Department of Informatics
King's College London
United Kingdom

7CCSMPRJ - Individual Project

Explainable Deep Learning for Portfolio Optimisation

Name: **Ingrid Pérez Aguilera**

Student Number: K24087939

Course: Computational Finance M.Sc.

Supervisor: **Riaz Ahmad**

Abstract

It is a précis of the report (normally in one page), which should include:

- A brief introduction to the project objectives
- A brief description of the main work of the project
- A brief description of the contributions, major findings, results achieved and principal conclusion of the project

Acknowledgements

It is a short paragraph to thank those who have contributed to the project work.

Table of Contents

1 Introduction **1**

1.1 Objectives . . . 3

1.2 Report Structure . . . 4

2 Background **6**

2.1 Portfolio Optimisation . . . 6

2.1.1 Modern Portfolio Theory . . . 7

3 Methodology **10**

4 Results **11**

5 Legal, Social, Ethical and Professional Issues **12**

6 Conclusion **13**

References **i**

A Appendix **iv**

List of Figures

2.1 Efficient Frontier in Risk-Return Space. [1] 9

List of Tables

Nomenclature

c Speed of light in a vacuum

h Planck constant

Glossary

Advantage Actor-Critic Algorithm that uses both an actor (policy) and a critic (value function) to learn optimal policies by estimating the advantage of actions taken. viii, 3

Algorithmic Trading Use of computer algorithms to automate trading decisions and execute trades in financial markets. 1

Artificial Intelligence Simulation of human intelligence processes by machines, especially computer systems, enabling them to perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation. viii, 2

Deep Deterministic Policy Gradient Algorithm that uses deep neural networks to learn policies for continuous action spaces, combining the benefits of deep learning and policy gradient methods. vii, viii, 3

Deep Learning Subset of machine learning that uses neural networks with many layers to learn from large amounts of data, enabling the model to automatically learn complex patterns and representations. viii, 2, 4, 6

Deep Reinforcement Learning Combination of deep learning and reinforcement learning, where deep neural networks are used to approximate the value function or policy in reinforcement learning tasks. viii, 3, 6

Efficient Frontier Set of optimal portfolios that offer the highest expected return for a given level of risk, or the lowest risk for a given level of expected return, in the context of portfolio optimisation. 8

Explainable Artificial Intelligence Methods and techniques in the application of artificial intelligence that make the results of the models understandable by humans, providing insights into how decisions are made. viii, 2

Feature Importance Technique for determining the contribution of each feature in a machine learning model to its predictions, helping to identify which features are most influential. 4, 6

Financial Markets Marketplaces where people trade financial securities, commodities, and other fungible items of value at low transaction costs and at prices that reflect supply and demand. 1

Hyper-parameter Parameter that is set before the learning process begins and control the learning process of a machine learning model, such as learning rate, batch size, and number of layers in a neural network. 2

Local Interpretable Model-agnostic Explanations Technique for explaining the predictions of any machine learning model by approximating it with a locally interpretable model, allowing users to understand the model's behaviour in a specific instance. viii, 4–6

Machine Learning Subset of artificial intelligence that enables systems to learn from data and improve their performance over time without being explicitly programmed. viii, 1, 6

Modern Portfolio Theory Investment theory that aims to construct a portfolio of assets that maximises expected return for a given level of risk, or minimises risk for a given level of expected return, through diversification. viii, 7

Portfolio Optimisation Process of selecting the best distribution of assets in a portfolio to achieve specific investment goals, such as maximising returns or minimising risk, while considering constraints and preferences. 1, 6

Proximal Policy Optimization Algorithm that optimises policies by ensuring that updates to the policy are not too large, maintaining a balance between exploration and exploitation. viii, 3

Reinforcement Learning Subset of machine learning where an agent learns to make decisions by taking actions in an environment to maximise cumulative reward. viii, 2, 4, 6

Reward Function Function that defines the feedback signal received by an agent in reinforcement learning, guiding the agent's learning process by providing rewards or penalties based on its actions. 2

SHapley Additive exPlanations Method for interpreting machine learning models by assigning each feature an importance value for a particular prediction, based on cooperative game theory. viii, 4, 6

Soft Actor-Critic Algorithm that combines the benefits of off-policy learning and entropy regularisation, allowing for more exploration and better stability in learning policies for continuous action spaces. viii, 3

Twin Delayed Deep Deterministic Policy Gradient Extension of Deep Deterministic Policy Gradient that addresses the overestimation bias in value function estimation by using two critic networks and delaying policy updates. viii, 3

Acronyms

A2C Advantage Actor-Critic. 3, 4

AI Artificial Intelligence. 2

DDPG Deep Deterministic Policy Gradient. 3, 4

DL Deep Learning. 2, 4, 6

DRL Deep Reinforcement Learning. 3–6

LIME Local Interpretable Model-agnostic Explanations. 4–6

ML Machine Learning. 1–6

MPT Modern Portfolio Theory. 7, 9

PPO Proximal Policy Optimization. 3, 4

RL Reinforcement Learning. 2, 4, 6

SAC Soft Actor-Critic. 3, 4

SHAP SHapley Additive exPlanations. 4–6

TD3 Twin Delayed Deep Deterministic Policy Gradient. 3, 4

XAI Explainable Artificial Intelligence. 2

Chapter 1

Introduction

Financial markets are highly complex systems influenced by numerous factors, including financial and political events, social trends and technological advancements. Moreover, their evolving and stochastic nature requires using the most advance computational developments to model the financial environment. The tasks of financial time series prediction and Portfolio optimisation are considerably intricate, due to the semi-strong form of market efficiency and the high level of noise. [2]

Algorithmic trading focuses on the application of analytical methods to automatically execute trading actions based on an algorithm without human intervention. In its early days, the field mainly studied the usage of a computer program to follow a predefined strategy [3]. Nonetheless, in recent years, algorithmic trading has evolved to a problem in which environment perception entails learning feature representation from highly non-stationary and noisy financial time series data and decision-making requires the algorithm to explore the environment and simultaneously make correct decisions in an online manner without supervision [4].

Machine Learning (ML) is at an advantage for the task given its capability to learn from historical data and make predictions about the future state of an environment. In the past years, research has explored the application of Deep Learning (DL) in future price prediction of financial assets [2, 5, 6, 7]. However, its main disadvantage is the inability to directly deal with trading, requiring an additional step to convert the predictions into actionable strategies. In contrast, Reinforcement Learning (RL) would allow the algorithm to learn a trading strategy directly from the environment, without the need for a separate step [8, 9]. In this case, there are two main approaches: first, the algorithm can learn the amount of assets to buy, sell or hold at each time step [10], or second, the algorithm can learn the optimal portfolio allocation and automatically rebalance the portfolio weights at each time step [11].

Despite the potential of RL in portfolio optimisation, its widespread adoption in the financial industry remains limited. This is primarily due to following challenges [12]:

1. difficulty in finding the appropriate algorithm with a suitable reward function and hyper-parameters to ensure efficiency and performance,

2. challenge of testing the algorithm in a real-world environment, and

3. lack of transparency of ML models, often referred to as black boxes, making it increasingly complex to interpret the algorithm's decisions.

In recent years, the rise in popularity of Artificial Intelligence (AI) and its widespread use have led to concerns regarding its decisions due to its black-box nature. The concept of explainability in AI, known as Explainable Artificial Intelligence (XAI), refers to a model's ability to provide details and reasons to make itself understandable [13]. The term was first coined in 2016 to describe the need for users to effectively understand, trust and manage artificial intelligence applications [14]. The need for explainability becomes particularly relevant in the context of financial usage, where the regulatory framework requires transparency and accountability in automated decision-making. Various relevant applications, including volatility models [15], credit risk assessment [16] and portfolio construction [12] have explored the concept of explainability in financial applications. This highlights its importance and the need to explore its advantages in more complex models, without inadvertently increasing the complexity of the overall methodology.

Consequently, this thesis will focus on addressing the aforementioned challenges by exploring the application of Deep Reinforcement Learning (DRL) to portfolio optimisation and implementing post-hoc explainability techniques.

1.1 Objectives

The objective of this thesis is to develop an explainable Deep Reinforcement Learning model for portfolio optimisation. A DRL model has the ability to leverage historical financial data to learn an investment strategy that efficiently allocates financial assets while maximising expected returns and minimising risk. Moreover, the incorporation of advanced explainability techniques enhances the interpretability and transparency of the model's decision-making. This project aims to bridge the gap between cutting-edge machine learning techniques and their practical application in finance by addressing the challenges of algorithm selection, simulation of real-world scenarios, and black box nature of ML models.

First, DRL models, such as Advantage Actor-Critic (A2C), Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradient (DDPG), Twin Delayed Deep Deterministic Policy Gradient (TD3) and Soft Actor-Critic (SAC), will be implemented to learn the optimal portfolio allocation from high-dimensional environment representations. The algorithms will be trained on historical financial data, including technical and macroeconomic indicators, with the goal of capturing the complex market dynamics. Each of the algorithms is better suited to a particular scenario, for instance, DDPG encourages maximum returns, while A2C reduces the variance.

Second, post-hoc explainability techniques: Feature Importance, SHapley Additive exPlanations (SHAP) values and Local Interpretable Model-agnostic Explanations (LIME) analysis, will be applied to interpret the model's decisions. The goal of these is to understand which market conditions, represented by financial data, lead to the actions/decisions, encoded as portfolio weights.

Finally, the performance of the DRL models will be analysed in different scenarios, including the impact of a larger financial environment representation, portfolio size and asset composition. The results will be compared with traditional portfolio optimisation methods to evaluate the effectiveness of the proposed approach.

1.2 Report Structure

This report is organised into six chapters, each of which focuses on a concrete area related to the problem. Additional material, together with source code, is included in the appendices.

The current chapter, 1, presents the motivation and the objectives of this thesis. It gives an overview of the potential of DRL in portfolio optimisation and its main challenges, particularly the lack of transparency of ML models.

Chapter 2 provides an overview of the theoretical background of the project, including financial markets and machine learning concepts. The problem of portfolio optimisation in the financial domain is outlined and the potential of DRL in this context is discussed. The chapter provides a comprehensive background explanation of the fundamentals of Deep Learning and Reinforcement Learning, including the main algorithms (A2C, PPO, DDPG, TD3, SAC) and techniques in their intersection, DRL. In addition, it gives an overview of the post-hoc explainability techniques: feature importance, SHapley Additive exPlanations (SHAP) values and Local Interpretable Model-agnostic Explanations (LIME) analysis, which will be used to interpret the model's decisions. Finally, it includes a comprehensive in-depth literature review on the topics of ML applied to portfolio optimisation and relevant applications of explainability techniques.

The methodology chapter 3 describes the techniques and methods used to solve the problem and outlines the implementation of the proposed solution. The chapter provides a detailed explanation of the architecture and components of the proposed DRL model, including the state representation, reward function, and training process.

The results of the experiments are presented in chapter 4, which analyses and evaluates the results obtained from the proposed implementation, while critically discussing the findings. It provides a detailed comparison of the proposed DRL strategies with traditional portfolio optimisation methods. Furthermore, it consists of an in-depth analysis of the model's decisions using post-hoc explainability techniques, in particular, SHAP values, feature importance and LIME analysis.

Chapter 5 discusses the legal, social, ethical and professional implications within the context of the project. By addressing these issues, the project aims to ensure that the proposed solution adheres to industry standards, while considering the implications of the technology.

Finally, the report concludes with a summary of the main points of the work, the contributions made, the results achieved as well as potential applications and future work in chapter 6.

Chapter 2

Background

This chapter provides an overview of the problem of portfolio optimisation in the financial domain, followed by a comprehensive explanation of the fundamentals of Deep Learning (DL) and Reinforcement Learning (RL), including the relevant algorithms in the field of Deep Reinforcement Learning (DRL). In addition, it discusses the need for explainability in Machine Learning (ML) and the main techniques used to achieve it: SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME) and Feature Importance. Finally, it presents the state of the art in portfolio optimisation using DRL and the recent advancements in explainability techniques in the field.

2.1 Portfolio Optimisation

Portfolio optimisation is the process of selecting optimal weights for a portfolio of assets in order to maximise expected returns for a given level of risk, or conversely, to minimise risk for a given level of expected returns [17]. In mathematical terms, the problem requires finding a solution to the specified objective function, which is typically a function of the expected returns and the risk associated with the portfolio [18]. The task becomes further complicated if a time dimension is introduced, as the portfolio weights need to be adjusted over time to capture the changes in market conditions and asset prices [19].

2.1.1 Modern Portfolio Theory

There exist several traditional frameworks that formalise the problem of portfolio allocation. Markowitz's Modern Portfolio Theory (MPT) was proposed in 1952 [20] and it provides a mathematical framework where investors choose optimal portfolios based on risk and return, by either minimising the risk given a specified return or, maximising the return given a specified risk [21]. The theory extends the concept of diversification by suggesting that owning financial assets of different kinds is less risky than owning assets of the same kind, due to the correlations between assets.

The main assumptions in MPT are:

- investors are risk-averse, rational, and seek to maximise return for a given risk;
- returns are normally distributed;
- markets are frictionless, meaning there are no transaction costs; and
- assets are infinitely divisible.

Under these assumptions, portfolio risk and return can be modelled as an optimisation problem. Let $\mathbf{w} = (w_1, w_2, \ldots, w_N)^T$ denote the portfolio weight vector, where each w_i indicates the proportion of capital allocated to asset i, subject to the budget constraint:

$$\sum_{i=1}^{N} w_i = 1 \quad \Leftrightarrow \quad \mathbf{w}^T\mathbf{1} = 1 \tag{2.1}$$

with $\mathbf{1} \in \mathbb{R}^N$ being a vector of ones, and subject to the non-negativity constraint, meaning that short-selling is not allowed:

$$w_i \geq 0 \quad \forall i = 1, 2, \ldots, N. \tag{2.2}$$

Let $\boldsymbol{\mu} = (R_1, R_2, \ldots, R_N)^T$ represent the vector of expected returns, and $\Sigma \in \mathbb{R}^{N\times N}$ the covariance matrix of asset returns. The expected return of the portfolio is then given by:

$$R_p = \mathbf{w}^T \boldsymbol{\mu}, \tag{2.3}$$

and the portfolio risk is quantified by the variance of returns:

$$\sigma_p^2 = \mathbf{w}^T \Sigma \mathbf{w}. \tag{2.4}$$

This formulation provides the foundation for solving the mean-variance optimisation problem, by either:

- minimising portfolio variance σ_p^2 subject to a target expected return R_p, or
- maximising expected return R_p subject to a risk constraint σ_p.

The Markowitz mean-variance optimisation problem can be expressed as:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^T \Sigma \mathbf{w} \\ \text{subject to} \quad & \begin{cases} \mathbf{w}^T \boldsymbol{\mu} = R_p \\ \mathbf{w}^T \mathbf{1} = 1 \\ \mathbf{w} \geq 0 \end{cases} \end{aligned} \tag{2.5}$$

Solving the mean-variance optimisation problem for varying levels of target return leads to a set of optimal portfolios that form the efficient frontier. It is typically visualised in a risk-return space, where the x-axis represents the risk (standard deviation) and the y-axis represents the expected return, as shown in Figure 2.1. Portfolios below the curve are suboptimal, while those on the frontier represent the best achievable combinations of risk and return.

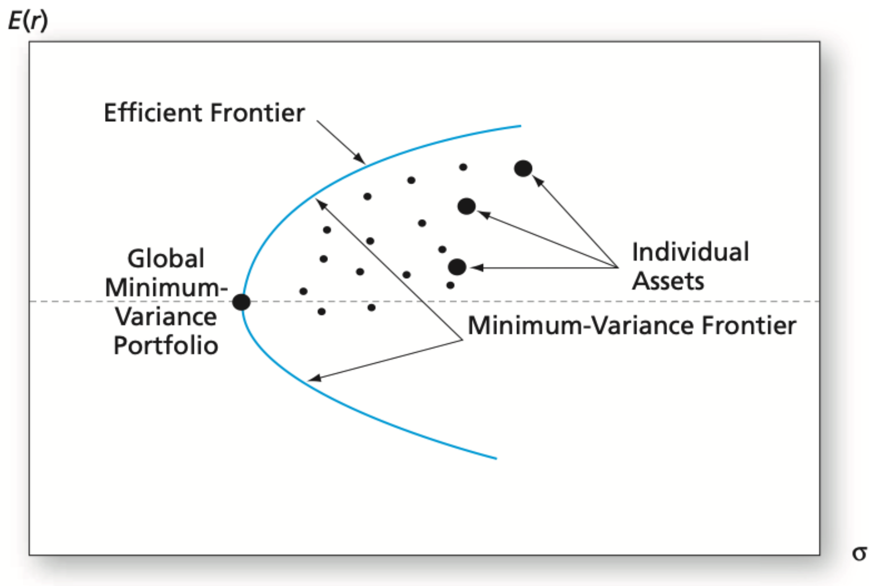

Figure 2.1: Efficient Frontier in Risk-Return Space. [1]

Despite the simplicity in the formulation of MPT, its assumptions do not reflect the behaviour of real markets. Moreover, modern markets are dynamic, non-stationary, and feature non-linear relationships, which have driven research into other approaches better suited to capture the complexities of modern financial markets.

Chapter 3

Methodology

Chapter 4

Results

Chapter 5

Legal, Social, Ethical and Professional Issues

Chapter 6

Conclusion

References

[1] Z. Bodie, A. Kane, and A. J. Marcus, *Investments*. McGraw-Hill Education, 10 ed., 2014.

[2] J. Shen and M. O. Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system," *Journal of Big Data*, vol. 7, pp. 1–33, 12 2020.

[3] K. Lei, B. Zhang, Y. Li, M. Yang, and Y. Shen, "Time-driven feature-aware jointly deep reinforcement learning for financial signal representation and algorithmic trading," *Expert Systems with Applications*, vol. 140, p. 112872, 2 2020.

[4] C. Ma, J. Zhang, J. Liu, L. Ji, and F. Gao, "A parallel multi-module deep reinforcement learning algorithm for stock trading," *Neurocomputing*, vol. 449, pp. 290–302, 8 2021.

[5] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *Journal of Big Data*, vol. 7, pp. 1–40, 12 2020.

[6] J. M. T. Wu, Z. Li, N. Herencsar, B. Vo, and J. C. W. Lin, "A graph-based cnn-lstm stock price prediction algorithm with leading indicators," *Multimedia Systems*, vol. 29, pp. 1751–1770, 6 2023.

[7] M. Hasan, M. Z. Abedin, P. Hajek, K. Coussement, M. N. Sultan, and B. Lucey, "A blending ensemble learning model for crude oil price forecasting," *Annals of Operations Research*, pp. 1–31, 1 2024.

[8] J. Moody and M. Saffell, "Learning to trade via direct reinforcement," *IEEE Transactions on Neural Networks*, vol. 12, pp. 875–889, 7 2001.

[9] H. Yang, X. Y. Liu, S. Zhong, and A. Walid, "Deep reinforcement learning for automated stock trading: An ensemble strategy," *ICAIF 2020 - 1st ACM International Conference on AI in Finance*, 10 2020.

[10] X.-Y. Liu, Z. Xiong, S. Zhong, H. Yang, and A. Walid, "Practical deep reinforcement learning approach for stock trading," *NeurIPS 2018 AI in Finance Workshop.*, 11 2018.

[11] M. Guan and X. Y. Liu, "Explainable deep reinforcement learning for portfolio management: An empirical approach," *ICAIF 2021 - 2nd ACM International Conference on AI in Finance*, 11 2021.

[12] D. G. Cortés, E. Onieva, I. Pastor, L. Trinchera, and J. Wu, "Portfolio construction using explainable reinforcement learning," *Expert Systems*, vol. 41, p. e13667, 11 2024.

[13] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 10 2019.

[14] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G. Z. Yang, "Xai—explainable artificial intelligence," *Science Robotics*, vol. 4, 12 2019.

[15] D. Brigo, X. Huang, A. Pallavicini, and H. S. de Ocariz Borde, "Interpretability in deep learning for finance: a case study for the heston model," *SSRN Electronic Journal*, 4 2021.

[16] R. García-Céspedes, F. J. Alias-Carrascosa, and M. Moreno, "On machine learning models explainability in the banking sector: the case of shap," *Journal of the Operational Research Society*, 2025.

[17] Y. Sato, "Model-free reinforcement learning for financial portfolios: A brief survey," *arXiv preprint arXiv:1904.04973*, 4 2019.

[18] B. Bruce and J. Greene, *Chapter 4 - Portfolio Construction*, pp. 133–178. Academic Press, 2014.

[19] X. Li, Y. Li, Y. Zhan, and X.-Y. Liu, "Optimistic bull or pessimistic bear: Adaptive deep reinforcement learning for stock portfolio allocation," *arXiv preprint arXiv:1907.01503*, 6 2019.

[20] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, pp. 77–91, 3 1952.

[21] F. M. Clinic, "Mean variance portfolio theory."

Appendix A

Appendix