi_{\boldsymbol{\theta}}}[\nabla_{\boldsymbol{\theta}} log\pi_{\boldsymbol{\theta}}(s,a) Q_{\pi}(s,a)]. \tag{4.26}$$

DPG adjusts the parameters $\boldsymbol{\theta}$ of the strategy towards the gradient direction of the objective function to maximize the objective function. The mathematical expression for parameter update is as follows:

$$\boldsymbol{\theta}' \leftarrow \boldsymbol{\theta} + \alpha \nabla_{\boldsymbol{\theta}} J(\pi_{\boldsymbol{\theta}}), \tag{4.27}$$

where α denotes the learning rate and $\nabla(\cdot)$is the first order partial derivative. It is necessary to sample the state and actions under the corresponding distribution.

The decision-making process is to evaluate the potential growth of assets in the near future and consider the portfolio weight vector of the investment based on the previous action a_t to obtain a new portfolio weight $\boldsymbol{\omega}_t$. Under this context, the investor will achieve the multi-period portfolio weight matrix $\boldsymbol{\omega} = (\boldsymbol{\omega}_{t+1}, \boldsymbol{\omega}_{t+2} \cdots, \boldsymbol{\omega}_{t+h})$ over h horizon periods. $\boldsymbol{\omega}$ captures the market investment behavior of intelligent agents, ultimately guiding asset portfolio selection action $\boldsymbol{a} = (a_{t+1}, \cdots, a_{t+h})$, achieving the goal of maximizing profits at certain investment risks.

4.4 Empirical results and analysis

This section provides empirical results to evaluate the performance of the proposed portfolio approach. Handling high-dimensional data is a complicated task that demands our attention. In this paper, I define a portfolio with more than 50 assets as a high-dimensional scenario.

4.4.1 Datasets and competing portfolio methods

This work mainly adopts three public data sets of daily closing prices, including the US S&P500 Index, Canadian S&P/TSX Composite Index, and US Dow Jones Industrial Average Index (DJIA) from 04/01/2010 until 12/07/2023. The datasets are divided into the training set and testing set, with the training period from 04/01/2010 to 31/12/2018 and the testing period from 02/01/2019 to 12/07/2023.

In the empirical evaluations, we compare the portfolio performance by using the following methods: 1) The proposed advanced multi-period DRL-based portfolio method combined WaveNet-enabled dependence information and CNN-enabled sequential information, denoted by MP-Adv-DRL-Cor; 2) The multi-period cost-sensitive portfolio selection method using CNN to extract the dynamic asset return features, temporal correlational convolution block (TCCB) to perform asset correlation and portfolio policy network (PPN) to obtain the portfolio selection decision-making, respectively, denoted by MP-CS-PPN-Cor (Zhang et al., 2022); 3) The multi-period DPG-based portfolio method with Ensemble of Identical Independent Evaluators (EIIE) algorithm (Jiang et al., 2017), denoted by MP-DPG; 4) The equal weight portfolio method, denoted by EW, 5) The single-period DRL-based portfolio method combined with WaveNet and CNN when investment period h=1, denoted by SP-Adv-DRL-Cor. Here, the MP-Adv-DRL-Cor, MP-CS-PPN-Cor, and MP-DPG methods optimize the objective reward function provided in Eq. (4.20), where the transaction cost is considered in portfolio management.

The related hyperparameter values are shown in Table 4.1.

Table 4.1. Hyperparameter values in empirical application.

Hyperparameter	**Value**	**Hyperparameter**	**Value**
Learning rate	2×10^{-4}	Hidden layer size	256
Optimizer	Adam	Decay rate	0.9999
Discount factor	0.98	Planning horizon	36
Mini-batch size	32	Look back window size	36
Hidden layers of CNN	2	Number of epochs	1000
Hidden layers of WaveNet	7	Parameter c_1, c_2	0.5

4.4.2 Performance measures

All the following empirical results are evaluated using the out-of-sample data ("test data"), and different metrics are adopted to measure the portfolio performances by utilizing different methods. Here, this study provides three primary metrics to measure portfolio performance. Firstly, as an indicator of the return on investment, the accumulated portfolio value (APV) is used to evaluate the increase in portfolio value over time. Here, the cumulative portfolio value considers the transaction cost in practice, and it is expressed as:

$$APV = p_0 \prod_{t=1}^{T}((1-\psi_{t+k})(1+\boldsymbol{\omega}_{t+k-1}^{\mathrm{T}}\boldsymbol{r}_{t+k})), \tag{4.28}$$

where p_0 is the initial value of the portfolio and ψ_t represents the transaction cost proportion.

However, APV typically focuses on total value without considering the risk factor. In some cases, the cumulative value of a portfolio may increase because the risk level of the portfolio has also increased. This implies that the portfolio may have taken more risk in an attempt to achieve higher overall value. To further evaluate the performance and risk of the portfolio, I employed the Sharpe ratio (SR) as the second indicator. Specifically, SR is utilized to account for risk and consider the volatility of the returns, which is calculated by dividing the average of risk-free returns by its deviation. A higher Sharpe ratio indicates a better risk-adjusted return on assets, which is given by:

$$SR = \frac{E_t(\hat{r}_t - r_{t,f})}{\sigma_t}, \tag{4.29}$$

where $\hat{r}_t$ is the return rate defined in Eq. (4.7) at the t-th time period and $r_{t,f}$ is the risk-free portfolio return rate.

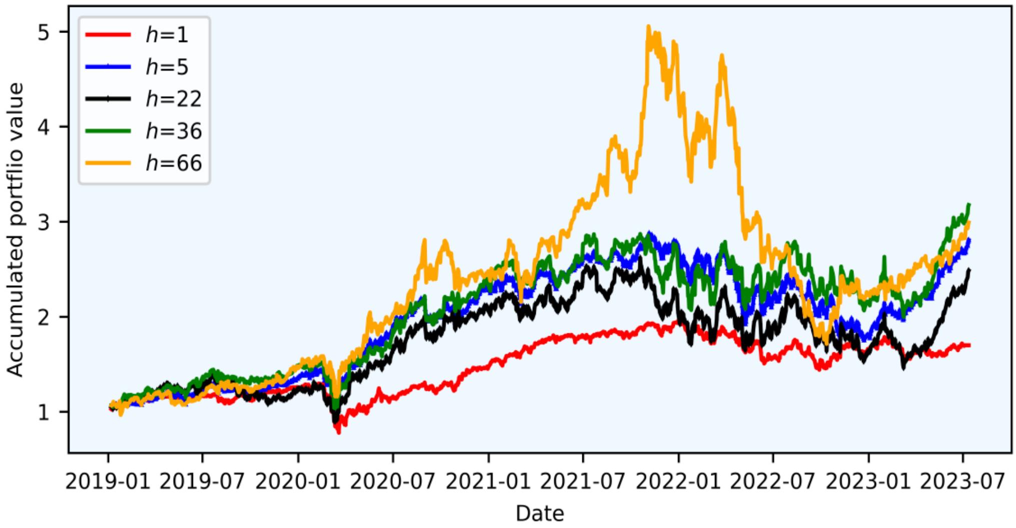
Sharpe ratio uses the standard deviation as a measure of risk without distinguishing upward and downward volatility, which means it may overly focus on short-term adverse fluctuations and overlook positive ones. Consequently, to better identify and evaluate a portfolio's downside risk, Maximum drawdown is also applied. MDD provides a measure of actual losses and reflects the maximum potential loss that investors may face, which is defined as:

$$\text{MDD}=\max_{t:j>t}\frac{(p_t^{'}-p_j^{'})}{p_t^{'}}, \tag{4.30}$$

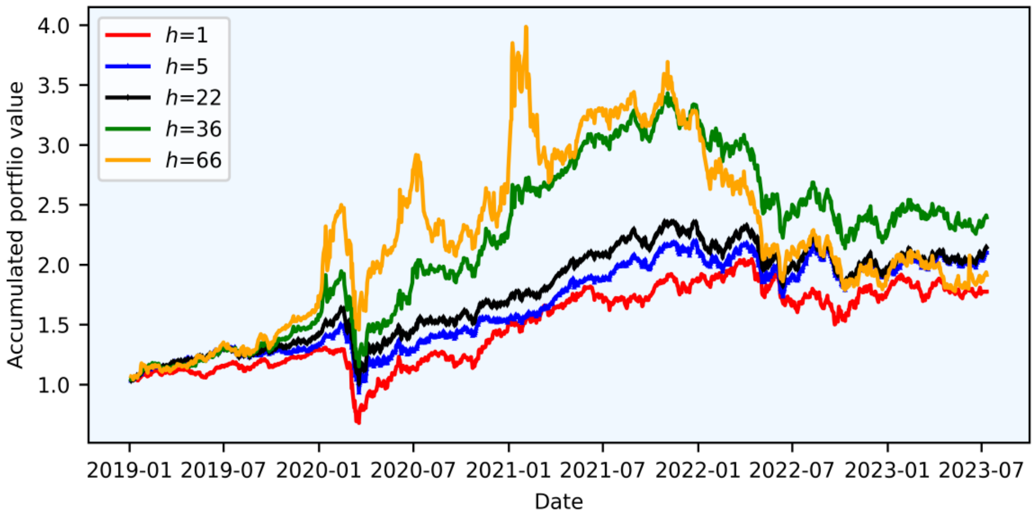
where j is a time period after t, and $j > t$; p_t is the total value of the portfolio on time period t as expressed in Eq. (4.2) and p_j is the aggregate value of the portfolio on day j. In general, the lower MDD is often considered a more stable and lower-risk investment.

4.4.3 Effects of investment horizon on portfolio performance

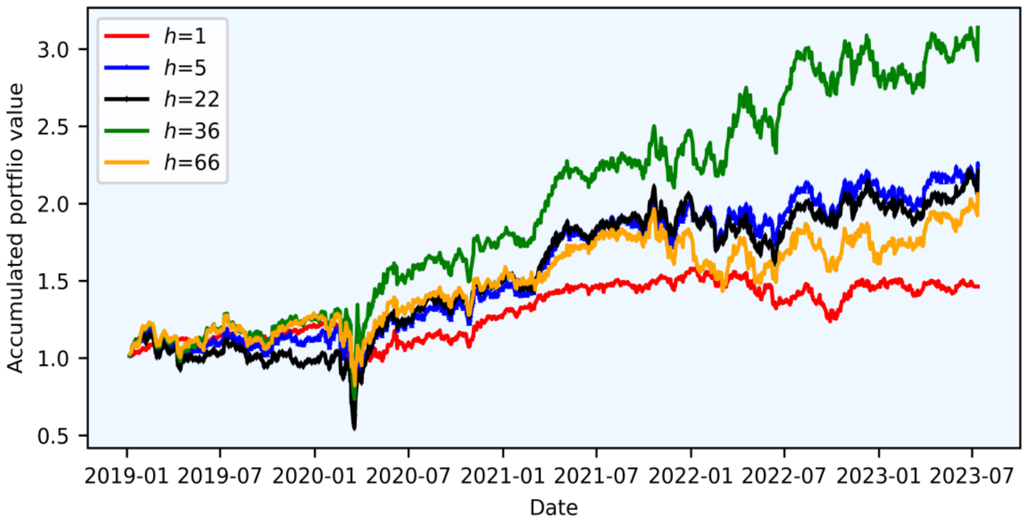
This work first evaluates empirically the influence of the holding period h on our proposed MP-Adv-DRL-Cor method for three different datasets and baseline parameters ξ=0.01% for the transaction costs and λ=0.01 for the risk aversion coefficient, respectively, where the results are shown in Fig. 4.3 and Table 4.2. In this study, assets are selected randomly from various financial indices to minimize selection bias. Here, 100 stocks from the S&P 500 Index, 50 stocks from the S&P/TSX Composite Index, and all 30 stocks from DJIA are chosen for constructing each portfolio.



(a) 100 assets of the S&P 500 dataset



(b) 50 assets of S&P/TSX dataset



(c) 30 assets of the DJIA dataset

Figure 4. 3. Accumulated portfolio value trajectories of MP-Adv-DRL-Cor under different horizon holding periods h on three datasets.

Table 4.2. Portfolio performances under different holding periods h on three datasets.

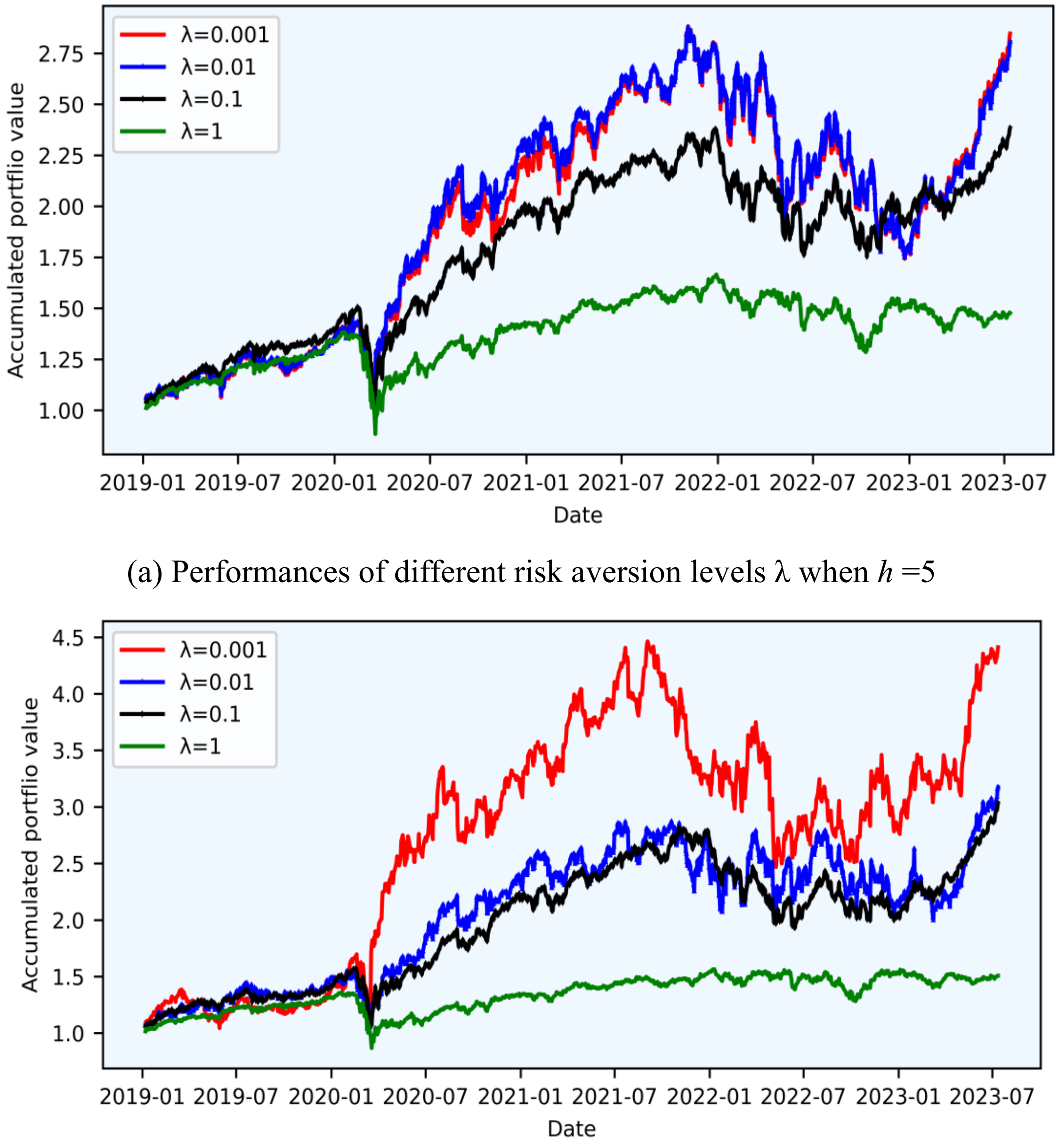
Holding period	Annual return(%)	Annual volatility(%)	Sharpe ratio	Max-drawdown(%)	Turnover
S&P 500 Index					
h=1	12.48	24.57	0.508	39.80	0.007
h=5	25.72	28.89	0.890	39.19	0.087
h=22	22.37	37.08	0.603	44.13	0.099
h=36	29.21	36.14	0.808	32.09	0.170
h=66	27.51	39.43	0.698	66.10	0.107
S&P/TSX Composite Index					
h=1	13.58	25.35	0.536	48.01	0.008
h=5	17.91	21.58	0.830	37.94	0.009
h=22	18.42	22.10	0.833	38.65	0.020
h=36	21.38	27.08	0.790	40.83	0.074
h=66	15.50	37.11	0.418	55.38	0.197
DJIA Index					
h=1	8.808	21.43	0.411	35.34	0.007
h=5	19.79	33.51	0.590	51.72	0.080
h=22	19.23	32.49	0.592	52.63	0.103
h=36	28.88	34.32	0.841	44.24	0.109
h=66	17.40	31.32	0.555	36.22	0.092

In general, the results from Fig. 4.3 and Table 4.2 show that extending the holding period (h=1, 5, 22, 36) leads to increased portfolio gains under the proposed MP-Adv-DRL-Cor method but also results in higher annual volatility. This is because when considering extended investment periods, investors can maximize the total utility over the long term instead of maximizing short-term utility, which typically avoids making myopic investment decisions in the short term. However, it is worth noting that continually extending the investment horizon period does not always yield better portfolio performance. For instance, when h=66, the investment strategy returns lower portfolio values than for h =36 across the three datasets and lower than for h=5 and 22 on the S&P/TSX and DJIA datasets. Financial market uncertainty can have a negative effect on optimal portfolio selection as the investment horizon increases. Thus, the horizon period h needs to be carefully selected. Interestingly, the empirical results from the three datasets indicate that the portfolio performance is optimal when the investment period h =36, based on the evaluation of annual volatility, Sharpe ratio, and Max-drawdown.

The results in Fig. 4.3 and Table 4.2 are broadly consistent with the literature on long-term portfolio allocation. The profitability of the portfolio increases with the investment horizon; annual volatility also increases. Interestingly, however, there is an inverted U shape from h =36 to h =66, suggesting that the cumulative return and risk decrease for longer investment horizons. These results are carried out with daily data, so the above figures correspond to portfolios of 36 and 66 trading days, respectively. These patterns are observed for most performance measures and the three datasets.

4.4.4 Portfolio performance under different risk aversion levels

Next, this work demonstrates the effect of the risk aversion coefficient λ on portfolio performance for 100 assets within S&P 500 dataset, over holding period h=5, and 36, at transaction cost rate ξ=0.01%. As shown in Fig. 4.4, the accumulative portfolio value tends to decrease with the increase of λ, and its trajectories also become flatter with fewer fluctuations during this process. Correspondingly, as indicated in Table 4.3, annual volatility declines with the increase of risk aversion coefficient λ. This downward trend is especially noticeable when λ rises from 0.1 to 1, reflecting an increased aversion to risk among investors. Such results stem from the significant impact of the risk aversion coefficient λ on portfolio risk. With the growth of λ, investors are more inclined to select conservative investment strategies to mitigate the risks in the portfolio. This preference leads to a decline in trading frequency and investment activity, as shown in Table 4.3.



(a) Performances of different risk aversion levels λ when h =5

(b) Performances of different risk aversion levels λ when h =36

Figure 4.4. Accumulated portfolio value trajectories of MP-Adv-DRL-Cor under different risk aversion levels λ when h =5 and 36.

Table 4.3. The portfolio performances under different risk aversion coefficients λ.

λ	Annual return(%)	Annual volatility(%)	Sharpe ratio	Max-drawdown(%)	Turnover
h=5					
λ=0.001	26.11	29.96	0.872	39.55	0.082
λ=0.01	25.71	28.89	0.890	39.19	0.087
λ=0.1	21.25	23.55	0.902	31.17	0.021
λ=1	9.037	19.70	0.459	36.08	0.017
h=36					
λ=0.001	39.99	38.44	1.014	44.66	0.333
λ=0.01	29.21	36.14	0.808	32.09	0.170
λ=0.1	27.91	25.99	1.074	31.64	0.064
λ=1	9.539	19.22	0.496	36.05	0.032

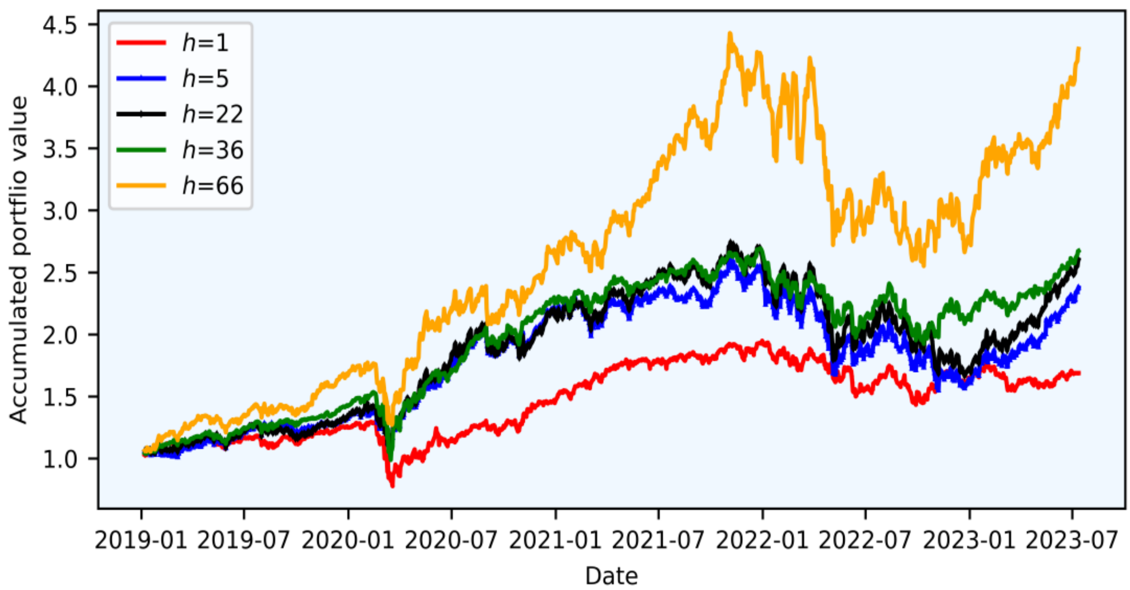
Specifically, when the risk aversion coefficient is very high, i.e., λ=1, the portfolio's volatility is significantly reduced. Consequently, the potential for substantial annual returns and a high Sharpe ratio is limited. For instance, the annual return is only 9.54% with λ=1 compared to 29.21% when λ=0.01 under *h*=36. The results indicate that the proposed MP-Adv-DRL-Cor method can flexibly adjust the portfolio risk level by setting the risk aversion coefficient, which is vital in managing downside risks. Notably, the risk aversion coefficient λ is set based on the investor's behavior. In the study of Zhang et al. (2022), the value λ was estimated, and they found that setting λ around 0.01 achieves a better balance between risk and return. Similar to the result from Zhang et al. (2022), the investors achieve a lower portfolio value when the risk aversion coefficient rises from 0.001 to 0.1, according to the results shown in Fig. 4.4 and Table 4.3.

The results in Fig. 4.4 and Table 4.3 demonstrate that increases in investor's risk aversion lead to more conservative portfolios reflected in lower volatilities and, hence, lower realized returns. This result is monotonic on the level of risk aversion coefficient and across investment horizons.

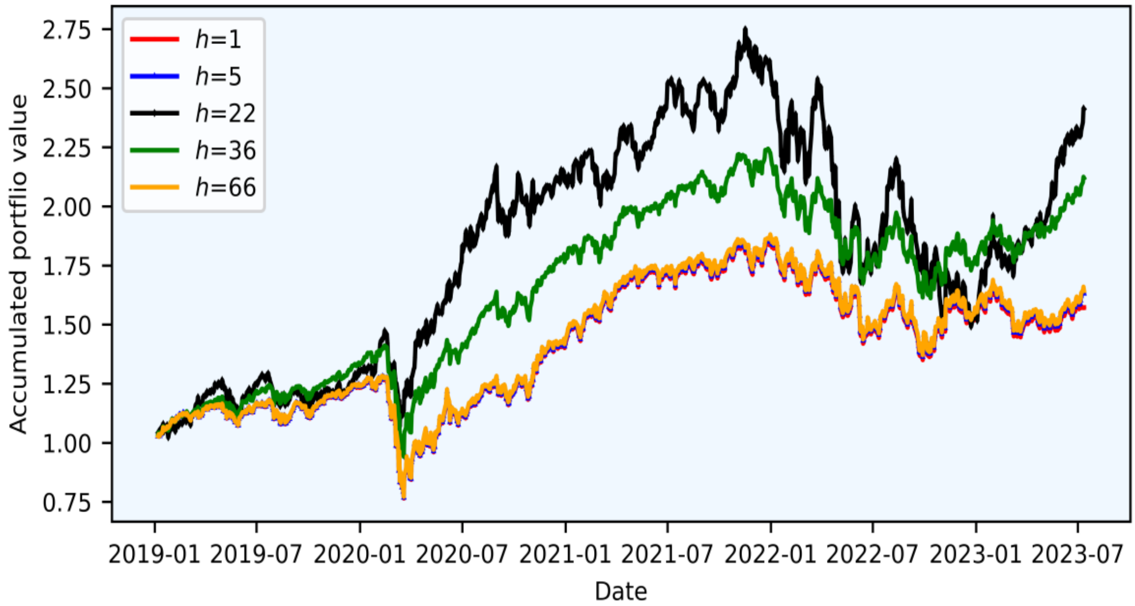
4.4.5 Portfolio performance under different transaction costs

This subsection evaluates the portfolio performances of the MP-Adv-DRL-Cor method under different transaction cost rates (ξ=0.05% and 0.5%) when λ=0.01 on 100 assets of the S&P 500 dataset. It is observed from Fig. 4.5 and Table 4.4 that the annual returns are higher when the transaction cost rate ξ is low at 0.05%, compared to a higher rate of 0.5%. Importantly, the MP-Adv-DRL-Cor can effectively reduce the portfolio turnover when the transaction cost rate ξ increases.

By employing the DRL framework, the proposed algorithm can predict market trends and optimal trade timing. Therefore, it reduces unnecessary trading activities and improves trading efficiency, ultimately leading to higher portfolio returns.



(a) Performances of different holding periods h when ξ=0.05%



(b) Performances of different holding periods h when ξ=0.5%

Figure 4.5. Accumulated portfolio value trajectories of MP-Adv-DRL-Cor under different holding periods h when ξ=0.05% and 0.5%.

Table 4.4. Portfolio performances under different holding periods h when ξ=0.05% and 0.5%.

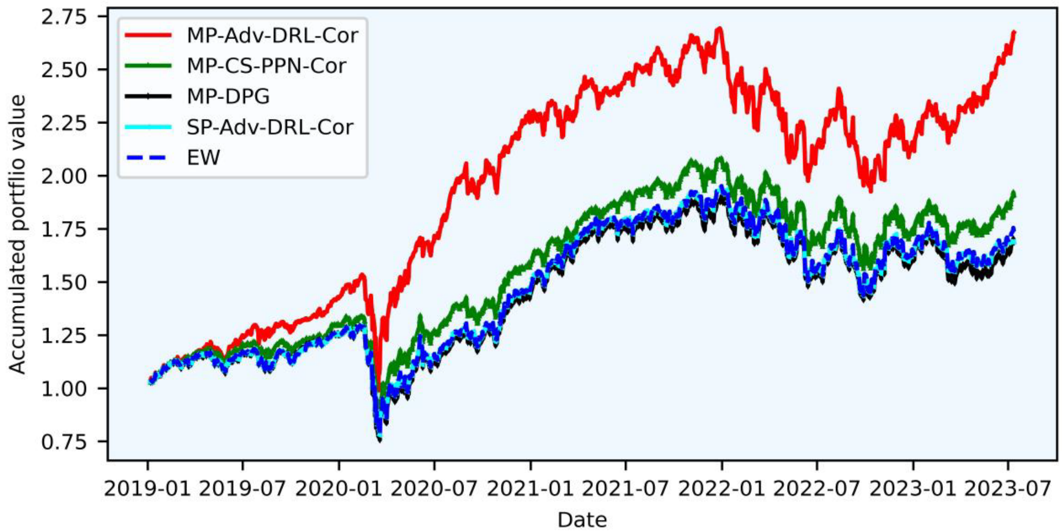
Holding period	Annual return(%)	Annual volatility(%)	Sharpe ratio	Max-drawdown(%)	Turnover
Transaction cost rate ξ=0.05%					
h=1	12.32	24.57	0.502	39.82	0.007
h=5	21.20	27.45	0.772	40.35	0.019
h=22	23.63	28.52	0.829	39.32	0.033
h=36	24.38	24.83	0.982	35.54	0.030
h=66	38.21	33.02	1.157	42.34	0.063
Transaction cost rate ξ=0.5%					
h=1	10.56	24.56	0.430	40.00	0.007
h=5	11.48	24.58	0.467	39.97	0.006
h=22	21.56	30.38	0.710	45.73	0.005
h=36	18.14	23.19	0.782	33.25	0.006
h=66	11.64	24.52	0.475	39.80	0.006

These results are qualitatively similar to the analysis of risk aversion. Transaction costs reduce the profitability of portfolios. Interestingly, this result is stronger as the investment horizon rises entailing sharper declines for h=66.

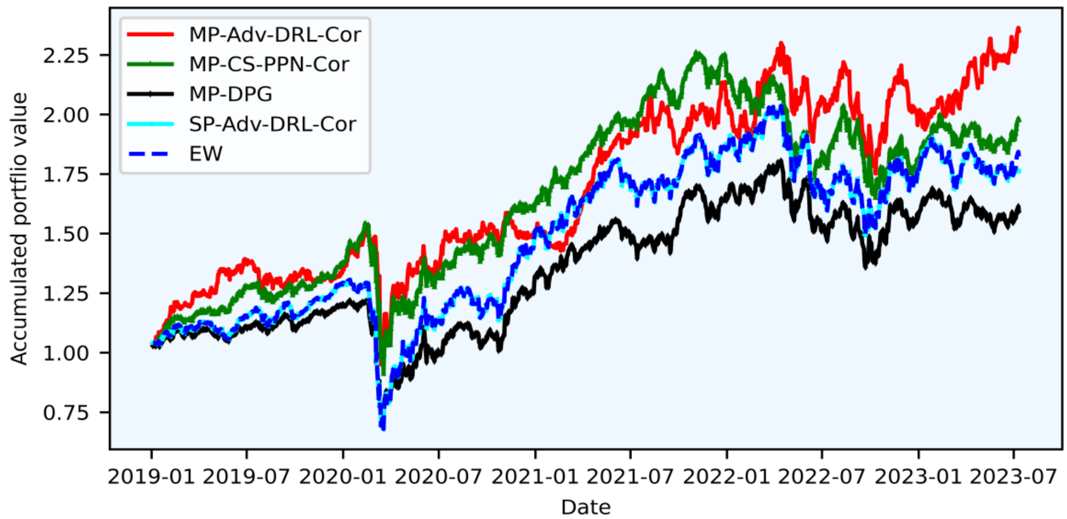
4.4.6 Portfolio performance comparisons

This subsection compares the portfolio performances of the five methods under different datasets when h=36. Here, 100 stocks randomly picked from the S&P500 Index, 50 stocks from the S&P/TSX Composite Index, and all 30 stocks from DJIA comprise each portfolio. The cumulative portfolio value and performance measures are presented in Fig. 4.6 and Table 4.5. The profitability results for the MP-Adv-DRL-Cor method are higher than for the other four approaches under three different benchmark portfolios. Interestingly, the MP-DPG method achieves the poorest performance on the three datasets. This is because the gradient descent optimization of DPG may easily converge into a local optimal point, resulting in a bad investment selection. Unlike MP-DPG, both MP-Adv-DRL-Cor and MP-CS-PPN-Cor strategies adopt learning methods and neural network techniques to extract asset price features and cross-asset dependence characteristics in dynamic financial markets. Thus, these two portfolio methods achieve better annual return and Sharpe ratio, which outperform other algorithms in balancing risk and return.

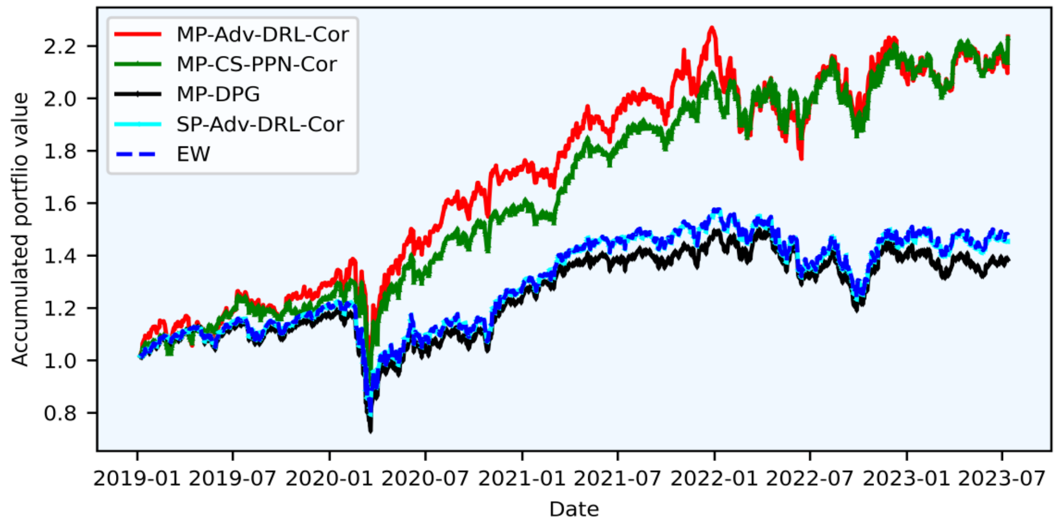
It is worth noting that the proposed learning method achieves the best portfolio performance in terms of annual return and Sharpe ratio on both the DJIA Index and S&P/TSX datasets. However, it has higher annual volatility and max-drawdown than other methods. The EW method performs poorly on both annual return and Sharpe ratio due to non-optimization. Moreover, the turnover of the DRL algorithms (MP-Adv-DRL-Cor, MP-CS-PPN-Cor, and MP-DPG) is much higher than that calculated by the EW method in Table 4.5. This is because the DRL algorithm maximizes returns by dynamically adjusting the weights, which may lead to more frequent trading decisions.



(a) 100 assets of S&P500 dataset



(b) 50 assets of S&P/TSX dataset



(c) 30 assets of the DJIA dataset

Figure 4.6. Accumulated portfolio value trajectories of five methods on three out-of-sample datasets (S&P500, S&P/TSX, and DJIA).

Table 4.5. The portfolio performances of five methods under different datasets.

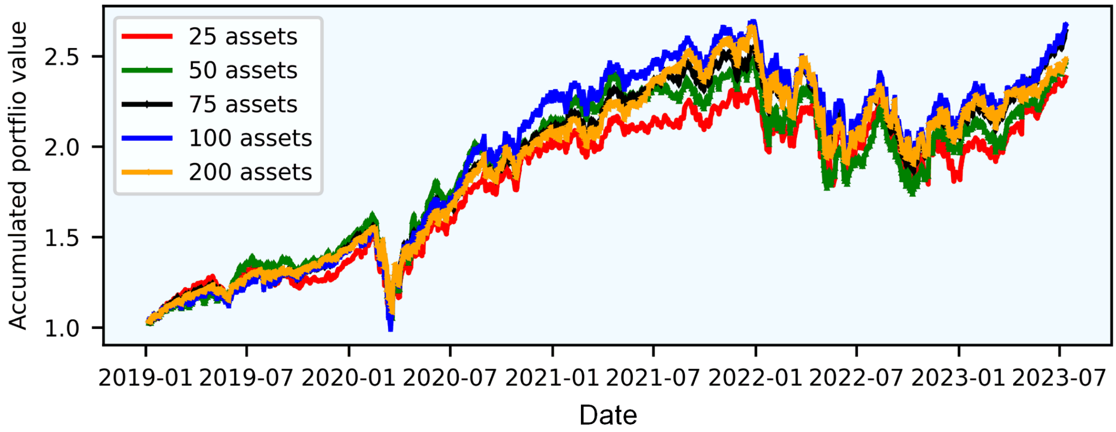
Method	Annual return(%)	Annual volatility(%)	Sharpe ratio	Max-drawdown(%)	Turnover
S&P100 Index					
MP-Adv-DRL-Cor	24.38	24.83	0.952	35.54	0.030
MP-CS-PPN-Cor	15.40	23.78	0.680	37.41	0.009
MP-DPG	12.28	25.78	0.477	41.40	0.034
SP-Adv-DRL-Cor	12.32	24.57	0.502	39.82	0.007
EW	13.02	24.59	0.530	39.80	0.006
S&P/TSX Composite Index					
MP-Adv-DRL-Cor	20.88	25.94	0.805	37.59	0.038
MP-CS-PPN-Cor	16.31	22.54	0.723	41.03	0.006
MP-DPG	10.90	24.05	0.453	43.59	0.118
SP-Adv-DRL-Cor	13.38	25.35	0.528	48.04	0.008
EW	14.26	25.39	0.562	48.19	0.007
DJIA Index					
MP-Adv-DRL-Cor	19.54	25.32	0.772	32.59	0.052
MP-CS-PPN-Cor	19.47	23.11	0.842	30.90	0.092
MP-DPG	7.48	22.26	0.336	38.64	0.057
SP-Adv-DRL-Cor	8.66	21.43	0.404	35.36	0.007
EW	9.05	21.43	0.422	35.38	0.005

The EW method assigns equal weights to the assets in the portfolio, which means the portfolio's construction is relatively simple and less likely to require frequent adjustments in response to market changes. Even though both MP-Adv-DRL-Cor and MP-CS-PPN-Cor methods adopt correlation networks, MP-CS-PPN-Cor with the TCCB framework does not follow the property of asset permutation invariance. Therefore, the portfolios can vary greatly when the ordering of assets changes. Unlike TCCB, Wavenet contains a simpler permutation invariant structure that can efficiently capture asset correlation, thus achieving higher portfolio performance in different datasets.

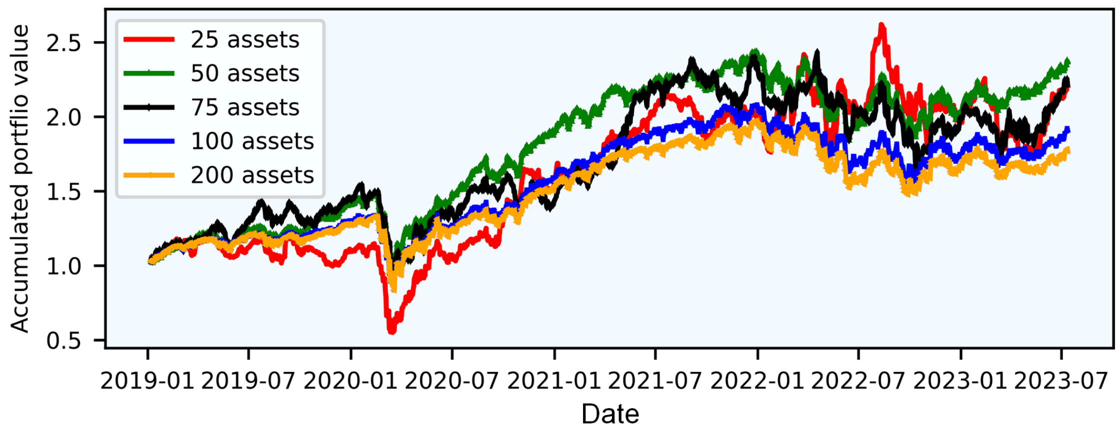
The results demonstrate the effectiveness and superiority of the MP-Adv-DRL-Cor portfolio with WaveNet compared to the other competing learning methods. As expected, the SP-Adv-DRL-Cor method achieves significantly lower cumulative portfolio gains, annual return, and Sharpe ratio than the multi-period portfolio method, similar to the MP-Adv-DRL-Cor and MP-CS-PPN-Cor strategies. This result is because the multi-period strategic investment employs a dynamic asset allocation to adapt the portfolio weights under time-varying financial conditions. Also, the multi-period strategy is more effective in managing the portfolio's overall risk by considering long-term risk and return. In contrast, a single-period (tactical investing) focuses more on short-term market volatility and takes advantage of market opportunities to maximize wealth in the short term while ignoring long-term investment opportunities.

4.4.7 Portfolio performance in high dimensions

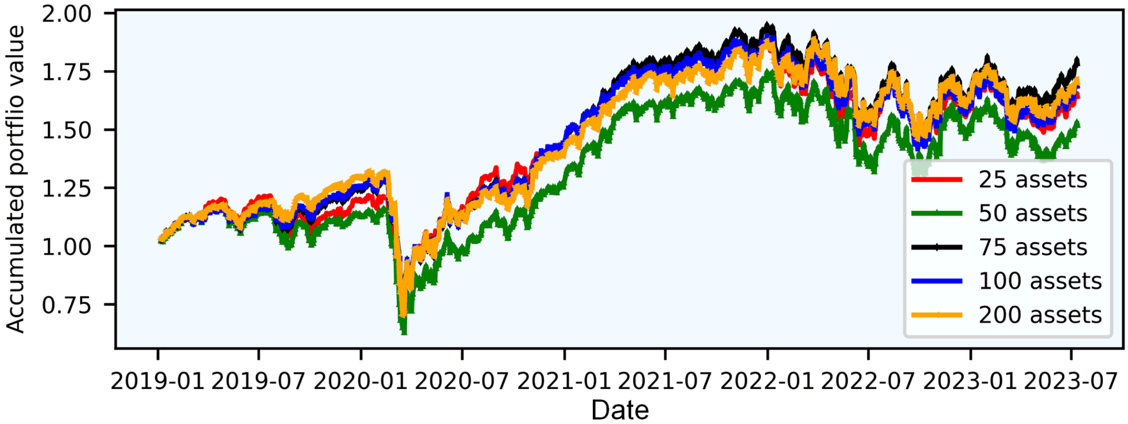
Finally, this subsection demonstrates the influence of the dimension of the portfolio on performance under five methods. I carry out the empirical application with increasing subsets of 25, 50, 75, 100, and 200 assets from S&P500 index datasets, and the results are provided in Fig. 4.7 and Table 4.6. The cumulative portfolio gains, annual return, and Sharpe ratio of the proposed MP-Adv-DRL-Cor portfolio method improve as the number of assets increases (except for 200 assets). These findings can be attributed to the higher dimension of the portfolio in the training set, which allows investors to access more financial market information and improve diversification (risk-return tradeoff).



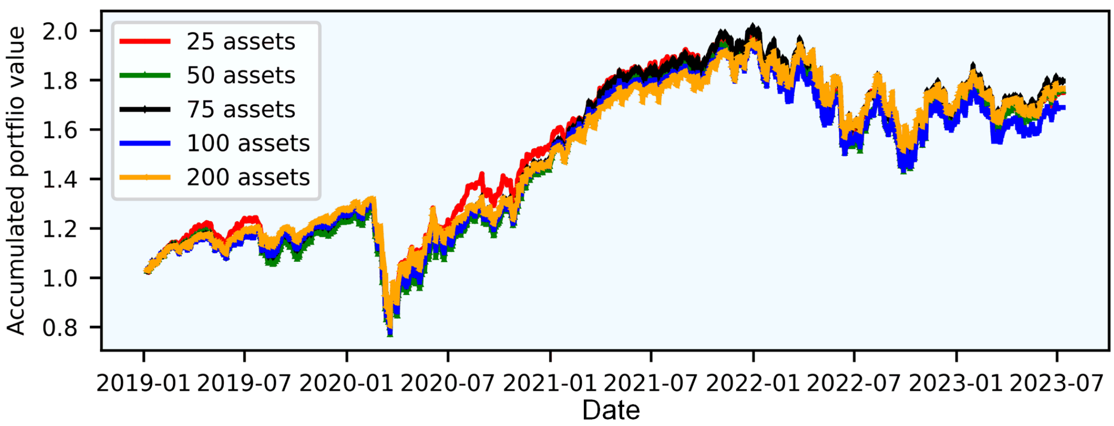
(a) MP-Adv-DRL-Cor



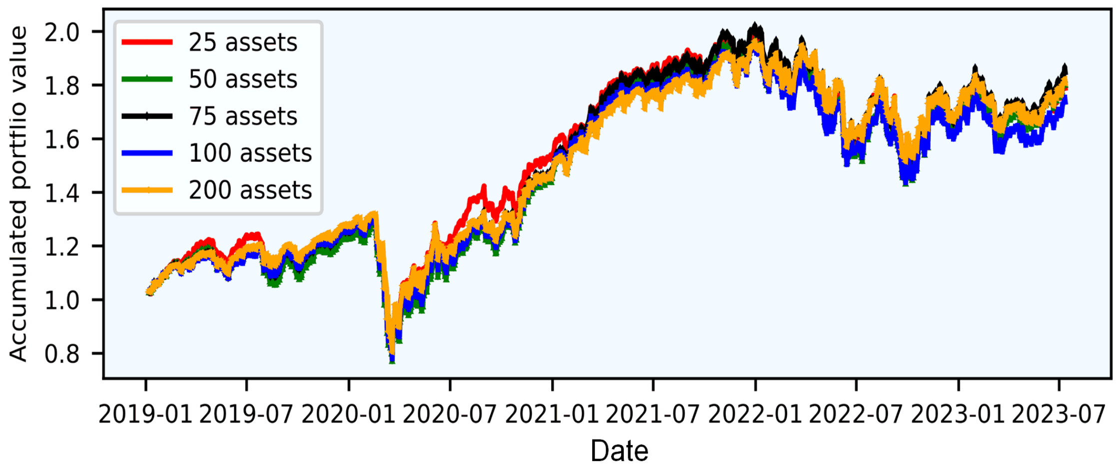
(b) MP-CS-PPN-Cor



(c) MP-DPG



(d) SP-Adv-DRL-Cor



(e) EW

Figure. 4.7. Accumulated portfolio value trajectories of five methods on the S&P500 dataset under different numbers of assets.

In this case, investors are more likely to assign lower weights to each asset and exploit further the benefits of diversification. By contrast, the portfolio performance in terms of annual return and Sharpe ratio of the MP-CS-PPN-Cor method improves when the number of assets is 75 and then declines as the number of assets increases further. This is particularly the case when the number of assets increases to 100 and then to 200.

In general, the MP-DPG and EW methods attain lower performances but are more stable (annual return, annual volatility, and Sharpe ratio) across different numbers of assets than MP-Adv-DRL-Cor, SP-Adv-DRL-Cor, and MP-CS-PPN-Cor portfolio methods. The results also show that the MP-Adv-DRL-Cor strategy is superior to the MP-CS-PPN-Cor approach, and the performance gap widens as the number of assets increases. For example, when the number of assets is 25, the annual return values of MP-Adv-DRL-Cor and MP-CS-PPN-Cor are 21.24% and 19.19%, respectively, having 10.68% enhancement. In contrast, the annual return values of these two methods are 24.38% and 15.40% over 100 assets, respectively, achieving 58.31% improvement. Such superiority mainly depends on the ability of the WaveNet approach to model asset dependence.

Table 4.6. Portfolio performances of five methods under different numbers of stocks.					
Asset numbers	Annual return(%)	Annual volatility(%)	Sharp ratio	Max-drawdown(%)	Turnover
MP-Adv-DRL-Cor					
25	21.24	24.47	0.868	28.30	0.061
50	22.86	23.24	0.984	32.72	0.018
75	24.01	23.04	1.042	29.23	0.045
100	24.38	24.83	0.952	35.54	0.030
200	22.34	23.47	0.952	30.17	0.027
MP-CS-PPN-Cor					
25	19.19	35.23	0.545	54.60	0.348
50	17.00	21.22	0.801	31.14	0.005
75	19.29	31.43	0.614	43.53	0.401
100	15.40	23.78	0.680	37.41	0.009
200	13.48	23.18	0.581	37.41	0.013
MP-DPG					
25	11.62	25.50	0.456	39.45	0.085
50	12.13	25.62	0.473	41.18	0.092
75	13.67	25.63	0.533	41.32	0.033
100	12.28	25.78	0.477	41.40	0.034
200	12.51	27.28	0.458	46.83	0.080
SP-Adv-DRL-Cor					
25	13.21	24.18	0.546	37.64	0.008
50	13.29	25.09	0.530	39.55	0.007
75	13.87	24.85	0.558	39.83	0.007
100	12.32	24.57	0.502	39.82	0.007
200	13.43	23.64	0.568	39.01	0.007
EW					
25	13.77	24.19	0.569	37.64	0.006
50	13.93	25.10	0.555	39.54	0.006
75	14.59	24.86	0.587	39.82	0.006
100	13.02	24.59	0.530	39.80	0.006
200	14.03	23.66	0.593	39.01	0.006

In conclusion, the above results confirm the superiority of the two-stream learning framework and the importance of extracting asset dependencies. In addition, the results also verify the learning ability of our method in addressing high-dimensional portfolio problems.

4.5 Conclusion

This chapter proposes an advanced multi-period portfolio selection that employs DRL for decision-making, convolutional neural networks to extract time-series price dynamic patterns, and WaveNet to identify asset dependence among a set of assets. The proposed advanced approach is capable of solving multi-period investment decisions in high-dimensional settings characterized by an investment pool of many stocks. An extensive empirical application to different datasets and under several investment horizons shows the superiority of the proposed approach in terms of cumulative returns, Sharpe ratio, and representation abilities in addressing constrained multi-period portfolio problems in different real-world settings.

CHAPTER 5

GENERAL CONCLUSIONS AND FURTHER RESEARCH

5.1 Conclusions

This thesis proposes three portfolio selection models by applying robust and DRL methods to perform investment in dynamic and complex financial markets. By implementing these strategies, portfolio performances (e.g., annual return, Sharpe ratio) can be efficiently optimized under different transaction cost rates and risk aversion coefficients. It is proven that the developed DRL-based portfolio method is particularly specialized in handling the complexity of high-dimensional data and improves portfolio performance not only in the short term but also in the long term portfolio conditions.

Chapter 2 proposed a robust portfolio selection model based on G-E-D-M-C-WCVaR to address the worst-case portfolio optimization problem. In detail, the GJR-GARCH model combined with EVT was designed to extract the conditional heteroscedasticity and extreme events of asset returns. The mixed copula model is also applied to capture dependency relationships of asset returns. The proposed robust portfolio method can shield against adverse movements of the model parameters and market observations and minimize the objective risk measure WCVAR. Finally, extensive empirical results were provided to evaluate the portfolio performances in the period during/after the COVID-19 pandemic outbreak. Results verified that the presented robust method outperforms the existing N-D-C-WCVaR and minimal variance portfolios regarding cumulative returns, Sharpe ratio, maximum drawdown, and portfolio risk, especially in the COVID-19 pandemic period.

In Chapter 3, to address the large-scale portfolio optimization problems, an advanced model-free DRL was applied to construct optimal portfolio selections. In particular, this chapter designed a TD3-based portfolio method to combine advanced exploration strategies and dynamic policy updates, where the investor's risk aversion and transaction cost constraints are embedded in an extended

Markowitz's mean-variance reward function. Empirical results demonstrated that the RTC-CNN-TD3 portfolio method has superiority against popular benchmarks and DRL alternatives under different datasets and different transaction cost rates, especially in high-dimensional market scenarios.

In Chapter 4, since investors may tend to seek continuous growth of their portfolios over multiple time periods, this work designed a multi-period portfolio optimization model based on DRL for long-term portfolio allocation under different transaction cost rates and risk aversion levels. This chapter proposed an advanced portfolio policy framework to extract the price dynamic patterns using CNN, capture asset dependence using WaveNet on high-dimensional assets, and perform optimal portfolio allocation using DRL. Then, these ML methods are embedded within a multi-period Bellman equation framework for multi-period portfolio optimization. Finally, numerous empirical findings support the proposed method's (MP-Adv-DRL-Cor) superiority and effectiveness in handling the constrained multi-period portfolio model on real-world datasets. In addition, the proposed method achieves better performance regarding higher portfolio return and Sharpe ratio than traditional benchmark methods and other DRL algorithms.

5.2 Further research

In addition to the current study's results, there are still more works worth exploring in future studies to advance further development in this field.

In Chapter 2, four asset classes (S&P500 index, gold, bitcoin, and U.S. 5-year treasury bonds) are considered. These asset classes represent different types of assets, including equities, commodities, cryptocurrencies, and fixed income. In future research, firstly, it will be interesting to expand more different asset classes for constructing a multi-dimensional asset portfolio. Meanwhile, more market factors, like incorporating macroeconomic factors, environmental, social, and governance (ESG) factors, and political events, may be taken into account in the model for a better understanding of the impact of market volatility on portfolios. Correspondingly, the mix copula model was applied in Chapter 2 to evaluate the tail dependence between four assets. This approach can be extended to value dependencies in high-dimensional assets. However, the computational complexity increases significantly with the number of assets increases. Facing these challenges, it may be feasible to categorize the assets into multiple sub-groups and apply the copula model to compute the correlations within each sub-group and then integrate the results. Moreover, other methods, such as factor modeling and neural networks, can also be explored to deal with correlations in high-dimensional assets. These methods may provide more flexibility and accuracy to help investors better understand and manage the risk of asset portfolios. Finally, the portfolio's performance has been evaluated by considering the effects of tail risk and extreme events in Chapter 2. The results show that the proposed strategy outperforms other widely used benchmark portfolios in terms of cumulative return and Sharpe ratio etc. Future research could explore more complex risk metrics, such as expected loss and tail risk, to improve the overall portfolio risk assessment and help investors better understand and manage their portfolio risk exposure.

In Chapter 3, DRL applications for investment strategies primarily focus on enhancing overall portfolio returns. While the chapter explored managing investment risk by enhancing portfolio diversity and improving the dynamics of asset allocation, it is worth exploring a comprehensive and in-depth analysis of investment risk in future work. This includes formulating more accurate investment risk metrics, as well as examining how to effectively control risk while maximizing returns within a reinforcement learning framework. Moreover, although this chapter provides an in-depth study of dynamic asset allocation, including the extraction of asset features and the improvement of the allocation strategies, it can still be extended by considering the impact of extreme events on asset prices before the feature extraction. The robustness of the investment strategy could be enhanced, and the model's adaptability to different market conditions can be improved by containing extreme events in the model. In addition, this chapter focuses on the DJIA and S&P stock markets, but other types of financial products, such as forwards, futures, and commodity markets, can be included in the subsequent research in order to enrich the results of financial market research. Given that various stock indices capture different aspects of its economy, it is valuable to select multiple market indices to represent specific economic entities accurately.

In Chapter 4, a DRL framework is applied to solve a multi-period and high-dimensional portfolio management problem. Future research could explore, firstly, several aspects of DRL algorithms applied to portfolio management for improving investment performance and strategy reliability. Secondly, it is beneficial to develop novel DRL strategies that integrate traditional portfolio theories and simulate different investors' behaviors to enhance investment strategies. Meanwhile, I could develop algorithmic reward functions for portfolio management and improve the outcomes of DRL algorithms based on large-scale financial data sets. Finally, ensuring compliance with algorithms and models will be an interesting area for future research, given the changing market conditions and financial rules. This involves not only designing models that can efficiently manage risks or maximize returns but also adapting to financial policy and regulatory shifts.

5.3 Conclusiones

Esta tesis propone tres modelos de selección de cartera mediante la aplicación de métodos robustos y DRL para realizar inversiones en mercados financieros dinámicos y complejos. Al implementar estas estrategias, el desempeño de la cartera (por ejemplo, rendimiento anual, índice de Sharpe) se puede optimizar eficientemente bajo diferentes tasas de costos de transacción y coeficientes de aversión al riesgo. Está demostrado que el método de cartera basado en DRL desarrollado está particularmente especializado en el manejo de la complejidad de datos de alta dimensión y mejora el rendimiento de la cartera no sólo en el corto plazo sino también en las condiciones de la cartera a largo plazo.

El Capítulo 2 propuso un modelo robusto de selección de cartera basado en G-E-D-M-C-WCVaR para abordar el problema de optimización de cartera en el peor de los casos. En detalle, el modelo GJR-GARCH combinado con EVT fue diseñado para extraer la heterocedasticidad condicional y los eventos extremos de los rendimientos de los activos. El modelo de cópula mixta también se aplica para capturar relaciones de dependencia de los rendimientos de los activos. El método de cartera robusta propuesto puede proteger contra movimientos adversos de los parámetros del modelo y las observaciones del mercado y minimizar la medida objetiva de riesgo WCVAR. Finalmente, se proporcionaron amplios resultados empíricos para evaluar el desempeño de la cartera en el período durante y después del brote de la pandemia de COVID-19. Los resultados verificaron que el método sólido presentado supera al N-D-C-WCVaR existente y a las carteras de varianza mínima en cuanto a rendimientos acumulativos, índice de Sharpe, reducción máxima y riesgo de cartera, especialmente en el período de la pandemia de COVID-19.

En el Capítulo 3, para abordar los problemas de optimización de carteras a gran escala, se aplicó un DRL avanzado sin modelo para construir selecciones óptimas de carteras. En particular, este capítulo diseñó un método de cartera basado en TD3 para combinar estrategias de exploración avanzadas y actualizaciones dinámicas de políticas, donde la aversión al riesgo del inversionista y las restricciones de costos de transacción están integradas en una función de recompensa de varianza media extendida de Markowitz. Los resultados empíricos demostraron que el método de cartera RTC-CNN-TD3 tiene superioridad frente a puntos de referencia populares y alternativas DRL bajo diferentes conjuntos de datos y diferentes tasas de costos de transacción, especialmente en escenarios de mercado de alta dimensión.

En el Capítulo 4, dado que los inversionistas pueden tender a buscar un crecimiento continuo de sus carteras durante múltiples períodos de tiempo, este trabajo diseñó un modelo de optimización de carteras de múltiples períodos basado en DRL para la asignación de carteras a largo plazo bajo diferentes tasas de costos de transacción y niveles de aversión al riesgo. Este capítulo propuso un marco de política de cartera avanzado para extraer los patrones dinámicos de precios utilizando CNN, capturar la dependencia de activos utilizando WaveNet en activos de alta dimensión y realizar una asignación óptima de cartera utilizando DRL. Luego, estos métodos de ML se integran dentro de un marco de ecuaciones de Bellman de períodos múltiples para la optimización de la cartera de períodos múltiples. Finalmente, numerosos hallazgos empíricos respaldan la superioridad y efectividad del método propuesto (MP-Adv-DRL-Cor) en el manejo del modelo de cartera restringida de períodos múltiples en conjuntos de datos del mundo real. Además, el método propuesto logra un mejor rendimiento con respecto a un mayor rendimiento de la cartera y un índice de Sharpe que los métodos de referencia tradicionales y otros algoritmos DRL.

5.4 Futuras investigaciones

Además de los resultados del estudio actual, todavía hay más trabajos que vale la pena explorar en estudios futuros para avanzar en el desarrollo en este campo.

En el Capítulo 2, se consideran cuatro clases de activos (índice S&P500, oro, bitcoin y bonos del Tesoro de Estados Unidos a 5 años). Estas clases de activos representan diferentes tipos de activos, incluidas acciones, materias primas, criptomonedas y renta fija. En investigaciones futuras, en primer lugar, será interesante ampliar más clases de activos diferentes para construir una cartera de activos multidimensional. Mientras tanto, en el modelo se pueden tener en cuenta más factores de mercado, como la incorporación de factores macroeconómicos, ambientales, sociales y de gobernanza (ESG), y acontecimientos políticos, para comprender mejor el impacto de la volatilidad del mercado en las carteras. En consecuencia, en el Capítulo 2 se aplicó el modelo de cópula mixta para evaluar la dependencia de cola entre cuatro activos. Este enfoque se puede ampliar para valorar las dependencias en activos de alta dimensión. Sin embargo, la complejidad computacional aumenta significativamente a medida que aumenta el número de activos. Ante estos desafíos, puede ser factible categorizar los activos en múltiples subgrupos y aplicar el modelo de cópula para calcular las correlaciones dentro de cada subgrupo y luego integrar los resultados. Además, también se pueden explorar otros métodos, como el modelado de factores y las redes neuronales, para abordar las correlaciones en activos de alta dimensión. Estos métodos pueden proporcionar más flexibilidad y precisión para ayudar a los inversores a comprender y gestionar mejor el riesgo de las carteras de activos. Finalmente, el desempeño de la cartera se evaluó considerando los efectos del riesgo de cola y eventos extremos en el Capítulo 2. Los resultados muestran que la estrategia propuesta supera a otras carteras de referencia ampliamente utilizadas en términos de rendimiento acumulado y índice de Sharpe, etc. La investigación futura podría explorar más métricas de riesgo complejas, como la pérdida esperada y el riesgo de cola, para mejorar la evaluación general del riesgo de la cartera y ayudar a los inversores a comprender y gestionar mejor la exposición al riesgo de su cartera.

En el Capítulo 3, las aplicaciones de DRL para estrategias de inversión se centran principalmente en mejorar los rendimientos generales de la cartera. Si bien el capítulo exploró la gestión del riesgo de inversión mejorando la diversidad de la cartera y la dinámica de asignación de activos, vale la pena explorar un análisis integral y profundo del riesgo de inversión en trabajos futuros. Esto incluye formular métricas de riesgo de inversión más precisas, así como examinar cómo controlar eficazmente el riesgo y al mismo tiempo maximizar los rendimientos dentro de un marco de aprendizaje reforzado. Además, aunque este capítulo proporciona un estudio en profundidad de la asignación dinámica de activos, incluida la extracción de características de los activos y la mejora de las estrategias de asignación, aún se puede ampliar considerando el impacto de eventos extremos en los precios de los activos antes de la extracción de características. Se podría mejorar la solidez de la estrategia de inversión y la adaptabilidad del modelo a diferentes condiciones del mercado al contener eventos extremos en el modelo. Además, este capítulo se centra en los mercados de valores DJIA y S&P, pero en la investigación posterior se pueden incluir otros tipos de productos financieros, como contratos a plazo, futuros y mercados de materias primas, para enriquecer los resultados de la investigación de mercados financieros. Dado que varios índices bursátiles capturan diferentes aspectos de su economía, es valioso seleccionar múltiples índices de mercado para representar entidades económicas específicas con precisión.

En el Capítulo 4, se aplica un marco DRL para resolver un problema de gestión de cartera de múltiples períodos y de alta dimensión. Las investigaciones futuras podrían explorar, en primer lugar, varios aspectos de los algoritmos DRL aplicados a la gestión de carteras para mejorar el rendimiento de la inversión y la confiabilidad de la estrategia. En segundo lugar, es beneficioso desarrollar nuevas estrategias DRL que integren las teorías de cartera tradicionales y simulen los comportamientos de diferentes inversores para mejorar las estrategias de inversión. Mientras tanto, podría desarrollar funciones algorítmicas de recompensa para la gestión de carteras y mejorar los resultados de los algoritmos DRL basados en conjuntos de datos financieros a gran escala. Por último, garantizar el cumplimiento de algoritmos y modelos será un área interesante para futuras investigaciones, dadas las cambiantes condiciones del mercado y las reglas financieras. Esto implica no sólo diseñar modelos que puedan gestionar eficientemente los riesgos o maximizar los retornos, sino también adaptarse a la política financiera y a los cambios regulatorios.

References

Aboussalah, A. M., & Lee, C. G. (2020). Continuous control with stacked deep dynamic recurrent reinforcement learning for portfolio optimization. *Expert Systems with Applications, 140*, 112891.

Aboussalah, A. M., Xu, Z., & Lee, C. G. (2022). What is the value of the cross-sectional approach to deep reinforcement learning? *Quantitative Finance, 22*(6), 1091-1111.

Aielli, G. P., (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business and Economic Statistics, 31*, 282-299.

Ait-Sahalia, Y., & Brandt, M. (2001). Variable selection for portfolio choice. *The Journal of Finance, 56*, 1297-1351.

Alexander, J. M., Rüdiger F. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance, 7* (3), 271-300.

Ali, F., Jiang, Y., & Sensoy, A. (2021). Downside risk in Dow Jones Islamic equity indices: Precious metals and portfolio diversification before and after the COVID-19 bear market. *Research in International Business and Finance, 58*, 101502.

Almahdi, S., & Yang, S. Y. (2017). An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems with Applications, 87*, 267-279.

Almahdi, S., & Yang, S. Y. (2019). A constrained portfolio trading system using particle swarm algorithm and recurrent reinforcement learning. *Expert Systems with Applications, 130*, 145-156.

Almgren, R., & Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk, 3*, 5-40.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath D. (1999). Coherent measures of risk. *Mathematical Finance, 9*(3), 203-228.

Barberis, N. (2000). Investing for the long run when returns are predictable. *The Journal of Finance, 55*, 225–264.

Basak, S., & Shapiro, A. (2001). Value-at-risk based risk management: Optimal policies and asset prices. *Review of Financial Studies, 14*, 371–405.

Bassett, G. W., Koenker, R., & Kordas G. (2004). Pessimistic portfolio allocation and choquet expected utility. *Journal of Financial Economics, 2*, 477–492.

Belhajjam, A., Belbachir, M., & El Ouardirhi, S. (2017). Robust multivairiate extreme value at risk allocation. *Finance Research Letters, 23*, 1-11

Bellman, R. (1957). A Markovian decision process. *Journal of mathematics and mechanics*, 679-684.

Ben Nasr, I., & Chebana, F. (2022). Estimation method for mixture copula models in hydrological context. *Journal of Hydrology, 615*, Part A, 128603.

Bernardi, M., & Catania, L. (2018). Portfolio optimisation under flexible dynamic dependence modelling. *Journal of Empirical Finance, 48*, 1-18.

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations Research, 52*(1), 35-53.

Betancourt, C., & Chen, W. H. (2021). Deep reinforcement learning for portfolio management of markets with a dynamic number of assets. *Expert Systems with Applications, 164*, 114002.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Economics, 31*(3), 307-327.

Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics, 72*, 498-505.

Brandt, M. (1999). Estimating portfolio and consumption choice: A conditional Euer equations approach. *The Journal of Finance, 54*, 1609-1646.

Brandt, M., & Clara, P. S. (2006). Dynamic portfolio selection by augmenting the asset space. *The Journal of Finance, 61*, 2187-2217.

Brennan, M. J., Schwartz, E. S., & Lagnado, R. (1997). Strategic asset allocation. *Journal of Economic dynamics and Control, 21*(8-9), 1377-1403.

Brennan, M. J., Schwartz, E. S., & Lagnado, R. (1999). The use of treasury bill futures in strategic asset allocation programs. In W. T. Ziemba, & J. Mulvey (Eds.), *World wide asset and liability modeling*. Cambridge University Press.

Bühler, H., Gonon, L., Teichmann, J., & Wood, B. (2018). Deep hedging, *Quantitative Finance, 19*(8), 1271-1291.

Campbell, J., Chan, Y., & Viceira, L. (2003). A multivariate model of strategic asset allocation. *Journal of Financial Economics. 67* (1), 41-80.

Campbell, R., Huisman, R., & Koedijk, K. (2001). Optimal portfolio selection in a value-at-risk framework. *Journal of Banking and Finance, 25*, 1789–1804.

Campbell, J., & Viceira, L. (1999). Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics. 114* (2), 433-495.

Campbell, J., & Viceira, L. (2001). Who should buy long-term bonds?. *American Economic Review, 91*, 99-127.

Campbell, J., & Viceira, L. (2002). Strategic asset allocation: Portfolio choice for long-term investors. New York, *NY: Oxford University Press*.

Chambers, R. G., & Quiggin, J. (2007). Dual approaches to the analysis of risk aversion. *Economica, 74*(294), 189-213.

Chaouki, A., Hardiman, S., Schmidt, C., Sérié, E., & Lataillade, J. (2020). Deep deterministic portfolio optimization. *The Journal of Finance and Data Science, 6*, 16-30.

Choi, J. H., Larsen, K., & Seppi, D. J. (2019). Information and trading targets in a dynamic market equilibrium. *Journal of Financial Economics, 132*(3), 22-49.

Constantinides, G. M. (1986). Capital market equilibrium with transaction costs. *Journal of Political Economy, 94*(4), 842-862.

Corsaro, S., De Simone, V., Marino, Z., & Scognamiglio, S. (2022). l1-Regularization in Portfolio Selection with Machine Learning. *Mathematics, 10*(4), 540.

Cui, T. X., Ding, S. S., Jin, H., & Zhang, Y. M. (2023). Portfolio constructions in cryptocurrency market: A CVaR-based deep reinforcement learning approach. *Economic Modelling, 119*, 106078.

Cui, T. X., Du, N. J., Yang, X. Y., & Ding, S. S. (2024). Multi-period portfolio optimization using a deep reinforcement learning hyper-heuristic approach. *Technological Forecasting and Social Change,198*,122944.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus Naive Diversification: How inefficient Is the 1/n Portfolio Strategy? *Review of Financial Studies, 22*, 1915-1953.

Deng, X., & Liang, Y. (2021). Robust portfolio optimization based on semi-parametric ARMA-TGARCH-EVT model with mixed copula using WCVaR. *Computer Economics*, https://doi.org/10.1007/s10614-021-10207-5.

Embrechts P., Lindskog, F., & McNeil, A. (2001). Modeling dependence with Copulas and applications to risk management. *Handbook of Heavy Tailed Distributions in Finance.* P Embrechts.

Engle, R. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics, 20* (3), 339-350.

Engle R.F., & Gonzalez-Rivera G. (2001) Semiparametric ARCH models. *Journal of Business and Economic Statistics, 9*(4), 345-359.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics, 22*, 367–381.

Eom, C., & Park, J. W. (2017). Effects of common factors on stock correlation networks and portfolio diversification. *International Review of Financial Analysis, 49*, 1-11.

Epstein, L., & Zin, S. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica, 57*, 937-969.

Epstein, L., & Zin, S. (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: an empirical investigation. *Journal of Political Economy, 99*, 263-286.

Fereydooni, A., & Mahootchi, M. (2023). An algorithmic trading system based on a stacked generalization model and hidden Markov model in the foreign exchange market. *Global Finance Journal, 56*, 100825.

Fernandez-Arjona, L., & Filipovic, D. (2022). A machine learning approach to portfolio pricing and risk management for high-dimensional problems. *Journal of Mathematical Economics, 32*(4), 982-1019.

Fishburn. P. C., (1977). Mean-risk analysis with risk associated with below-target returns. *American Economic Review, 67*, 116-126.

Gaivoronski, A. A., & Pflug, G. C. (2005). Value-at-risk in portfolio optimization: properties and computational approach. *Journal of Risk, 7*(2), 1-31.

Garcia, R., & Tsafack, G. (2011). Dependence structure and extreme comovements in international equity and bond markets. *Journal of Banking and Finance, 35* (8), 1954–1970.

Garcia-Galicia, M., Carsteanu, A. A., & Clempner, J. B. (2019). Continuous-time reinforcement learning approach for portfolio management with time penalization. *Expert Systems with Applications, 129*, 27-36.

Ghaoui, L.E., Oks, M., & Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operational Research, 51* (4), 543–556.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance, 48*, 1779–1801.

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance, 32*, 100577.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang X., Wang G., Cai J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern recognition, 77*, 354-377.

Guo, X., Chan, R. H., Wong, W.K., & Zhu, L.X. (2019). Mean–variance, mean–VaR, and mean–CVaR models for portfolio selection with background risk. *Risk Management, 21*, 73–98.

Halperin, I. (2019). The QLBS Q-learner goes NuQLear: Fitted Q iteration, inverse RL, and option portfolios. *Quantitative Finance, 19*(9), 1543-1553.

Hambly, B., Xu, R., & Yang, H. (2023). Recent advances in reinforcement learning in finance. *Mathematical Finance, 33*(3), 437-503.

Han, Y., Li, P., Xia, Y., (2017). Dynamic robust portfolio selection with copulas. *Finance Research Letters, 21*, 190–200.

Han, Y.W., Li, P., Li, J., Wu, S.M., (2020). Robust portfolio selection based on copula change analysis. *Emerging Markets Finance and Trade, 56*, 3635–3645.

Hellmich, M., and Kassberger, S., (2011). Efficient and robust portfolio optimization in the multivariate Generalized Hyperbolic framework. *Quantitative Finance, 11*(10), 1503-1516.

Henriques, I., & Sadorsky, P. (2023). Forecasting NFT coin prices using machine learning: Insights into feature significance and portfolio strategies. *Global Finance Journal, 23*, 100904.

Hibiki, N. (2006). Multi-period stochastic optimization models for dynamic asset allocation. *Journal of Banking & Finance, 30*(2), 365-390.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics, 3*, 1163-1174.

Huang, Z., & Tanaka, F. (2022). MSPM: A modularized and scalable multi-agent reinforcement learning-based system for financial portfolio management. *Plos One, 17*(2), e0263689.

Ibragimov, R., & Walden, D. (2007). The limits of diversification when losses may be large. *Journal of Banking and Finance, 31*, 2551–2569.

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance, 58*, 1651–1684.

Jaisson, T. (2022). Deep differentiable reinforcement learning and optimal trading. *Quantitative Finance, 22*(8),1429-1443.

James, W, & Stein, C. (1961). Estimation with quadratic loss. Proceedings 4th Berkeley Symposium Probability and Statistics. *University of California Press, Berkeley, 452*.

Jang J., & Seong, N. Y. (2023). Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory. *Expert Systems with Applications, 218*, 119556.

Jiang, Z. Y., Xu, D. X., & Liang, J. J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *Available at: https://arxiv.org/abs/1706.10059*

Kakouris, I., & Rustem, B. (2014). Robust portfolio optimization with copulas. *European Journal of Operational Research, 235* (1), 28–37.

Kamali, R., Mahmoodi, S., & Jahandideh, M.T. (2019). Optimization of multi-period portfolio model after fitting best distribution. *Finance Research Letters, 30*, 44-50.

Karmakar, M., & Paul, S. (2019). Intraday portfolio risk management using VaR and CVaR: A CGARCH-EVT-Copula approach. *International Journal of Forecasting, 35*, 699–709.

Kim, T. S., & Omberg, E. (1996). Dynamic nonmyopic portfolio behavior. *Review of Financial Studies, 9*, 141-161.

Kirby, C., & Ostdiek, B. (2012). Optimizing the performance of sample mean-variance efficient portfolios. *AFA 2013 San Diego Meetings Paper,* Available at SSRN: https://ssrn.com/abstract=1821284.

Kircher, F., & Rsch, D. (2021). A shrinkage approach for sharpe ratio optimal portfolios with estimation risks. *Journal of Banking and Finance, 133*(7), 106281.

Krokhmal, P., Uryasev, S., & Palmquist, J. (2001). Portfolio optimization with conditional value- at-risk objective and constraints. *Journal of Risk, 4*(2), 43–68.

Laborda, R., & Olmo, J. (2017). Optimal asset allocation for strategic investors. *International Journal of Forecasting, 33(*4), 970-987.

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, 10*, 603-621.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88*, 365-411.

Ledoit, O., & Wolf, M. (2012). Nonlinear Shrinkage Estimation of Large Dimensional Covariance Matrices. *Annals of Statistics, 40*, 1024-1060.

Ledoit, O., & Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies, 30*, 4349-4388.

Ledoit, O., & Wolf, M. (2022). The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. *Journal of Financial Econometrics, 20*, 187-218.

Li, B., Wang, J., Huang, D., & Hoi, S. C. (2018). Transaction cost optimization for online portfolio selection. *Quantitative Finance, 18*(8), 1411-1424.

Li, Y., Jiang, S., Wei, Y., & Wang S. (2021a). Take Bitcoin into your portfolio: A novel ensemble portfolio optimization framework for broad commodity assets. *Financial Innovation, 7*(63), https://doi.org/10.1186/s40854-021-00281-x.

Li, Z., Liu, X. Y., Zheng, J., Wang, Z., Walid, A., & Guo, J. (2021b). FinRL-podracer: high performance and scalable deep reinforcement learning for quantitative finance. *Proceedings of the Second ACM International Conference on AI in Finance.* New York, NY, USA, 48, 1-9.

Lillicrap, T. P., Hunt, J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning, *International Conference on Learning Representations*, San Diego, CA, USA, May 7-9.

Lin, F., Wang, M., Liu, R., & Hong, Q. (2020). A DDPG algorithm for portfolio management, *2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science*, Xuzhou, China, 222-225.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance, 20*(4), 587-615.

Liu, H., & Loewenstein, M. (2002). Optimal portfolio selection with transaction costs and finite horizons. *The Review of Financial Studies, 15*(3), 805-835.

Liu, Y. J., & Zhang, W. G. (2019). Possibilistic Moment Models for Multi-period Portfolio Selection with Fuzzy Returns. *Computational Economics, 53*(4), 1657-1686.

Low, R.K.Y. (2018). Vine copulas: Modelling systemic risk and enhancing higher-moment portfolio optimisation. *Accounting Finance, 58*, 423–463.

Lucey, B. M., & Muckley, C. (2011). Robust global stock market interdependencies. *International Review of Financial Analysis, 20*(4), 215-224.

Ma, G., Siu, C. C., & Zhu, S. P. (2019). Dynamic portfolio choice with return predictability and transaction costs. *European Journal of Operational Research, 278*(3), 976-988.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77-91.

Marzban, S., Delage, E., Li, J. Y. M., Desgagne-Bouchard, J., & Dussault, C. (2023). WaveCorr: Deep reinforcement learning with permutation invariant convolutional policy networks for portfolio management. *Operations Research Letters, 51*(6), 680-686.

Mavruk, T. (2022). Analysis of herding behavior in individual investor portfolios using machine learning algorithms. *Research in International Business and Finance, 62*, 101740. https://doi.org/10.1016/j.ribaf.2022.101740.

McNeil, A. J., & Frey R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of Empirical Finance, 7*(3), 271-300.

Merton, R. (1969). Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics, 51*, 247-257.

Merton, R. (1971). Optimum consumption and portfolio rules in a continuous time model. *Journal of Economic Theory, 3*, 373-413.

Moallemi, C. C., & Saglam, M. (2015). Dynamic portfolio choice with linear rebalancing rules. *Journal of Financial and Quantitative Analysis, 52*(3), 1247-1278.

Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks and Learning Systems, 12*(4), 875-889.

Moody, J., Wu, L., Liao, Y., & Saffell, M. (1998). Performance functions and reinforcement learning for trading systems and portfolios. *Journal of forecasting, 17*(5-6), 441-470.

Ngo, V. M., Nguyen, H. H., & Van Nguyen, P. (2023). Does reinforcement learning outperform deep learning and traditional portfolio optimization models in frontier and developed financial markets? *Research in International Business and Finance, 65*, 101936.

Nikoloulopoulos, A.K., Joe, H., & Li, H. (2012). Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics and Data Analysis, 56* (11), 3659–3673.

Olschewski, S., Diao, L., & Rieskamp, J. (2021). Reinforcement learning about asset variability and correlation in repeated portfolio decisions. *Journal of Behavioral and Experimental Finance, 32*, 100559.

Park, H., Min, K. S., & Dong, G. C. (2020). An intelligent financial portfolio trading strategy using deep q-learning. *Expert Systems with Applications, 158*, 113573.

Park, S., Song, H., & Lee, S. (2019). Linear programing models for portfolio optimization using a benchmark. *The European Journal of Finance, 25*(5), 435-457.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review (Philadelphia), 47* (2), 527–556.

Peck, J., & Yang, H. (2011). Investment cycles, strategic delay, and self-reversing cascades. *International Economic Review, 52*(1), 259.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics, 3*, 119–131.

Pigorsch, U., & Schafer, S. (2022). High-dimensional stock portfolio trading with deep reinforcement learning. *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr), 1-8*. https://arxiv.org/abs/2112.04755.

Qureshi, F., Kutan, A. M., Ismail, I., & Gee, C. S. (2017). Mutual funds and stock market volatility: An empirical analysis of Asian emerging markets. *Emerging Markets Review, 31*, 176-192.

Rachev, S. T., Stoyanov, S., & Fabozzi F. J. (2007). Advanced stochastic models, Risk assessment, and portfolio optimization: The ideal risk, uncertainty, and performance measures, *John Wiley*, Finance.

Raj, A., Mukherjee, A. A., Jabbour, A., & Srivastava, S. K. (2022). Supply chain management during and post-COVID-19 pandemic: Mitigation strategies and practical lessons learned. *Journal of Business Research, 142*, 1125-1139.

Rockafellar, R.T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance, 26* (7), 1443–1471.

Roni, M., & Jean-Luc, V. (1996). Trading volume with private valuation: evidence from the ex-dividend day. *The Review of Financial Studies, 2*, 471-509.

Rubesam, A. (2022). Machine learning portfolios with equal risk contributions: Evidence from the Brazilian market. *Emerging Markets Review, 51*, 100891. https://doi.org/10.1016/j.ememar.2022.100891.

Sahamkhadam, M., Stephan, A., & Östermark, R. (2018). Portfolio optimization based on GARCH-EVT-Copula forecasting models. International *Journal of Forecasting, 34* (3), 497-506.

Samuelson, P. (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics, 51*, 239-246.

Schroder, M., & Skiadas, C. (1999). Optimal consumption and portfolio selection with stochastic differential utility. *Journal of Economic Theory, 21*, 68- 126.

Sebastian, O., Linan, D., & Jörg, R. (2021). Reinforcement learning about asset variability and correlation in repeated portfolio decisions. *Journal of Behavioral and Experimental Finance, 32*, 100559.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance, 19*(3), 425-442.

Shavandia, A., & Khedmati, M. (2022). A multi-agent deep reinforcement learning framework for algorithmic trading in financial markets. *Expert Systems with Applications, 208*, 118124.

Shen W., Wang J., & Ma S. (2014). Doubly regularized portfolio with risk minimization. *Proceedings of the AAAI Conference on Artifificial Intelligence, 28*, 1286–1292.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP* (Vol. 8, No. 3, pp. 229-231).

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 1, pp. 197-206).

Su, X., Bai, M., & Han, Y. (2021). Robust portfolio selection with regime switching and asymmetric dependence. *Economic Modelling, 99*, 105492.

Sun, X.L., Liu, C., Wang, J., & Li, J.P. (2020). Assessing the extreme risk spillovers of international commodities on maritime markets: A GARCH-Copula-CoVaR approach. *International Review of Financial Analysis, 68* (C).

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. *MIT press*. http://www.scholarpedia.org/article/Reinforcement_learning.

Ta, V. D., Liu, C. M., & Tadesse, D. A. (2020). Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. *Applied Sciences, 10*(2), 437.

Topaloglou, N., Vladimirou, H., & Zenios, S. A. (2002). Cvar models with selective hedging for international asset allocation. *Journal of Banking and Finance, 26* (7), 1535-1561.

Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 12.

Vieira, E. B. F., & Filomena, T. P. (2020). Liquidity constraints for portfolio selection based on financial volume. *Computational Economics, 56*(4), 1055-1077.

Wang, H., & Zhou, X. Y. (2020). Continuous Time Mean-Variance Portfolio Selection: A Reinforcement Learning Framework. *Mathematical Finance, 30*(4),1273-1308.

Wang, M., & Ku, H. (2022). Risk-sensitive policies for portfolio management. *Expert Systems with Applications, 198*(15), 11680.

Watcher, J. (2002). Portfolio and consumption decisions under meanreverting returns: an exact solution for complete markets. *Journal of Financial and Quantitative Analysis, 37*, 63-91.

Wei, J., Yang, Y. X., Jiang, M., & Liu, J.G. (2021). Dynamic multi-period sparse portfolio selection model with asymmetric investors' sentiments. *Expert Systems with Applications, 177*,114945.

Weiß, G. N., & Scheffer, M. (2015). Mixture pair-copula-constructions. *Journal of Banking and Finance, 54*, 175–191.

Weiß, G.N. (2013). Copula–GARCH versus dynamic conditional correlation: an empirical study on var and ES forecasting accuracy. *Review of Quantitative Finance and Accounting, 41* (2), 179–202.

Wiering, M., & Otterlo, M. (2012). Reinforcement Learning: State of the Art. Computational Intelligence and Complexity. *Springer.*

Winkelmann, Kurt. D. (2004). Improving portfolio efficiency - Risk budgeting, implied confidence levels, and changing allocations. *Journal of Portfolio Management, 30*(2), 23-38.

Wu, G., & Xiao, Z. (2002). An analysis of risk measures. *Journal of Risk, 4*(4), 53–75.

Xu, K., Zhang, Y. F., Ye, D. H., Zhao, P. L., & Tan, M. K. (2020). Relation-aware transformer for portfolio policy learning. *International Joint Conference on Artificial Intelligence,* https://doi.org/10.24963/ijcai.2020/641

Xu, W., & Dai, B. (2022). Delta-gamma-like hedging with transaction cost under reinforcement learning technique. *Journal of Derivatives, 29*(5), 60-82.

Yang, H., Liu, X. Y., Zhong, S., & Walid, A. (2020). Deep reinforcement learning for automated stock trading: An ensemble strategy. *Proceedings of the First ACM International Conference on AI in Finance*, 1-8.

Zakoian, J. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control, 18*(5), 931-955.

Zeng, Y., & Klabjan, D. (2018). Portfolio optimization for American options. *Journal of Computational Finance, 22*(3), 37-64.

Zhang, J., & Maringer, D. (2016). Using a genetic algorithm to improve recurrent reinforcement learning for equity trading. *Computational Economics, 47*, 551-567.

Zhang, W. G., Zhang, X., & Chen, Y. (2011). Portfolio adjusting optimization with added assets and transaction costs based on credibility measures. *Insurance Mathematics & Economics, 49*(3), 353-360.

Zhang, X. L., Zhang, W. G., & Xiao, W. L. (2013). Multi-period portfolio optimization under possibility measures. *Economic Modelling, 35*(5), 401-408.

Zhang, Y., Zhao, P., Wu, Q., Li, B., Huang, J., & Tan, M. (2022). Cost-sensitive portfolio selection via deep reinforcement learning. *IEEE Transactions on Knowledge and Data Engineering, 34*(1), 236-248.

Zhao, T. L., Ma, X., Li, X. M., & Zhang, C. M. (2023). Asset correlation based deep reinforcement learning for the portfolio selection. *Expert Systems with Applications, 221*, 119707.

Zhu, S., & Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operational Research, 57* (5), 1155–1168.