Scalable, Robust and Unbiased Reinforcement Learning-based System Designs for Financial Portfolio Management

March 2023

Zhenhan Huang

Scalable, Robust and Unbiased Reinforcement Learning-based System Designs for Financial Portfolio Management

Graduate School of Science and Technology

Degree Programs in Systems and Information Engineering

University of Tsukuba

March 2023

Zhenhan Huang

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

ACKNOWLEDGMENTS

I am indebted to my supervisor, Dr. Fumihide Tanaka, for his steadfast support and valuable guidance throughout my PhD project. His expertise and advice have been invaluable in this journey of mine. I would also like to express my sincere gratitude to the esteemed members of the thesis defense committee for providing me with the opportunity to present my work and for their helpful and constructive feedback. Their insights and suggestions have allowed me to further refine and improve my thesis. I am extremely grateful for the opportunity to benefit from their knowledge.

I would also like to extend my appreciation to the current and former members of the Fumihide Tanaka Lab for their valuable help. In particular, I would like to thank Denis Pena, Yijie Guo, and Yohei Noguchi for their support and for the enlightening discussions of ours.

Lastly, I am eternally thankful to my family, especially my parents, for their unwavering support and encouragement throughout my PhD journey. Their love and understanding played a fundamental role in the successful completion of this work. I would also like to thank my friends from Washington University in St. Louis and Rensselaer Polytechnic Institute for their emotional support, which has been indispensable to me.

This work was supported by following funding source:

- JST SPRING, Grant Number JPMJSP2124

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

RL reinforcement learning

PM financial portfolio management

MDP Markov Decision Process

crypto cryptocurrency

MSPM a modularized and scalable multi-agent RL-based system for PM

CryptoRLPM a novel end-to-end scalable RL-based system incorporating on-chain data for crypto PM

CEX centralized exchange

DEX decentralized exchange

QT quantitative trading

DQN Deep Q-Network

PPO Proximal Policy Optimization

SARL Reinforcement-Learning Based Portfolio Management with Augmented Asset Movement Prediction States [121]

DE disposition effect

NF narrow framing

ARL Adversarial Deep Reinforcement Learning in Portfolio Management [65]

CRP (Uniform) Constant Rebalanced Portfolio

FAIB an end-to-end framework for appraising investment biases in PM systems

CM Crypto Module of CryptoRLPM

1 INTRODUCTION

1.1 RESEARCH BACKGROUND

Applications of artificial intelligence (AI) in quantitative trading (QT) has always been a fascinating topic in academia and industry. Among the approaches, reinforcement learning (RL) is particularly wished to be implemented by researchers due to the resemblant nature of the problem settings.

In the early days, many pioneers in RL-based QT focused on single-asset trading problems due to the constrained computing power. These pioneers have proposed various novel-at-the-time reward designs in RL for financial portfolio management (PM), such as [31, 78, 79]. With the increase in computing power and the development of CUDA [53], artificial neural networks (ANN), i.e., deep learning (DL), have become dominant in the AI fields during the past decade. By incorporating DL into RL, multi-asset management (namely financial portfolio management), becomes more feasible and relevant than ever. Financial portfolio management (PM), regarding the allocation of capital into multiple assets, aims to maximize accumulated profits with an option to minimize the overall risks of the portfolio.

However, while inspiring, the existing approaches have certain limitations. When designing and developing RL-based systems for PM, researchers seldom focus on the scalability and reusability of the systems to accommodate the ever-changing markets. RL agents in the existing RL-based systems are ad-hoc trained and rarely reusable for different portfolios. Also, the existing systems are barely scalable to answer the increasing need for varying numbers of assets in portfolios and heterogeneous data input. Furthermore, these systems lack a modular design to be compatible with different RL agents for different assets.

Moreover, another critical reason why nowadays, autonomous algorithms are largely implemented in financial trading is that it is expected there should be no issue of investment biases in the case of QT. Certain cognitive biases which human investors often have may lead to irrational decision-making in investment, like over-extrapolation or frequent trading, and eventually prevent investors from profit-making. Algorithms are believed to overcome the weaknesses in human cognition. However, the researchers who have proposed successful QT (in our case, RL-based) systems that achieve superior capital return performance neglected the fact that a high-performance system does not sufficiently indicate that it can overcome investment biases. The presence of these biases in the existing RL-based systems for PM is ignored and rarely

appraised. It is not guaranteed that those high-performance systems also outperform human investors over certain proxies of investment bias.

Meanwhile, with the emergence of new technologies in recent years, such as blockchain, comes new categories of financial vehicles, e.g., cryptocurrency (crypto), and researchers are enthusiastic about the designs and implementations of RL onto these new assets. As regarded as highly volatile and majorly driven by the sentiments flowing across social media platforms, more attention should be paid to the potential biases in the RL-based system targeting the trading of these emerging assets. Yet, it is rarely the case, not even mentioning that a number of existing research has relied on subjective feedback to the questionnaires or surveys distributed for the investigation of behavioral biases of crypto investors, despite the fact that the blockchains provide comprehensive and easily accessible data of investors.

1.2 MOTIVATIONS

In light of the aforementioned issues and insufficiencies in related research, this thesis endeavors to provide valid and robust answers to the following questions through **a series of empirical studies**:

- Q1. Can we build a modularized, scalable, and robust RL-based system for PM which outperforms the existing baselines?
- Q2. Can such a system solve the long-standing issues of lack of reusability and inefficient learning in RL-base PM?
- Q3. Is this system unbiased compared to human investors? Moreover, can we build a general framework for the quantification and appraisal of biases in any given RL-based PM system?
- Q4. Can crypto investors' degrees of investment biases be appraised by directly using on-chain (on blockchain) wallet information without subjective feedback or inaccessible information?
- Q5. Can we integrate on-chain information into a novel RL-based crypto PM system for outperformance? Will this system also stand unbiased?

By answering the above-mentioned research questions and achieving the research goals in each study, this thesis establishes structured connections among reinforcement learning, financial portfolio management, behavioral finance, and blockchain technologies.

1.3 CONTRIBUTIONS

The core topics of this thesis are reinforcement learning, financial portfolio management, behavioral finance, and blockchain technologies. This

thesis makes the **contributions** the under the following **three themes** to address the issues and insufficiencies of the exiting research and to answer the research questions raised in Section 1.2.

1. RL-BASED PM We are the first to bring the concepts of scalability and reusability into the system designs of multi-agent reinforcement learning-based portfolio management.

- In Chapter 6, we propose the first multi-agent RL-based system for PM (MSPM) with a scalable and modularized design to address the issues of ad-hoc, fixed, and inefficient model training in the RL-based methods in the existing research.
- This study, together with a novel end-to-end scalable RL-based system incorporating on-chain data for crypto PM (CryptoRLPM) proposed in Chapter 10, provides a robust answer to the **Q1** and **Q2** in Section 1.2. In fact, MSPM becomes a stepping stone to inspire more creative system designs in multi-agent reinforcement learning-based financial portfolio management.
- Additionally, the scalability of two RL-based PM methods proposed, MSPM and CryptoRLPM, are discussed in Section 6.5.2.5 and Section 10.2.2.2, respectively.

2. INVESTMENT BIASES The evaluation of investment biases in RL-based PM systems is ignored and never appraised in the existing research until ours.

- In Chapter 7, as the first of this kind, we investigate the existence and degrees of investment biases in MSPM, an RL-based PM system.
- Not only that, in Chapter 9 we design and develop FAIB, the first appraisal framework for evaluating investment biases in any given RL-based portfolio management systems of heterogeneous types of financial assets.
- In Chapter 10, CryptoRLPM, is also proved to overcome and outperform human investors in terms of the two biases when investing cryptocurrencies and stand robust and unbiased by the appraisal using FAIB.
- Thanks to these studies, investment bias now becomes a new category of metrics measuring the performance of RL-based PM systems. Additionally, for the first time, a general framework is provided to researchers to appraise the bias proxies in PM systems directly. Hence, the **Q3** in Section 1.2 is answered.

3. CRYPTOCURRENCIES AND ON-CHAIN METRICS The third contribution is that this thesis is the first to introduce on-chain data into the

investigation of investors' behavioral biases and incorporate on-chain data into the RL-based PM system.

- On-chain metrics are like vitals to human, or fundamentals of a company. On-chain data reflect the state, running details, and measurements of a blockchain and its crypto.
- In Chapter 8, it is the first time the behavioral biases of crypto investors are evaluated directly through the utilization of on-chain records. Before that, researchers who evaluated crypto investors' behaviors rely on the subjective, uncertain, and dubious feedback to the surveys distributed or opaque and inaccessible data from centralized exchanges.
- Further, in Chapter 10, we build CryptoRLPM, the first RL-based system incorporating on-chain data with a scalable and modularized design for crypto PM.
- More importantly, by using FAIB proposed in Chapter 9, we inspect and validate that CryptoRLPM stands unbiased after the incorporation of on-chain metrics and sentiments, which is aligned with the goal of the thesis to build a scalable, robust and unbiased RL-based PM system.
- By achieving the research goals in studies covered in the three chapters mentioned above (Chapter 8-Chapter 10), the **Q4** and **Q5** in Section 1.2 are practically answered.

By contributing to the three themes, and providing well-grounded answers to the research questions, this thesis aims to provide an avant-garde and unprecedented roadmap towards more modularized, scalable, unbiased, and robust system designs for RL-based crypto PM.

1.4 THESIS STRUCTURE

The thesis chapters, excluding this chapter, are organized as follows. From Chapter 2 to Chapter 5, we introduce and discuss the essential background knowledge and concepts of the topics covered in the thesis. In particular, Chapter 2 introduces the fundamentals of reinforcement learning (RL) related to the studies of the thesis. Chapter 3 focuses on the essential concepts of financial portfolio management (PM) related to this thesis. This includes the introduction to i). two financial markets, U.S. stock markets and crypto markets, ii). the latest trends in RL-based methods for PM and their limitations, and iii). performance measurement and metrics, e.g., daily rate of return (DRR). Chapter 4 details the background of blockchain and crypto related to the studies in this thesis, along with an introduction to on-chain metrics. In Chapter 5, we focus on the introduction and discussion of the topics of behavioral finance

related to this thesis. The topics include the heretics and investment bias and their definitions, a potential explanation of the origin of the biases, and related studies in the thesis.

In Chapter 6, we propose a multi-agent RL-based system for PM (MSPM). By bringing scalability and reusability, MSPM aims to address the issue of ad-hoc, fixed, and inefficient model training in the existing RL-based methods.

In Chapter 7, we investigate the proxies of two well-known biases in financial investment, disposition effect (DE) and narrow framing (NF), in a cutting-edge RL-based system for PM (MSPM). We aim to examine the RL-based system's capacity to overcome the biases which human investors often have in financial investment.

Furthermore, in Chapter 8, by utilizing on-chain data, we study investors' biases when trading cryptocurrencies. Cryptocurrency, as an emerging financial asset, is alluring investors globally. The on-chain metrics of a crypto/blockchain are like a company's fundamentals. This study marks the first attempt to fully reveal the behavioral biases of crypto traders directly through on-chain records, without using any inaccessible nor indirect data sources like centralized exchange databases or questionnaires/surveys.

Covered in Chapter 9, we formalize an end-to-end framework for appraising investment biases in PM systems (FAIB). FAIB shall be employed to answer if certain investment biases, e.g., disposition effect, exist in the decision-making of any given heterogeneous-asset RL-based PM system.

Combining and emanating from our findings from Chapter 7 to Chapter 9, we propose a novel RL-based system utilizing on-chain data with a scalable design for financial PM in Chapter 10: CryptoRLPM. As far as we know, this is the first time the utilization of on-chain data and the design of the scalability are incorporated, with investment biases appraised, in an RL-based system for crypto PM.

Figure 1.1 shows the diagram of the thesis's structure, and the chapters are highlighted by the attributed contributions. With the organic connections among the three major themes of contributions, this diagram can be deemed a roadmap towards more scalable, robust and unbiased system designs for RL-based portfolio management.

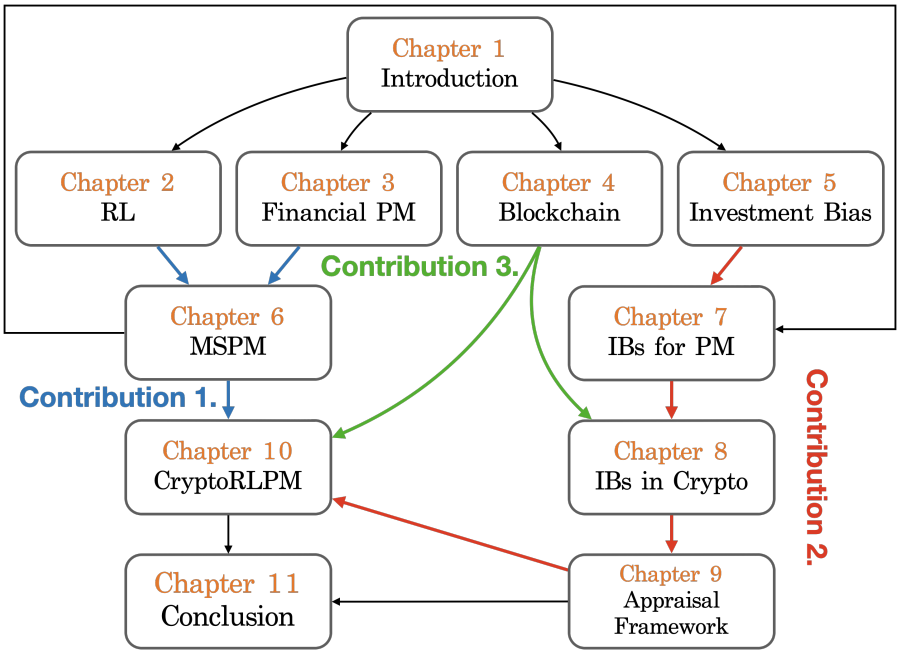


Figure 1.1: Dependency diagram of the thesis's structure with the contribution of each chapter highlighted. This diagram not only indicates the routes to read the thesis but also provides a roadmap towards more scalable, robust, and unbiased system designs for RL-based PM. This diagram was inspired by [35, 104].

2

REINFORCEMENT LEARNING (RL)

This chapter introduces the fundamentals of reinforcement learning (RL) related to the studies covered in the thesis. These fundamentals include the concepts of and mathematics behind Markov decision process, policies, value functions, temporal-difference methods, Q-learning, and a policy gradient method. To a certain extent, the contents of this chapter follow [108].

2.1 MARKOV DECISION PROCESS (MDP)

Markov Decision Process (MDP) relates to a mathematical framework for the decision-making procedure in which at each time step t an agent observes a state s from an environment it interacts with. Then, the agent takes an action a accordingly to maximize a reward r and receives a new state s_{t+1} at the next time step $t+1$.

A Markov decision process is defined by a tuple $(\mathcal{S}, \mathcal{A}, P, r, \gamma)$, where

- $\mathcal{S}$ denotes set of all states (state space).
- $\mathcal{A}$ denotes set of all actions (action space), and $\mathcal{A}(s) \subseteq \mathcal{A}$ denotes state-dependent action space.
- $p : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$ denotes state-transition probability.

  Specifically,

  $$p(s' \mid s, a) \doteq \Pr\{S_t = s' \mid S_{t-1} = s, A_{t-1} = a\} \tag{2.1}$$

  for all $s', s \in \mathcal{S}$, and $a \in \mathcal{A}$, is the probability of transitioning from state s_t to state s_{t+1} in the region of $S_{t+1} \subseteq S$ by taking an action a_t [108], and with Markov property,

  $$p(s' \mid s, a) \doteq \Pr\{S_t = s' \mid S_{t-1}, A_{t-1}, S_{t-2}, A_{t-2}, ..., S_0, A_0\},$$

  denoting that the probability of transitioning into a new state s' is independent of the previous states given present state s.
- $r : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ defines the reward function, which is the expected reward subject to certain state s, action a, and subsequent state s'.
- γ denotes the discount rate, where $0 \leq \gamma \leq 1$.

2.1.1 *Policy and Value Functions*

POLICY: A deterministic policy is a mapping $\pi : \mathcal{S} \to \mathcal{A}$, in which an action $a \in \mathcal{A}$ will be certainly chosen by an agent given a state $s \in \mathcal{S}$.

On the other hand, a stochastic policy is a mapping $\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$, which yields a set of conditional probabilities over an action space $\mathcal{A}$ given a state $s \in \mathcal{S}$. Specifically,

$$\pi(a|s) = Pr\{S_{t-1} = s \mid A_{t-1} = a\}$$

denoting the conditional probability an agent chooses action $a \in \mathcal{A}$ given state $s \in \mathcal{S}$.

STATE-VALUE FUNCTION: The state-value function $v_\pi : \mathcal{S} \to \mathbb{R}$ is the expected total discounted reward of state s under policy π:

$$v_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s], \tag{2.2}$$

where G_t is the total discounted return starting from time step t, and

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

where the discount rate $\gamma \in [0, 1]$ determines the relative importance of future rewards compared to immediate rewards [108].

ACTION-VALUE FUNCTION: The action-value function q_π is the expected total discounted reward of taking action a in state s under policy π:

$$q_\pi(s, a) = \mathbb{E}_\pi[G_t|S_t = s, A_t = a] \tag{2.3}$$

BELLMAN EXPECTATION EQUATIONS: The state-value function $v_\pi(s)$ can be decomposed into immediate reward plus discounted value of successor state, and satisfies the Bellman expectation equation [7]:

$$\begin{aligned} v_\pi(s) &= \mathbb{E}_\pi\left[R_{t+1} + \gamma v_\pi\left(S_{t+1}\right) \mid S_t = s\right] \\ &= \sum_{a \in \mathcal{A}(s)} \pi(a \mid s) \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p\left(s', r \mid s, a\right)\left(r + \gamma v_\pi\left(s'\right)\right), \end{aligned}$$

where $p\left(s', r \mid s, a\right)$ is the *dynamics* of MDP, and $p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$.

Similarly, the action-value function $q_\pi(s, a)$ can be decomposed, and satisfies the Bellman expectation equation:

$$\begin{aligned} q_\pi(s, a) &= \mathbb{E}_\pi\left[R_{t+1} + \gamma q_\pi\left(S_{t+1}, A_{t+1}\right) \mid S_t = s, A_t = a\right] \\ &= \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p\left(s', r \mid s, a\right)\left(r + \gamma \sum_{a' \in \mathcal{A}(s')} \pi\left(a' \mid s'\right) q_\pi\left(s', a'\right)\right) \end{aligned}$$

2.1.2 *Optimality*

OPTIMAL POLICY: We define a partial ordering over policies as $\pi \geq \pi'$ if $v_\pi(s) \geq v_{\pi'}(s)$, for all $s \in S$

indicating that there exists an optimal policy π_* that satisfies $\pi_* \geq \pi, \forall \pi$.

Therefore, all optimal policies achieve the optimal state- and action-value functions:

$$v_*(s) = v_{\pi_*}(s) = \max_\pi v_\pi(s) \tag{2.4}$$

and

$$q_*(s,a) = q_{\pi_*}(s,a) = \max_\pi q_\pi(s,a), \tag{2.5}$$

where optimal state-value function $v_*(s)$, and optimal action-value function $q_*(s,a)$, are the maximum state- and action-value functions, respectively, over all the policies. Once the optimal action-value function $q_*(s,a)$ is determined, the deterministic optimal policy π_* can be obtained by $\pi_*(s) = \arg\max_{a \in \mathcal{A}(s)} q_*(s,a)$. An optimal policy $\pi_*(s)$ is guaranteed but may not be unique.

BELLMAN OPTIMALITY EQUATIONS: The optimal value functions $v_*(s)$ and $q_*(s,a)$ are recursively related by the Bellman optimality equations:

$$v_*(s) = \max_{a \in \mathcal{A}(s)} \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p\left(s', r \mid s, a\right)\left(r + \gamma v_*\left(s'\right)\right)$$

and

$$q_*(s,a) = \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p\left(s', r \mid s, a\right)\left(r + \gamma \max_{a' \in \mathcal{A}(s')} q_*\left(s', a'\right)\right).$$

2.2 RL METHODS IMPLEMENTED

Two RL methods are implemented to learn the optimal policies for PM in the studies covered in the thesis:

1. Deep Q-learning [76]
2. Proximal Policy Optimization (PPO) [97]

Deep Q-learning falls under the category of value-based methods, referring to the class of algorithms which only use value function approximation to obtain an estimate of the optimal action-value function to learn the optimal policy for a given task. Value-based methods, with ϵ-greedy used for action space exploration, are more likely to learn

deterministic policies $\pi : \mathcal{S} \to \mathcal{A}$, and are primarily implemented for tasks with *discrete* action space.

Two representative groups of value-based methods are Monte Carlo (MC) and Temporal-difference (TD). In RL, *episodic* tasks denote tasks that have a start and end (*terminal state*). In an episodic task, a complete sequence of the agent's interaction with the environment is an *episode*. For MC methods, the estimate to state-value function (Equation 2.2) or action-value function (Equation 2.3) is not updated until the end of an episode. Whereas TD methods only need to wait until the next time step to update the value function [108].

Q-learning is an off-policy TD method and an early breakthrough in RL [108], which is outlined in Algorithm 1:

Algorithm 1 Q-learning: $Q(s,a) \approx q_*(s,a)$ [108]

Parameters: step size $\alpha \in (0,1]$, small $\epsilon > 0$
Initialize $Q(s,a)$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$, arbitrarily except that $Q(terminal,\cdot) = 0$
for each episode **do**
  Initialize $S_{t=0} \in \mathcal{S}$
  while S_t is not terminal **do**
    $A \leftarrow \pi(S_t)$ (e.g., ϵ-greedy), given $\pi(s) \leftarrow \text{argmax}_a Q(s,a)$
    $S_{t+1}, R_{t+1} \leftarrow p(\cdot,\cdot \mid S, A)$
    $Q(S_t,A_t) \leftarrow Q(S_t,A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1},a) - Q(S_t,A_t)]$
    $S_t \leftarrow S_{t+1}$
  end while
end for
return $Q(s,a)$

In Algorithm 1,

$$Q(S_t,A_t) \leftarrow Q(S_t,A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1},a) - Q(S_t,A_t)]$$

denotes how Q-learning updates its estimate $Q(s,a)$ to the optimal value function at time step t, where α is the step size.

In this thesis, regarding the length of the datasets, the RL tasks are defined episodic, of which majority are with discrete action spaces for trading-signal generation, e.g., an action space of {*Buying, Selling, and Skipping*}. Moreover, these RL tasks are with high-dimensional state spaces. Targeting the tasks, we focus on the usage of Deep Q-learning with Experience Reply [76, 77], in which the $Q(s,a)$ from Q-learning can be represented by a neural network, namely Deep Q-Network (DQN). At each iteration i, DQN is updated with the following loss function:

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s')\sim U(D)}\left[\left(\underbrace{r + \gamma \max_{a'} Q\left(s', a'; \theta_i^-\right)}_{\text{target}} - \underbrace{Q\left(s,a;\theta_i\right)}_{\text{prediction}}\right)^2\right]$$

(2.6)

where θ^- are the network parameters used to compute the target, and θ are the parameters of the DQN.

Pseudocode of Deep Q-learning with Experience Reply is outlined in Algorithm 2 [77]:

Algorithm 2 Deep Q-learning with Experience Reply [77]

Initialize replay memory D
Initialize action-value function Q with random weight θ
Initialize target action-value function $\hat{Q}$ with weights $\theta^- = \theta$
for episode=1 to M **do**
  Initialize sequence s_1 $\{x_1\}$ and
  preprocessed sequence $\phi = \phi(s_1)$
  for $t = 1$ to T **do**
    With probability ε select a random action a_t
    otherwise select $a_t = \text{argmax}_a\, Q\left(\phi\left(s_t\right), a; \theta\right)$
    Execute action a_t in emulator and observe reward r_t
and image x_{t+1}
    Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi\left(s_{t+1}\right)$
    Store transition $\left(\phi_t, a_t, r_t, \phi_{t+1}\right)$ in D
    Sample random minibatch of transitions $\left(\phi_j, a_j, r_j, \phi_{j+1}\right)$
from D
    Set $y_j = \begin{cases} r_j & \text{if episode terminates at step j + 1} \\ r_j + \gamma \max_{a'} \hat{Q}\left(\phi_{j+1}, a'; \theta^-\right) & \text{otherwise} \end{cases}$.
    Perform a gradient descent step on $\left(y_j - Q\left(\phi_j, a_j; \theta\right)\right)^2$ with
respect to the network parameters θ
    Every C steps reset $\hat{Q} = Q$, i.e., $\theta^- = \theta$
  end for
end for

In Algorithm 2, at each time step t, $D_t = e_1, ..., e_t$ denotes dataset where the experiences $e_t = (s_t, a_t, r_t, s_{t+1})$ of the agents are stored, which are pooled over episodes into a *reply memory*. This technique is called *experience replay* [67]. The experience $e \sim U(D)$ are sampled in the inner loop of Algorithm 2.

Different from Q-learning or other value-based methods, PPO is able to handle RL tasks with *continuous* action spaces, and to learn stochastic or deterministic policies. PPO is implemented for solving the RL task of sampling from the probability distribution for generating reallocation weights in Chapter 6.

PPO [97] is an actor-critic style policy gradient method that has been widely used on continuous action space problems, due to its desirable performance and ease of implementation. A policy π_θ is a parametrized mapping: $\mathcal{S} \times \mathcal{A} \to [0, 1]$ from state space to action space. Among the

different objective functions of PPO, the clipped surrogate objective [97] is implemented:

$$L(\theta) = \hat{\mathbb{E}}\pi_{\theta'}[min(r_t(\theta)A_t^{\theta'}, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t^{\theta'})] \tag{2.7}$$

where

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta'}(a_t|s_t)}$$

and $A_t^{\theta'}$, the advantage function, is expressed as:

$$A_t^{\theta'} = Q^{\theta'}(s_t, a_t) - V^{\theta'}(s_t)$$

in which, the state-action value function $Q^{\theta'}(s_t, a_t)$ is:

$$Q^{\theta'}(s_t, a_t) = \mathbb{E}_{\pi_{\theta'}}[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a]$$

and the value function $V^{\theta'}(s_t)$ is:

$$V^{\theta'}(s_t) = \mathbb{E}_{\pi_{\theta'}}[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s]$$

Pseudocode of PPO is outlined in Algorithm 3 [97, 118]:

Algorithm 3 PPO-Clip [97, 118]

Initialize policy parameters θ_0
Initialize value function parameter ϕ_0
for k=0,1,2,... **do**
Collect set of trajectories $\mathcal{D}_k = \{\tau_i\}$ by running policy $\pi_k = \pi(\theta_k)$ in the environment.
Compute rewards-to-go $\hat{R}_t$
Compute advantage estimates, $\hat{A}_t$ (using any method of advantage estimation) based on the current value function V_{ϕ_k}.
Update the policy by maximizing the PPO-Clip objective: $\theta_{k+1} = \arg\max_\theta \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \min\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} A^{\pi_{\theta_k}}(s_t, a_t), \quad g\left(\epsilon, A^{\pi_{\theta_k}}(s_t, a_t)\right)\right)$ typically via stochastic gradient ascent with Adam [52].
Fit value function by regression on mean-squared error:

$$\phi_{k+1} = \arg\min_\phi \frac{1}{|\mathcal{D}_k| T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \left(V_\phi(s_t) - \hat{R}_t\right)^2$$

typically via some gradient descent algorithm.
end for

In Algorithm 3, $g(\epsilon, A) = \begin{cases} (1+\epsilon)A & A \geq 0 \\ (1-\epsilon)A & A < 0 \end{cases}$ which brings simplified version of Equation 2.7.

3

FINANCIAL PORTFOLIO MANAGEMENT (PM)

3.1 STOCK AND CRYPTOCURRENCY MARKETS

Common stock (stock) consists of shares (fractional ownership) of a listed (publicly-traded) company, and stock exchanges are where the transactions of the shares are taken place. As of December 2022, the U.S. tops all other countries, with the market capitalization of its listed domestic companies exceeding $40 trillion [4], and has two largest stock exchanges in the world: New York Stock Exchange (NYSE) and Nasdaq, by market capitalization [106]. The stock-related studies covered in the thesis focus on the stocks offered and traded in U.S. stock markets.

Cryptocurrency (crypto), on the other hand, as an emerging financial vehicle, is alluring investors globally. Cryptocurrency is a digital currency that uses blockchain as a decentralized ledger for secure and transparent transactions without intermediaries [11]. According to the latest survey by Pew Research Center [25], by July 2022, 16% of U.S. adults have invested in, traded, or used a cryptocurrency. According to The Block Research and CoinGecko.com [16, 90], the annual trading volumes of the entire crypto market in 2021, summing centralized and decentralized exchanges, are about 20 trillion USD. Akin to stocks, cryptocurrencies are also traded at exchanges. There are two types of crypto exchanges: i). centralized exchange (CEX), e.g., Binance, and ii). decentralized exchange (DEX), e.g., Uniswap. The three main differences between CEXs and DEXs are:

1. CEXs are usually controlled by an entity and have higher liquidity, whereas DEXs are deployed on blockchain networks facilitating peer-to-peer (P2P) transactions and have lower liquidity,
2. CEXs take custody of users' assets, whereas DEXs do not, and
3. By regulations, CEXs require KYC (Know-Your-Customer) process to verify their users' identity, whereas DEXs are not subject to any regulatory requirements and do not require KYC.

In contrast to stocks whose prices, to a great extent, reflect the financial performance of the companies in a long-term horizon, cryptocurrencies tend to be more volatile, whose prices are deemed to be highly influenced and driven by investors' sentiments flowing across social media platforms.

3.2 FINANCIAL PORTFOLIO MANAGEMENT (PM)

Quantitative trading (QT) refers to trading financial assets utilizing mathematical models, statistical analysis, and computational algorithms. There are four major practices of QT: 1). Portfolio management, 2). Single-asset trading (algorithmic trading), 3). Order execution, and 4). Market making. We focus on the first practice of the QT.

Portfolio management (PM) is a continuous process of reallocating capital into multiple assets [73], and it aims to maximize accumulated profits with an option to minimize the overall risks of the portfolio. To perform such a practice, portfolio managers who focus on stock markets conventionally read financial statements and balance sheets, follow the news from media and announcements from financial institutions and analyze stock price trends.

As a classic task in QT, PM is often modeled as a problem of sequential weight reallocation of multiple assets on multiple time series, for which we have the followiing settings:

3.2.1 *Weights of Portfolio*

At the beginning of each time step t, we can reallocate a given portfolio with the a vector of weights:

$$a_t = (a_{1,t}, a_{2,t}, ..., a_{m^*,t})^T \tag{3.1}$$

where $m^* = m + 1$ is the number of assets m in the portfolio plus one risk-free asset, namely cash, and $\sum_{i=1}^{m^*} a_{i,t} = 1$.

Then, at the end of time step t, the weights of the portfolio become

$$w_t = \frac{y_t \odot a_t}{y_t \cdot a_t} \tag{3.2}$$

due to the fluctuation of assets' prices during the time step period of t, where

$$y_t = \frac{v_t^{+(close)}}{v_{t-1}^{+(close)}} = (1, \frac{v_{2,t}^{+(close)}}{v_{2,t-1}^{+(close)}}, ..., \frac{v_{m^*,t}^{+(close)}}{v_{m^*,t-1}^{+(close)}})^T$$

is the vector of assets' relative prices referring to the changes of asset prices during the period of t. $v_{i,t}^{+(close)}$ denotes the closing price of the i-th asset at the end of time t, where $i = \{2, ..., m^*\}$, excluding the risk-free asset whose closing price should always be 1.

Figure 3.1 shows the details of price fluctuations.

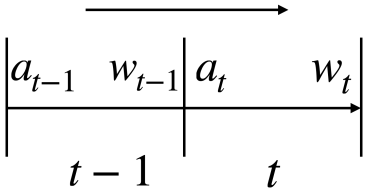


Figure 3.1: Transformation of portfolio's weights due to the fluctuation in assets' prices.

3.2.2 *Periodical Return*

At the end of each time step t, the portfolio's rate of return is

$$r_t = a_t \cdot y_t - \beta \sum_{i=1}^{m^*} |a_{i,t} - w_{i,t}| \tag{3.3}$$

where

m^*	the number of assets
w_t	the allocation weights of the assets
$\beta \sum_{i=0}^{n} \lvert a_{i,t} - w_{i,t} \rvert$	the transaction cost for a n-asset portfolio
β	the commission rate

3.3 WHY RL FOR PM?

Compared to other domains of machine learning, e.g., supervised learning, RL presents multiple advantages for QT, especially for PM:

- RL provides an end-to-end solution without hassle.
  - As an RL agent directly and sequentially learns from market states to make decisions and maximize long-term returns, PM is naturally suited to RL approaches.
  - Whereas supervised learning requires correctly-labeled datasets, which may not be available in the case of PM.
- The objectives of RL can be set flexibly and directly.
  - Constraints, like transaction cost and slippage, can be incorporated into the reward function designing of RL for maximizing portfolio's capital return.
  - Risks can also be incorporated into the objective of RL. For example, in [79], differentiable Sharpe ratio is set as the reward the agent receives. Moreover, differentiable downside deviation risk is set as the reward in [78].
  - Thus, RL methods have more adaptability to generalize to various conditions of assets and markets.

  - However, it is not easy to incorporate the above settings in supervised learning. Supervised learning is often rigid and usually targets on trading-signal generation, which emphasizes on short-term profits, without consideration of transaction costs and risks.

- Importantly, for PM, RL agent can directly produce and reallocate weights of portfolios, but supervise learning cannot and majorly relies on the man-made rules and labels.

3.4 EXISTING RL METHODS FOR PM

By the resemblant nature of the problem, researchers expectedly wish to incorporate RL methods in PM. In early days, due to the limitation of the computational resources, the research of QT majorly targets single-asset trading [31, 78, 79]. Recently, with the increase in computing power and available data, researchers have started to combine deep learning (DL) and RL, as Deep RL (DRL), for multiple-asset trading, namely PM. The authors of [47] propose a PM framework for cryptocurrencies using Deep Deterministic Policy Gradient (DDPG) [66, 101]. [65] proposes a method called Adversarial Training for portfolio optimization with the implementation of three different RL methods: DDPG, Proximal Policy Optimization (PPO)[97] and Policy Gradient (PG). Akin to receiving information from various sources as portfolio managers generally do, existing approaches incorporate heterogeneous data [121]. Recently, multi-agent reinforcement learning (MARL) approaches are also proposed by researchers[63, 69, 109]. In[63], the authors propose MAPS, a system involving a group of Deep Q-network [76] (DQN)-based agents corresponding to individual investors, to make investment decisions and create a diversified portfolio. MAPS can be recognized as a reinforcement-learning implementation of ensemble learning [91] by its very nature. In addition,[68] proposes iRDPG to generate adaptive quantitative trading strategies by using DRL and imitation learning. However, while inspiring, the existing approaches seldom focus on scalability and reusability to accommodate the ever-changing markets. RL agents in the existing multi-agent-based systems are ad-hoc trained and rarely reusable for different portfolios. Also, the existing systems are barely scalable to answer the need for scaled number of assets in portfolios and increasing heterogeneous data input. For example, in SARL [121], the encoder's intake is either financial news data for embedding or stock prices for trading signals generation, but can not be both of them, and this issue prevents the encoder from efficiently producing holistic information and eventually limits the RL-based agents' learning. Furthermore, the existing systems lack a modular design to be compatible with different RL agents for different assets. A literature review of RL methods for PM and other QT tasks can be found in [107].

3.5 PERFORMANCE METRICS

We use the following performance metrics to measure and benchmark the performances of the baselines and RL methods proposed in the thesis. Among the following metrics, for DRR, ARR, and SR, we want them to be as high as possible, whereas we want MD to be as low as possible.

- Daily rate of return (DRR)

$$DRR_T = \frac{1}{T}\sum_{t=1}^{T} \exp(R_t), \tag{3.4}$$

where T is the terminal time step, and

$$R_t = \ln\left(a_t \cdot y_t - \beta \sum_{i=1}^{m^*} |a_{i,t} - w_{i,t}|\right) \tag{3.5}$$

is the risk-unadjusted periodic (daily) rate of return obtained at every time step, where $\beta \sum_{i=1}^{m^*} |a_{i,t} - w_{i,t}|$ is the transaction cost and $\beta = 0.0025$ is the commission rate.

- Accumulated rate of return (ARR) [86]

$$ARR_T = \frac{p_T}{p_0}, \tag{3.6}$$

where T is the terminal time step, p_0 is the portfolio value at the initial time step, and

$$p_T = p_0 \exp\left(\sum_{t=1}^{T} R_t\right) \tag{3.7}$$

which stands for the portfolio value at the terminal time step.

- Sortino ratio (SR) [103] is often referred to as a risk-adjusted return, which measures the portfolio performance compared to a risk-free return, adjusted by the portfolio's downside risk. In our case, Sortino ratio is calculated as

$$SR = \frac{\frac{1}{T}\sum_{t=1}^{T} \exp(R_t) - R_f}{\sigma^{downside}} \tag{3.8}$$

where R_t is the risk-unadjusted periodic (daily) rate of return. Portfolio's downside risk $\sigma^{downside}$ is calculated as

$$\sigma^{downside} = \sqrt{\mathrm{Var}(R_l - R_f)}, \tag{3.9}$$

where R_f is the risk-free return and conventionally equals zero, R_l are the less-than-zero returns in R_t for all t, and $t = T$ is the terminal time step.

- Max drawdown (MD) is the biggest drop (in %) between the highest (peak) and lowest (valley) of the accumulated rate of return of a certain period of time.

4

BLOCKCHAIN AND CRYPTOCURRENCY

Blockchain refers to a distributed ledger stored on computers around the world. The advantages of using blockchain technology for digital transactions and data storage are four-fold: i). decentralized, ii). distributed, and iii). tamper-proof, and iv). trustless. The bitcoin blockchain is the first and most famous blockchain [82]. Bitcoin blockchain uses a proof-of-work (PoW) consensus mechanism to verify new transactions to the ledger. Using a decentralized ledger and PoW, Bitcoin became the first to solve the double-spending and elusive Byzantine General problems.

Bitcoin is the cryptocurrency of the Bitcoin blockchain. Cryptocurrency (crypto) is a digital currency that uses blockchain as a ledger for storing records of cryptocurrency transactions so that anyone can send and receive without intermediaries. Since the introduction and success of Bitcoin, different blockchain networks, with their native cryptocurrencies, have been developed, e.g., the Ethereum network. Ethereum is the first blockchain introducing the concept of smart contracts [10], which allows developers to build and deploy decentralized applications (DApps). Among the existing blockchains, Ethereum is the second largest blockchain after Bitcoin. According to [14], by December 2022, the market capitalization of Bitcoin is more than $324 billion, and that of Ethereum is more than $148 billion, which can be calculated by [15]:

$$\text{Market Capitalization} = \text{Circulating Supply} \times \text{Market Price}$$

Cryptocurrencies other than Bitcoin are called altcoins or alternative coins. As of December 2022, there are more than 22,000 altcoins listed on CoinMarketCap.com, and new ones are kept being created continuously [112]. However, many of these altcoins turned out to be scams eventually [20, 114].

We can easily compare blockchains with public companies: as public companies issue their stocks, blockchains offer cryptocurrencies through initial coin offerings (ICOs). To make an analogy, blockchain can be referred to as a public company, and cryptocurrency as its publicly-traded shares.

Table 4.1 shows a brief comparison between public companies and blockchains.

4.1 CRYPTOCURRENCY: AN EMERGING FINANCIAL VEHICLE

Cryptocurrency, or crypto, as an emerging financial vehicle, is alluring investors globally. According to a survey by Pew Research Center [25],

CHARACTERISTICS	PUBLIC COMPANY	BLOCKCHAIN
1. Organization	Centralized	Decentralized
2. Offering	Stock	Cryptocurrency
3. Ownership	Shareholders	N/A
4. Key Information	Fundamentals	On-chain metrics
5. Regulation	By government entities	Not regulated

Table 4.1: A brief comparison between the characteristics of public company and blockchain.

by July 2022, 16% of U.S. adults have invested in, traded, or used a cryptocurrency. According to The Block Research and CoinGecko.com [16, 90], the annual trading volumes of the entire crypto market in 2021, summing centralized and decentralized exchanges, are about 20 trillion USD.

One of the most well-known characteristics of crypto is its high volatility which is likely driven by the sentiments flowing across social media platforms [55], e.g., Twitter.

There could be many factors that influence cryptocurrencies' prices, to name a few, i). Technological innovations [119], ii). Social sentiments, iii). Regulations, and iv). Celebrity effect. The factors like iii) and iv) could further influence ii). Figure 4.1 [122] visualizes Bitcoin's price trends and linked major events from 2008 to March 2020.

Different factors can affect the choice of trading, using, or mining a cryptocurrency. In [100], the results indicate that more than half of the participants believe that the name and logo affect their choice of cryptocurrency. In fact, several studies [30, 74, 75] discuss the similarities between crypto trading and online gambling, and [21] suggest that the former could be a form of the latter, and potentially leads to excessive behavior and harm in some individuals.

Consequently, researchers are attracted to investigate investors' behavioral traits in investing in cryptocurrencies. The existing related research majorly focuses on the investigation with the implementation of questionnaires and surveys [1, 60, 70]. However, to what extent the feedbacks to these questionnaires or surveys truthfully reflect the investors' actual practices in investing in cryptocurrencies remains uncertain and dubious.

Therefore, in Chapter 8 of this thesis, we inspect and appraise the behavioral biases and portfolio properties of cryptocurrency investors by utilizing the on-chain information of wallet records directly from the Ethereum network. Particularly, we directly evaluate three behavioral bias proxies and four different wallet properties of investors by analyzing the related transactions of each unique wallet address.

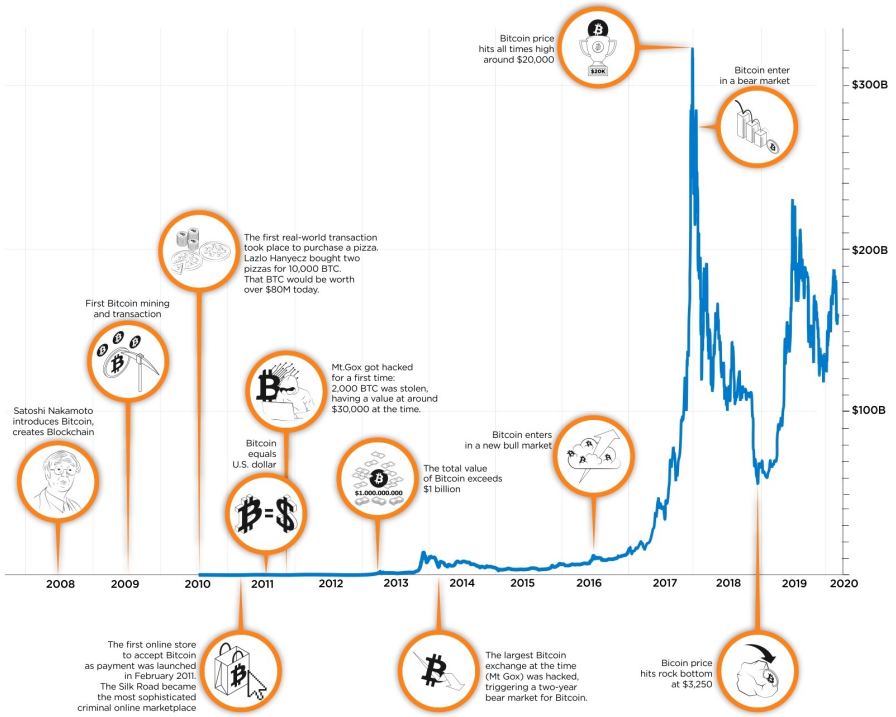


Figure 4.1: A visualization of Bitcoin's price trends and linked major events from 2008 to March 2020 [122].

4.2 ON-CHAIN METRICS: AS FUNDAMENTALS OF BLOCKCHAIN

On-chain metrics record the information generated by the blockchain network, like hash rate, circulating supplies, and exchange flows, whereas the trading records of wallet addresses are merely a small subset of the data on blockchains.

On-chain metrics of a blockchain are naturally comparable to the fundamentals, i.e., financial and operational details, of a public company. Just as the stock price fluctuates around the value of the company, it can be inferred that the price of crypto should also fluctuate around "some" intrinsic values. However, whether if cryptocurrencies have intrinsic values has always been a controversial and inconclusive topic [5, 32, 71, 92, 113], of which the discussion is beyond of the scope of this thesis.

With that being said, since on-chain metrics provide insights into the state, activities, and health of a blockchain and its crypto, in this thesis, we assume that on-chain metrics contribute to the approximation of the intrinsic values of cryptocurrencies, if such values exist.

Compared to the fundamentals, which disclose most of the information about a company, on-chain data are more precise and complete records of everything there is to know about a blockchain, without the possibility of falsification thanks to the decentralized, transparent, and tamper-resistant nature of blockchain networks. Most on-chain data are

recorded in a real-time and temporal-sequence manner, reflecting all the running details and measurements of a blockchain.

4.3 RELATED TOPICS COVERED IN THE THESIS

Due to the aforementioned nature of on-chain data, people expectedly wish to utilize and incorporate on-chain data into their systems for price prediction and quantitative trading [44–46, 94].

However, despite the fact that on-chain data are informative and beneficial, utilization of on-chain data has not been implemented in an RL-based system for PM so far. Moreover, the metrics found effective may not be applicable to blockchains other than Bitcoin or Ether, and this factor has been barely considered in the system design of the existing studies.

Thus, in Chapter 10, we propose CryptoRLPM, a novel end-to-end scalable RL-based system incorporating on-chain data for cryptocurrency portfolio management. With CryptoRLPM, we aim to answer two intriguing, yet, not answered, questions:

1. If and to what extent can the utilization of on-chain data improve the performance of a novel RL-based system compared to the baselines?

2. Does this system stand unbiased?

CryptoRLPM is a mid-frequency (10 to 30-minute interval) PM system consisting of five different units covering the process from information comprehension to trading order execution. In CryptoRLPM, the on-chain metrics are tested and specified for each crypto to solve the ineffectiveness of metrics.

Akin to the settings of MSPM in Chapter 7, each Crypto Module (CM) is constructed separately instead of jointly. That is, each CM reallocates a single-asset portfolio with a risk-free asset (i.e., cash) in it, and hence n CM will be required for an n-asset portfolio to be actually reallocated. By this setting, once a CM is trained, it becomes reusable and can be combined with other Crypto Modules for any given portfolio's weighted reallocation. Moreover, this setting of CryptoRLPM allows the portfolios to become scalable, with the underlying cryptocurrencies of the portfolios can be changed anytime at will. As metrics like sentiments from social media are incorporated, CryptoRLPM is also appraised regarding its robustness in terms of investment biases using FAIB.

To the best of our knowledge, CryptoRLPM is the first RL-based system using on-chain metrics for cryptocurrency PM. The benchmarking results indicate that CryptoRLPM robustly outperforms the baselines. Furthermore, we also prove that CryptoRLPM stands unbiased by testing with FAIB.

5 INVESTMENT BIASES AND BEHAVIORAL FINANCE

Behavioral finance falls into the intersection between psychology and finance. As opposed to conventional finance and economics, behavioral finance believes investors are not always rational, instead often have and are affected by their cognitive biases, resulting in irrational or sub-optimal decision-making [51, 105]. Moreover, such heuristic behaviors occasionally lead to misplacing of financial assets.

In this chapter, we focus on introducing the topics of behavioral finance related to this thesis. The first topic is about heuristics and investment bias, and after that, there will be types and cases of bias in investment. Then, we introduce one explanation of the origin of the biases. Since we will implement the biases as performance metrics in the later chapters, we also define the proxies of the biases that will be examined and implemented. In the last part of this chapter, we discuss the recent case studies of the investigation into investors' behavioral traits and biases in stock markets.

5.1 BEHAVIORAL FINANCE: HEURISTICS

Behavioral finance assumes that financial markets are informationally inefficient, in contrast with the Efficient Market Hypothesis (EMH) [59]. Behavioral finance aims to provide explanations for why and how people make irrational financial decisions [59, 99].

One key concept in behavioral finance, or cognitive science in general, is heuristics, referring to "a simple procedure that helps find adequate, though often imperfect, answers to difficult questions " [49].

In practice, people often lack enough cognitive abilities to comprehend the available information to make decisions. In this situation, individuals tend to make satisfactory rather than optimal decisions [98], and this idea is called bounded rationality [102]. A case of bounded rationality is "conjunction fallacy" [116], which occurs when people estimate a conjunction is more probable than at least one of its conjuncts. For example, the classic "Linda Problem" [117]:

> *Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.*

Which of the following is more probable?

1. Linda is a bank teller.

2. Linda is a bank teller and is active in the feminist movement.

Most people choose option 1, given the description and traits of Linda. However, people neglect the fact that the probability of two independent events occurring together can never be higher than the occurrence of one of the events.

As a result of bounded rationality, people may become more dependent on heuristics which may lead to various cognitive biases and more fallacies, as "...cognitive biases stem from the reliance on judgmental heuristics " [115]. Also, [34] states that "heuristics are the 'shortcuts' that humans use to reduce task complexity in judgment and choice, and biases are the resulting gaps between normative behavior and the heuristically determined behavior." The embodiment of heuristics in behavioral finance and investment results in a series of biases of investors.

Figure 5.1 [42], created by John Manoogian III and Buster Benson, visualizes and organizes the definition and categorization of more than 180 cognitive biases and heuristics.

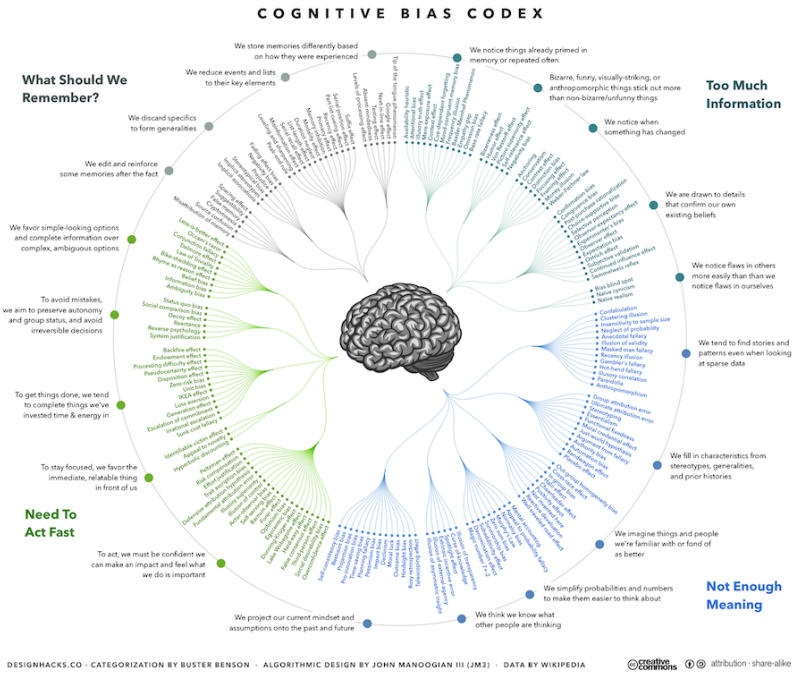


Figure 5.1: The Cognitive Bias Codex - 180+ biases, designed by John Manoogian III and Buster Benson [42]

5.2 INVESTMENT BIAS

Financial decisions are among the most important life-shaping decisions that people make. However, due to various cognitive biases and a low average degree of financial literacy, decision-making in investment

by personal investors often violates financial principles, which may result in irrational behaviors like over-extrapolation from past returns or frequent trading [18, 56, 57]. Such irrational behaviors can be interpreted by cognitive biases, which are systematic deviations from optimal reasoning [105]. Even top portfolio managers cannot overcome these biases when making investment decisions [29].

Here are three examples of common biases people may have when making investment decisions and how their proxies are measured:

- **Disposition effect (DE)** [58, 83, 99, 110] is about investors' tendency to realize profits too soon and keep losses too long. The proxy of disposition effect (DE) is measured by the difference between the proportions of gains realized and losses realized. A positive DE indicates that the investor has the disposition effect as the proportion of profits realized is larger than the proportion of losses realized. The greater the DE, the greater the degree of the disposition effect. On the other hand, a negative DE indicates no disposition effect.

- **Narrow framing (NF)** [48, 50, 58] relates to investors' tendency to make isolated and sub-optimal decisions and to trade assets without considering the holistic picture of their portfolios. The proxy of narrow framing (i.e., trade cluster, TC, or NF) is measured by the difference between 1 and the division of the number of trading days and the number of trades made during the same period. The lower the NF, the greater the degree of narrow framing as the investor more likely tends to execute trades separately instead of collectively.

- **Overconfidence (OC)** [84] is about an investor's tendency to trade frequently but unsuccessfully. In this study, the proxy of overconfidence (OC) of an investor equals 1 (indicating the existence of overconfidence) if the investor is in the highest portfolio turnover decile and in the lowest performance decile and 0 otherwise (indicating nonexistence of overconfidence).

In [18], the authors argue from a quantitative behavioral genetics perspective that the biases in investment root in human genes and are manifestations of innate and evolutionary features of human behavior. The author claim that genetic differences explain up to 45% of the remaining variation across individual investors. The authors also claim that general education is not found to reduce the relative importance of genetic factors in explaining investment biases, but financial work experience may help to reduce investment biases. This study provides empirical support for evolutionary arguments that biases are results of natural selection, which have survived because they were advantageous in evolutionary ancient times [9, 89]. The findings in this study are

lined up with the statement that "a bias is a source of error which is systematic rather than random," in [24].

5.3 RELATED TOPICS COVERED IN THE THESIS

As biases are unavoidable for humans, both individual and institutional investors want to utilize algorithms to avoid biases in investment since it is expected that there should be no such issue in the case of algorithmic trading. It then naturally becomes necessary to ensure that algorithmic trading (reinforcement learning (RL)-based systems in our case) can indeed outperform human investors on the discovered investment biases that are formularized as bias proxies or metrics. The existing research [40, 47, 63] has proposed successful RL-based systems for PM which achieve superior capital return performance. However, a high-performance system does not sufficiently indicate that it can overcome investment biases, or in order words, outperforms human investors over certain proxies of investment bias. The presence of these biases in the existing RL-based methods for PM is ignored and rarely appraised. In Chapter 7 and Chapter 10, we investigate the existence and degrees of common investment biases in two cutting-edge RL-based systems for PM: MSPM and CryptoRLPM, targeting stock and crypto PM, respectively. We prove that both PM systems overcome and outperform human investors over the investment biases. Moreover, in Chapter 9, we discuss the design of an end-to-end framework for appraising investment biases in portfolio management systems of heterogeneous types of financial assets (FAIB). FAIB can be considered, and utilized as, a guideline on how an appraisal framework shall be designed to answer if certain investment biases exist in the decision-making of any given heterogeneous-asset RL-based PM system, and to what degrees it has the biases if such biases exist.

Also, in Chapter 8, we inspect and appraise the behavioral biases and portfolio properties of cryptocurrency investors by utilizing the on-chain information of wallet records directly from the Ethereum network. By retrieving and analyzing the unique wallet addresses and related transactions, we have obtained three behavioral bias proxies of the investors behind the wallets and five different properties of the wallets. Furthermore, we distinguish and analyze the wallets of human investors and trading bots. The results of statistical tests indicate the significant differences between human investors and trading bots on all behavioral biases and wallet properties.

6

A MODULARIZED AND SCALABLE RL-BASED SYSTEM FOR PORTFOLIO MANAGEMENT

In this chapter, we propose a modularized and scalable multi-agent RL-based system for PM (MSPM). To bring scalability and reusability to attain dynamic, adaptive, and efficient RL-based PM, MSPM has been designed to have two types of asynchronously-updated modules: Evolving Agent Module (EAM) and Strategic Agent Module (SAM). An EAM is an information-generating module with a Deep Q-network (DQN) agent, and it receives heterogeneous data and generates signal-comprised information for a particular asset. An SAM is a decision-making module with a Proximal Policy Optimization (PPO) agent for portfolio optimization, and it connects to multiple EAMs to reallocate the corresponding assets in a financial portfolio. Once been trained, EAMs can be connected to any SAM at will, like assembling LEGO blocks, for multi-asset portfolio allocation. With its scalable and reusable design, MSPM aims to address the issue of ad-hoc, fixed, and inefficient model training in the existing RL-based methods. By experimenting on 8-year U.S. stock market data, we confirm that MSPM outperforms five different baselines in terms of the accumulated rate of return (ARR), daily rate of return (DRR), and Sortino ratio (SR). We back-test and compare MSPMs on four different portfolios to validate the indispensability of EAM. The merit of MSPM's architecture are reflected in the reusability and scalability brought by settings of EAM and SAM, as well as MSPM's outperformance over the existing baselines, including a cutting-edge RL-based PM system.

The contents of this chapter are based on the following publication:

[40] Huang, Zhenhan, and Fumihide Tanaka. "MSPM: A modularized and scalable multi-agent reinforcement learning-based system for financial portfolio management." PLoS ONE 17.2 (2022): e0263689.

6.1 RELATED WORK

In the early years, researchers and professionals believe that certain behaviors of price and volume will repeat periodically and consistently. Based on this recognition, the technical indicators (TI) are invented by using historical price and volume data to predict the movement of asset prices[27]. TIs are mostly formulas or particular patterns, and the trading strategies that utilize TIs are referred to as technical analysis (TA)[80]. However, as pre-defined formulas and patterns cannot cover

all market movements, it is getting harder and harder for TA to adapt to the fast-changing market. With the increase in computing power and available data, researchers have started to use deep learning (DL) to predict stock price movements. DL uses high-dimensional data to train complex and non-linear neural network models as trading strategies. Fortunately, DL's adaptability to the market is promisingly improved compared to TA. Recently, deep reinforcement learning (DRL) has emerged rapidly as the combination of DL and reinforcement learning (RL). By utilizing neural networks (NN), a DRL-based agent is particularly good at extracting useful information from high-dimensional data and taking sequential actions based on rewarding. DRL methods have led to many breakthroughs in multiple fields. For instance,[76] successfully utilizes Deep Q-learning agents to learn directly from high-dimensional raw pixel input to play video games. Due to the sequential decision-making nature of financial investment, researchers naturally attempt to solve stock trading problems using DRL methods. [47] designed a cryptocurrencies portfolio management (PM) framework using Deep Deterministic Policy Gradient (DDPG)[66, 101] which is a model-free DRL algorithm. [65] proposes the Adversarial Training method to improve training efficiency using three different RL methods: DDPG, Proximal Policy Optimization (PPO)[97] and Policy Gradient (PG). Although these approaches have presented potential performance, the data input of these approaches is still traditional historical data, namely opening-high-low-closing prices (OHLC) and trading volumes. Unlike preceding research, [121] proposes SARL, an RL framework that can incorporate heterogeneous data to generate PM strategies. Moreover, to address the challenge of balancing between exploration and exploitation, [68] proposes iRDPG for developing trading strategies by DRL and imitation learning. Multi-agent systems have also been proposed. In[63], the authors propose MAPS, a cooperative system containing multiple agents, to create diversified portfolios and to adapt to the continuously changing market conditions. However, while the existing approaches tackle PM problems with promising methods and techniques, these systems, with the strategies generated, are mostly fixed and ad-hoc. The existing systems or frameworks lack a modular design to be compatible with different trained RL agents. The RL agents trained for one portfolio can hardly be reused for different portfolios. These systems also lack scalability to accommodate the increasing number of assets and profundity of market information. In this chapter, we propose MSPM for solving the problems.

6.2 CONTRIBUTIONS

This chapter proposes a novel multi-agent reinforcement learning-based system with a modularized and scalable architecture for PM (MSPM). In MSPM, assets are vital and organic building blocks. This vitalness is

reflected in that each asset has its dedicated module: Evolving Agent Module (EAM). An EAM takes heterogeneous data and utilizes a DQN-based agent to produce signal-comprised information. After we set up and trained the EAMs corresponding to the assets in a portfolio, we connected them to a decision-making module: Strategic Agent Module (SAM). An SAM represents a portfolio and uses the profound information from the connected EAMs for asset reallocation. EAM and SAM are asynchronously updated, and EAMs' reusability allows themselves to be combined and connected to multiple SAMs discretionarily. With the power of parallel computing, we can perform capital reallocation for various portfolios at scale, simultaneously.

The contribution of this chapter can be concluded as the following:

- To the best of our knowledge, MSPM is the first approach that formalizes a modularized and scalable multi-agent reinforcement learning system using signal-comprised information for financial portfolio management.
- MSPM with its modularized and reusable design addresses the issue of ad-hoc, fixed, and inefficient model training in the existing RL-based methods.
- By experiment and comparison, we confirm that our MSPM system outperforms five different baselines under extreme market conditions of U.S. stock markets during the global pandemic, from January to December 2020.
- EAM-enabled MSPM systems improve accumulated rate of return of two different portfolios by 49.3% and 426.6% compared to Adversarial PG[65], a state-of-the-art RL-based method, and by 186.5% and 369.8% compared to (Uniform) Constant Rebalanced Portfolio (CRP)[19], a conventional PM strategy. In addition, the average winning rate of the EAMs in the two portfolios achieves 80%.
- Furthermore, we validate the indispensability of Evolving Agent Module (EAM) by back-testing MSPM on four different investment portfolios. Among the portfolios, EAM-enabled MSPMs improve accumulated rate of return by at least 1341.8% compared to the EAM-disabled MSPMs.

6.3 DATA

6.3.1 *Data acquisition*

The historical price data used in this chapter are QuoteMedia's End of Day US Stock Prices (EOD) [88] from Jan 2013 to Dec 2020 obtained using Nasdaq Data Link's API, which can be accessed by subscribing at:

https://data.nasdaq.com/data/EOD-end-of-day-us-stock-prices. We also use web news sentiment data (FinSentS)[43] from Nasdaq Data Link provided by InfoTrie, which can be accessed by subscribing at: https://data.nasdaq.com/databases/NS1/data.

6.3.2 *Feature selection and data curation*

We select the adjusted- close, open, high, and low prices and volumes features from QuoteMedia's EOD data as the historical price data. We also select the sentiment and news_buzz from InfoTrie's FinSentS Web News Sentiment. Each feature in EOD data is normalized by dividing the first (day-one) value of that feature, and there is no missing value in any of these features. For FinSentS data, we use original values of the sentiment feature in FinSentS data, and we fill the missing values (accounting for 9.51% of the total data) prior year 2013 with a neutral sentiment: zero (0). Since the FinSentS data are not as straightforward as EOD data, we put the description of the selected features of FinSentS data in Table 6.1.

Table 6.1: Feature selection of FinSentS data

Feature	Description
sentiment	A measure of bullishness and bearishness of equity prices calculated as a statistical index of the news corpus. Sentiment scores are defined on a scale of -5 to 5 indicating from the most bearish to the most bullish.
news_buzz	Normalized value of change in standard deviations of periodic number of news items (news volume) used for generating sentiment. Buzz scores reflect a sharp change in news volume thus serving as a risk alert indicator. Defined on a scale of 1-10 high buzz score reflects higher volatility.

6.4 METHODOLOGY

Our MSPM system consists of two types of modules: EAM and SAM. The relationship between EAMs and SAMs is illustrated in Figure 6.1. Figure 6.2 illustrates a even more intuitive overview of MSPM's architecture. To accommodate MSPM in the sequential decision-making problems financial portfolio management, we configured the specific settings for EAM and SAM. An EAM contains a DQN **agent** and **acts** to generate signal-comprised information (historical prices with buy/-closing/skip labels) for a designated asset. To train the agent in EAM, we constructed a sequential decision-making problem with designated asset's historical prices and financial news as the **state** that the agent observes at each time step. An DQN agent acts to buy or close a position, or simply to skip at every time step based on the latest prices

and financial news data input, in order to maximize its total reward. The actions (signals) then will be matched and stacked back to the corresponding price data to formalize the signal-comprised information. EAM's architecture is illustrated in Figure 6.3. On the other hand, an SAM manages an investment portfolio and contains a PPO **agent** that reallocates the assets in that portfolio. SAMs are connected to multiple EAMs as an investment portfolio often has more than one asset. In the decision-making process of SAM, the state that the PPO agent **observes** at each time step is the combination of the signal-comprised information which the connected EAMs generate. Further, the PPO agent **acts** to generate the reallocation weights for the assets in the portfolio, which total up to 1.0. Figure 6.4 provides an overview of the SAM's architecture. For both EAM and SAM, the composition of the assets' historical prices and financial news or news sentiments is the **environment** their agents interact with. Each EAM is reusable. Once an EAM is set up and trained, it can be effortlessly connected to any SAM. An SAM connects to at least one EAM. EAMs are retrained periodically using the latest information from the market, media, financial institutions, etc., and we implemented the former two data sources in this study. In the following sections, we explain the technical details of EAM and SAM.

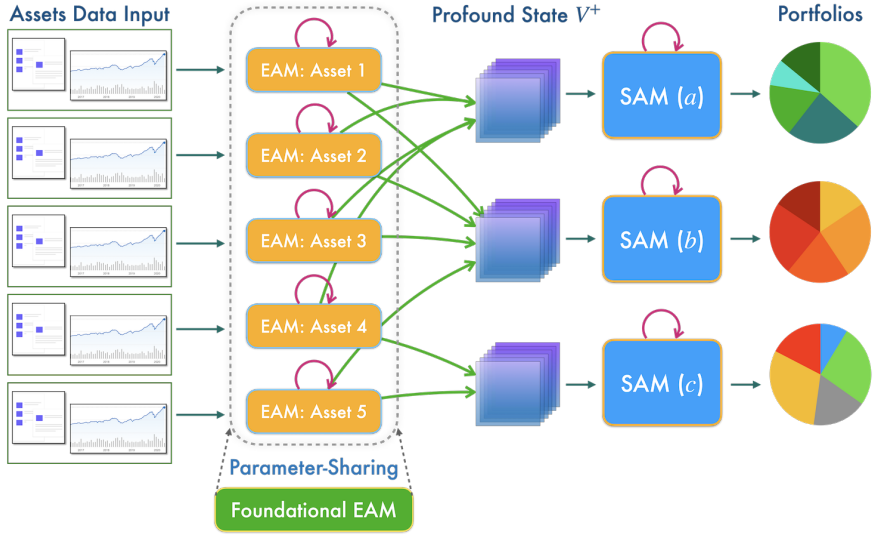


Figure 6.1: Overview of the surjection relationship between Evolving Agent Modules (EAMs) and Strategic Agent Modules (SAMs). Each EAM is responsible for a single asset and employs a DQN agent, and it utilizes heterogeneous data to produce signal-comprised information. Each SAM is a module for a portfolio that employs a PPO agent to reallocate the assets with stacked signal-comprised 3-D tensor ***profound state*** V^+ from EAMs connected. Moreover, trained EAMs are reusable for different portfolios and therefore can be combined and connected to any SAMs at will. By parallel computing, capital reallocation may be performed for various portfolios at scale simultaneously.

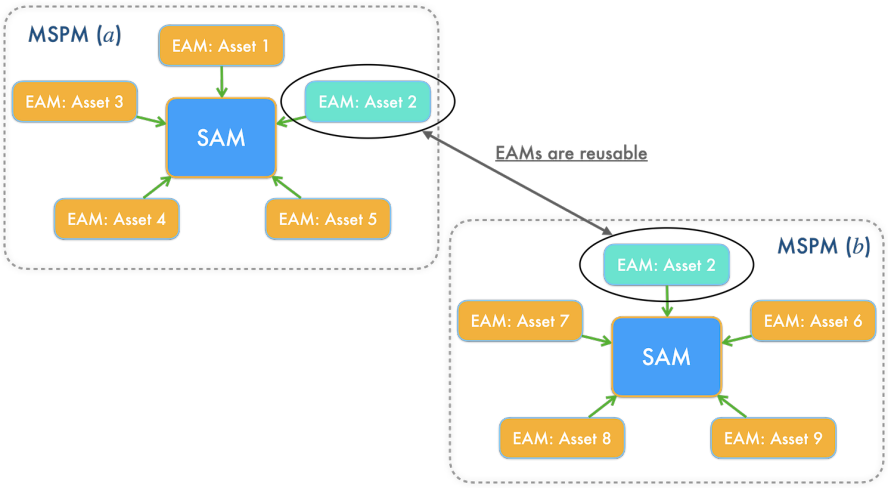


Figure 6.2: A more intuitive illustration of MSPM's architecture. EAMs are reusable for different portfolios. EAMs can be combined and connected to any SAMs at will, like assembling LEGO blocks.

6.4.1 *Evolving Agent Module (EAM)*

6.4.1.1 *State*

At any given periodic (daily) time-step t, the agent in EAM observes state v_t, which consists of the designated asset's recent n-day historical prices s_t and sentiment scores ρ_t. Specifically,

$$v_t = (s_t, \rho_t), \tag{6.1}$$

where s includes the designated asset's n-day close, open, high and low prices and volumes. ρ includes the predicted and averaged news sentiments, using a pre-trained FinBERT classifier [2, 22] for asset-related financial news, which ranges continuously from -5.0 to 5.0, indicating bearishness (-5.0) or bullishness (5.0). Furthermore, ρ also includes *news_buzz*. This attribute is an attempt to alleviate the unbalanced-news issue in the existing research [121]. Instead of restarting from the beginning after every episodic reset of the environment, the environment resets at a random time point of the data [61].
Because the news sentiments from FinSentS data and the sentiments generated by FinBERT are similar, and due to the restriction of APIs and web scraping, we only utilize FinSentS data as the sentiments input for the experiments in this chapter.

6.4.1.2 *Deep Q-Network*

For an EAM, we train a Deep Q-network (DQN) agent and follow the sequential decision-making of Deep Q-learning [76]. Deep Q-learning is a value-based method that derives a deterministic policy $\pi(\theta)$, which is a mapping: $S \rightarrow A$ from state space to discrete action space. We use

a Residual Network with 1-D convolution [37] to represent Q^θ, the estimate of action-value function, which the agent acts based on:

$$Q^\theta(s_t, a_t) = \mathbb{E}_{\pi_\theta}[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a] \tag{6.2}$$

For model selection, we have tested different architectures of neural network models for the DQN agent in EAM. Among them, we chose Residual Network with 1-D convolution since it performed the best on the validation dataset described in Table 6.2 in Data ranges section. We also performed numerous experiments for hyperparameter tuning on the validation dataset to make sure the hyperparameters implemented and stated in the article are the optimized for the use cases in this research.

DQN EXTENSIONS: We implement three extensions [61] of the original DQN, namely dueling architecture [120], Double DQN [36] and two-step Bellman unrolling.

TRANSFER LEARNING: Instead of training every EAM from scratch, we initiate and train a **foundational EAM**, using historical prices of AAPL (Apple Inc.), and then train all other EAMs based on this pre-trained EAM. By doing so, the foundational EAM shares its parameters with other EAMs which obtains prior knowledge of the pattern of stock trends. This transfer learning approach may help to tackle the **data-shortage** issue of newly-listed stocks due to the limited historical prices and news data available for training purposes.

6.4.1.3 *Action*

The DQN agent in EAM acts to trade the designated asset with an action of either buying, selling, or skipping, at every time step t. The choice of an action, a_t = {buying, closing, or skipping}, is called an ***asset trading signal***. As indicated in the actions, there is no short (selling) position, and a new position will be opened only after an existing position has been closed.

6.4.1.4 *Reward*

The reward, r_t, received by the DQN agent at each time step t is:

$$r_t(s_t, \iota_t) = \begin{cases} 100(\sum_{i=t_\iota}^{t} \frac{v_t^{(close)}}{v_{t-1}^{(close)}} - 1 - \beta), \text{ if } \iota_t \\ 0, \text{ if not } \iota_t \end{cases} \tag{6.3}$$

where $v_t^{(close)}$ is the close price of the given asset at time step t. t_l is the time step when a long position is opened and commissions are deducted, β stands for the commission of 0.0025 and ι_t is the indicator of an opening position (i.e., a position is still open).

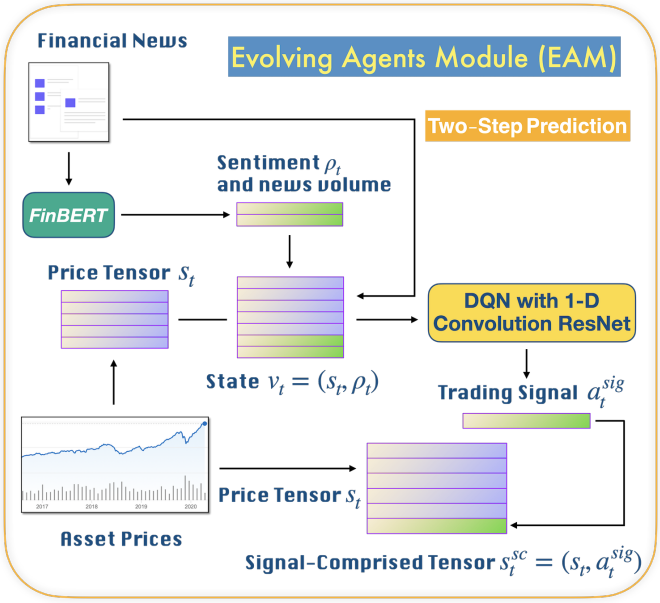


Figure 6.3: Abstract of EAM's architecture.
An EAM is a module for a designated asset. Each EAM takes two types of heterogeneous data: 1. designated asset's historical prices and 2. asset-related financial news. At the center of an EAM is an extended DQN agent using a 1-D convolution ResNet for sequential decision making. Instead of training every EAM from scratch, we train EAMs by transfer learning using a foundational EAM. At every time step t, the DQN agent in EAM observes state v_t of historical prices s_t and news sentiments ρ_t of the designated asset, and acts to trade with an action a_t^{sig} of either buying, selling, or skipping, and eventually generates a 2-D signal-comprised tensor s_t^{sc} using new prices s_t and signals a_t^{sig}.

6.4.2 *Strategic agent module (SAM)*

6.4.2.1 *State (stacked signal-comprised tensor)*

Once EAMs have been trained, we feed new historical prices, s_t, and financial news of the designated assets, to generate predictive trading signals a_t^{sig}. Then we stack the same new historical prices to a_t^{sig} to formalize a 2-D signal-comprised tensor s_t^{sc} as the data source to train SAM. Because an SAM is connected to multiple EAMs, the 2-D signal-comprised tensors from all connected EAMs are stacked and transformed into a 3-D signal-comprised tensor called ***profound state*** v_t^+, which is the state that SAM observes at each time step t.

6.4.2.2 *Proximal policy optimization*

A PPO [97] agent is at the center of SAM to reallocate assets. PPO is an actor-critic style policy gradient method that has been widely used on continuous action space problems, due to its desirable performance and

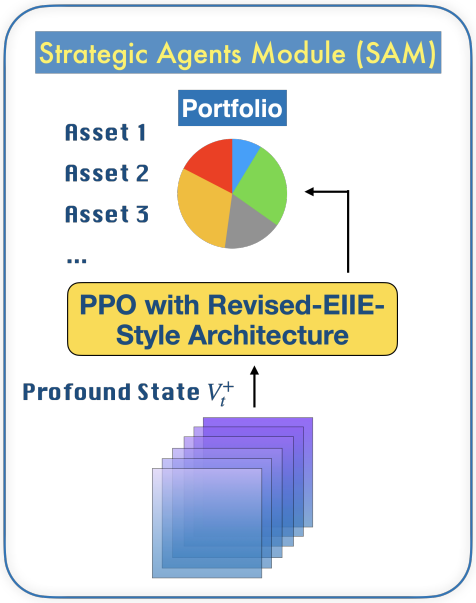


Figure 6.4: Abstract of SAM's architecture.
An SAM is a module for an investment portfolio. The input of SAM, profound state $V_t^+ \in \mathbb{R}^{f \times m^* \times n}$, is a 3-D tensor, where f is the number of features, $m^* = m + 1$ is the number of assets m in the portfolio plus cash and n is the fixed rolling-window length. Each SAM takes the profound state V_t^+ which is stacked and transformed from 2-D tensors from connected EAMs, and further generates the reallocation weights for the assets in the portfolio.

ease of implementation. A policy π_θ is a parametrized mapping: $S \times A \rightarrow [0, 1]$ from state space to action space. Among the different objective functions of PPO, we implement the clipped surrogate objective [97]:

$$L(\theta) = \hat{\mathbb{E}}\pi_{\theta'}[min(r_t(\theta)A_t^{\theta'}, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t^{\theta'})] \tag{6.4}$$

where

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta'}(a_t|s_t)}$$

and $A_t^{\theta'}$, the advantage function, is expressed as:

$$A_t^{\theta'} = Q^{\theta'}(s_t, a_t) - V^{\theta'}(s_t)$$

in which, the state-action value function $Q^{\theta'}(s_t, a_t)$ is:

$$Q^{\theta'}(s_t, a_t) = \mathbb{E}_{\pi_{\theta'}}[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a]$$

and the value function $V^{\theta'}(s_t)$ is:

$$V^{\theta'}(s_t) = \mathbb{E}_{\pi_{\theta'}}[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s]$$

For the PPO agent, we design a policy network architecture targeting the uniqueness of continuous action space in financial portfolio management problems, inspired by the EIIE topology [47]. Because assets' reallocated weights at time step t are strictly required to total up to 1.0, we set m^* normal distributions $N_1(\mu_t^1, \sigma), ..., N_{m^*}(\mu_t^{m^*}, \sigma)$, and we sample $x_t \in \mathbb{R}^{m^* \times 1}$ from the distributions, where $m^* = m + 1$ and $\mu_t \in \mathbb{R}^{1 \times m^* \times 1}$ is the linear output of the last layer of the neural network and with standard deviation $\sigma = 0$. We eventually obtain the reallocation weights $a_t = Softmax(x_t)$ and the log probability of x_t for the PPO agent to learn.

Figure 6.5 shows the details of the policy network (actor) of SAM, denoted by θ'. Due to the resemblance and equivalence, architectures of the value network (critic) and target policy network, denoted by θ, are not illustrated.

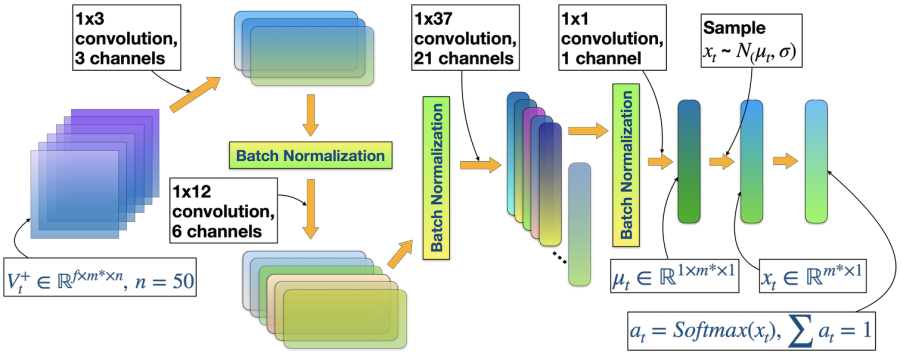


Figure 6.5: Policy network (θ') of SAM to accommodate PPO algorithm. Profound state V_t^+ is the input of the network. f is the number of features, m^* is the number of assets in the portfolio, and $n = 50$ is the fixed rolling-window length. After $x_t \in \mathbb{R}^{m^* \times 1}$ are sampled from the normal distributions $N_1(\mu_t^1, \sigma), ..., N_{m^*}(\mu_t^{m^*}, \sigma)$, we calculate log probability of x_t and obtained the reallocation weights $a_t = Softmax(x_t)$. ReLu activation function [81] is set after every convolutional layer, except the last one.

6.4.2.3 *Action*

The action the PPO agent takes at each time step t is

$$a_t = (a_{1,t}, a_{2,t}, ..., a_{m^*,t})^T \tag{6.5}$$

which is the vector of reallocating weights at each time step t, and $\sum_{i=1}^{m^*} a_{i,t} = 1$. Figure 6.6 shows the details of price fluctuations.

Once the assets are reallocated by a_t, the allocation weights of the portfolio eventually become

$$w_t = \frac{y_t \odot a_t}{y_t \cdot a_t} \tag{6.6}$$

at the end of time step t due to the price fluctuation during the time step period; where,

$$y_t = \frac{v_t^{+(close)}}{v_{t-1}^{+(close)}} = (1, \frac{v_{2,t}^{+(close)}}{v_{2,t-1}^{+(close)}}, ..., \frac{v_{m^*,t}^{+(close)}}{v_{m^*,t-1}^{+(close)}})^T \tag{6.7}$$

is the relative price vector, that is, the changes of asset prices over time, including the prices of assets and cash. $v_{i,t}^{+(close)}$ denotes the closing price of the i-th asset at time t, where $i = \{2, ..., m^*\}$, excluding cash (risk-free asset) whose closing price should always be 1.

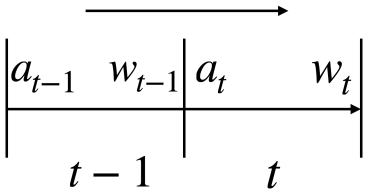


Figure 6.6: Transformed allocation weights due to the fluctuation in assets' prices.

6.4.2.4 *Reward*

Inspired by [47] in which the agent maximizes the sum of the logarithmic value, and [65] in which the authors try to cluster the periodic portfolio risk to alleviate the biases in training data and to prevent exposure to highly-volatile assets, we set the reward to be a risk-adjusted rate of return, r_t^*, which PPO agent receives at each time step t:

$$r_t^*(s_t, a_t) = \ln(a_t \cdot y_t - \beta \sum_{i=1}^{m^*} |a_{i,t} - w_{i,t}| - \phi\sigma_t^2) \tag{6.8}$$

where m^* is the number of assets, w_t represents the allocation weights of the assets at the end of time step t.

$$\beta \sum_{i=0}^{n} |a_{i,t} - w_{i,t}| \tag{6.9}$$

is the transaction cost, where $\beta = 0.0025$ is the commission rate, and $\phi = 0.001$ is the risk discount which can be fine-tuned as a hyperparameter.

$$\sigma_t^2 = \frac{1}{n} \sum_{t-n+1}^{t} \sum_{i=1}^{m^*} (y_{i,t-n+1}, \overline{y_{i,t-n+1}})^2 \tag{6.10}$$

measuring the volatility of fluctuation in assets' prices during the last n days.

$$\overline{y_{i,t-n+1}} = \frac{1}{n} \sum_{t-n+1}^{t} y_{i,t-n+1} \tag{6.11}$$

is the volatility of the profit of an individual asset. We expect the agent to secure a maximum risk-adjusted rate of return (capital gain) every time step, as what is expected from human portfolio managers.

6.5 EXPERIMENTS

In this section, we build different portfolios, and train MSPM to periodically reallocate the assets in each portfolio.

To validate the advantages and viability of MSPM and its settings, we follow the conventional scheme of back-testing which is the way to evaluate financial trading strategies in the research area of QT [47, 63, 65, 68, 121]. Beyond that, we further perform statistical tests to compare the stability of daily rate of return (DRR) between our system: MSPM and the state-of-the-art RL-based method: ARL. The results of the tests demonstrate the robustness and flexibility of MSPM to adapt to the ever-changing financial market.

The portfolios, datasets, and performance metrics for benchmarking will be introduced and described. After that, we explain and discuss the experimental results and examine MSPM's stability of daily rate of return. We also inspect the signal generation and position-holding of EAMs. In the end, we validate the necessity of EAM by back-testing four different portfolios. The back-testing performance of MSPM will be compared with the existing baselines.

6.5.1 *Preliminaries*

6.5.1.1 *Portfolios*

We first propose two portfolios: (a) and (b) to compare back-testing performance. Portfolio(a) includes three stocks: Apple, AMD, and Alphabet (symbol codes: [AAPL, AMD, GOOGL]), and Portfolio(b) includes three other stocks: Alphabet, Nvidia, and Tesla (symbol codes: [GOOGL, NVDA, TSLA]). To build portfolio(a) and portfolio(b), we trained two SAM/MSPMs: SAM/MSPM(a) and SAM/MSPM(b). Additionally, the two SAMs shared the same EAM for the stock in common: Alphabet (GOOGL). Later, we propose two other portfolios (c) and (d), which make four portfolios in total, to validate the necessity of EAM. Details can be found in the Validation of EAM section. For all these four portfolios, we set initial portfolio value to be $p_0 = 10,000$.

6.5.1.2 *Data ranges*

Among the EAMs to be trained, the foundational EAM (AAPL) is trained initially, and its parameters are shared with other EAMs as their foundation for transfer learning. As shown in Table 6.2, EAM-training data, ranging from January 2009 to December 2015, contains the historical prices (s_t) and news sentiments (ρ_t) of the stocks, including AAPL, in portfolios (a) and (b). EAM-predicting data, with the same data structure as EAM-training and ranging from January 2016 to December 2020, is used for EAMs to predict and generate trading signals (actions

of DQN agents). Then, EAM-predicting data along with the generated trading signals became the signal-comprised data for SAM/MSPMs. There are three datasets of signal-comprised data: SAM/MSPM-training and SAM/MSPM-validating to train and validate SAMs, respectively; and SAM/MSPM-experiment, from January 2020 to December 2020, for back-testing and other experiments. Details can be found in Table 6.2. It is worth noting that a low percentage (9.51%) of missing values from the alternative data (sentiments) shall not affect MSPM's scalability nor reusability since, as a general framework, MSPM is neutral on the structures, types, or sources of the data input.

Table 6.2: Date ranges of the data

Purpose	Data Range
EAM-training	Jan 2009~Dec 2015
EAM-predicting	Jan 2016~Dec 2020
SAM/MSPM-training	Jan 2016~Dec 2018
SAM/MSPM-validation	Jan 2019~Dec 2019
SAM/MSPM-experiment	Jan 2020~Dec 2020

EAM-training dataset includes the historical prices (s_t) and news sentiments (ρ_t) of all assets in the portfolios constructed, and is used to train the AAPL-based foundational EAM, and transfer learning for the four other assets. EAM-predicting dataset includes new historical prices (s_t) and news sentiments (ρ_t^*) for EAMs to generate signal-comprised tensors ($s_t^{sc} = (s_t, a_t^{sig})$) to formalize the SAM/MSPM-training (v^+) data. SAM/MSPM-validation and SAM-back-testing data have the same structure as SAM/MSPM-training but are used solely for the purposes of validation and back-testing.

6.5.1.3 *Performance metrics*

We use the following performance metrics to measure the performances of the baselines and MSPM system.

- **Daily rate of return (DRR)**

$$DRR_T = \frac{1}{T}\sum_{t=1}^{T} \exp(R_t), \tag{6.12}$$

where T is the terminal time step, and

$$R_t = \ln\left(a_t \cdot y_t - \beta \sum_{i=1}^{m^*} |a_{i,t} - w_{i,t}|\right) \tag{6.13}$$

is the risk-unadjusted periodic (daily) rate of return obtained at every time step, where $\beta \sum_{i=1}^{m^*} |a_{i,t} - w_{i,t}|$ is the transaction cost and $\beta = 0.0025$ is the commission rate.

- **Accumulated rate of return (ARR) [86]**

$$ARR_T = \frac{p_T}{p_0}, \tag{6.14}$$

where T is the terminal time step, p_0 is the portfolio value at the initial time step, and

$$p_T = p_0 \exp\left(\sum_{t=1}^{T} R_t\right) \tag{6.15}$$

which stands for the portfolio value at the terminal time step.

- **Sortino ratio (SR) [103]**is often referred to as a risk-adjusted return, which measures the portfolio performance compared to a risk-free return, adjusted by the portfolio's downside risk. In our case, Sortino ratio is calculated as

$$SR = \frac{\frac{1}{T}\sum_{t=1}^{T} \exp(R_t) - R_f}{\sigma^{downside}} \tag{6.16}$$

where R_t is the risk-unadjusted periodic (daily) rate of return. Portfolio's downside risk $\sigma^{downside}$ is calculated as

$$\sigma^{downside} = \sqrt{\text{Var}(R_l - R_f)}, \tag{6.17}$$

where R_f is the risk-free return and conventionally equals zero, R_l are the less-than-zero returns in R_t for all t, and $t = T$ is the terminal time step.

- **Max drawdown (MD)** is the biggest drop (in %) between the highest (peak) and lowest (valley) of the accumulated rate of return of a certain period of time.

For DRR, ARR and SR, we want them to be as high as possible, whereas we want MD to be as low as possible.

6.5.2 *Results and discussion*

6.5.2.1 *Back-testing performance*

We back-test and compare the performance of our MSPM system to different baselines, including the traditional and cuttings-edge RL-based portfolio management strategies [41, 64]. The baselines are listed as follows:

- CRP stands for (Uniform) Constant Rebalanced Portfolio, which involves investing an equal proportion of capital in each asset, namely 1/N, which seems simple but, in fact, challenging to beat [19].
- **Buy and hold (BAH)** strategy involves investing without rebalancing. Once the capital is invested, no further allocation will be made.
- **Exponential gradient portfolio (EG)** strategy involves investing capital into the latest stock with the best performance and uses a regularization term to maintain the portfolio information.
- **Follow the regularized leader (FTRL)** strategy tracks the Best Constant Rebalanced Portfolio until the previous period, with an additional regularization term. This strategy reweights based on the entire history of the data with an expectation to obtain maximum returns.
- **ARL** refers to the adversarial deep reinforcement learning in portfolio management (Adversarial PG) [65], which is a state-of-the-art (SOTA) RL-based portfolio management method.

As shown in Figure 6.7 and Figure 6.8, for both portfolios (a) and (b), MSPM system improves ARR, by at least 49.3% and 426.6% compared to ARL, a SOTA RL-based PM method, and by 186.5% and 369.8% compared to CRP, a traditional PM strategy, during the year of 2020. The result demonstrates the advantage of MSPM at gaining capital returns. Table 6.3 gives details about MSPM's outperformance over existing baselines in terms of the ARR and DRR. Further, MSPM's superior performance on SR indicates that MSPM takes better consideration of harmful volatility and achieves higher risk-adjusted returns.

Table 6.3: Comparison of back-testing performance of the baselines and MSPM

	Portfolio (a)						Portfolio (b)					
Metric	CRP	BAH	EG	FTRL	ARL	MSPM	CRP	BAH	EG	FTRL	ARL	MSPM
DRR (%)	0.175	0.173	0.173	0.174	0.333	**0.404**	0.350	0.460	0.440	0.395	0.367	**0.938**
ARR (%)	45.9	44.7	44.9	45.4	88.1	**131.5**	120.6	175.4	164.0	140.7	107.6	**566.6**
MD (%)	**-23.3**	-23.6	-23.5	-23.4	-34.3	-31.3	**-33.6**	-35.7	-35.3	-34.5	-37.6	-60.6
SR	1.95	1.88	1.89	1.92	2.13	**2.86**	3.24	3.54	3.50	3.38	2.35	**4.18**

It is worth noting that for portfolio (a), both MSPM and ARL achieve promising SR, but for portfolio (b), only MSPM has a much better Sortino ratio than ARL, which indicate MSPM's higher adaptability to the ever-changing market compared to not only the traditional strategies but also the preceding RL-based method.

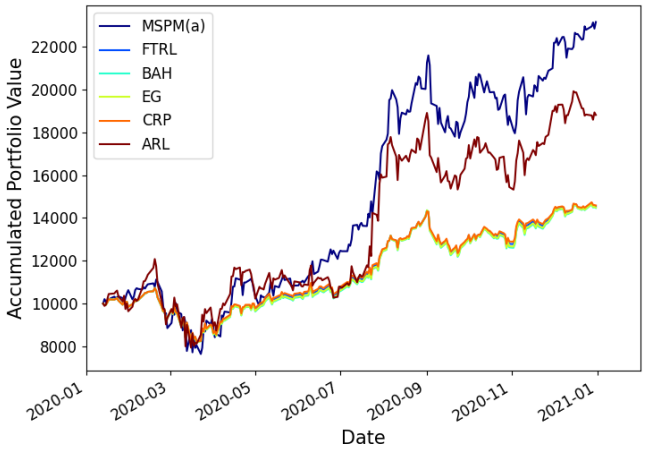


Figure 6.7: MSPM(a) outperforms all baselines on Portfolio(a) in terms of the accumulated portfolio value in back-testing.

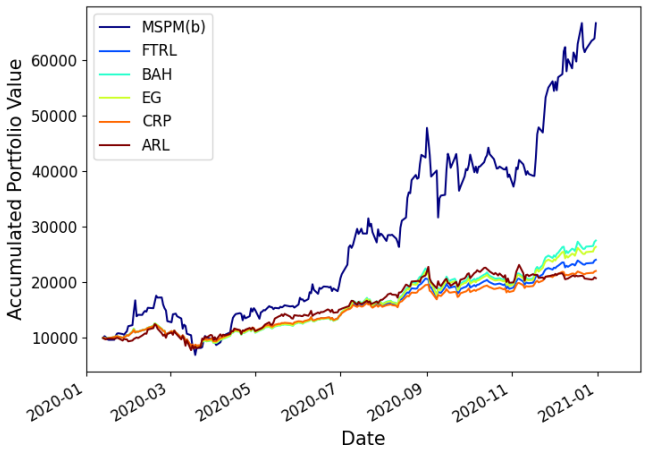


Figure 6.8: MSPM(b) outperforms all baselines on Portfolio(b) in terms of the accumulated portfolio value in back-testing.

6.5.2.2 *Stability of daily rate of return (DRR)*

Due to the high max drawdown (MD) of MSPM for portfolio(b) (60.6%), we want to examine and compare the general stability of DRR between MSPM and the state-of-the-art RL-based method: ARL. For this purpose, we first calculate DRR's 5-day rolling standard deviation (RstdDRR) as the proxy of the stability of DRR. Higher RstdDRR indicates lower stability of DRR.

To calculate the RstdDRR, we first calculate the simple moving average (SMA) [62] of DRR$\in \mathbb{R}^k$ for the past n data-points (days) by the following formula:

$$SMA_i = \frac{DRR_{i-n+1} + DRR_{i-n+2} + ... + DRR_i}{n} \tag{6.18}$$

for $i = n, ..., k$. Then, we subtract SAM_i from the 5-day DRRs used in the calculation, and then take the square root of the squared summation to have the rolling standard deviation: RstdDRR$\in \mathbb{R}^{k-n}$:

$$RstdDRR_i = \sqrt{\frac{(DRR_{i-n+1}-SMA_i)^2+(DRR_{i-n+2}-SMA_i)^2+...+(DRR_i-SMA_i)^2}{n}} \tag{6.19}$$

where $i = n, ..., k$.

Figure 6.9 shows the histograms of MSPM and ARL's RstdDRR for portfolio(a), and histograms in Figure 6.10 are for portfolio(b). According to Figure 6.9, the right tail of ARL's RstdDRR is fatter than that of MSPM's RstdDRR, and MSPM has a lower average RstdDRR ($M_{(a)} = 0.031$, $SD_a = 0.019$) than ARL ($M_{(a)} = 0.034$, $SD_a = 0.020$), indicating MSPM has higher stability of DRR on portfolio(a). However, Figure 6.10 depicts that the right tail of MSPM's RstdDRR is fatter than that of ARL's RstdDRR, and the mean of MSPM's RstdDRR ($M_{(b)} = 0.049$, $SD_b = 0.027$) is larger than the mean of ARL's RstdDRR ($M_{(b)} = 0.032$, $SD_b = 0.022$). For more information, figures in Figure 6.11 and Figure 6.12 give the comparison between MSPM and ARL's RstdDRR for portfolio (a) and (b). As shown in Figure 6.11, the RstdDRR of MSPM is less volatile than that of ARL, but it is the opposite case in Figure 6.12.

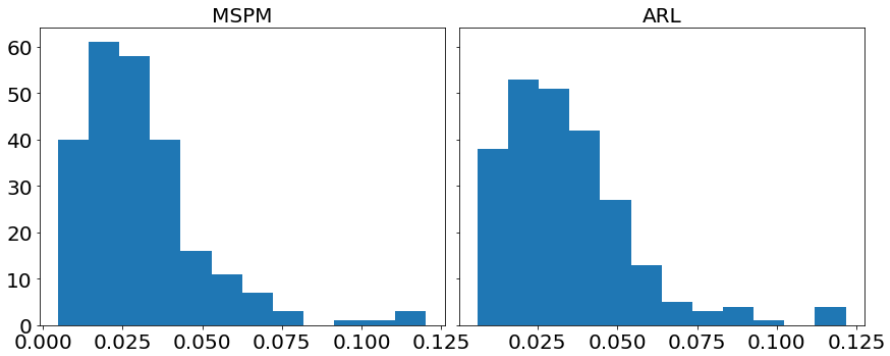


Figure 6.9: For portfolio(a), histograms of MSPM and ARL's 5-day RstdDRR depict right-skewed distributions.

Since the histograms in Figure 6.9 and Figure 6.10 show skewed bell shapes, we use Shapiro-Wilk test [93] to confirm the normality

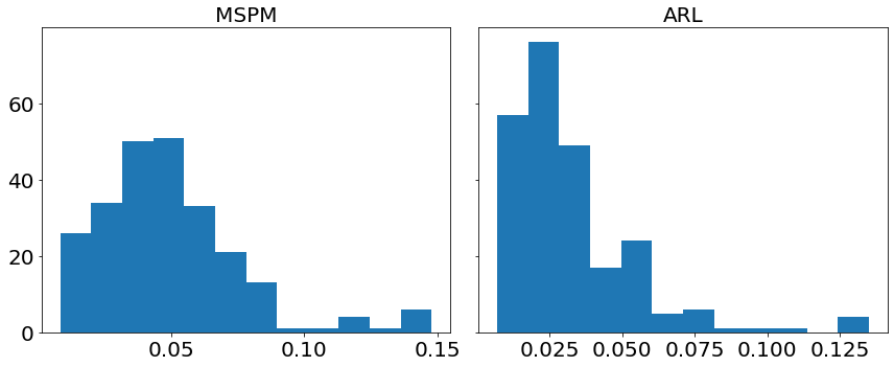


Figure 6.10: For portfolio(b), histograms of MSPM and ARL's 5-day RstdDRR depict right-skewed distributions.

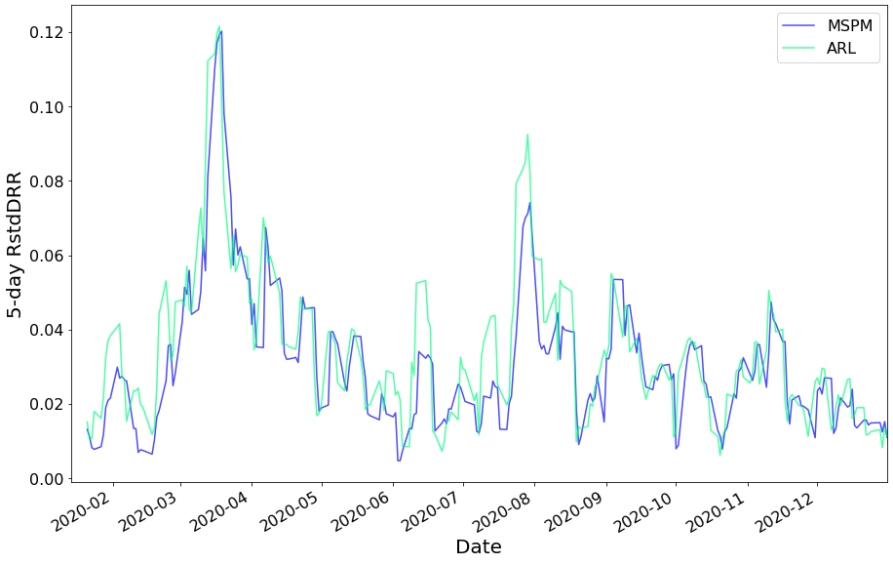


Figure 6.11: 5-day RstdDRR of Portfolio(a): MSPM versus ARL.

of the distributions. After that, we use Levene's test [85] to examine the variance equality. We use Python's SciPy library to perform these two tests. By implementing Shapiro–Wilk test, we find that MSPM and ARL's RstdDRR are not statistically from normal distributions for both portfolios (p-values are less than 0.05). Moreover, according to Levene's test, MSPM and ARL's RstdDRR do not always have homogeneity of variance: for portfolio (a) they do, whereas for portfolio(b) they do not. With the assumptions verified, we perform the one-tail and two-sample Mann–Whitney U test [72] (a non-parametric version of unpaired t-test) to rigorously compare MSPM and ARL's stability of DRR, also using Python's SciPy library. For portfolio(a), because the mean RstDRR of MSPM is less than the mean RstDRR of ARL, the hypothesis H_0 is that MSPM has a lower or same stability than ARL (the group mean of RstdDRR of MSPM is greater or equal to that of ARL), and the alternative hypothesis H_a is that MSPM has higher stability than ARL

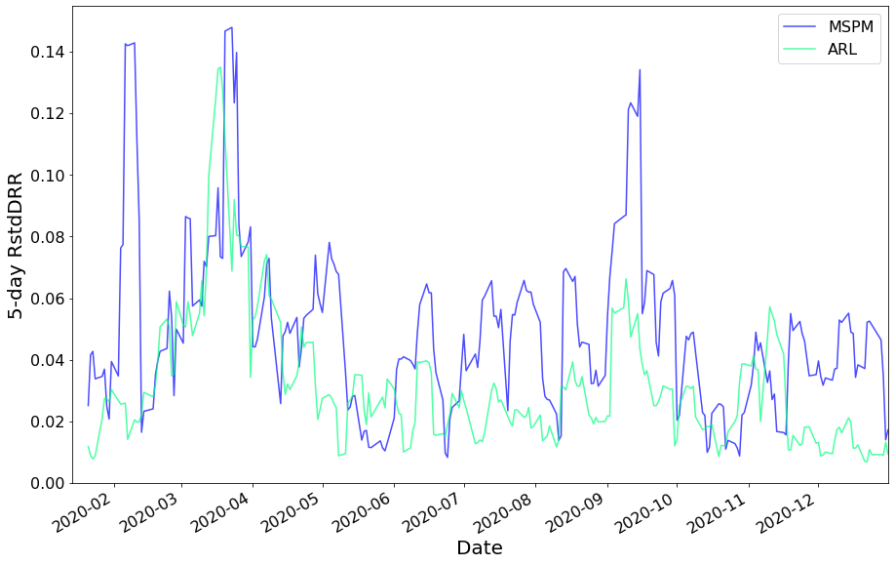


Figure 6.12: 5-day RstdDRR of Portfolio(b): MSPM versus ARL.

(the group mean of RstdDRR of MSPM is less than that of ARL). For portfolio(b), because the mean RstDRR of MSPM is higher than the mean RstDRR of ARL, the hypothesis H_0 is that MSPM has higher or same stability than ARL (the group mean of RstdDRR of MSPM is less or equal to that of ARL), and the alternative hypothesis H_a is that MSPM has a lower stability than ARL(the group mean of RstdDRR of MSPM is greater than that of ARL). We set the significance level to be .05. If the p-value from the test is less than 0.05, we reject H_0 and accept H_a; otherwise, we accept the null hypothesis H_0. The detailed settings of the statistical test are:

- Statistical test: one-tail and two-sample Mann–Whitney U test
- For portfolio (a), null hypothesis $H_0^{Pa} : \mu_{MSPM}^{RstdDRR} - \mu_{ARL}^{RstdDRR} \geq 0$
- For portfolio (a), alternative hypothesis $H_a^{Pa} : \mu_{MSPM}^{RstdDRR} - \mu_{ARL}^{RstdDRR} < 0$
- For portfolio (b), null hypothesis $H_0^{Pb} : \mu_{MSPM}^{RstdDRR} - \mu_{ARL}^{RstdDRR} \leq 0$
- For portfolio (b), alternative hypothesis $H_a^{Pb} : \mu_{MSPM}^{RstdDRR} - \mu_{ARL}^{RstdDRR} > 0$
- Significance level: .05

As the results represented in Table 6.4, MSPM has significantly higher stability of DRR than ARL for portfolio(a) by rejecting H_0 and accepting H_a ($U_a = 25426.0, p - value = .005$). For portfolio(b), because H_0 is accepted ($U_b = 16209.0, p - value < .001$), we confirm that MSPM has lower stability of DRR than ARL. The conclusions are aligned with the MD in Table 6.3 and the underwater plots in Figure 6.13, Figure 6.14,

Figure 6.15 and Figure 6.16 which illustrate the drawdowns during year 2020. It is clear in Figure 6.13 and Figure 6.14 that ARL has more frequent and intensive drawdowns for portfolio(a) compared to MSPM, but MSPM becomes the more volatile one for portfolio(b) according to Figure 6.15 and Figure 6.16. The results indicate that although MSPM achieves an outstanding performance in gaining capital returns, it does not naturally come with higher stability. However, low stability (or high risk) does not necessarily refer to danger. Since for both portfolio (a) and (b), MSPM has the highest Sortino ratios, which consider only the downside risk, MSPM's lower stability for portfolio (b) may come from a higher upside risk. In conclusion, there should be a trade-off between performance and stability, and this can be further investigated and considered in future studies.

Table 6.4: Results of statistical test on the RstdDRR of MSPM and ARL

	MSPM (n=241)		ARL (n=241)				
Portfolio	M(SD)	Normality	M(SD)	Normality	EV	U	p-value
(a)	0.031(0.019)	$6.98 * 10^{-15}$	0.034(0.020)	$3.39 * 10^{-13}$	0.324	25426.0	.005
(b)	0.049(0.027)	$1.48 * 10^{-11}$	0.032(0.022)	$1.19 * 10^{-16}$	10.070	16209.0	$< .001$

M: Mean; SD: Standard deviation; Normality: Shapiro–Wilk test; EV: Levene's test; U: U-Statistics

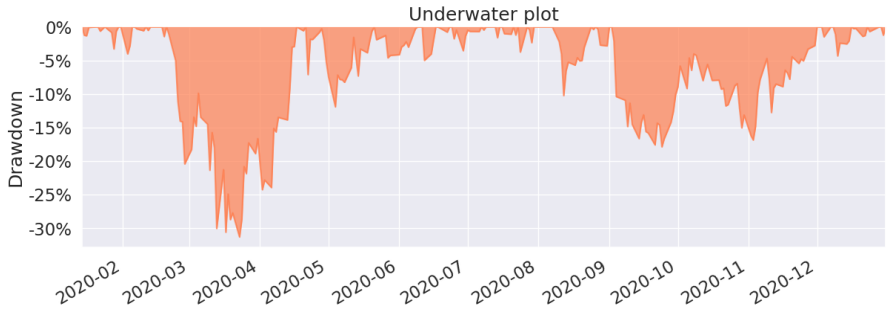


Figure 6.13: Underwater plot of MSPM for Portfolio(a).

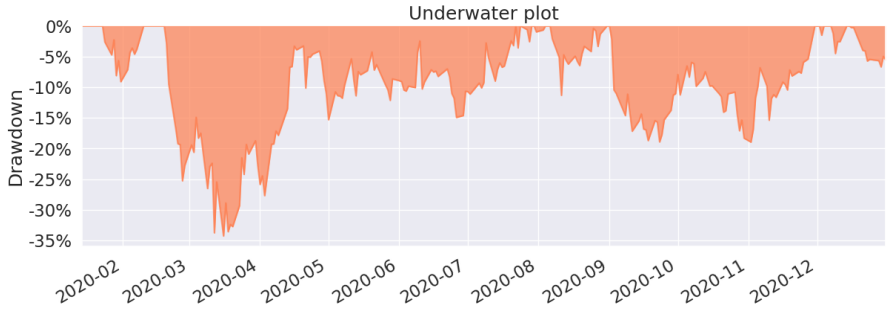


Figure 6.14: Underwater plot of ARL for Portfolio(a).

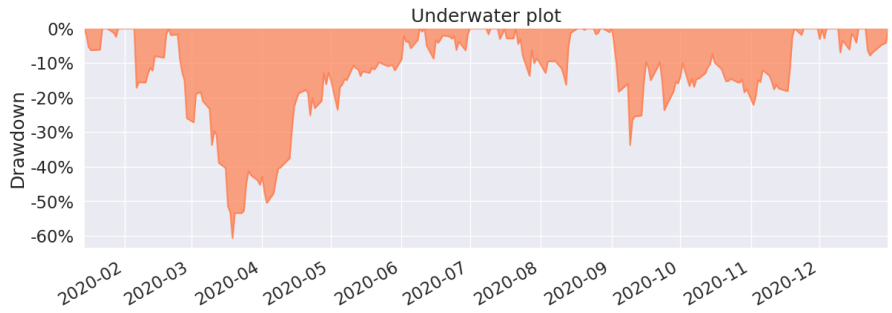


Figure 6.15: Underwater plot of MSPM for Portfolio(b).

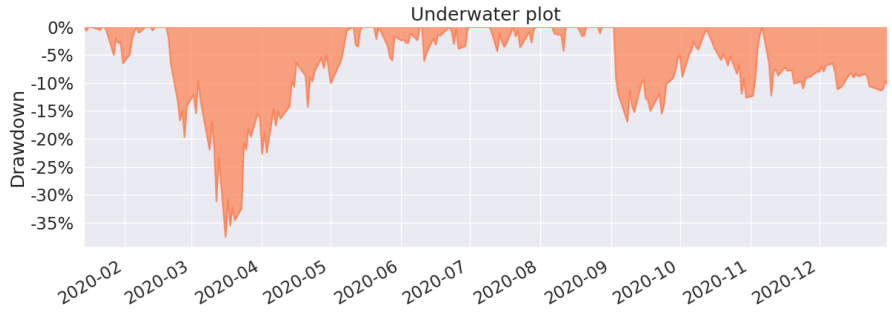


Figure 6.16: Underwater plot of ARL for Portfolio(b).

6.5.2.3 *EAM: Case study*

To better understand how EAM contributes to SAM, we illustrate the position-holding information using the signals generated by the EAMs of portfolio (a) and (b) in Figure 6.17, Figure 6.18, Figure 6.19, Figure 6.20, and Figure 6.21. The figures represent the five underlying assets: AAPL, AMD, GOOGL, NVDA, TSLA. In each plot, signals of Buying and Skipping are marked with cyan and orange circles, and the positions opened or closed are marked with either star or square symbols. The grey line is the normalized price movement. A position is opened when the first Buying signal is generated after the latest position has been closed. A position is closed when the first Closing signal is generated after a position has been opened and not yet been closed. We use dashed lines to divide different position-holding periods. If a position is profit-making based on the opening and closing prices, we color the period as light green (winning position), otherwise light red. Period of no-position will be left as blank. According to the results illustrated in the figures, the positions are opened and closed at just the right timings by the corresponding EAMs for most assets.

As shown in Table 6.5, the number of positions opened by any EAM is less than ten, and the highest is NVDA and TSLA's eight opened positions. The most profit-making EAM is TSLA, with ARR of 799%. These results exemplify the high quality and reliability of the signals generated by the EAMs. The winning rates of all the five EAMs are more than 50%. Since averaged winning rate is 80%, it indicates that even

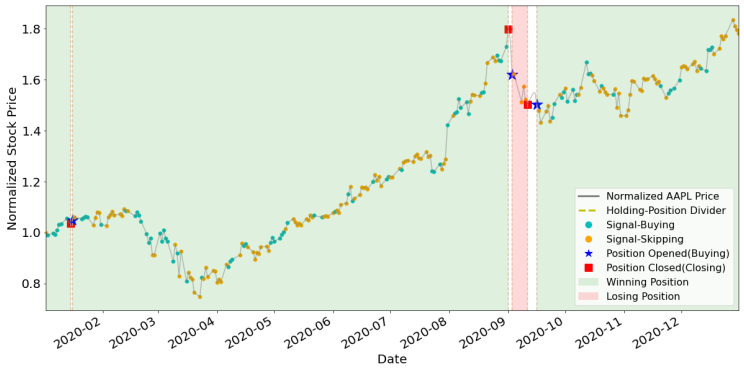


Figure 6.17: Signals and position-holding of AAPL's EAM.

Figure 6.18: Signals and position-holding of AMD's EAM.

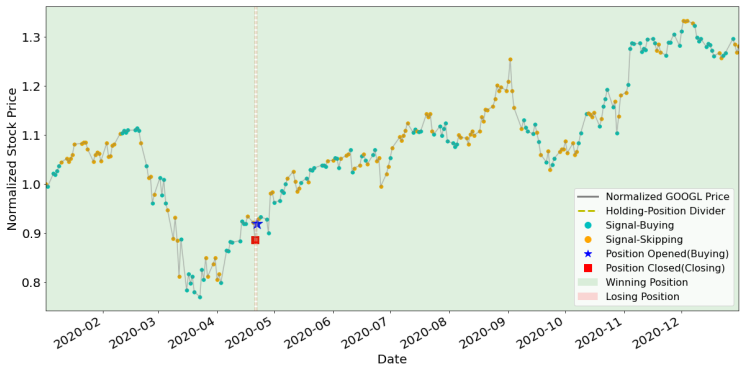


Figure 6.19: Signals and position-holding of GOOGL's EAM.

with a mediocre averaged winning rate, SAM still can efficiently utilize the information generated by the EAMs and has the outperformance

Figure 6.20: Signals and position-holding of NVDA's EAM.

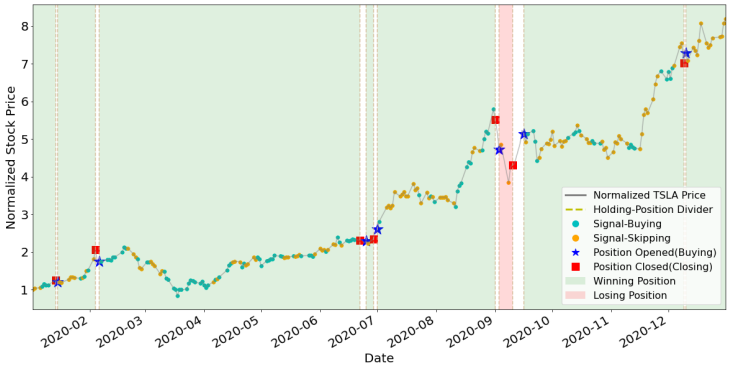


Figure 6.21: Signals and position-holding of TSLA's EAM.

compared to ARL. The results also indicate that the MSPM can perform even better if we improve the winning rate of EAMs.

Table 6.5: Statistics of EAMs' position-holding during year 2020

Asset/EAM	# of positions	# of winning positions	# of losing positions	Wining rate (%)	ARR (%)
AAPL	4	3	1	75%	124%
AMD	4	3	1	75%	125%
GOOGL	2	2	0	100%	44%
NVDA	8	5	3	62%	115%
TSLA	8	7	1	88%	799%
				Averaged: 80%	

6.5.2.4 *Validation of EAM*

As EAMs provide the trading signal-comprised information to SAMs, we intend to verify the indispensability of EAM by comparing the performance of MSPMs with and without EAMs. For this purpose, we set four different portfolios: (a), (b), (c), and (d), in which (c) and (d)

are newly introduced. Portfolio(c) consists of three stocks: Alphabet, Nvidia, and Amazon (symbol codes: [GOOGL, NVDA, AMZN]), and portfolio(d) consists of three other stocks: Nvidia, Facebook, and Microsoft (symbol codes: [NVDA, FB, MSFT]). Two MSPMs/SAMs share the same EAM for the common stocks, which are NVDA and AMZN. The initial portfolio values are still set to be 10,000. Figure 6.22 shows EAM-enabled and EAM-disabled MSPMs' the accumulated returns of different portfolios. As shown in the figure, EAM-enabled MSPMs always perform better than EAM-disabled MSPMs, and this conclusion can be reconfirmed by Table 6.6. As listed in the table, EAM-enabled MSPMs largely outperform EAM-disabled MSPMs in terms of DRR, ARR, and SR. In terms of portfolio (d), EAM-enabled MSPM achieves ARR and SR of 115.6% and 2.45, whereas EAM-disabled MSPM's ARR and SR is -5.9% and 0.01. The results validate that the SAMs can only have an ideal performance with the trading signal-comprised information from EAMs.

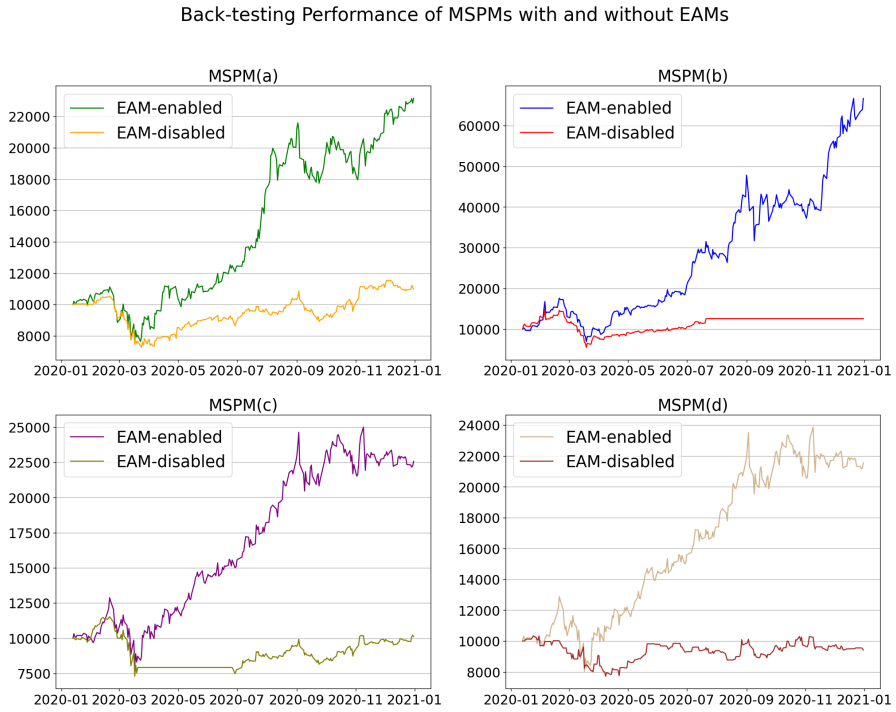


Figure 6.22: Accumulated portfolio values of MSPMs, with and without EAMs, from back-testing for portfolio (a), (b), (c) and (d). For all the four portfolios, EAM-enabled MSPMs perform significantly better than EAM-disabled MSPMs.

6.5.2.5 *Discussion on scalability and reusability of MSPM*

To address the issue of inefficient model training in RL-based PM, EAMs are designed to be independent and reusable. Once an EAM has been trained, it can be added to any SAM without retraining. For example,

Table 6.6: Comparison of back-testing performance of EAM-enabled and EAM-disabled MSPMs

	Portfolio (a)		Portfolio (b)		Portfolio (c)		Portfolio (d)	
Metric	MSPM	MSPM(w/o)	MSPM	MSPM(w/o)	MSPM	MSPM(w/o)	MSPM	MSPM(w/o)
DRR (%)	**0.404**	0.065	**0.938**	0.163	**0.403**	0.040	**0.383**	0.002
ARR (%)	**131.5**	9.8	**566.6**	25.8	**125.8**	1.1	**115.6**	-5.9
MD (%)	-31.3	**-31.0**	**-60.6**	-62.8	-37.6	**-36.7**	-37.6	**-25.3**
SR	**2.86**	0.61	**4.18**	1.00	**2.57**	0.32	**2.45**	0.01
MSPM: EAM-enabled MSPM; MSPM(w/o): EAM-disabled MSPM								

in the previous sections, portfolio(a) and portfolio(b) share one EAM in common: GOOGL, and it saves time and resources from redundant model training. On the other hand, to address the issues of ad-hoc and fixed model training in RL-based PM, MSPM allows the number of EAMs connected to any single SAM to be scaled up. In the EAM: Case study section, each EAM represents a single asset, and since these EAMs are trained, they are ready to be connected to any SAM. For example, to build a portfolio containing two assets, e.g., AAPL and TSLA, we can connect the corresponding two EAMs to an SAM to train and build the portfolio. Meanwhile, the rest of the EAMs can also be used in other portfolios. If later we want to scale up the volume of this portfolio to four assets, we simply add two more EAMs, e.g., GOOGL and NVDA, to the SAM without wasting time for training the EAMs again. Although SAM needs to be retrained once its volume is scaled up, the benefits brought by the EAMs are considerable since it has been validated in the previous section that the performance of an EAM-enabled SAM is largely improved compared to an EAM-disabled EAMs. Moreover, MSPM's scalability allows EAMs to accommodate the need for heterogeneous and alternative data input, like the sentiments data utilized in our research. As an opposite example, once a model of SARL [121] targeting a specific portfolio has been trained, this model becomes rigid and therefore, cannot be scaled or modified any more. If one or more assets in that model/portfolio is expected to be added, removed, or replaced, a new model has to be trained for the purpose. This is also the case in any other RL-based systems for PM before MSPM, such as ARL [65] and [47]. Therefore, with MSPM's scalability and reusability to create dynamic and adaptive portfolios, researchers and portfolio managers can simultaneously perform capital reallocation for various portfolios of a large volume of assets at scale by parallel computing.

6.6 LIMITATIONS AND FUTURE WORK

In this chapter, to accommodate MSPM in sequential decision-making problems of PM, we only implement DQN and PPO to formalize the agents in EAM and SAM modules. We left the implementation of other algorithms in MSPM to future studies. Additionally, the trade-off be-

tween the stability of DRR and the performance metrics (ARR, DRR, or SR) may be further considered when designing the reward functions in future studies. We only implement the historical prices and sentiments data in this research, and we plan to utilize more heterogeneous data, e.g., satellite images, in the future studies.

7

INVESTMENT BIASES AS ALTERNATIVE PERFORMANCE METRICS FOR RL-BASED PM

Due to cognitive biases and a low average degree of financial literacy, individual investors' decision-making in investment often violates financial principles, which may lead to irrational behaviors and result in capital losses. As for humans, biases are unavoidable, and utilization of algorithms to avoid biases in investment is demanded since it is expected that there should be no such issue in the case of algorithmic trading. A considerable number of successful RL-based systems for PM have been proposed, which are proven to achieve superior capital return performance. However, a high-performance system does not sufficiently indicate that it can overcome investment biases, or in order words, outperforms human investors over certain proxies of investment bias.

Hence, in this chapter, we appraise the existence and degrees of two common investment biases, disposition effect (DE) and narrow framing (NF), in MSPM with variant settings. By experiments on and comparisons among 135 portfolios of a variety of compositions, we prove this system's outperformance over human investors on the proxies of the two investment biases during the year 2021. Moreover, we demonstrate MSPM's adaptability to accommodate various RL methods for PM as a general framework by introducing and applying new settings and extensions to MSPM. The experimental results show our study as an initial step closer to an unbiased and more robust RL-based system design for PM.

The contents of this chapter are based on the following published paper:

[39] Huang, Zhenhan, and Fumihide Tanaka. "Investment Biases in Reinforcement Learning-based Financial Portfolio Management." 2022 61st Annual Conference of the Society of Instrument and Control Engineers (SICE). IEEE, 2022, pp. 494-501.

7.1 METHODOLOGY

In this section, we first describe the abstract and settings of a cutting-edge multi-agent RL-based method for PM (MSPM). Then, we further discuss the composition and construction of the portfolios which will be examined in the experiments. After that, we introduce the investment biases involved in the inspection and discussion of the experiments.

7.1.1 *Settings of MSPM*

7.1.1.1 *Module*

MSPM is a multi-agent RL-based framework for financial portfolio management [40]. There are two types of modules in MSPM system: Evolving Agent Module (EAM) and Strategic Agent Module (SAM). Figure 7.1 illustrates the overview of MSPM's architecture with EAMs and SAMs. An EAM is a signal-generating module which receives heterogeneous data input, including historical prices and news sentiments, and produces signal-comprised information for a designated asset. As a portfolio-optimizing module, an SAM reallocates the portfolio's assets by using the information from the connected EAMs. EAMs are reusable and can be combined with any given different SAMs (portfolios). To conduct the experiments in this chapter, we modify the architectures and settings of the original EAM and SAM of MPSM. Figure 7.2 illustrates the after-modified architectures of EAM and SAM, and the details will be described and discussed in the following sections.

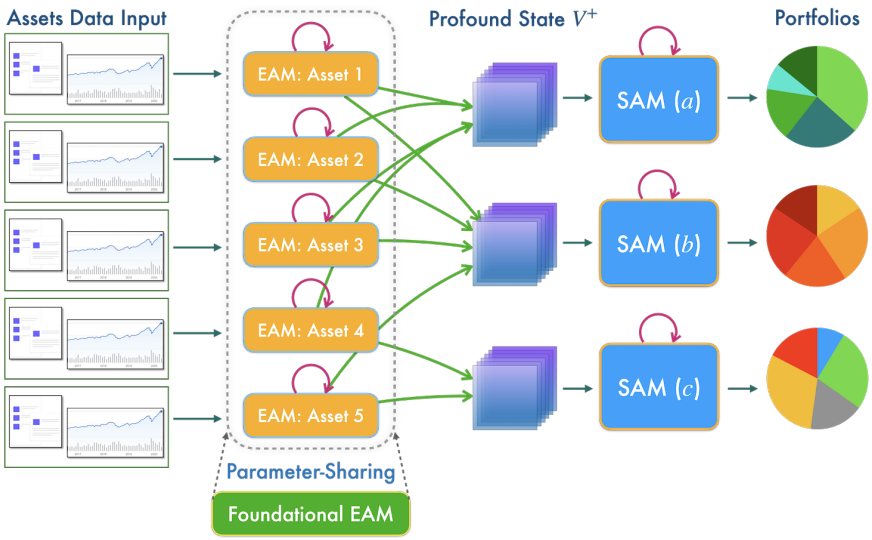


Figure 7.1: Overview of MSPM's architecture [40]

7.1.1.2 *Environment*

The historical prices and related news sentiments of the designated asset together formalize the environment which an EAM's RL-based agent interacts with. Each EAM is reusable, and periodically retrained. An EAM will be effortlessly connected and feed information to any SAM when the EAM has been trained. The environment which SAMs' RL-based agents interact with is the combination of the assets' historical prices and signals generated by the connected EAMs.

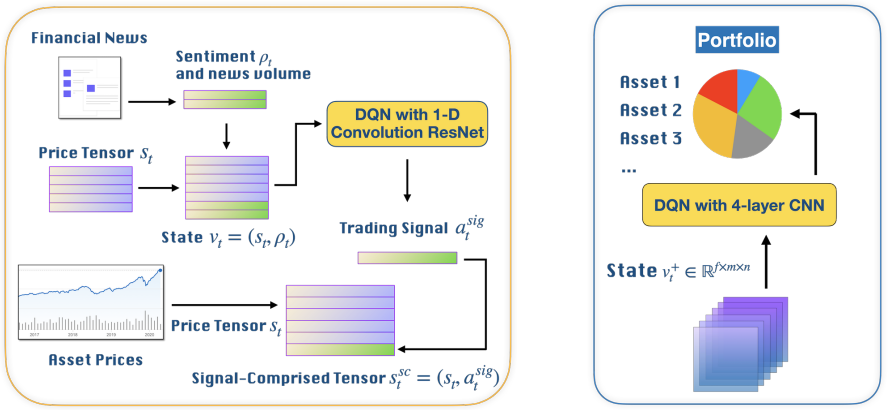


Figure 7.2: The architectures of EAM and SAM of MSPM after the modifications [40]

7.1.1.3 *State*

The state v_t which an EAM observes, contains the recent n-day prices and related news sentiments of the designated asset, at every time step t. Explicitly, $v_t = (s_t, \rho_t)$, where s_t is the n-day close, open, high and low prices and volumes (OHLCV). ρ_t involves two features: *news sentiments* and *news buzz* from FinSents data [43].
Then, an SAM connects to the EAMs of the underlying assets to reallocates the portfolio. The state v_t^+, SAM observes at time step t, is a 3-D tensor which involves the stacked new historical OHLCV s_t and the trading signals a_t^{sig} generated and provided by the EAMs of all the assets. Specifically, $v_t^+ \in \mathbb{R}^{f \times m \times n}$, where f is the number of features (OHLCV+sentiments), m is the number of assets and cash, and n stands for recent n days.

7.1.1.4 *Deep Q-network Agent*

Different from the original settings of MSPM, we utilize deep Q-network (DQN) [76] agents in EAM and SAM to interact with their environments. DQN, as a value-based method, derives a parametrized deterministic policy $\pi(\theta)$ mapping state space S to discrete action space A, with the estimate of action-value function

$$Q^{\theta}(s_t, a_t) = \mathbb{E}_{\pi_\theta}[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a] \tag{7.1}$$

on which the agent acts based, where a_t represents the action that the agent takes at time step t. For EAM, following the original settings of MSPM, we use a 1-D convolutional neural network (CNN) to represent $Q^{\theta}_{EAM}(s_t, a_t)$. For SAM, we use a plain 4-layer CNN architecture to represent its DQN agent's $Q^{\theta}_{SAM}(s_t, a_t)$.

7.1.1.5 *Action Space: EAM*

An EAM utilizes a DQN agent to choose an action from {buying, selling, or skipping} to buy and sell the asset at time step t, and this action is the asset's ***trading signal*** which will be stacked with new prices and feed to SAMs later. There is no selling action.

7.1.1.6 *Action Space: SAM*

Different from the setting in the original MSPM paper, to reallocate multiple assets for conducting experiments in this chapter, we design a new discrete action space of SAM which is the m-ary Cartesian power of the set of **discreteness** D with power m equal to the number of assets in the portfolio. Specifically,

$$D^m = \{(d_1, \dots, d_m) \mid d_i \in D \text{ for every } i \in \{1, \dots, m\}\}, \tag{7.2}$$

where only the sets with summation of at least one are kept, i.e., $\sum (d_1, \dots, d_m) \geq 1$.

For instance, if an SAM is responsible for a portfolio of three assets ($m = 3$), and the discreetness $D = 2$, then the number of qualified sets is $D^m - 1 = 2^3 - 1 = 7$, including $\{(0,0,1), (0,1,0), \dots, (1,1,1)\}$, excluding $(0,0,0)$ which cannot formalize a valid reallocation weight of portfolio.

Eventually, the reallocation weight of the portfolio at time step t is derived as

$$w_t = \frac{d_t}{\sum d_t}, \tag{7.3}$$

and w_t represents the **voted weight**. Discreteness D is a hyper-parameter which can be further tuned, and $D = 2$ is set for the experiments in this chapter.

7.1.1.7 *Reward functions*

We follow the original settings of the reward functions in MSPM. The reward r_t which EAM received at time step t is:

$$r_t(s_t, \iota_t) = \begin{cases} 100(\sum_{i=t_\iota}^{t} \frac{v_t^{(close)}}{v_{t-1}^{(close)}} - 1 - \beta), \text{ if } \iota_t \\ 0, \text{ if not } \iota_t \end{cases} \tag{7.4}$$

where $v_t^{(close)}$ is asset's close price at time step t. t_l indicates the starting time step of an opened long position, and $\beta = 0.0025$ is the commission rate. ι_t indicates if the long position is still open. The reward function of SAM also follows the original setting of MSPM, and therefore its description will be skipped.

7.1.2 *Portfolios*

135 different portfolios are built for the experiments. The compositions of stocks in the portfolios are obtained by sampling from 18 different stocks without replacement. Each portfolio consists of 2, 3, or 4 different stocks, and no two portfolios have the same stock composition. Moreover, SAMs are constructed separately instead of jointly. In other words, each SAM reallocates only one asset with cash (a risk-free asset), and the number of SAMs in a portfolio equals the number of assets of that portfolio. With this setting, it is guaranteed that the decision-making of MSPM on every individual asset is consistent and independent. The 18 stocks are large-cap stocks from 6 different sectors in U.S. stock market. The 18 stocks are: [AAPL ABT AMD BAC COST CRM DIS JNJ JPM KO NFLX NKE NVDA PFE PG TSLA V WMT]. Table 7.1 depicts the 16 (out 18) assets and the composition of the first 6 portfolios. Figure 7.3 displays the stock symbols and the number of portfolios which each stock is in. Among the stocks, NVDA appears in 17 portfolios, the lowest number of portfolios, and BAC appears in 31 portfolios, the highest number of portfolios.

Table 7.1: Composition of the first six portfolios

Portfolio	AAPL	ABT	AMD	BAC	COST	CRM	DIS	JNJ	JPM	KO	NFLX	NVDA	PFE	TSLA	V	WMT
(a)			*									*				
(b)	*									*		*				
(c)			*						*				*			
(d)						*					*					*
(e)				*	*									*	*	
(f)		*	*				*	*								

Diversification of portfolios: As shown in Figure 7.4, there are 6 different sectors which the 18 stocks exclusively belong to: Technology, Consumer Defensive, Financial Services, Healthcare, Communication Services, and Consumer Cyclical. Since the only investment vehicle (asset) discussed in this chapter is stock, it is expected that the stocks in each portfolio are from various sectors so that the overall risk of that portfolio can be more or less diversified [54, 73]. Therefore, we define four diversification levels of the portfolios: **fully-diversified, well-diversified, semi-diversified and undiversified**. Specifically, for a 2-assets portfolio, it is fully-diversified if both assets are from different sectors, otherwise it is undiversified. For the case of 2-assets portfolios, there shall be no well-diversified nor semi-diversified portfolio. For a 3-assets portfolio, it is fully-diversified if all three assets are from different sectors, well-diversified if the portfolio has two different sectors, or undiversified if there is only one sector. For the case of 3-assets portfolios, there shall be no semi-diversified portfolio. Finally, for a 4-assets portfolio, it is fully-diversified if all assets are from different

sectors, well-diversified if the portfolio has three different sectors, semi-diversified if it has two different sectors, or undiversified if only one sector. Table 7.2 provides the detailed categorization of portfolios' diversification levels. Table 7.3 displays the statistics of the portfolios of different number of assets (stocks) by their diversification levels. In addition, Figure 7.5 shows the overall distribution of the diversification levels in terms of their portfolios, and 124 out of a total 135 (nearly 92%) portfolios are at least well-diversified. There are merely 7 undiversified portfolios in total, and all of these undiversified portfolios are 2-assets portfolios (in which both stocks are from the same sector). Thus, according to their distribution of sectors, nearly 92% of the portfolios to be inspected in the experiments are diversified at certain levels as expected.

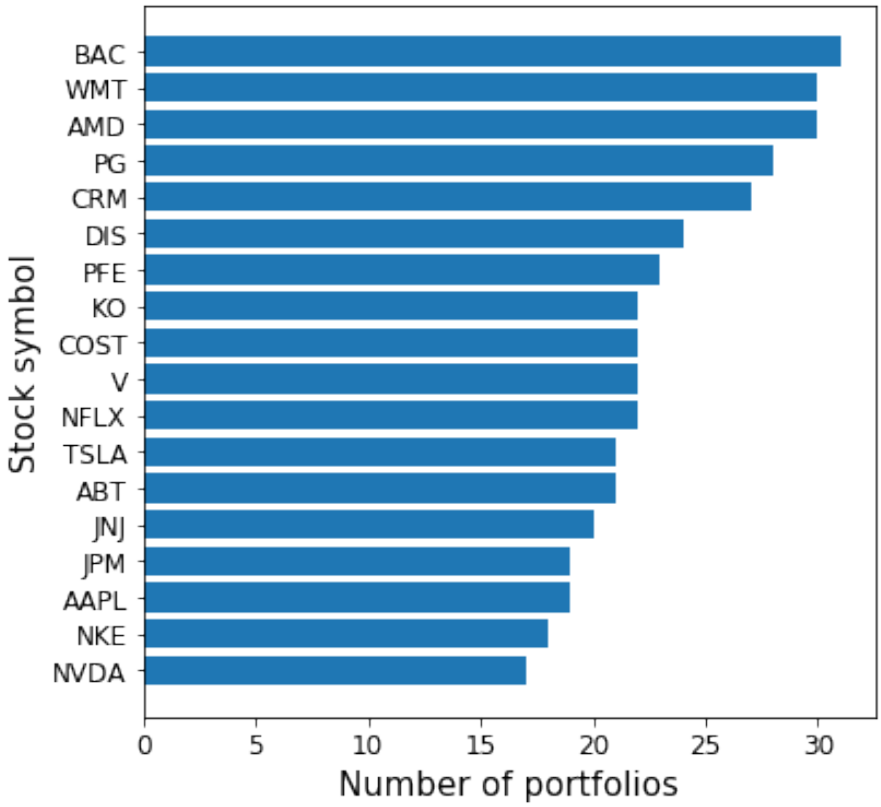


Figure 7.3: Distribution of the stocks in terms of the number of portfolios in descending order

Table 7.2: Categorization of portfolios' diversification levels

	Diversification Level			
Portfolio Type	Fully-diversified	Well-diversified	Semi-diversified	Undiversified
2-assets	assets from 2 sectors	-	-	assets from same sector
3-assets	assets from 3 sectors	assets from 2 sectors	-	same as above
4-assets	assets from 4 sectors	assets from 3 sectors	assets from 2 sectors	same as above

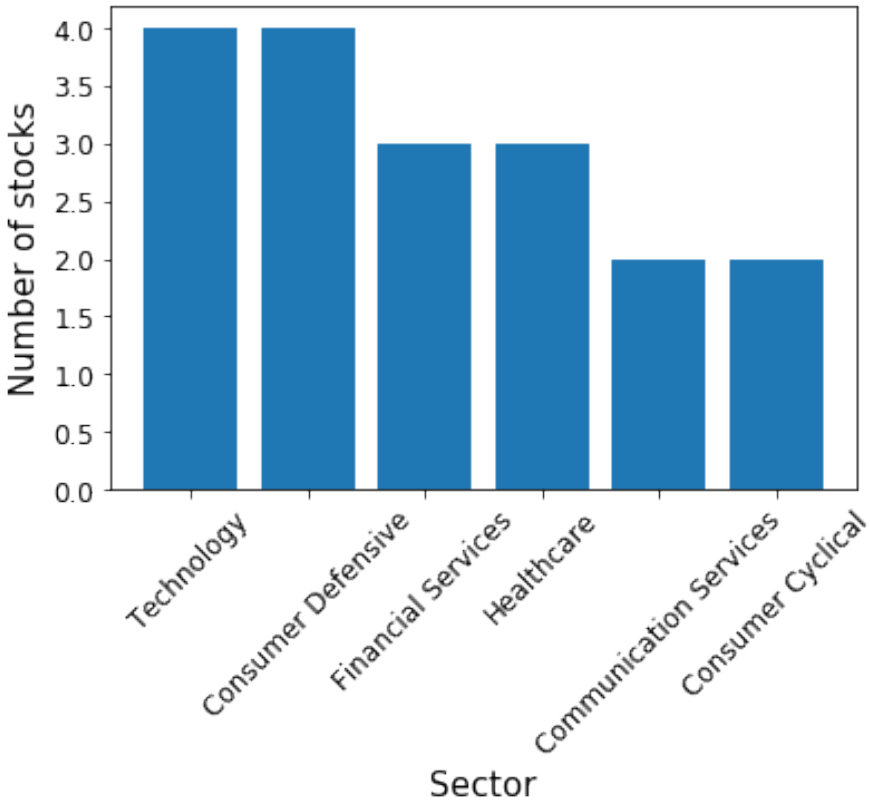


Figure 7.4: Distribution of the stocks' sectors in descending order

Table 7.3: Statistics of portfolios by the number of assets(stocks) and diversification levels

	Diversification Level			
Portfolio Type	Fully-diversified	Well-diversified	Semi-diversified	Undiversified
2-assets (n=38)	82% (n=31)	-	-	18% (n=7)
3-assets (n=48)	65% (n=31)	35% (n=17)	-	0% (n=0)
4-assets (n=49)	39% (n=19)	53% (n=26)	8% (n=4)	0% (n=0)

7.1.3 *Bias proxies*

In this chapter, we focus on two investment biases: disposition effect (DE) and narrow framing (NF). As the stocks are randomly sampled for creating the portfolios, the biases related to the selection of assets, e.g., lottery stock preference, will not be inspected or discussed. The description and proxies of the investment biases are listed as the following:

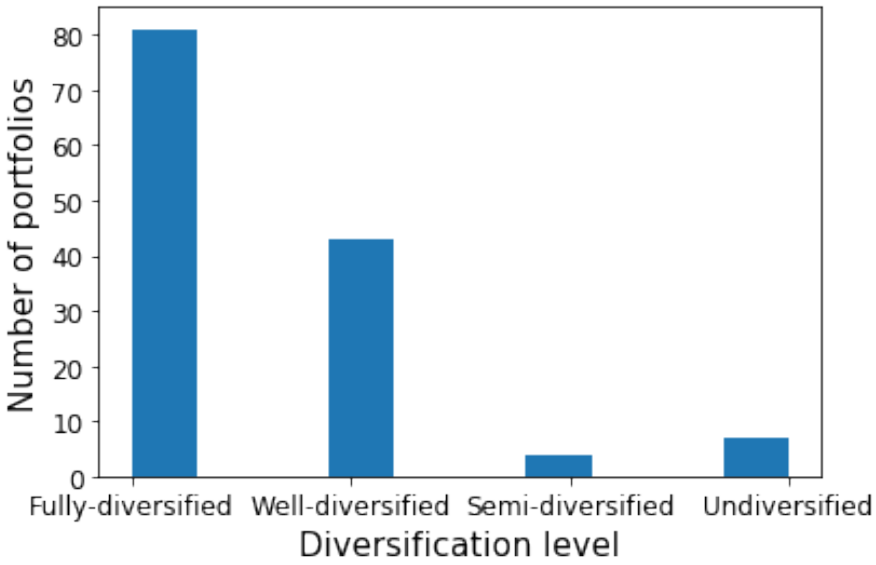


Figure 7.5: Distribution of the diversification levels in terms of the number of portfolios

- Disposition effect (DE) [58, 83, 99, 110] is about investors' tendency to realize profits too soon and keep losses too long. The proxy of disposition effect (DE) is measured by the different between the proportion of gains realized (PGR) and the proportion of losses realized (PLR). Specifically,

$$\mathrm{DE} = \mathrm{PGR} - \mathrm{PLR}, \tag{7.5}$$

where

$$\mathrm{PGR} = \frac{\mathrm{RLG}}{(\mathrm{RLG} + \mathrm{PG})}, \mathrm{PLR} = \frac{\mathrm{RLL}}{(\mathrm{RLL} + \mathrm{PL})}.$$

RLG is the number of trades with gains realized, RLL is the number of trades with losses realized, PG is the number of trades with gains on paper, and PL is the number of trades with losses on paper. A positive DE indicates that the investor has the disposition effect as the proportion of profits realized is larger than the proportion of losses realized. Therefore, it is expected that MSPM has a low or negative DE, indicating a low degree of or no disposition effect.

- Narrow framing (NF) [48, 50, 58] is about investors' tendency to make isolated and sub-optimal decisions and to trade assets without considering the holistic picture of their portfolios. The proxy of narrow framing is measured by trade cluster (TC). Specifically,

$$\mathrm{TC} = 1 - \frac{\mathrm{TDAY}}{\mathrm{TRADE}} \tag{7.6}$$

TDAY is the number of days when the investor trades assets, and TRADE is the number of trades the investor executes in those days. A low TC often indicates a high degree of narrow framing since the investor tends to execute trades separately instead of collectively. It is expected that MSPM has a high proxy of NF (TC) indicating the inclination of trading collectively, i.e., a low degree of NF.

7.2 EXPERIMENT

We build 135 portfolios in which the stocks are randomly sampled from the pool consists 18 different stocks. We train and backtest MSPM using the historical data of prices, news sentiments and trading signals. Then, we calculate and discuss the biases proxies from the reallocation records of the portfolios managed by MSPM. After that, we perform statistical tests to compare the degrees of investment biases between MSPM and human investors. Since the goal of the study is to inspect the existence and degrees of the investment biases, the performance in terms of capital return will not be discussed.

7.2.1 *Data Ranges*

Table 7.4 lists the ranges of the datasets. EAM(training) data include the historical daily OHLCV data (s_t) from [87], news sentiments and buzz (ρ_t) and EAM-generated trading signals (a_t^{sig}) for the stocks in the portfolios. Among them, the news sentiments and buzz are from FinSents data [43] which contain the news media and social networks, and range continuously from -5.0 (most bearish) to 5.0 (most bullish). For the purpose of generating trading signal-comprised data for SAM/MSPMs, EAM(predicting) data range from January 2016 to December 2021, and have the same data structure as EAM(training). Once obtained, the signal-comprised data will then be split into three subsets, for training, validating and back-testing purposes, respectively.

7.2.2 *Results and Discussion*

We backtest the 135 portfolios and compare the bias proxies of MSPM with human investors. The summary statistics of the portfolios are described and discussed in this section.

Table 7.4: Description of Data Ranges

Purpose	Range
EAM(training)	2009 Jan~2015 Dec
EAM(predicting)	2016 Jan~2021 Dec
SAM/MSPM(training)	2016 Jan~2019 Dec
SAM/MSPM(validation)	2020 Jan~2020 Dec
SAM/MSPM(backtesting)	2021 Jan~2021 Dec

7.2.2.1 *Proxies*

The bias proxies of human investors for comparison in this chapter are from [3]. Table 7.5 reveals the statistics of the bias proxies. The average DE of MSPM across all portfolios is -0.1477 (SD=0.1435), which is negative and much less than the average DE of human investors (HM), which is 0.0372 (SD=1.1220), indicating that MSPM averagely has a lower degree of disposition effect than human investors. On the other hand, MSPM's average proxy of NF, across all portfolios, is 0.0904 (SD=0.0725), which is higher than that of human investors (HM), which is 0.0100 (SD=0.1550), indicating that MSPM averagely has a lower degree of narrow framing than human investors. In addition, Table 7.6 shows that the portfolios across all diversification levels have lower degrees of disposition effect and narrow framing than human investors. Figure 7.6 and Figure 7.7 reveal more details of DE. In Figure 7.6, we observe that the average monthly DE of MSPM across all portfolios is lower than that of human investors (HM), except in January, February, and August. If we take a look at each portfolio separately, which is depicted in Figure 7.7, the monthly DE of the first six portfolios are below the average DE of human investors (HM) for the most of the time. The same conclusion can also be obtained from Figure 7.8 which illustrates the monthly DE of the portfolios of different diversification levels from Jan to Dec 2021.

Table 7.5: Summary statistics of the bias proxies

	DE		NF	
	M	SD	M	SD
MSPM (n=135)	**-0.1477**	0.1435	**0.0904**	0.0725
HM (n=21542)	0.0372	1.1220	0.0100	0.1550
M: mean; SD: standard deviation				

Table 7.6: Bias proxies of MSPM by portfolios' diversification levels

	DE		NF	
Diversification Level	M	SD	M	SD
Fully-diversified (n=81)	-0.1270	0.1436	0.0719	0.0648
Well-diversified (n=43)	-0.1880	0.1381	0.1337	0.0701
Semi-diversified (n=4)	-0.1578	0.1951	0.0457	0.0383
Undiversified (n=7)	-0.1260	0.1271	0.0512	0.0562
M: mean; SD: standard deviation				

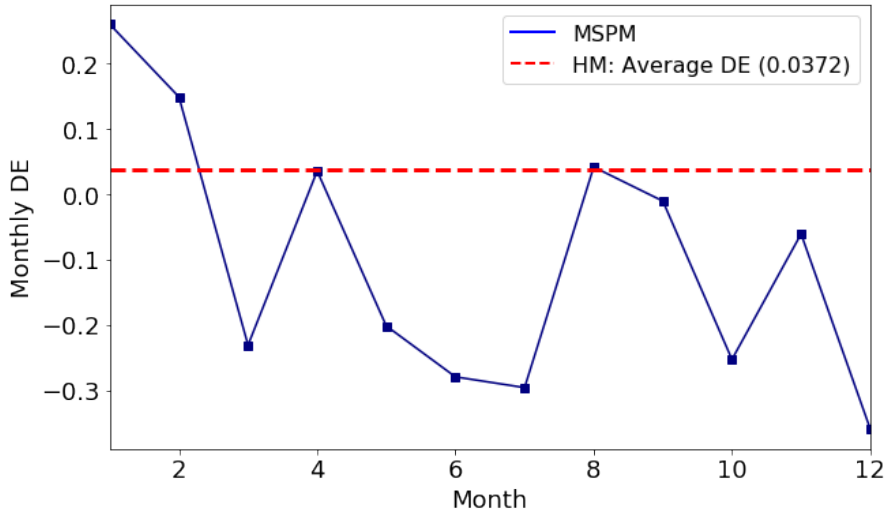


Figure 7.6: Monthly DE of portfolios from Jan to Dec 2021

7.2.2.2 *Statistical Tests*

Since the statistics of the bias proxies have been obtained, we perform the unpaired t-tests comparing the proxies between MSPM and human investors (HM). For both DE and NF, hypothesis H_0 is that MSPM has a lower degree of investment bias than human investors: the group mean of DE of MSPM is lower than DE of human investors, and the group mean of NF of MSPM is higher than or equal to NF of human investors. The alternative hypothesis H_a is that MSPM has a higher degree of investment bias than human investors: the group mean of DE of MSPM is higher than or equal to DE of human investors, and the group mean of NF of MSPM is less than or equal to NF of human investors. The significance level is set to be .05. We reject H_0 and accept H_a if the p-value is less than 0.05; otherwise, we accept H_0. Specifically:

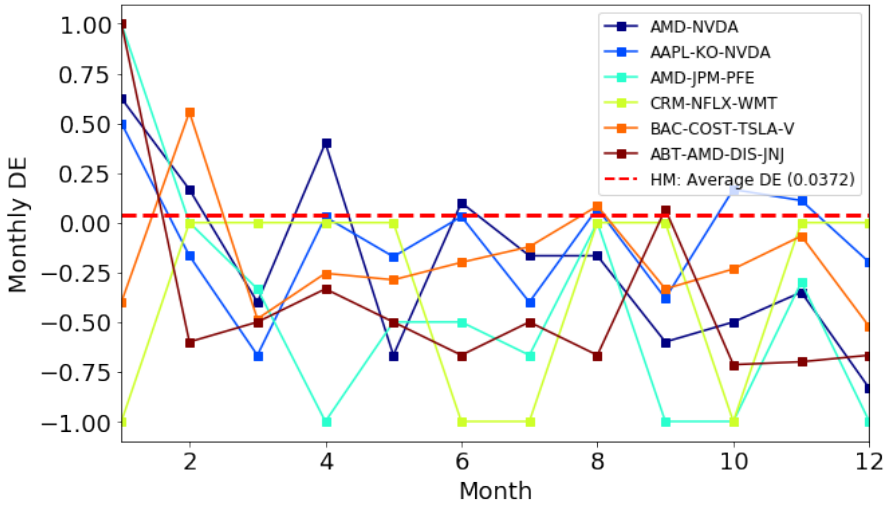


Figure 7.7: Monthly DE of the first six portfolios from Jan to Dec 2021

- Statistical test: unpaired t-test
- DE: null hypothesis $H_0^{DE} : \mu_{MSPM}^{DE} - \mu_{HM}^{DE} < 0$
- DE: alt hypothesis $H_a^{DE} : \mu_{MSPM}^{DE} - \mu_{HM}^{DE} \geq 0$
- NF: null hypothesis $H_0^{NF} : \mu_{MSPM}^{NF} - \mu_{HM}^{NF} > 0$
- NF: alt hypothesis $H_a^{NF} : \mu_{MSPM}^{NF} - \mu_{HM}^{NF} \leq 0$

As the results shown in Table 7.7, by rejecting H_0^{DE} and accepting H_a^{DE} $(t(21675) = -1.9146, p - value = .028)$, and rejecting H_0^{NF} and accepting H_a^{NF} $(t(21675) = 6.0226, p - value < .001)$, it is confirmed that MSPM has significantly lower degrees of both DE and NF than human investors. Therefore, as an RL-based system for PM, MSPM is proved to be able to overcomes the two biases which human investors often have in investing.

Table 7.7: Statistical tests on the proxies of disposition effect (DE) and narrow framing (NF)

	MSPM (n=135)	HM (n=21542)		
Proxy	M(SD)	M(SD)	t(21675)	p-value
DE	-0.1477(0.1435)	0.0372(1.1220)	-1.9146	0.028
NF	0.0904(0.0725)	0.0100(0.1550)	6.0226	<0.001
M: mean; SD: standard deviation;				

Since diversification is one important factor in conventional PM, we also want to inspect the DE and NF among portfolios that MSPM reallocates under different diversification levels. Due to the limited number of portfolios under each diversification level, to obtain a statistically

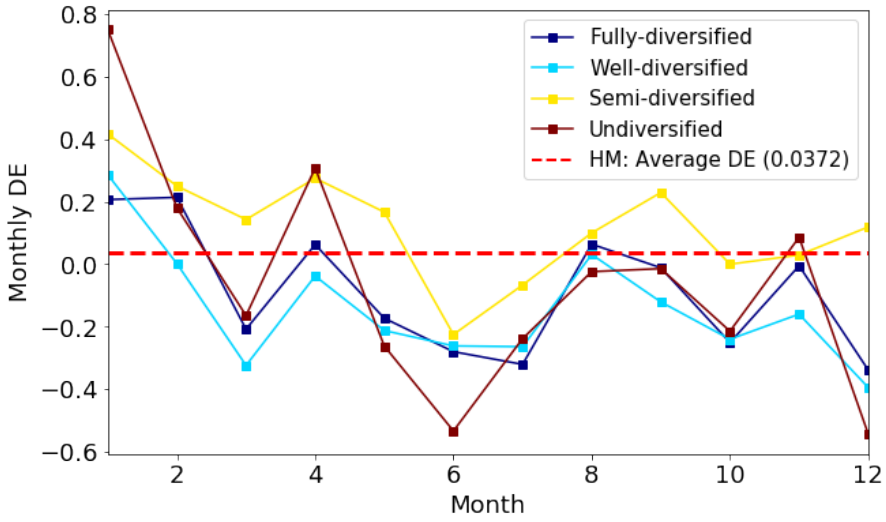


Figure 7.8: Monthly DE of the portfolios of different diversification levels from Jan to Dec 2021

valid result, the portfolios being tested are fully-diversified and well-diversified, which involve 81 and 43 portfolios, respectively. The hypotheses being tested are similar to those in the previous tests, but are two-tailed because we are more interested in examining if DE or NF is different between the two diversification levels under the management by MSPM. Specifically:

- Statistical test: unpaired t-test
- DE: null hypothesis $H_0^{DE} : \mu_{FD}^{DE} - \mu_{WD}^{DE} = 0$
- DE: alt hypothesis $H_a^{DE} : \mu_{FD}^{DE} - \mu_{WD}^{DE} \neq 0$
- NF: null hypothesis $H_0^{NF} : \mu_{FD}^{NF} - \mu_{WD}^{NF} = 0$
- NF: alt hypothesis $H_a^{NF} : \mu_{FD}^{NF} - \mu_{WD}^{NF} \neq 0$

The subscripts FD and WD represent fully-diversified and well-diversified portfolios. As the results presented in Table 7.8, well-diversified portfolios have significantly lower degrees of both DE and NF than fully-diversified portfolios by rejecting H_0^{DE} and accepting H_a^{DE} ($t(122) = 2.2810, p\text{−}value = .024$), and rejecting H_0^{NF} and accepting H_a^{NF} ($t(122) = -4.9126, p - value =< 0.001$). It is an intuitively interesting result since portfolios with a higher diversification level are supposed to have lower degrees of investment biases than portfolios with a lower diversification level. Further probe into the cause will be implemented in future studies.

7.2.2.3 *Case Study: AAPL and NFLX*

To provide a more intuitive interpretation of how MSPM reallocates stocks in portfolios, Figure 7.9 depicts the decision-making of MSPM

Table 7.8: Statistical tests on the bias proxies in the portfolios of two types of diversifications (MSPM)

	FD (n=81)	WD (n=43)		
Proxy	M(SD)	M(SD)	t(122)	p-value
DE	-0.1270(0.1436)	-0.1880(0.1381)	2.2810	0.024
NF	0.0719(0.0648)	0.1337(0.0701)	-4.9126	<0.001
FD: fully-diversified; WD: well-diversified;				

for rebalancing two individual stocks, AAPL and NFLX, in a portfolio. Table 7.9 provides the detailed statistics behind the DE of the portfolio of AAPL-NFLX. As shown in the Figure 7.9, the green mark B indicates buying (a position is opened) and the red mark S indicates selling (an opened position is now closed). Whenever a loss happens, MSPM does not hesitate to sell, and when there is a gain, MSPM keeps the holding with discretion. The line in yellow indicates the monthly DE of the portfolios of AAPL and NFLX, which is also lower than the human investors' average DE, except only for August. Table 7.9 clearly displays that the proportion of gains realized (PGR) is much lower than the proportion of losses realized (PLR), which leads to a low and negative DE. This result aligns with MSPM's low degree of DE: -0.2893 on this portfolio and -0.1477 averagely on all portfolios, evidencing that MSPM overcomes the disposition effect. This case study may further help to reveal the internal mechanism that drives RL-based systems to overcome the investment biases in future studies.

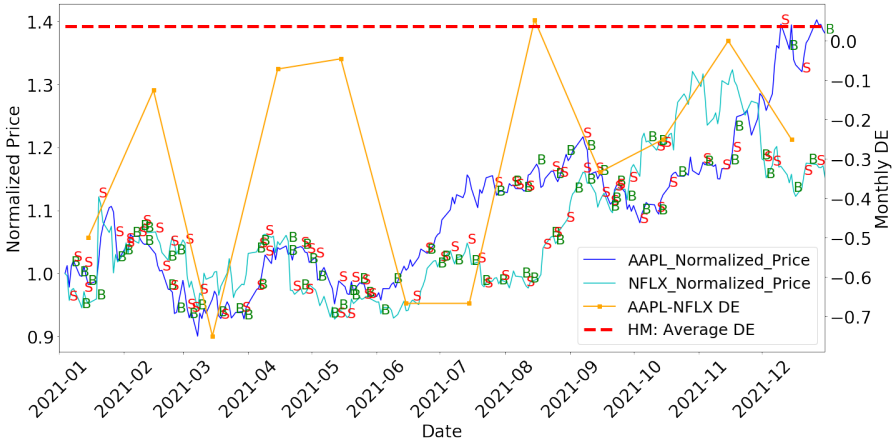


Figure 7.9: Decision-making and monthly DE of AAPL-NFLX by MSPM

7.3 LIMITATIONS AND FUTURE WORK

As various biases exist in the investing behaviors of human investors, we plan to investigate more bias proxies. We also want to recruit partici-

Table 7.9: Detailed statistics behind AAPL-NFLX's DE (MSPM)

Gain		Loss	
Realized	Paper	Realized	Paper
29	45	47	22
PGR: 0.3919		PLR: 0.6812	
DE = PGR - PLR = -0.2893			

pants in future studies to obtain human investors' data with higher granularity so that the inspection of human investors' monthly or weekly decision-making will be viable. Moreover, it becomes natural to ask if an RL-based PM system has any undiscovered unique biases. On the other hand, it is worth probing into the detailed mechanism that drives the system to overcome the biases. It is expected that other existing RL-based PM systems can overcome the investment biases as well, and we plan to investigate and provide comparison experiments with the related RL methods in future studies. Additionally, the relationship between portfolios' diversification levels and degrees of bias proxies will be further examined.

8

INVESTMENT BIASES OF CRYPTO TRADERS ON ETHEREUM NETWORK: AN EMPIRICAL STUDY

In Chapter 8, we turn the spotlight to cryptocurrencies and investors' behavioral biases when trading cryptocurrencies.

Cryptocurrency (crypto), as an emerging financial vehicle, is alluring investors globally. However, attributable to an ambiguousness of fundamental values and, to a great extent, being driven by investors' sentiments flowing across social media platforms like Twitter.com and Telegram, cryptocurrencies are deemed highly volatile assets. Consequently, researchers are attracted to investigate investors' behavioral biases in investing in cryptocurrency.

So far, existing research majorly focuses on the investigation with the implementation of questionnaires and surveys. However, to what extent the feedbacks to these questionnaires or surveys truthfully reflect the investors' actual practices in investing in cryptocurrencies is uncertain and dubious.

In contrast to stock or other conventional financial assets for which investors' trading records are barely accessible from brokerages, investors' records of cryptocurrency trading can be effortlessly obtained owing to the openness and transparency nature of decentralized blockchains.

Among the existing blockchains, Ethereum is the second largest blockchain platform after Bitcoin. According to Glassnode.com [33], as of 08/31/2022, there are more than 400,000 active wallet addresses (addresses mapped to individual investors or trading bots in a non-injective surjective manner) with their trading information publicly updated in real-time in Ethereum. By examining the data from the publicly viewable on-chain wallet addresses, we can possess direct information regarding investors' behaviors in trading cryptocurrencies without struggles to recruit ideal participants or the likelihood of receiving inaccurate feedback from the surveys distributed.

Yet, few existing research studies the behavioral traits of cryptocurrency traders by inspecting information directly from on-chain wallets. Thus, this chapter analyzes how cryptocurrency investors behave when investing in the Ethereum network. Specifically, we investigate and evaluate cryptocurrency investors' particular behavioral bias proxies by utilizing the on-chain information of wallet addresses directly from Ethereum. In this chapter, we aim to reveal the degrees of behavioral bias proxies of cryptocurrency investors in their investment decision-making.

By analyzing more than 1,086 unique wallet addresses and related transactions, we have obtained three behavioral bias proxies of the in-

vestors behind the wallets and four different properties of the wallets. We find that wealthier investors tend to trade more often and appear to be more confident. We also find that the higher the trading frequencies of the investors, the less the returns from their transactions. Furthermore, we distinguish and analyze the wallets of human investors and trading bots. The statistical tests' results indicate the significant differences between human investors and trading bots on all behavioral biases and wallet properties.

It is the first time the behavioral biases of cryptocurrency investors are fully revealed directly through on-chain records, without using any inaccessible nor indirect data sources like centralized exchange databases or questionnaires/surveys.

The contents of this chapter are based on the following preprint:

[38] Huang, Zhenhan, and Fumihide Tanaka. "Behavioral Biases of Cryptocurrency Investors." Available at SSRN 4280610 (2022).

8.1 METHODOLOGY

This section begins with how the data of wallet addresses and trading records are obtained and processed by proposing and establishing a pipeline for the data acquisition and pricing-matching of on-chain wallet records. Then, the proxies of the behavioral biases and the portfolio properties to be appraised and inspected in this chapter will be introduced and described.

8.1.1 *Data Acquisition Pipeline*

8.1.1.1 *Wallet Addresses*

We select the wallet addresses to be obtained and inspected in this chapter by considering the following two facts: 1. We want to ensure the statistics and experimental results are as current and informative as possible; 2. Further, we target and inspect the trading activities and investment biases of the investor behind each wallet address. Thus, we focus on the wallet addresses recently actively trading at decentralized exchanges. Etherscan provides DEX Tracker [23], a tool for tracking the latest 10,000 trading transactions across multiple decentralized exchanges (DEXs). These transactions can also be exported and downloaded in .CSV format. However, since practically no more than 5,000 transactions starting from the date chosen can be exported due to the restriction set by Etherscan, we retrieve the first 5,000 daily transactions from September 01, 2022, to September 15, 2022, resulting in 75,000 trading transactions in total. Although each transaction comes with a unique transaction hash (Txn Hash), there are transac-

tions with duplicated transaction hash since one single transaction may involve multiple ERC-20 tokens as well as multiple DEXs if the transaction occurred at a DEX aggregator. Therefore, for the transactions with duplicated transaction hash, we keep only the first transaction. After that, each of these transactions is matched with the wallet (an investor who initiated the transaction) and contract addresses (DEX where the transaction occurred).

8.1.1.2 *Price-matched Transactions*

We assume active DEX investors are represented by the wallet addresses obtained, and we fetch recent 500 transactions from each of the unique wallet addresses obtained. Those transactions that are not swap transactions or have not occurred at DEXs are filtered out. Once the recent transactions of each wallet address are fetched, we match the token pairs swapped in each transaction with the trading prices.

Since not every ERC-20 token is with high liquidity for its prices to be easily retrieved, to possibly match the valid trading prices, we retrieve token prices with the finest granularity for the time the transaction occurred by implementing three different data sources: CoinAPI.io, CoinGecko.com, and Binance.com [8, 12, 13]. Since these three sources can cover the majority of ERC-20 tokens, a small number of swap transactions, which have non-retrievable tokens, are removed from the trading history of the affiliated wallets. Figure 8.1 illustrates the data acquisition and price-matching process.

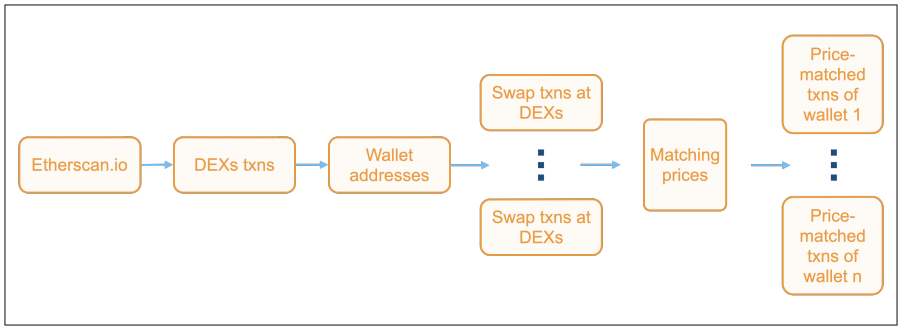


Figure 8.1: Pipeline of the process of wallet data acquisition and price-matching

8.1.2 *Bias Proxies and Wallet Properties*

In this chapter, we inspect three behavioral biases: disposition effect (DE), narrow framing (NF), and overconfidence (OC), along with three different wallet (portfolio) properties: size (SIZE), performance (PERF), and risk (RISK) and average number of daily trades (DTrades). The description and proxies of the biases are listed as the following:

- **Disposition effect (DE)** [58, 83, 99, 110] is about investors' tendency to realize profits too soon and keep losses too long. The proxy of disposition effect (DE) is measured by the difference between the proportions of gains realized and losses realized. A positive DE indicates that the investor has the disposition effect as the proportion of profits realized is larger than the proportion of losses realized. The greater the DE, the greater the degree of the disposition effect. On the other hand, a negative DE indicates no disposition effect.

- **Narrow framing (-NF)** [48, 50, 58] relates to investors' tendency to make isolated and sub-optimal decisions and to trade assets without considering the holistic picture of their portfolios. The proxy of narrow framing (i.e., trade cluster, TC, or NF) is measured by the difference between 1 and the division of the number of trading days and the number of trades made during the same period. In this chapter, negative NF (-NF) is taken for the appraisal. The higher the -NF, the greater the degree of narrow framing as the investor more likely tends to execute trades separately instead of collectively.

- **Overconfidence (OC)** [84] is about an investor's tendency to trade frequently but unsuccessfully. In this chapter, the proxy of overconfidence (OC) of an investor equals 1 (indicating the existence of overconfidence) if the investor is in the highest portfolio turnover decile and in the lowest performance decile and 0 otherwise (indicating nonexistence of overconfidence).

- **Size (SIZE)** is the average of the asset under management (AUM) of an investor's wallet (portfolio) in U.S. Dollar (USD) between the first and last day of the transaction period.

- **Performance (PERF)** is about the average of an investor's relative realized returns specified to each transaction, as the capitalization of the wallet (portfolio) varies indefinitely due to token transfers during the transaction period.

- **Risk (RISK)** is about the standard deviation of an investor's relative realized returns during the transaction period.

- **Daily Trades (DTrades)** relate to the average number of daily trades made by a wallet.

8.2 EMPIRICAL RESULTS

By following the data acquisition pipeline, prices and trading positions are matched with the transactions fetched for each wallet address,

resulting in the complete data of 1,086 unique wallets and their transactions. The dates of the transaction history across the 1,086 wallets range from June 2020 to Sept 2022.

For each wallet, we quantify the bias proxies based on historical transactions and wallet properties. Then, we inspect and discuss the relationship among the bias proxies and portfolio properties. After that, we conduct statistical tests to compare the decision-making in investing between two groups: human investors and trading bots.

8.2.0.1 *Wallets and Transactions*

Figure 8.2 is the distribution of trading hours in a day in Pacific time (PST). From the plot, it can be assumed that most transactions occurred in line with the work schedule in PST, indicating that most of these investors are living in the western region of North America.

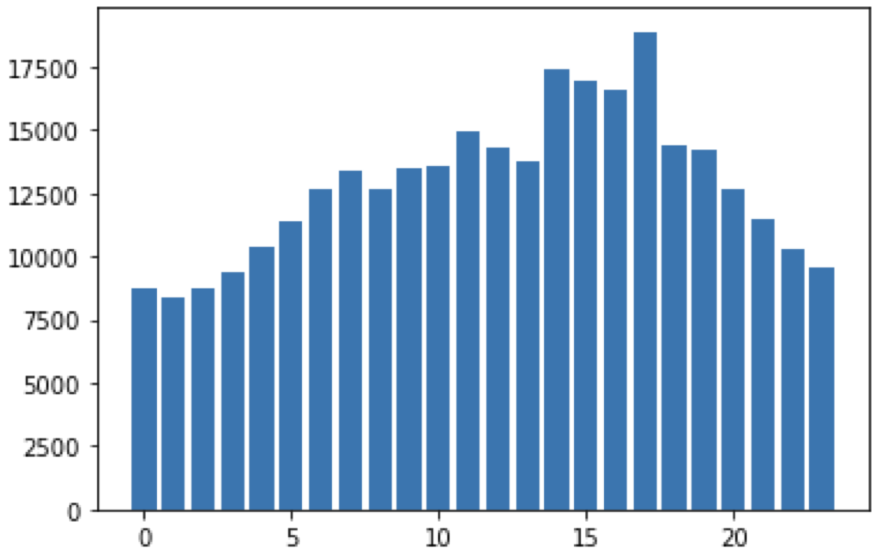


Figure 8.2: Distribution of investors' daily trading hours (ranging from 00:00 to 24:00) in Pacific time (PST)

Figure 8.3 shows the distribution of average daily transactions. Fortunately, the average daily transactions of most wallets fall into the range of below 50 trades, indicating that the controllers behind the majority of these wallets are more like to be human investors instead of trading bots.

8.2.1 *Proxies and Properties*

There are 952 unique wallets after the proxies and properties are calculated. The summary statistics of the bias proxies and wallet properties are revealed in Table 8.1.

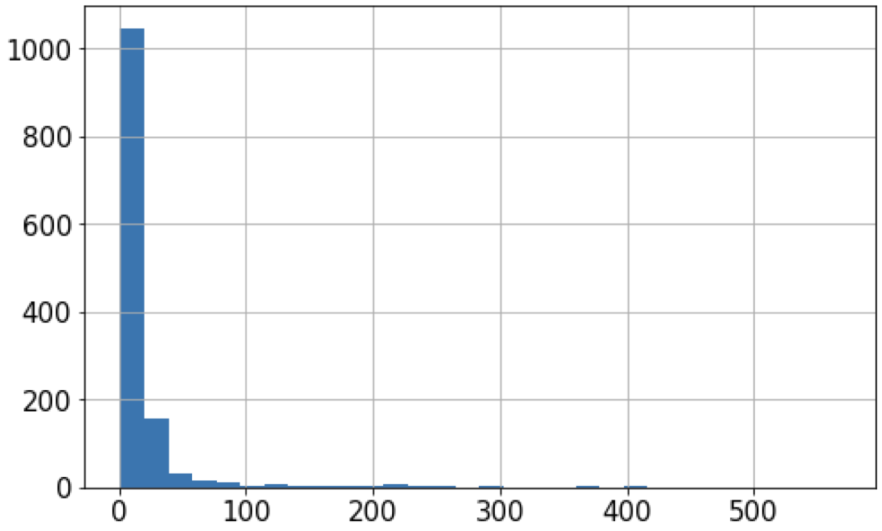


Figure 8.3: Distribution of average daily transactions made by the wallets

Table 8.1: Summary statistics of the bias proxies and wallet properties

	Mean	Standard deviation	Minimum	25th percentile	Median	75th percentile	Maximum
DE	-0.016366	0.326777	-1.000000	-0.167192	-0.041655	0.025300	1.000000
-NF	-0.022279	0.069404	-0.500000	0.000000	0.000000	0.000000	0.000000
OC	0.005252	0.072319	0.000000	0.000000	0.000000	0.000000	1.000000
SIZE	3535.425410	10993.323465	0.173100	61.407632	375.118898	2854.350948	139623.792165
PERF	0.053217	0.196584	-0.818505	-0.000124	0.004579	0.037062	1.962978
RISK	0.123927	0.206175	0.000000	0.020254	0.044879	0.133654	2.045084
DTrades	15.004324	35.001915	1.000000	2.961829	5.462987	15.636538	481.000000

As shown in Table 8.1, the average number of daily transactions across all wallets is 15. The highest number of average daily transactions of a single wallet is 481. One assumption for such a large number of transactions is that it is a trading bot instead of a human investor behind this wallet since it is hard to imagine humans can trade nearly 500 times a day. Surprisingly, the average DE proxy of the wallets is negative, indicating no disposition effect. The average return performance is about 5%, and the median is also above 0, indicating that cryptocurrency investors are generally making profits, even though it is a bear market during the period of 2020 to 2022. This finding may provide evidence that cryptocurrency investors are acumen.

Furthermore, it can be observed from Figure 8.4, which shows the linear correlations (Pearson's r) of bias proxies and wallet properties, that disposition effect (DE) and negative narrow framing (-NF) has

a positive correlation, so do portfolio risk (RISK) and performance (PERF), which can be more or less expected. The positive correlation between portfolio size (SIZE) and daily trades (DTrades), and the positive correlation between SIZE and overconfidence (OC), assume that wealthier investors prone to trade more often, and appear to be more confident. However, disposition effect (DE) and performance (PERF), are positively correlated with a 0.35 correlation coefficient. Meanwhile, both disposition effect (DE) and negative narrow framing (-NF) are negatively correlated with daily trades (DTrades). Also, the negative correlation between performance (PERF) (as well as portfolio risk (RISK)) and daily trades (DTrades) assumes the higher the trading frequencies, the lower the returns and the higher the risks.

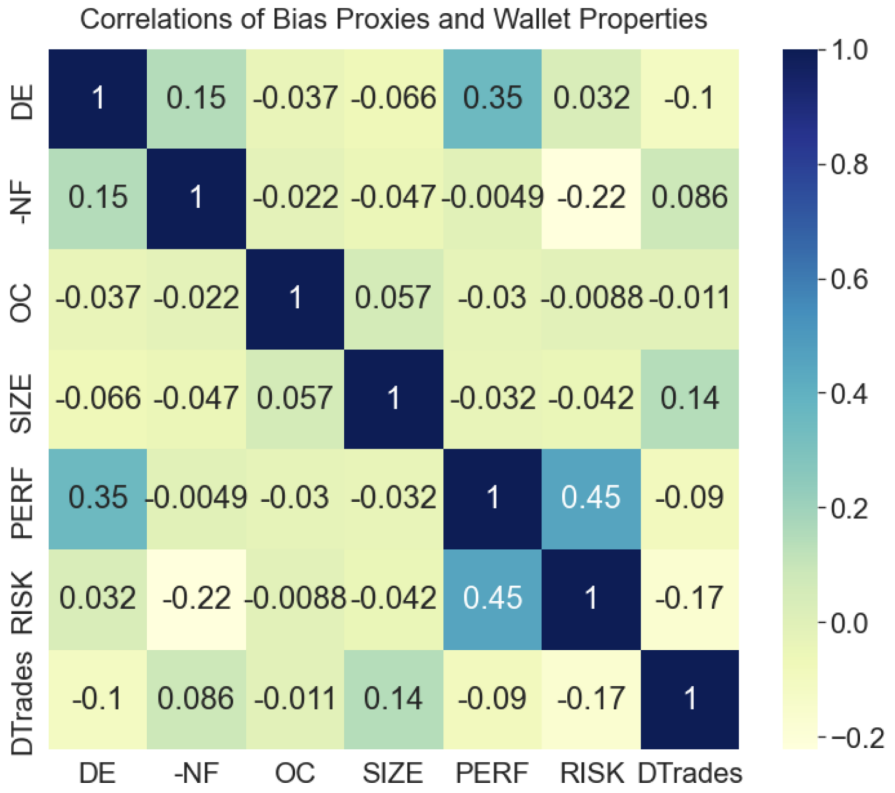


Figure 8.4: Linear correlations (Pearson's r) of bias proxies and wallet properties

8.2.2 *Statistical Tests*

To come by insights into the distinguished behaviors of human investors and trading bots on the Ethereum network, we assume trading bots are those which make more than 50 trades per day. Then, we perform statistical tests to inspect if there is any difference in the bias proxies and wallet properties between human investors and trading bots. Shapiro-Wilk test [93] is implemented to examine the normality of the

distributions of the proxies and properties. After that, we implement Levene's test [85] to confirm their variance equality. With the assumptions verified, we perform the one-tail and two-sample Mann–Whitney U test [72] to examine the differences of bias proxies and wallet properties between the human investors and trading bots, using Python's SciPy library. The results shown in Table 8.2 indicate that the proxies and properties between humans and bots differentiate in all aspects except for overconfidence(OC). Unexpectedly, human investors outperform trading bots on portfolio performance. However, due to the lack of available data on trading bots, the statistics of -NF and OC of trading bots cannot be more accurately revealed. More data will be retrieved, and further analysis will be conducted in the next version of the pre-print.

Table 8.2: Statistical Test: one-tail and two-sample Mann-Whitney U Test

	Human	Trading Bot		
Proxy	Mean(SD)	Mean(SD)	Statistics	P-value
DE	-0.0111(0.3313)	-0.1361(0.1558)	22517.0	1.195527e-02
-NF	-0.0233(0.0708)	0.0(0.0)	15020.0	3.771129e-03
OC	0.0055(0.0739)	0.0(0.0)	18340.0	6.405468e-01
SIZE	3134.5843(10391.2021)	12674.6028(18311.2116)	7491.0	2.703835e-10
PERF	0.0555(0.2005)	0.0017(0.0042)	24489.0	2.415347e-04
RISK	0.1287(0.2094)	0.0162(0.011)	29772.0	1.237686e-11
DTrades	9.7976(9.9626)	133.7185(111.6583)	0.0	8.555053e-27

8.3 LIMITATIONS AND FUTURE WORK

As various biases exist in the behaviors of human investors in investing, we plan to investigate more bias proxies. We also plan to perform clustering of the wallet data by implementing the K-means algorithm and characterize multiple groups of investors based on the results of clustering. Moreover, we will summarize and report investors' proxies and properties and attribute the statistics to different clustered groups in future studies.

9

AN END-TO-END FRAMEWORK DESIGN FOR APPRAISING INVESTMENT BIASES IN PM SYSTEMS

As MSPM's degrees of investment bias in reallocating both stock and cryptocurrency portfolios have been analyzed in the previous chapters, in this chapter, we discuss the design of an end-to-end framework for appraising investment biases in portfolio management systems of heterogeneous types of financial assets (FAIB).

FAIB can be considered as a guideline on how an appraisal framework shall be designed to answer if certain investment biases exist in the decision-making of any given heterogeneous-asset RL-based PM system. FAIB can also be utilized to answer questions like to what degrees it has the biases if such biases exist.

FAIB consists of five sequential modules where the latter module holds a dependency on the former module, whereas each module has its independent functionality.

The design and development of the appraisal framework covered in this chapter follow the sequential scheme of the five modules, and shall be deemed as **one particular case** of many potential implementations of FAIB. Therefore, this study not only introduces the concept of FAIB but also presents a specific realization of FAIB.

9.1 METHODOLOGY

FAIB consists of five modules which are listed below in the sequential order of the procedure:

1. `Portfolio-Construction Module (PCM)` builds PM-system-targeted portfolios with the underlying assets sampled from three different pools.
2. `Portfolio-Backtesting Model (PBM)` backtests the PM system on the reallocation of the portfolios built.
3. `Proxy-Estimation Module (PEM)` evaluates the PM system's degrees of the investment bias proxies using the backtesting results from PBM.
4. `Proxy-Inspection Module (PIM)` conducts hypothesis testing to inspect the existence of biases.
5. `Proxy-Summarization Module (PSM)` visualizes and summarizes the experimental results with a comprehensive report generated.

Figure 9.1 shows the directed acyclic diagram of the five modules. The detailed composition and functionality of each module will be described in this section.

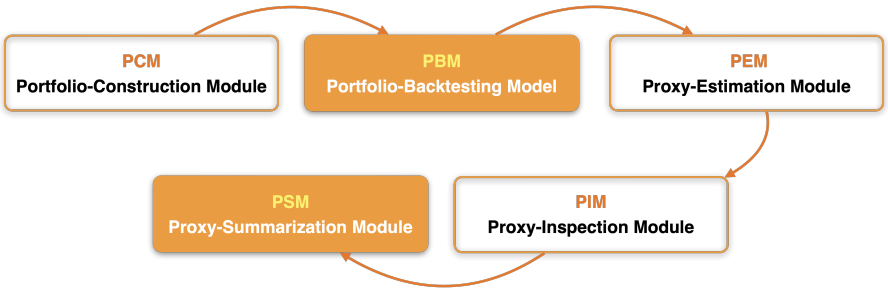


Figure 9.1: Directed acyclic diagram of FAIB: Consisting of five modules

9.2 BACKTESTING

9.2.1 *Portfolio-Construction Module (PCM)*

9.2.1.1 *Asset Pools*

PCM builds portfolios according to the types of assets designated by the target of any given PM system. In order to build portfolios at different diversification levels, the types of assets currently supported by FAIB are 1) common stock, 2) cryptocurrency, and 3) ETF. Three different pools are formed with these assets. Moreover, the available assets in each pool are limited and assorted with specific criterion so that there will be sufficient data for the construction of portfolios and the appraisal of investment biases. The information regarding each pool is revealed as the following:

- Common Securities (stocks) Pool
  - `Portfolio Type`: stock-based portfolio
  - `Source`: Nasdaq-100 Index
  - `Pool Size`: 100
- Cryptocurrencies (cryptos) Pool
  - `Portfolio Type`: crypto-based portfolio
  - `Criterion`: market capitalization
  - `Source`: Binance.com
  - `Pool Size`: 100
- Exchange-Traded Funds (ETFs) Pool
  - `Portfolio Type`: ETF-based portfolio
  - `Criterion`: asset under management (AuM)

- `Source`: ETF Database (etfdb.com)
- `Pool Size`: 100

9.2.1.2 *Portfolio Construction*

The portfolios are built by randomly sampling assets from the designated pool, with the consideration of the number of assets, diversification levels, and sectors in the case of stock-based portfolios. No portfolios will have the same composition of assets. To construct the portfolios for the PM system to perform allocation, PCM requires settings of four parameters. First and foremost, we need to designate the type of asset (stock, crypto, or ETF), and the portfolios will be built with the assets sampled from the corresponding pool. Then, we set the range of the number of portfolios to be built, $[PT_{min}, PT_{max}]$, as well as the range of the number of assets in each portfolio, $[AT_{min}, AT_{max}]$. The range of well-diversified level in percentage, $[WD_{min}, WD_{max}]$, is also required to be set. The available and default settings of PCM's parameters can be found in Table 9.1.

Table 9.1: Available and default settings of PCM's parameters.

Parameter	Min	Max	Range ($[min,max]$)	Option	Remark
Type of Asset	N/A	N/A	N/A	Common Stock; Cryptocurrency; Exchange-Traded Fund	N/A
Portfolios (PT)	1	N/A	[50, 300]	N/A	Min and Max are for the reference. Range is for the practice in this study.
Asset (AT)	1	N/A	[2, 5]	N/A	Same as above
Well-diversified Level (WD)	0%	100%	[10%, 100%]	N/A	Same as above

9.2.2 *Portfolio Backtesting Module (PBM)*

After qualified portfolios have been built (and the corresponding models have been trained in case of RL-based system, e.g., MSPM) by PCM, PBM backtests the PM system on the reallocation of the portfolios. Certain parameters are also required to be set for PBM to work properly, and the parameters are 1). **data source**, 2). **time interval**, 3). **date range**, and 4). **PM system**. The settings of each parameter are going to be introduced in this section.

9.2.2.1 *Parameters*

DATA SOURCE: The data source is automatically set by PBM subject to the asset type. There are two data sources in PBM:

1. `Nasdaq Data Link` provides end of day US stocks and ETFs prices.
2. `Binance REST API` provides multi-interval Binance.com-listed cryptocurrency prices.

Depending on the asset type, PBM decides which source the historical data feed into the PM system for backtesting are from.

TIME INTERVAL AND DATE RANGE: PBM also allows the time interval of historical data to be set. Due to the limitation of `Nasdaq Data Link`, stock and ETF data can only be retrieved at daily intervals, whereas crypto data can be retrieved at multiple intervals. The third parameter of PBM is Date Range, which can be any subset of range from 01/01/2020 to 10/01/2022. The data range should be set as wide as possible to guarantee enough data for backtesting. The data retrieved will be split into three data subsets 1). **training set: 50%**, 2). **validation set: 10%**, and 3). **prediction set: 40%**. Detailed information about the three parameters of PBM can be found in Table 9.2.

Table 9.2: Detailed information about the three parameters of PBM.

Parameter	Range	Source	Option	Remark
Data Source	N/A	Stock: Nasdaq Crypto: Binance ETF: Nasdaq	N/A	N/A
Time Interval	N/A	N/A	Stock: 1Day Crypto: 15m, 30m, 1H,6H12H,1Day ETF: 1Day	N/A
Data Range	01/01/2020 ~10/01/2022	N/A	N/A	Training: 50% Validation: 10% Prediction: 40%

9.2.2.2 *PM System and Model Training*

PBM is responsible for model training and backtesting of the selected PM system. PBM is suggested to be designed in a way that any type of PM system for bias inspection can be accepted. In the case of appraising ML-based PM systems, FAIB triggers model training immediately after the qualified portfolios are constructed to feed into PBM. For MSPM, or CryptoRLPM, in the next chapter, the trained models are reusable and can be combined with other trained models for reallocating different portfolios. However, for most of the existing ML or RL-based PM systems, due to a lack of reusability in their designs, a specific model is required to be trained when there is a new portfolio to be reallocated.

9.2.2.3 *Backtesting*

In this chapter, we follow the backtesting scheme proposed and covered in Chapter 7 and continue using MSPM in the implementation and case study of FAIB. However, subject to different PM systems, the model training and backtesting in PBM may require adjustments and extra design accordingly. With that being said, as long as it is a ML-based PM system to be appraised by FAIB, no changes nor different settings shall be required compared to our settings of PBM.

9.3 APPRAISAL

9.3.1 *Proxy Estimation Module (PEM)*

Using backtesting results generated by PBM, PEM evaluates degrees of the investment bias proxies of the PM system. Since FAIB is a guideline for the system design, any given quantifiable bias proxies can be included and inspected in PEM, depending on the specific realization and practical implementation of the appraisal framework. As a particular implementation of FAIB, this study inspects two bias proxies which have already been introduced in the previous chapters: disposition effect (DE) and narrow framing (NF).

9.3.2 *Proxy-Inspection Module (PIM)*

PIM conducts hypothesis testing to inspect the existence of biases. Once quantified and obtained by PEM, the investment bias proxies of the PM system will be fed into PIM for statistical testing. The normality of the proxy data will first be checked. Then, depending on the asset types, human investor data will be examined jointly to test homogeneity of variance. By completing these two steps, a parametric or non-parametric test will be determined to be performed to compare the degrees of certain investment biases between the PM system and human investors.

9.3.3 *Proxy-Summarization Module (PSM)*

PSM visualizes and summarizes the experimental results with a comprehensive report generated. There can be different options for visualization. For example, investment bias proxies of the PM system can be visualized using line plots by PSM for a given time interval, e.g., weekly interval, and the corresponding human investor's proxies can also be added for comparison.

9.4 CASE STUDY: CRYPTOCURRENCY-BASED PORTFOLIO

In Chapter 7, we examine and discuss the existence and degrees of two investment biases in an RL-based PM system: MSPM. In Chapter 8, we inspect the investment biases of cryptocurrency traders on the Ethereum network. It becomes appealing to investigate the existence and degrees of the same investment biases in MSPM when reallocating cryptocurrency-based portfolios.

9.4.1 *Parameter Settings*

The requirement of parameter settings applies to the first two modules: PCM and PBM. Figure 9.2 displays the specified parameters of PCM and PBM in the case study.

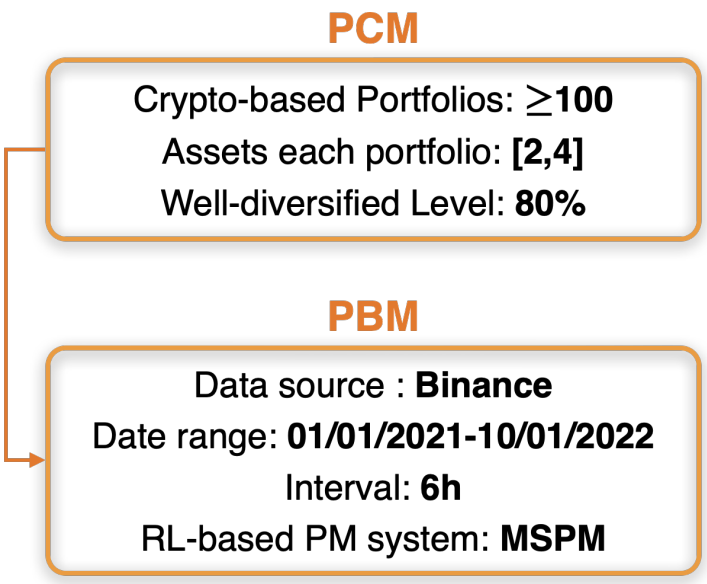


Figure 9.2: The specified parameters of PCM and PBM in the case study.

9.4.2 *Bias Proxies*

There are two investment biases to be appraised in this case study:

- **Disposition effect (DE)** [58, 83, 99, 110] is about investors' tendency to realize profits too soon and keep losses too long. DE has been introduced and inspected in previous chapters.
- **Narrow framing (-NF)** [48, 50, 58] relates to investors' tendency to make isolated and sub-optimal decisions and to trade assets without considering the holistic picture of their portfolios. -NF has been introduced and inspected in previous chapters.

9.4.3 *Results and Discussion*

The results will be revealed and discussed in the sequential order of the five modules of FAIB.

9.4.3.1 *PCM: Portfolio Construction*

By the parameter settings, 135 different portfolios are built. Each of the portfolios consists of 2 $\sim$ 4 different cryptocurrencies. The compositions of the portfolios involve the cryptos randomly sampled from the Cryptocurrencies Pool. To guarantee sufficient data for model training and backtesting, the size of the pool is limited to 18 different cryptos from 5 categories (similar to the concept of the sector in stock market), and cryptos are sampled without replacement for the construction of the portfolios. Figure 9.3 shows the symbols of the 18 different cryptos and their distribution in the portfolios, and Figure 9.4 plots the 5 categories to which the 18 cryptos exclusively belong to, and their distribution.

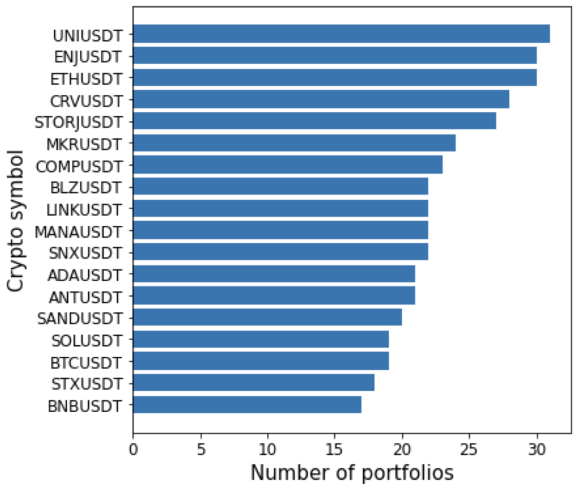


Figure 9.3: The symbols of the 18 different cryptos and their distribution in the portfolios.

Similar to what we expect from the stock portfolios in Chapter 7, cryptocurrency portfolios are also expected to be diversified and from different sectors. Therefore, we continue to follow the setting of the four

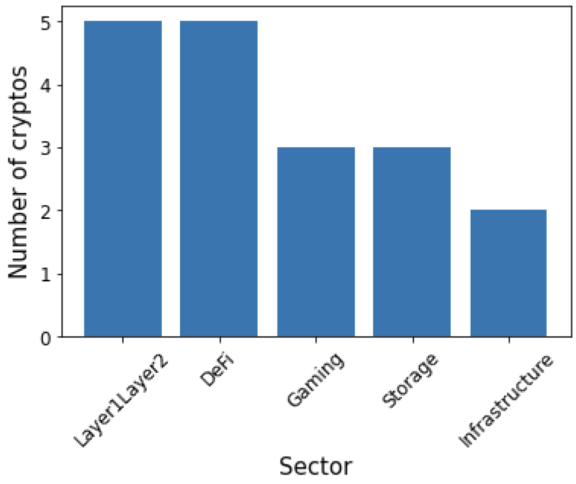


Figure 9.4: The categories which the cryptos exclusively belong to, and their distribution.

diversification levels of portfolios in Chapter 7. Table 9.3 reveals the categorization of the crypto portfolios' diversification levels, and Figure 9.5 plots the overall distribution of the diversification levels. Among the 135 crypto portfolios, 120 portfolios are at least well-diversified, resulting 89% of the portfolios to be appraised in this study being diversified at expected levels.

Table 9.3: Categorization of the crypto portfolios' diversification levels.

	Diversification Level			
Portfolio Type	Fully-diversified	Well-diversified	Semi-diversified	Undiversified
2-assets	assets from 2 sectors	-	-	assets from same sector
3-assets	assets from 3 sectors	assets from 2 sectors	-	same as above
4-assets	assets from 4 sectors	assets from 3 sectors	assets from 2 sectors	same as above

9.4.3.2 *PBM: Backtesting MSPM*

In this case study, we continue utilizing MSPM as the RL-based PM system to be appraised. We want to validate if MSPM remains robust as an unbiased PM system in the case of cryptocurrency investing (PM). In addition, to ensure an even stricter testing, the portfolio reallocation in MSPM in this case study will rely solely on SAMs, without the usage of the high-quality trading signals generated by EAMs [40] (see Section 6.5.2.3 and Section 6.5.2.4). The corresponding models are trained and backtested after the 135 portfolios have been constructed in PCM and fed into PBM.

To fetch the historical crypto data, PBM automatically sets the data source as Binance. For training, validating, and backtesting purposes,

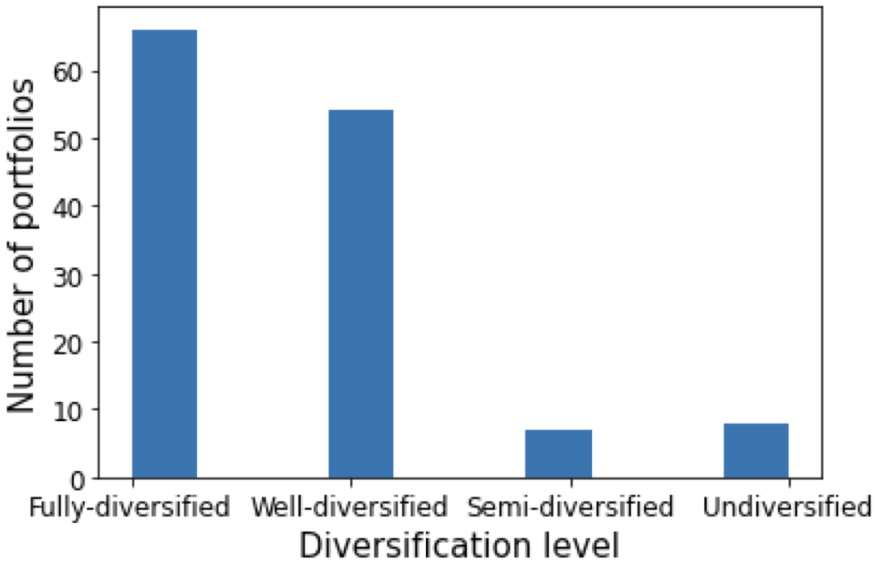


Figure 9.5: Overall distribution of the diversification levels.

the data fetched are split into three subsets following the Data Range setting of PBM, which can be found in Table 9.2.

9.4.3.3 *PEM: Estimation of Bias Proxies*

There are two bias proxies going to be examined. The first one is disposition effect (DE) which is about an investor's propensity to sell winning assets too early and hold losing assets for too long. As a positive DE measurement indicates the existence of disposition effect, we want DE to be negative and to be as low as possible. On the other hand, narrow framing (NF) is Investors' tendency to trade assets without considering the holistic picture of their portfolios. Since a higher -NF indicates a higher degree of narrow framing as the investor tends to execute trades separately instead of collectively. We also want -NF to be as low as possible. The formula of how DE is calculated can be found in Chapter 7. The formula of TC, the proxy of NF can be also found in Chapter 7.

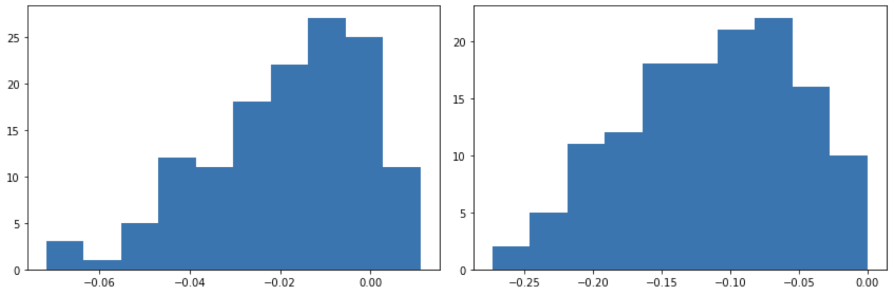
Table 9.4 shows the summary statistics of MSPM's bias proxies. It is very clear that MSPM has a negative average measurement of DE, indicating no disposition effect. Furthermore, the DE of MSPM at its 75 percentile is still negative, strongly evidencing MSPM's robustness in overcoming disposition effect in cryptocurrency PM. It is also clear that MSPM has obtained a very low -NF measurement.

Additionally, we want to perform statistical tests to further compare these two proxies to human investors. Figure 9.6a and Figure 9.6b show the histograms of DE and -NF measurements of MSPM, respectively. Both histograms depict right-skewed distributions. The tests regarding normality and variance homogeneity of DE and -NF measurements will

Table 9.4: Summary statistics of DE and -NF of MSPM.

Proxy	Count	Mean	SD	Min	25%	50%	75%	Max
DE	135.0	-0.018485	0.018103	-0.071815	-0.028613	-0.015314	-0.003665	0.011065
-NF	135.0	-0.112693	0.064562	-0.273567	-0.161390	-0.107456	-0.061008	-0.000000

be performed to determine which type of statistical test to be performed later.



(a) Histogram of MSPM's DE measurement.

(b) Histogram of MSPM's -NF measurement.

Figure 9.6: Histograms of MSPM's DE and -NF depict right-skewed distributions.

9.4.3.4 *PIM: Inspection of Bias Proxies*

Since the statistics of the bias proxies have been obtained, PIM conducts hypothesis testing to compare the proxies between MSPM and human crypto investors (HM) from the Chapter 8, and to inspect the existence of biases.

We implement Shapiro-Wilk test [93] to check the normality of the distributions, and Levene's test [85] to inspect the variance homogeneity. The results of the normality test revealed in Table 9.5, show that none of MSPM or HM's DE and -NF is statistically from normal distributions, with p-values less than 0.05. Also, according to Levene's test (EV), MSPM and HM's DE and -NF do not have homogeneity of variance. With the assumptions verified, we perform the one-tail and two-sample Mann-Whitney U test [72] (a non-parametric version of unpaired t-test) to rigorously compare MSPM and HM's degrees of disposition effect (DE) and narrow framing (-NF).

For either dispositions effect and narrow framing, the hypothesis H_0 is that MSPM has a lower degree of measurement than human crypto investors. In other words, the group mean of DE or -NF of MSPM should be lower than those of human crypto investors. The alternative hypothesis H_a is that MSPM has a higher degree of measurement than

Table 9.5: Results of the normality and equality of variance tests on the DE and -NF of MSPM and HM indicate no normal distributions nor homogeneity of variance.

Normality			
	Human	MSPM	EV
DE	7.622e-32	5.167e-05	1.888e-15
-NF	0.0	0.034	1.629e-06

human crypto investors, or, the group mean of DE or -NF of MSPM is higher than or equal to those of human crypto investors. The significance level is set to be 0.05. We reject H_0 and accept H_a if the p-value is less than 0.05; otherwise, we accept H_0. Specifically:

- Statistical test: one-tail and two-sample Mann-Whitney U test
- Null hypothesis:
  - $H_0^{DE} : \mu_{MSPM}^{DE} - \mu_{HM}^{DE} < 0$
  - $H_0^{-NF} : \mu_{MSPM}^{-NF} - \mu_{HM}^{-NF} < 0$
- Alternative hypothesis:
  - $H_a^{DE} : \mu_{MSPM}^{DE} - \mu_{HM}^{DE} \geq 0$
  - $H_a^{-NF} : \mu_{MSPM}^{-NF} - \mu_{HM}^{-NF} \geq 0$

As the results shown in Table 9.6, by rejecting H_0^{DE} and accepting H_a^{DE} with p-value<.05, and rejecting H_0^{-NF} and accepting H_a^{-NF} with p-value=.011), it is confirmed that MSPM has significantly lower degrees of both DE and -NF than human crypto investors. Therefore, MSPM is proven to overcome and outperform human investors in terms of the two biases when investing in cryptos.

Table 9.6: One-tail and two-sample Mann-Whitney U test confirms that MSPM has significantly lower degrees of both DE and -NF than human crypto investors.

	MSPM (n=135)	HM (n=912)		
Proxy	Mean(SD)	Mean(SD)	Statistics	p-value
DE	-0.018485(0.0181)	-0.011112(0.3313)	110796.5	2.836e-05
-NF	-0.112693(0.1127)	-0.023256(0.0708)	53176.0	3.944e-102

9.4.3.5 *PSM: Visualized Monthly DEs*

To give an example of how PSM may work, visualizations of monthly DEs are generated to summarize the changes in MSPM's DE measurements over time. Particularly, Figure 9.7 shows the monthly DE of MSPM with the human crypto investor's average DE also indicated.

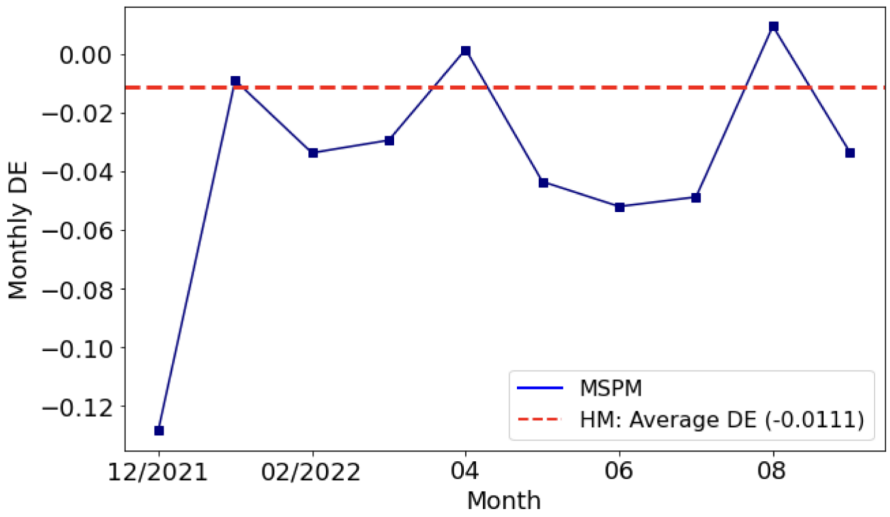


Figure 9.7: Monthly DE of MSPM for crypto PM from Dec 2021 to Sept 2022, with the human crypto investor's average DE also indicated.

It can be observed in Figure 9.7 that MSPM across all portfolios has a lower average monthly DE than human investors, except the months of January, February, and August. If we take a closer look at individual portfolios separately, which is depicted in Figure 9.8, the monthly DEs of the first six portfolios are also below the average DE of human crypto investors most of the time. This observation also applies to the monthly DE of MSPM across different diversification levels illustrated in Figure 9.9 from Dec 2021 to Oct 2022.

9.5 LIMITATIONS AND FUTURE WORK

This chapter aims to introduce the concept of FAIB and its five modules, and to present a specific implementation of FAIB in a case study following the guidelines proposed. Yet, as the implementation in this chapter is a special case of FAIB, it certainly cannot cover all potential functionalities or variations of FAIB and its five modules. Therefore, we plan to design and develop more implementations of FAIB. For example, we may add more bias proxies into the estimation and inspection in PEM and PIM, in other implementations of FAIB in future studies.

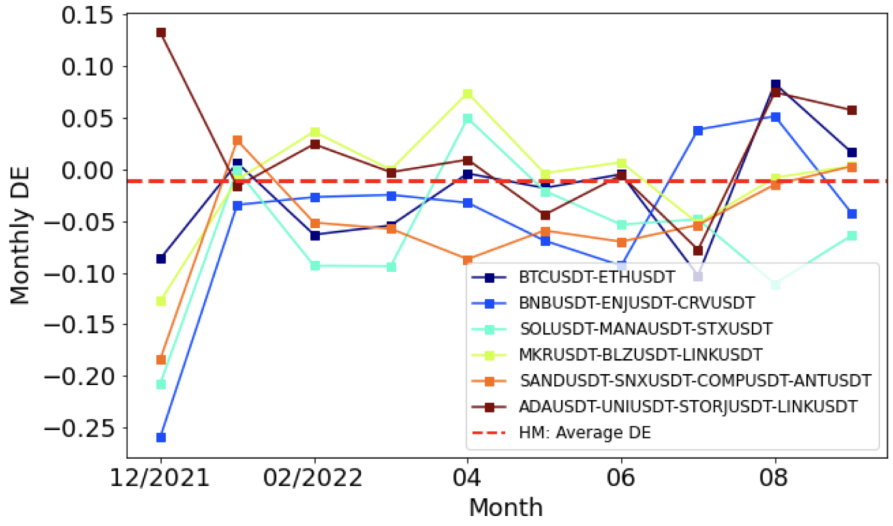


Figure 9.8: Monthly DE of the first six crypto portfolios from Dec 2021 to Sept 2022 are also below the average DE of human crypto investors most of the time.

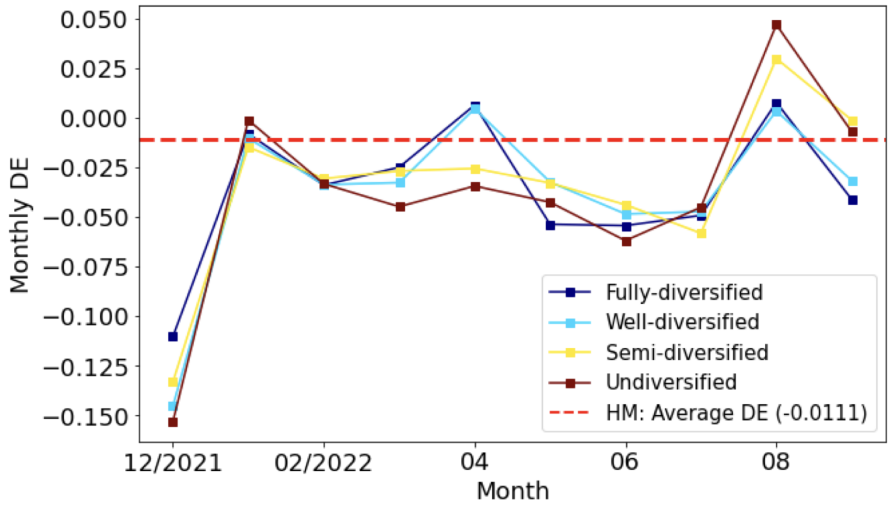


Figure 9.9: Monthly DE of MSPM for crypto PM across different diversification levels from Dec 2021 to Sept 2022.

10

A SCALABLE RL-BASED SYSTEM USING ON-CHAIN DATA FOR CRYPTOCURRENCY PM

In Chapter 8, we investigate the behavioral traits of cryptocurrency traders by utilizing the on-chain data from the Ethereum network. Nevertheless, trading records of wallet addresses are merely a small subset of the data on blockchains. Nowadays, there are many blockchain networks or platforms, and most of these blockchains have their native cryptocurrencies (cryptos). To make an analogy, blockchain can be referred to as a company, and cryptocurrency as its publicly-traded shares. The on-chain data (or on-chain metrics) of a blockchain network can be compared to the fundamentals of a company.

Just as fundamentals disclose most of the information about a company, on-chain data are more precise and complete records of everything there is to know about the blockchain network. More importantly, the valuation of cryptocurrencies can be determined or reflected by multiple factors, including typical on-chain metrics, like hash rate, circulating supplies, exchange flows, or balance on exchanges. Most on-chain data are recorded in a real-time and temporal-sequence manner, and reflect all the running details and measurements of a specific blockchain network and its native cryptocurrency.

Due to the aforementioned nature of on-chain data, people wish to utilize and incorporate on-chain data into their systems for price prediction and quantitative trading [44–46, 94], since the price of crypto can be determined by multiple factors, e.g., hash rate, a typical on-chain metric. Therefore, the incorporation of on-chain data into quantitative trading systems is naturally expected.

However, despite the fact that on-chain data are informative and beneficial, utilization of on-chain data has not been implemented in an RL-based system for PM so far. To what extent this utilization helps the system to outperform the baselines in terms of capital performance and if the system stands unbiased are two intriguing subjects to be answered.

Thus, in this chapter, we propose CryptoRLPM, a novel end-to-end scalable RL-based system incorporating on-chain data for cryptocurrency portfolio management. CryptoRLPM is a mid-frequency (10 to 30-minute interval) PM system consisting of five different units covering the process from information comprehension to trading order execution. In CryptoRLPM, the on-chain metrics are tested and specified for each crypto to solve the ineffectiveness of metrics. Akin to the settings of MSPM in Chapter 7, each Crypto Module of CryptoRLPM (CM) is constructed separately instead of jointly. That is, each CM reallocates a

single-asset portfolio with a risk-free asset (i.e., cash) in it, and hence n CM will be required for an n-asset portfolio to be actually reallocated. By this setting, once a CM is trained, it becomes reusable and can be combined with other CMs for any given portfolio's weighted reallocation. Moreover, this setting of CryptoRLPM allows the portfolios to become scalable, with the underlying cryptos of the portfolios can be changed anytime at will. By backtesting with three portfolios constructed in this study, CryptoRLPM achieves positive ARR, DRR, and SR, at which all the baselines are negative. CryptoRLPM achieves at least 46.79% improvement in ARR, at least 0.8724% improvement in DRR, and at least 1.0181 improvement in SR, compared to the baseline Bitcoin. As metrics like sentiments from social media are incorporated, CryptoRLPM is also appraised regarding its robustness in terms of investment biases via using FAIB.

To the best of our knowledge, CryptoRLPM is the first RL-based system using on-chain metrics for cryptocurrency PM. The benchmarking results indicate that CryptoRLPM robustly outperforms the baselines. Furthermore, we also prove that CryptoRLPM stands unbiased by testing with FAIB.

10.1 METHODOLOGY

CryptoRLPM consists of five main units, covering the process from information comprehension to trading order execution:

- Data Feed Unit (DFU)
- Data Refinement Unit (DRU)
- Portfolio Management Unit (PMU)
- Live Trading Unit (LTU)
- Agent Updating Unit (AUU)

The architecture of CryptoRLPM with the compositions of each of the five units is illustrated in Figure 10.1. The five units are interrelated, and each is responsible for handling at least one task. From a holistic perspective, Data Feed Unit (DFU) and Data Refinement Unit (DRU) are base units which relate to the **data generation**, Portfolio Management Unit (PMU) is responsible for the initial model training of RL agents for a single or multiple portfolios. Live Trading Unit (LTU) and Agent Updating Unit (AUU) are the units responsible for the living trading functionality and the maintenance of the agent and the reallocation of the portfolios. In the following sections, we break down and explain the technical details and tasks of each of the units. Yet, the introduction to LTU and AUU will be rather conceptual since the purpose of the study is to validate the outperformance and to appraise the investment

biases of CryptoRLPM. Although we have no intention to conduct live trading using CryptoRLPM in this study, we do plan to present the implementation of live trading functionality of CryptoRLPM in future studies.

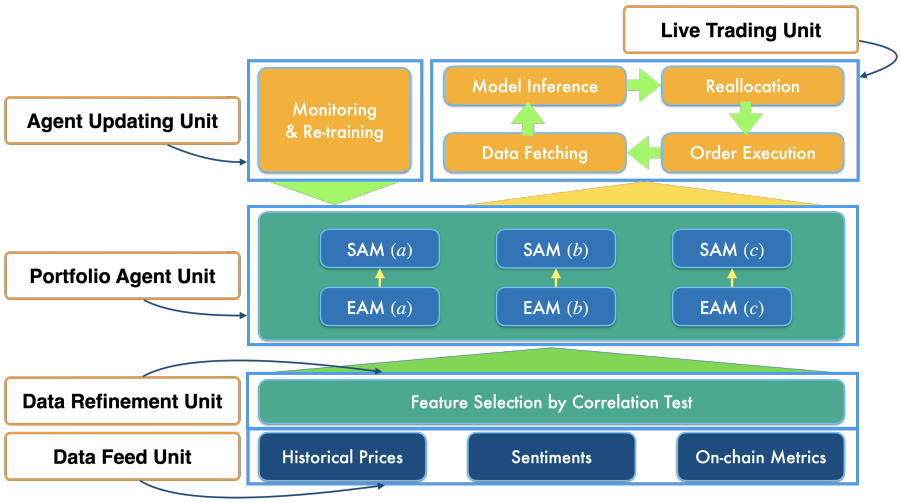


Figure 10.1: The architecture of CryptoRLPM which depicts the abstract of the compositions of each of its five units.

10.1.1 *Data Feed Unit (DFU)*

Data Feed Unit (DFU) is the most fundamental unit of CryptoRLPM which controls the acquisition of data not only for initial model training but also for subsequent and ongoing data feed requirements during live trading and model re-training. Figure 10.2 displays the system design of DFU. The following sections explain the data retrieval process, and introduce the on-chain metrics utilized in the DFU of CryptoRLPM.

10.1.1.1 *Data Retrieval*

Once the underlying cryptos of the portfolio are confirmed, DFU triggers the data retrieval for historical price data and on-chain metrics by calling the Binance REST API and Santiment SanAPI, respectively. The historical crypto price data are obtained by calling Binance's REST API [8], and the on-chain metrics used in this study are obtained by calling the Public REST API (SanAPI) of Santiments.net, which can be accessed by purchasing the SanAPI Basic Subscription at [95]. The data retrieved will be stored in two separate SQLite databases. The stored data will then be fetched by Data Refinement Unit (DRU) and subsequently fed into the Portfolio Agnet Unit (PAU) for initial model training, Live Trading Unit (LTU) for live trading, and Agent Updating Unit (AUU) for ongoing model re-training.

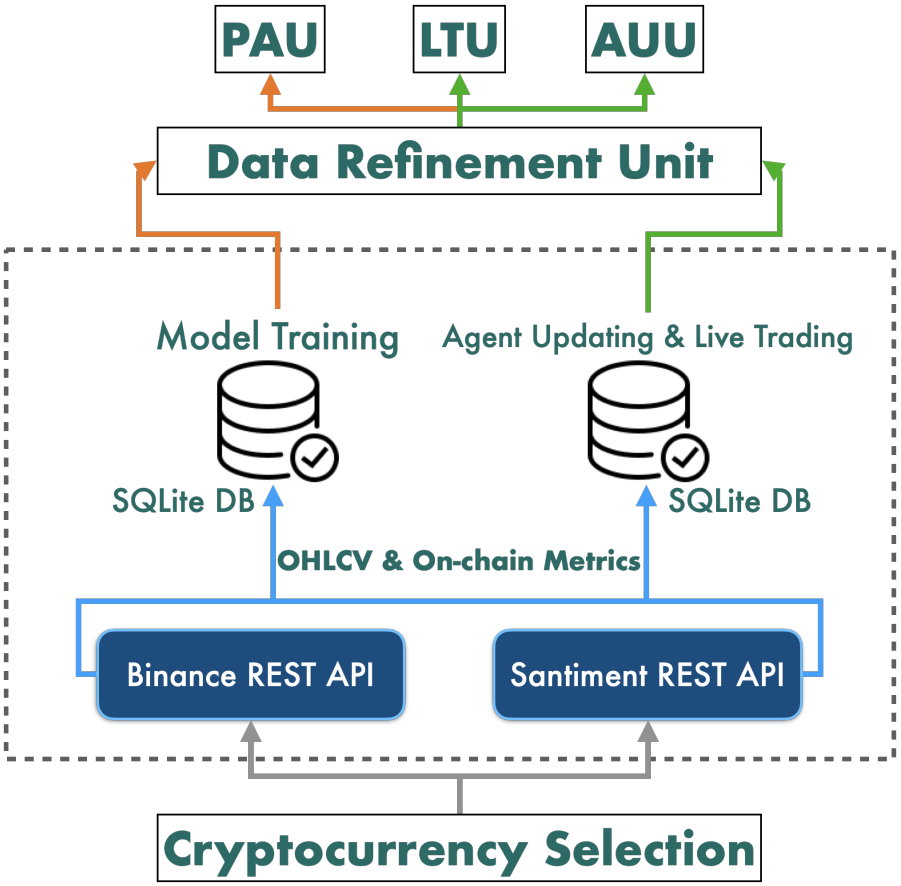


Figure 10.2: The system design of DFU, with the data flow indicated and components illustrated.

10.1.1.2 *On-chain Metrics*

On-chain metrics refer to the information generated from the decentralized ledger of blockchains. For example, Daily Active Addresses, the number of distinct addresses that participated in a transfer for the given crypto on any given day, indicates the daily level of crowd interaction (or speculation) with a crypto [96]. As most of the blockchains have their own native cryptocurrencies (cryptos), the on-chain metrics of a specific blockchain reflect the real-time status, ongoing details, and measurements of that blockchain.

Therefore, if we recognize a blockchain as a publicly-traded company, then the blockchain's cryptocurrency can be referred to as the stock of that company, and the on-chain metrics of a blockchain are like the fundamentals of the company. In fact, thanks to the decentralized nature of blockchains, on-chain metrics become more accurate and transparent measurements of blockchain networks, compared to the fundamentals of companies. Moreover, different from the fundamentals of companies which are disclosed periodically, on-chain metrics are always public and instant, and are recorded in real time.

According to the efficient market hypothesis (EMH) [26], we may assume that the valuation of the blockchain's crypto reflects all the existing available information which includes the on-chain metrics particularly. Thus, the incorporation of on-chain data into quantitative trading systems is naturally expected. However, to the best of our knowledge, such incorporation of on-chain data has not been implemented in an RL-based PM system so far.

Moreover, the metrics found effective on Bitcoin's price prediction may not be applicable to other cryptos, not even mentioning that not every crypto has the same group of available metrics. This **ineffectiveness of metrics** has been barely considered in existing studies.

Thus, in this study, we implement and incorporate on-chain metrics into the environment of CryptoRLPM, an RL-based PM system. The on-chain metrics are tested and specified for each crypto to solve the ineffectiveness of metrics.

AVAILABLE METRICS: The on-chain metrics implemented in this study are the metrics available in SanAPI Basic Subscription Plan, and are subject to different cryptos. As the on-chain and social metrics are intertwined on API platforms and at practical levels, we do not intentionally differentiate between them; they are both on-chain metrics in this study.

10.1.2 *Data Refinement Unit (DRU)*

For any given crypto (e.g. Bitcoin), the correlation tests between the on-chain metrics and three-period returns are performed. Figure 10.3 displays the system design of DRU which is framed by the dashed line. Three-period returns refer to the percentage change (returns) of a crypto's prices of periods of 12, 24, and 48. For example, if we implement the daily OHLCV of Bitcoin, then the three-period returns are the percentage changes of Bitcoin's daily close prices for every 12, 24, and 48 days. The detailed process of DRU will be introduced in this section.

10.1.2.1 *Correlation Test for Feature Selection*

We want to select the valid ones from a large pool of on-chain metrics for the construction of the environment which the RL agents interact with. To achieve this goal, the linear relationship between each of the three-period returns and the on-chain metrics will be inspected for a particular crypto. We obtain the Pearson's correlation coefficients between the returns and metrics. There will be three groups of coefficients corresponding to the three-period returns. For each group, the metrics are sorted in terms of the coefficients, and the qualified metrics in the groups of the highest and the lowest five will be sorted out. Once the qualified metrics have been sorted out from all three groups, we

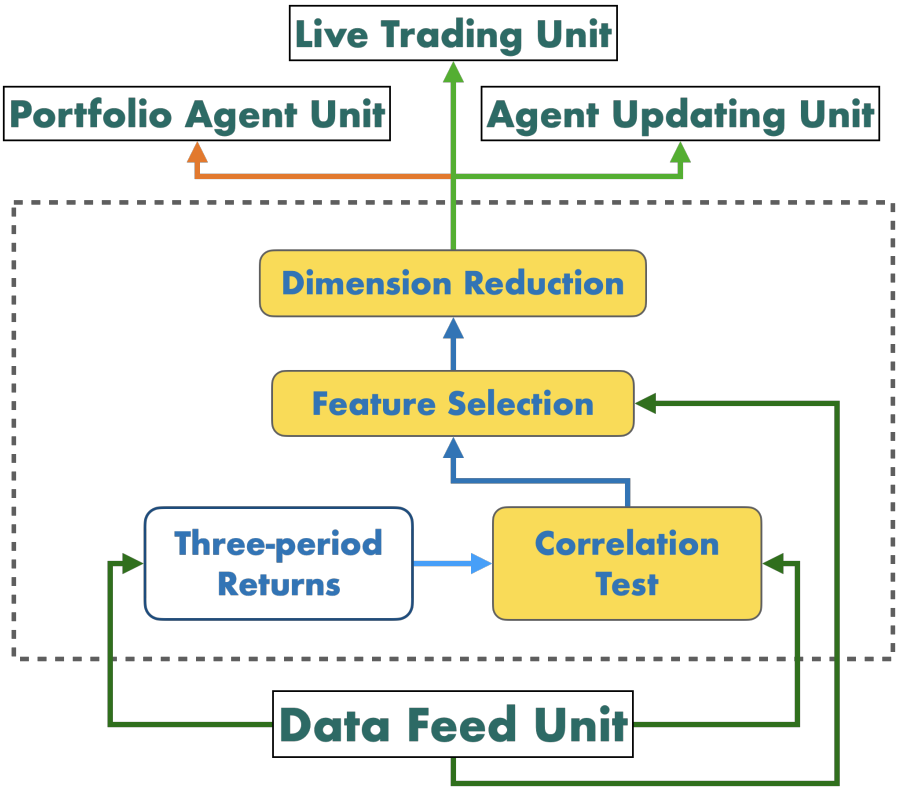


Figure 10.3: The system design of DRU, with the data flow indicated and components illustrated.

rank these metrics by their frequencies of appearance. Then, the top 10 metrics ranked will be selected as valid features for constructing the environment for the agents in PAU.

DIMENSION REDUCTION: As every time the selected metrics are tested and specified for a particular crypto, the issue of ineffective metrics is solved. Yet, to promote more efficient agent learning, the selected metrics will be applied with rolling normalization and rolling PCA for dimension reduction for representation learning, before being fed into the subsequent units. The principle components explain at least 80% of variance will be extracted as the representation of the top-10 metrics, and be fed into PAU, LTU, and AUU.

10.1.3 *Portfolio Agent Unit (PAU)*

In Chapter 7, the Strategic Agent Modules (SAMs) of MSPM [40] are constructed separately instead of jointly. That is, each SAM reallocates a single-asset portfolio with a risk-free asset (i.e., cash) in it. Akin to those settings, we define a Crypto Module (CM) in CryptoRLPM as a bonded module consisting of a pair of Evolving Agent Module (EAM) and SAM for trading a single crypto dedicatedly. Hence, for example, n

CMs will be required for a n-asset portfolio to be actually reallocated. By this setting, once a CM is trained, it can be incorporated into any given portfolio's weighted reallocation, with other CMs. Moreover, to make the design of CM modularized for more efficient training, the EAM in a CM can be optional in certain cases, for example, when the sentiment-included on-chain metrics are fed directly from DRU into the SAM in that CM. This setting of CryptoRLPM allows the PAU to become scalable, with any given portfolio's underlying cryptos being variable anytime.

10.1.3.1 *Settings of PAU*

Figure 10.4 shows the system design of PAU which is framed by the dashed line. For the agent training of crypto x, on-chain metrics will be fed into DRU from DFU for selection and dimension reduction. Then, the refined metrics, along with OHLCV data, will be fed from DRU into the dedicated EAM of crypto x in PAU. Alternatively, for more efficient training, the refined metrics can also be fed directly into the SAM as directed by the orange dashed line, in which case EAM becomes optional, but the high-quality trading signals from EAM [40] (see Section 6.5.2.3 and Section 6.5.2.4) will not be utilized. The signals produced by EAM and the new OHLCV data together formalize the signal-comprised information which will be fed into the SAM of crypto x for decision-making. The trained models will be registered separately in Model Storage. PAU will continue interacting with AUU for model updating and LTU for and live trading. The settings of EAM and SAM are inherited from [40] and [39] with modifications, and the details are described and discussed as the following:

ENVIRONMENT: A crypto-dedicated CM consists of a pair of an EAM (optional) and an SAM. The historical OHLCV and refined on-chain metrics of the designated crypto together formalize the environment which the EAM's RL-based agent interacts with. The combination of the signals produced by the trained EAM and new OHLCV (or signal-comprised information) formalizes the environment which the SAM' RL-based agent interacts with. Each CM is reusable, and periodically re-trained by AUU.

STATE: Within the dedicated CM, the SAM is connected with EAM for producing the weight of the specific crypto. The state v_t, which an EAM observes at every time step t, involves the recent n-interval (e.g. 30-minute) OHLCV and refined on-chain metrics of the designated crypto. $v_t = (s_t, \rho_t)$, where s_t is the n-interval OHLCV, and ρ_t consists of the refined on-chain metrics from DRU. Following the setting of the original SAM in MSPM, the state v_t^+ which the SAM observes, at time step t, involves the stacked new historical OHLCV s_t and the trading

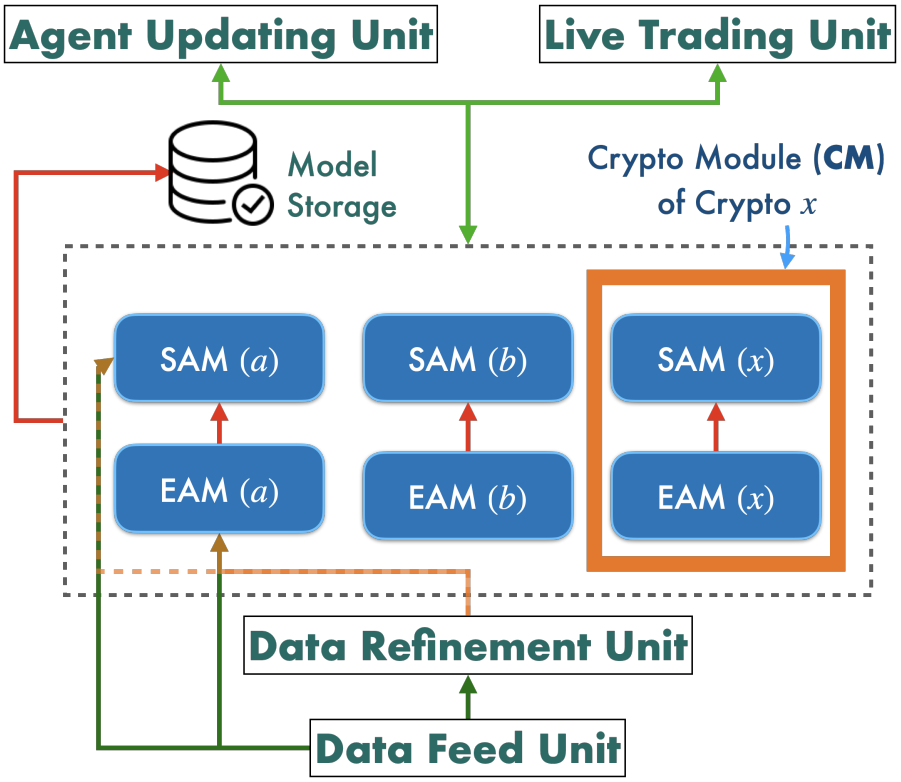


Figure 10.4: The system design of PAU, with the data flow indicated and components illustrated.

signals a_t^{sig}. Nevertheless, since SAM in CryptoRLPM is responsible for one crypto, for $v_t^+ \in \mathbb{R}^{f \times m \times n}$, f is the number of features (OHLCV and on-chain metrics), $m = 2$ indicates the designated crypto and cash, and n stands for recent n intervals.

DEEP Q-NETWORK AGENT: Both EAM and SAM utilize Deep Q-network (DQN) agent to interact with their environments, which has been introduced in both Chapter 6 and Chapter 7. Also, for the estimates of action-value functions of EAM and SAM, $Q^{\theta}_{EAM}(s_t, a_t)$ and $Q^{\theta}_{SAM}(s_t, a_t)$, we follow the settings in Chapter 7 and use a 1-D convolutional neural network (CNN) and a plain 4-layer CNN architecture to represent, respectively.

ACTION SPACE OF EAM: We continue to apply the action space of EAM of MSPM from Chapter 7. At every time step t, the DQN agent in EAM takes an action a_t of either {buying, selling, or skipping} the designated crypto. The actions that EAM chooses to take formalize the crypto's ***trading signal***. Stacked with new OHLCV, the actions will be fed to the SAM within the same CM later.

ACTION SPACE OF SAM: In CryptoRLPM, each CM itself represents a portfolio of the designated crypto and the risk-free asset (cash), which is

reallocated by the SAM within it. SAM of CryptoRLPM assigns the full weight to either the risk-free asset or the crypto. Simply, at every time step t, the action a_t which SAM of CryptoRLPM take is a choice from $\{[0., 1.]$ or $[1., 0.]\}$ indicating the reallocation weight of the portfolio of designated crypto and cash, which can also be deemed a special case of the action space of SAM in Chapter 7. By this setting, once an SAM is trained, it can be combined with other CMs, and then, be integrated into the voted-weight reallocation of any given multi-crypto portfolio.

REWARD FUNCTION: The reward functions of both EAM and SAM of CryptoRLPM follow the settings in the original MSPM in Chapter 6 and Chapter 7.

10.1.4 *Live Trading Unit (LTU)*

Since CryptoRLPM is expected to be an end-to-end system design for cryptocurrency portfolio management, it is natural to integrate and introduce a live trading functionality. Thus, in this section, we discuss Live Trading Unit (LTU) of CryptoRLPM, which is for the live reallocation of the portfolio every 10-30 minutes. The realization of LTU depends on the APIs of specific exchanges, and the implementation of LTU will not be further discussed. Figure 10.5 shows the system design of LTU, which is framed by a dashed line.

For every n interval, new data are fetched and refined following the schemes of the first two units. Then, the newly-fetched and refined data are fed into PAU for new weight inference for the CMs (each corresponds to a designated crypto) in the portfolio, and the set P_t consists of the reallocation weights obtained from all m CMs (cryptos) of the portfolio at time step t

$$P_t = \left\{ (p_t^1, \dots, p_t^m) \mid p_t^i \in \mathbb{R}^2 \text{ for every } i \in \{1, \dots, m\} \right\}, \tag{10.1}$$

and the voted weight w_t will be formalized as the reallocation weight of the portfolio at time step t

$$w_t = \frac{\sum d_t^i}{m}, \text{ for } i \in \{1, \dots, m\} \tag{10.2}$$

The formalized reallocation weight w_t of the portfolio will be wrapped into the format of the designated exchange's API (e.g., Binance). Whenever the portfolio's weight is updated and formatted, a reallocation request will be sent to exchange through their APIs.

10.1.5 *Agent Updating Unit (AUU)*

Agent Updating Unit is responsible for the scheduled model re-training, as well as the unscheduled updating of CMs. After each fixed time

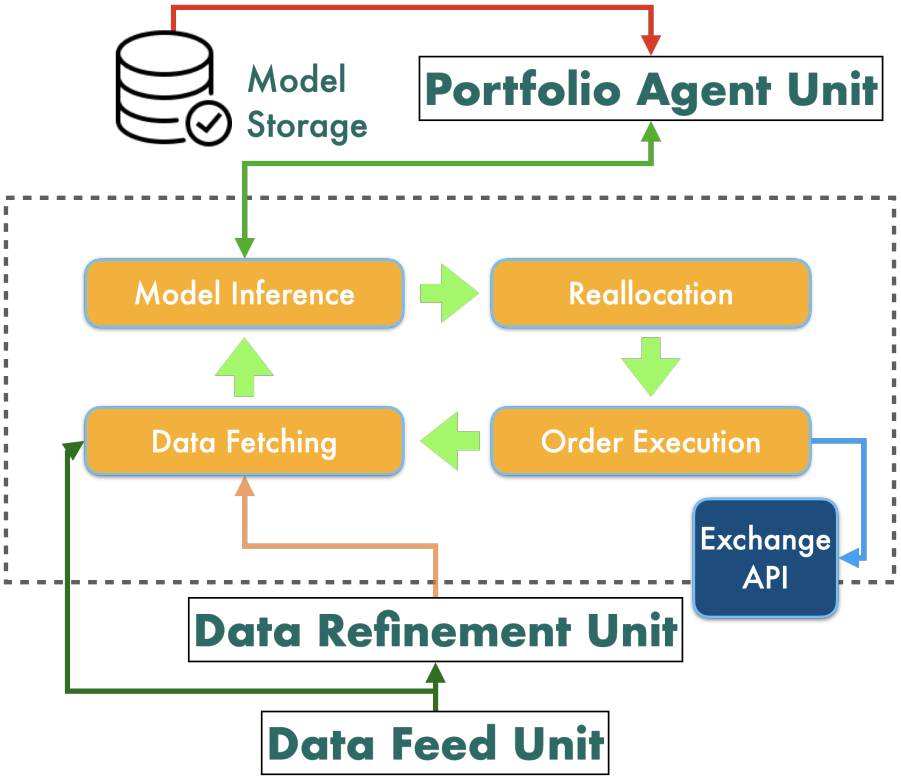


Figure 10.5: The system design of LTU, with the data flow indicated and components illustrated.

interval set in days, the agent models will be re-trained, and the portfolio will be updated if there is any change to the underlying cryptos, such as scaling or replacing.

10.2 EXPERIMENTS

10.2.1 *Preliminaries*

10.2.1.1 *Portfolios*

Three portfolios are proposed for the experiments to benchmark the backtesting performance between CryptoRLPM and the baselines:

1. Portfolio(a) includes two cryptos:
   - Names: Bitcoin and Storj
   - Crypto symbols: [BTC, STORJ]
2. Portfolio(b) includes three cryptos:
   - Names: Bitcoin, Storj and Bluzelle
   - Crypto symbols: [BTC, STORJ, BLZ]
   - Portfolio(b) shares two cryptos in common with Portfolio(a)

3. Portfolio(c) includes two cryptos:
    - Names: Bitcoin, Storj, Bluzelle and Chainlink
    - Crypto symbols: [BTC, STORJ, BLZ, LINK]
    - Portfolio(c) shares three cryptos in common with Portfolio(b)

There are four distinct cryptos among all three portfolios, whose prices are denominated by USDT [111], a stablecoin with U.S. dollar equivalency. Thanks to the reusability of CM and the scalability of PAU, the trained crypto-designated CMs can be applied to different portfolio-designated PAUs for the cryptos in common, which is considerably time-saving and energy-saving and promotes more efficient model training.

Hence, to build the three portfolios, all we need to do is to train the CMs for the four distinct cryptos: CM(BTC), CM(STORJ), CM(BLZ), and CM(LINK), and organize the trained CMs into the PAUs which represent and reallocate the three portfolios. Additionally, based on the cryptos in common, we will discuss the scalability of CryptoRLPM and PAU. The initial portfolio value is set $p_0 = 10,000$, for all these three portfolios.

10.2.1.2 *Data Ranges*

For the purpose of the experiments, DFU obtains the historical 6-hour OHLCV data (s_t) from [8], and the on-chain metrics (ρ_t) originated from [95] and refined by DRU later. In this study, the sentiments and on-chain metrics are refined and feed directly into SAMs of CMs from DRU by leveraging the modularized design of CM (Section 10.1.3.1), to guarantee a more efficient training, and stricter appraisal of investment biases by FAIB. After the refined metrics feed into CMs, they will be split into three subsets: 1). CM(training) ranging from 2020 Oct to 2021 Dec; 2). CM(validation) ranging from 2022 Jan to 2022 Feb; and 3). CM(backtesting) ranging from March 2022 to 2022 Sept. It is worth noting that as the underlying cryptos vary, the data ranges of different portfolios slightly vary. The ranges of the datasets are listed in Table 10.1.

Table 10.1: Description of Data Ranges

Purpose	Range
CM(training)	2020 Oct~2021 Dec
CM(validation)	2022 Jan~2022 Feb
CM(backtesting)	2022 Mar~2022 Sept

10.2.1.3 *Performance metrics*

We continue to use the performance metrics from Chapter 6 to measure the performances of the baselines and CryptoRLPM system. The performance metrics are 1. Daily Rate of Return (DRR), 2. Accumulated rate of return (ARR), and 3. Sortino ratio (SR). All three metrics have been introduced in Chapter 6, and higher measurements of these metrics indicate higher performance.

10.2.2 *Results and Discussion*

10.2.2.1 *Backtesting performance*

Since the intention of the study is to validate the viability of the system design and its features, the baselines to be benchmarked are simply rates of return of the underlying cryptos of each portfolio. We backtest and compare the performance of our CryptoRLPM system to these baselines.

As shown in Figure 10.6, Figure 10.7, and Figure 10.8, for all three portfolios, CryptoRLPM achieves positive ARR, DRR, and SR, at which all baselines at negative. CryptoRLPM achieves at least 46.79% improvement on ARR, at least 0.8724% improvement on DRR, and at least 1.0181 improvement on SR, compared to the baseline Bitcoin.

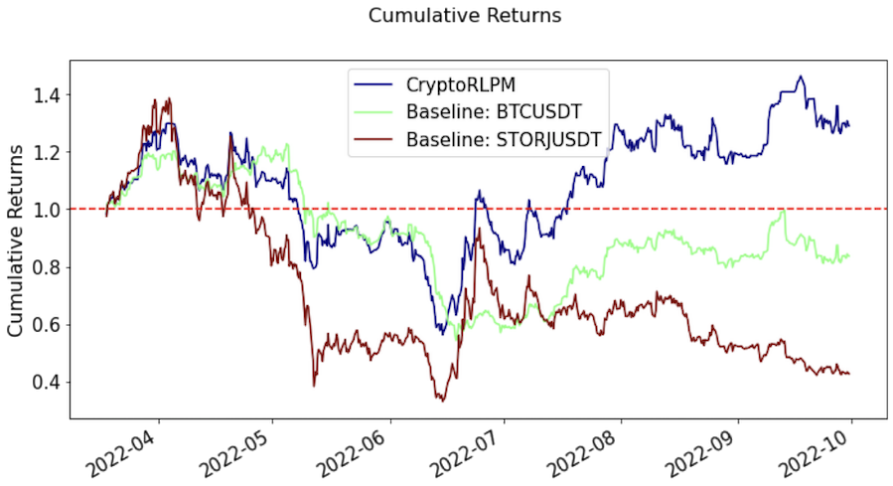


Figure 10.6: CryptoRLPM outperforms all baselines on Portfolio(a) in terms of the accumulated portfolio value in backtesting.

The result demonstrates and validates the viability of CryptoRLPM in gaining capital returns. Table 10.2 details CryptoRLPM's outperformance over the baselines in terms of the ARR, DRR, and SR.

The negligible variation of the baselines' performance across different portfolios is brought by the varying data ranges due to the varying underlying cryptos. It is worth noting that CryptoRLPM achieves

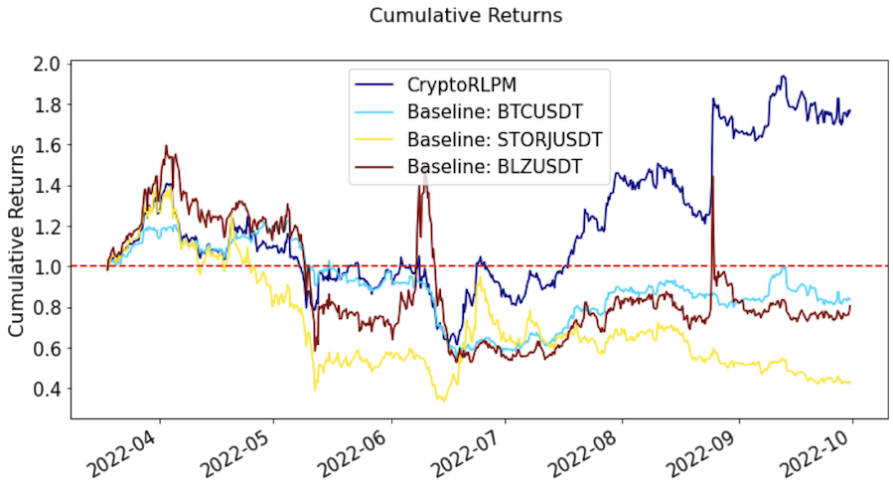


Figure 10.7: CryptoRLPM outperforms all baselines on Portfolio(b) in terms of the accumulated portfolio value in backtesting.

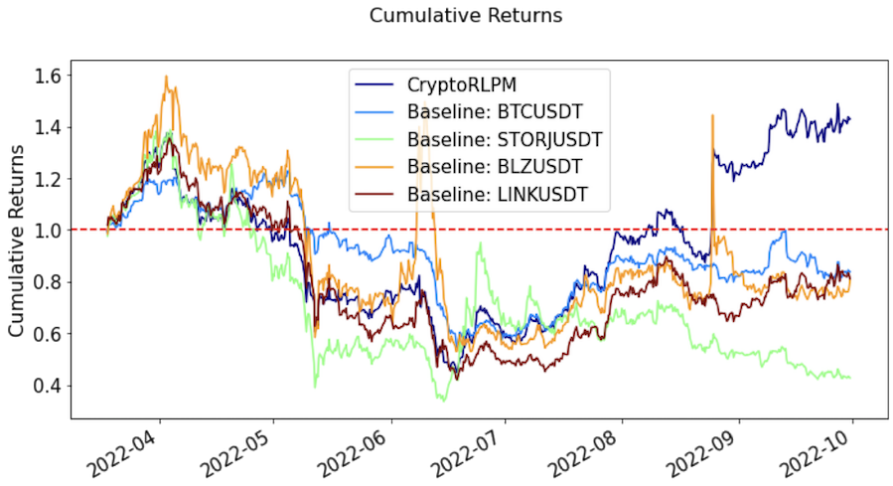


Figure 10.8: CryptoRLPM outperforms all baselines on Portfolio(c) in terms of the accumulated portfolio value in backtesting.

promising SR for all portfolios, which indicates CryptoRLPM's robust ability at profit-making and adaptability to the ever-changing market.

10.2.2.2 *Scalability of CryptoRLPM and PAU*

Each crypto in CryptoRLPM is reallocated by a dedicated and decentralized Crypto Module (CM), which enables CrytoRLPM to be a scalable PM system. Such scalability is reflected in that once the CMs of the underlying cryptos are trained for any given portfolio, the trained CMs become reusable and changeable.

For example, given a portfolio $P_{example}$ consisting of three trained CMs/cryptos: $[a, b, c]$, how can we adjust the portfolio in a way that crypto c changes to a new crypto x? We can simply train a new CM(x)

Table 10.2: Comparison of backtesting performance of the baselines and CryptoRLPM.

			Baselines			
		CryptoRLPM	BTC	STORJ	BLZ	LINK
Portfolio (a)	ARR (%)	**31.26**	-15.53	-56.29	-	-
	DRR (%)	**0.0547**	-0.0227	-0.0800	-	-
	SR (%)	**0.8497**	-0.1684	-0.4345	-	-
Portfolio (b)	ARR (%)	**79.87**	-15.45	-56.25	-18.22	-
	DRR (%)	**0.106**	-0.0229	-0.0783	0.026	-
	SR (%)	**1.3604**	-0.1667	-0.4343	0.4302	-
Portfolio (c)	ARR (%)	**43.71**	-15.45	-56.25	-18.22	-20.35
	DRR (%)	**0.0726**	-0.0229	-0.0783	0.026	-0.0048
	SR (%)	**0.922**	-0.1667	-0.4343	0.4302	0.0655

of crypto x, unplug the CM(c) of crypto x from the PAU of portfolio $P_{example}$, and plug the trained CM(x) to the PAU. Scaling up, or scaling down, a portfolio is at will and even easier. Again, for $P_{example}$, if we decide to exclude a crypto, e.g., b, we just unplug CM(b) from the PAU. We can also add a new crypto y into $P_{example}$ by plugging a trained CM(y).

Figure 10.9 provides an intuitive illustration featuring the scalability of PAU's architecture. Trained CMs of any cryptos are reusable for different PAUs/portfolios. Trained CMs can be added/plugged to, or removed/unplugged from, any PAUs at will.

10.2.2.3 *Bias Proxies Inspection using FAIB*

As metrics like sentiments from social media are incorporated, CryptoRLPM is also appraised regarding its robustness in terms of investment biases using FAIB. The settings of each module of FAIB for inspecting CryptoRLPM are listed as the following:

- `Portfolio-Construction Module (PCM)`
  - The portfolio construction follows the setting in Chapter 9 for inspecting MSPM, involving the same 135 crypto portfolios.
  - Each of the 135 different portfolios consists of $2 \sim 4$ different cryptos which are sampled without replacement from a pool of 18 unique cryptos from 5 different categories.

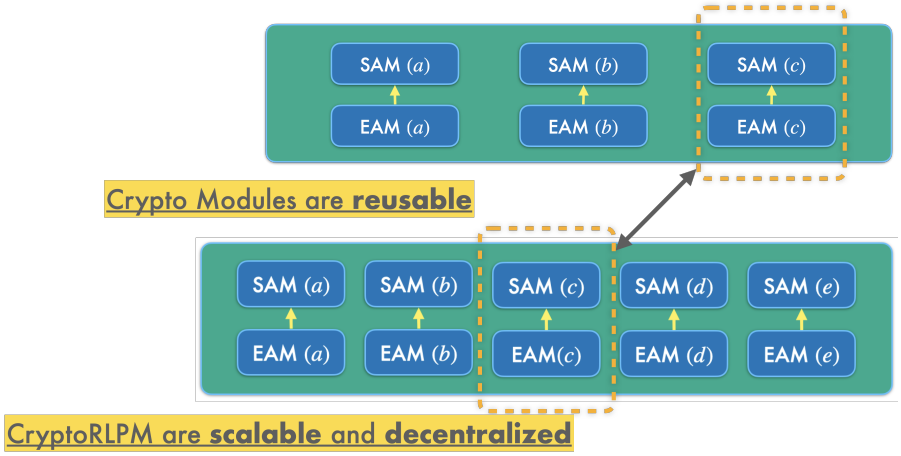


Figure 10.9: An intuitive illustration featuring the scalability of PAU's architecture. Trained CMs of any cryptos are reusable for different PAUs/portfolios. Trained CMs can be added/plugged to, or removed/unplugged from, any PAUs at will.

  - There are 120 out of the 135 portfolios (89%) at least well-diversified.

- `Portfolio-Backtesting Model (PBM)`
  - PM system: CryptoRLPM
  - The historical OHLCV are fetched from [8], and the on-chain metrics are fetched from [95]. To fetch historical crypto data, PBM automatically sets the data source as Binance.
  - Date ranges follow Table 10.1.
- `Proxy-Estimation Module (PEM)`
  - Bias proxies 1): disposition effect (DE), referring to investors' propensity to sell winning assets too early and hold losing assets for too long.
  - Bias proxies 2): narrow framing (-NF), referring to investors' tendency to trade assets without considering the holistic picture of their portfolios.
  - Both bias proxies have been introduced in Chapter 9.
  - Table 10.3 shows the summary statistics of CryptoRLPM's bias proxies. It is very clear that CryptoRLPM has a negative average measurement of DE, indicating no disposition effect. Furthermore, the DE of CryptoRLPM at its 50 percentile is still negative, strongly evidencing CryptoRLPM's robustness in overcoming disposition effect in cryptocurrency PM. It is also clear that CryptoRLPM has obtained a very low -NF measurement even at its 75 percentile.
- `Proxy-Inspection Module (PIM)`

  - CryptoRLPM will be compared with the human crypto investors (HM) from Chapter 8, in terms of the degrees of their bias proxies.
  - Assumption Check 1): Shapiro-Wilk test [93] to check normality of the distributions.
  - Assumption Check 2): Levene's test [85] to inspect the variance homogeneity.
  - As revealed in Table 10.4, none of CryptoRLPM or HM's DE and -NF are statistically from normal distributions (p-values are less than 0.05).
  - Also shown in Table 10.4, CryptoRLPM and HM's DE and -NF do not have homogeneity of variance according to Levene's test.
  - With the assumptions checked, one-tail and two-sample Mann–Whitney U test [72] is performed to compare CryptoRLPM and HM's degrees of both bias proxies.

- `Proxy-Summarization Module (PSM)` is for visualizing the monthly DEs that summarize the changes of CryptoRLPM's DE measurements over time.

Table 10.3: Summary statistics of DE and -NF of CryptoRLPM.

Proxy	Count	Mean	SD	Min	25%	50%	75%	Max
DE	135.0	-0.006905	0.016061	-0.049232	-0.017534	-0.007921	0.002656	0.054597
-NF	135.0	-0.204671	0.071678	-0.338983	-0.265461	-0.213710	-0.136821	-0.048000

Table 10.4: Results of the normality and equality of variance tests on the DE and -NF of CryptoRLPM and HM indicate no normal distributions nor homogeneity of variance.

	Normality		
	Human	CryptoRLPM	EV
DE	7.622e-32	0.075	9.735e-16
-NF	0.000	0.000	9.156e-10

Following the hypotheses in PIM in Chapter 9, the null hypothesis H_0 is that CryptoRLPM has a lower degree of the measurement than human crypto investors, or, the group mean of DE or -NF of CryptoRLPM should be lower than those of human crypto investors. The

alternative hypothesis H_a is that CryptoRLPM has a higher degree of measurement than human crypto investors, or, the group mean of DE or -NF of CryptoRLPM is higher than or equal to those of human crypto investors.

With the significance level set at 0.05, we reject H_0 and accept H_a if the p-value is less than 0.05; otherwise, we accept H_0. Specifically:

- Statistical test: one-tail and two-sample Mann–Whitney U test
- Null hypothesis:
  - $H_0^{DE} : \mu_{CryptoRLPM}^{DE} - \mu_{HM}^{DE} < 0$
  - $H_0^{-NF} : \mu_{CryptoRLPM}^{-NF} - \mu_{HM}^{-NF} < 0$
- Alternative hypothesis:
  - $H_a^{DE} : \mu_{CryptoRLPM}^{DE} - \mu_{HM}^{DE} \geq 0$
  - $H_a^{-NF} : \mu_{CryptoRLPM}^{-NF} - \mu_{HM}^{-NF} \geq 0$

As the results shown in Table 10.5, by rejecting H_0^{DE} and accepting H_a^{DE}, and rejecting H_0^{-NF} and accepting H_a^{-NF}, with p-values<.05, it is confirmed that CryptoRLPM has significantly lower degrees of both DE and -NF than human crypto investors. Therefore, CryptoRLPM is proven to overcome and outperform human investors in terms of the two biases when investing cryptocurrencies, and stand robust and unbiased by the appraisal using FAIB.

Table 10.5: One-tail and two-sample Mann-Whitney U test confirms that CryptoRLPM has significantly lower degrees of both DE and -NF than human crypto investors.

	CryptoRLPM (n=135)	HM (n=912)		
Proxy	Mean(SD)	Mean(SD)	Statistics	p-value
DE	-0.006905(0.0161)	-0.011112(0.3313)	110796.5	2.836e-05
-NF	-0.204671(0.0717)	-0.023256(0.0708)	53176.0	3.944e-102

10.3 LIMITATIONS AND FUTURE WORK

In this chapter, we introduce the concepts of CryptoRLPM. To ensure more efficient training for backtesting, we feed the refined metrics directly into PAU from DRU, without utilizing the trading signals from EAMs of CMs. We leave the utilization of trading signals of EAMs to future studies, and believe that such utilization will surely improve the outperformance (in terms of ARR, DRR, and SR) of CryptoRLPM even more. Also, we plan to include more baselines for benchmarking,

which can be some conventional PM strategies, such as CRP, or RL-based methods, such as ARL. Additionally, in this chapter, we focus on validating the outperformance of CryptoRLPM through backtesting and benchmarking, and appraising the investment biases of CryptoRLPM. We plan to present the implementation of the live trading functionality of CryptoRLPM in future studies.

11 CONCLUSION

By achieving the research goals of the study in each chapter, this thesis aims to establish structured and organic connections among reinforcement learning, financial portfolio management, behavioral finance, and blockchain technologies. More importantly, by contributing to the three themes in Section 1.3, this thesis provides sound and robust answers to the research questions raised in Section 1.2. Specifically, through the research and development of two novel RL-based PM systems (Chapter 6 and Chapter 10), investment biases and a general appraisal framework (Chapter 7 - Chapter 9), and utilization of on-chain wallet information and metrics (Chapter 8 and Chapter 10), we aim to provide an avant-garde and unprecedented roadmap towards more scalable, robust and unbiased system designs for RL-based PM. Such a roadmap can be presented by the diagram of the thesis's structure (Fig 1.1 in Section 1.4).

This chapter reviews and summarizes the studies covered in the previous chapters and provides insights into future studies. The summarization is organized by the three themes in Section 1.3. To better facilitate the summarization, the roadmap (diagram of the thesis's structure) is revisited in Fig 11.1.

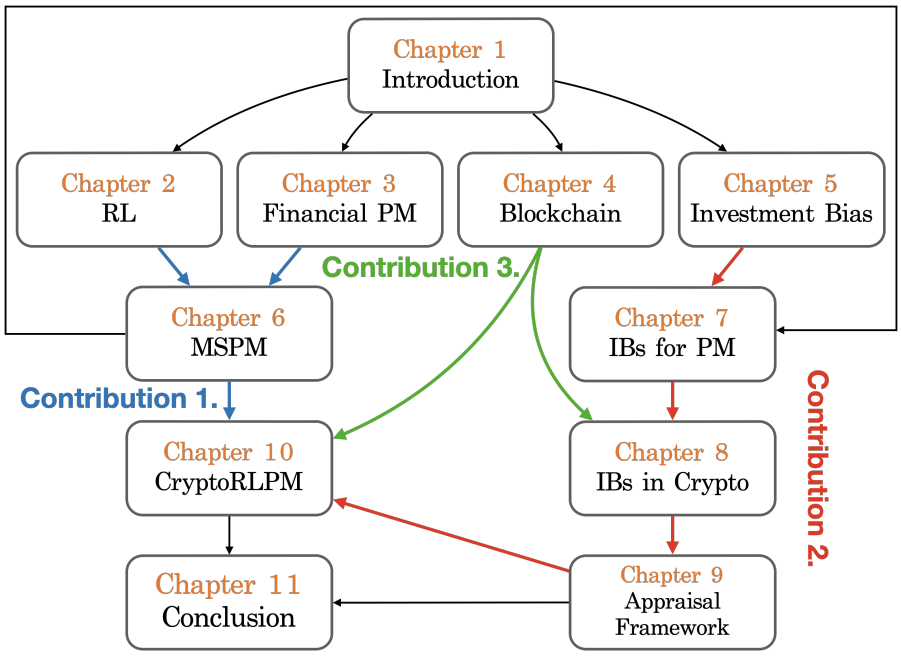


Figure 11.1: Dependency diagram of the thesis's structure with the contribution of each chapter highlighted.

11.1 CONTRIBUTION 1. (RL-BASED PM)

In Chapter 6, we propose MSPM, the first modularized multi-agent RL-based system that brings scalability and reusability to PM. MSPM consists of two types of modules in MSPM: EAM (an asset-dedicated module) and SAM (a decision-making module). With its modularized and reusable design, MSPM addresses the issue of ad-hoc, fixed, and inefficient model training in the existing RL-based methods. By experiment and benchmarking, we have confirmed that our MSPM system outperforms five different baselines under extreme market conditions of U.S. stock markets during the global pandemic, from January to December 2020. Particularly, EAM-enabled MSPM systems improve the accumulated rate of return of two different portfolios by 49.3% and 426.6% compared to Adversarial PG[65], a state-of-the-art RL-based method, and by 186.5% and 369.8% compared to Constant Rebalanced Portfolio (CRP)[19], a conventional PM strategy. In addition, the average winning rate of the EAMs in the two portfolios achieves 80%. We also exemplified the high quality and reliability of the signals generated by EAM, and validated the necessity of EAM. Furthermore, we validate the indispensability of Evolving Agent Module (EAM) by backtesting MSPM on four different investment portfolios. Among the portfolios, EAM-enabled MSPMs outperforms the EAM-disabled MSPMs. The experimental results prove that MSPM is qualified as a stepping stone to inspire more creative system designs in RL-based PM. In Chapter 10, we propose CryptoRLPM, a novel end-to-end scalable RL-based system incorporating on-chain data for cryptocurrency portfolio management. The benchmarking results indicate that CryptoRLPM robustly outperforms the baselines. Particularly, by backtesting with three portfolios constructed, CryptoRLPM achieves positive ARR, DRR, and SR, at which all baselines are negative. CryptoRLPM achieves at least 46.79% improvement in ARR, at least 0.8724% improvement in DRR, and at least 1.0181 improvement in SR, compared to the baseline Bitcoin. Additionally, CryptoRLPM achieves promising SR for all portfolios, which indicates CryptoRLPM's robust ability at profit-making and adaptability to the ever-changing market. MSPM, together with CryptoRLPM, provide a valid and robust answer to the **Q1** and **Q2** in Section 1.2.

11.2 CONTRIBUTION 2. (INVESTMENT BIASES)

The study covered in Chapter 7, as the first of this kind, investigates the existence and degrees of two investment biases, disposition effect, and narrow framing, in a cutting-edge multi-agent RL-based system for PM (MSPM) proposed in Chapter 6. By experimenting with 135 different portfolios of different diversification levels, we prove that MSPM overcomes and outperforms human investors over the two investment biases. Also, by applying new settings of and extensions to

MSPM, we validate the adaptability of MSPM as a general framework to accommodate various RL methods for financial portfolio management. Furthermore, we demonstrate the RL-based system's decision-making in portfolio reallocation and further validate the outperformance of MSPM on the disposition effect in a case study. The experimental results prove our study as an initial step closer to an **unbiased** and more **robust** reinforcement learning-based system design for financial portfolio management.

To take a step forward, in Chapter 9, we design and develop FAIB, the first appraisal framework for evaluating investment biases proxies in PM systems. FAIB shall be employed to answer if certain investment biases, e.g., disposition effect, exist in the decision-making of any given heterogeneous-asset RL-based PM system, and to what degrees it has the biases if such biases exist. We constructed 135 unique crypto portfolios with 18 cryptos from 5 different sectors. By experimenting with MSPM in a case study of FAIB, we demonstrate the compatibility and functionality of the five modules of FAIB. We also prove that MSPM remains robust as an unbiased PM system for cryptocurrency investing in terms of two bias proxies, disposition effect, and narrow framing.

Moreover, in Chapter 10, by constructing and experimenting with the same 135 crypto portfolios, we also prove CryptoRLPM, a novel RL-based PM system for cryptocurrency trading, to overcome and outperform human investors in terms of the two biases when investing cryptos, and stand robust and unbiased by the appraisal using FAIB.

As a delightful result, the **Q3** in Section 1.2 has been answered by the contributions made by the studies covered in the above-mentioned chapters, and more importantly, investment bias now becomes a new category of metrics measuring the performance of RL-based PM systems.

11.3 CONTRIBUTION 3. (CRYPTOS AND ON-CHAIN METRICS)

As covered in Chapter 8, this is the first study that analyzes cryptocurrency investors' portfolio properties and biased behaviors by utilizing the on-chain wallet records on blockchains, without using any inaccessible nor indirect data sources like centralized exchange databases or questionnaires (surveys). This study proposes a complete pipeline for the acquisition and pricing-matching of on-chain wallet records. This study quantifies and reveals the relationship between three behavioral bias proxies and four portfolio properties of more than 952 cryptocurrency investors in their investment decision-making by using on-chain wallet records from the Ethereum network and correlation tests performed. This study also distinguishes and analyzes the wallets of human investors and trading bots. We find that wealthier investors appear to be more confident and tend to trade more often. We find that the higher the trading frequencies of the investors, the less the

returns from their transactions. The statistical tests' results indicate the significant differences between human investors and trading bots on all behavioral biases and wallet properties, and human investors outperform trading bots on certain biases.

Further, in Chapter 10, we build CryptoRLPM, which, to the best of our knowledge, is the first RL-based system PM incorporating on-chain data with a scalable and modularized design for cryptocurrency PM. CryptoRLPM consists of five modularized units. In Data Refinement Unit (DRU) of CryptoRLPM, the on-chain metrics are tested and specified for each crypto, and this setting solves the ineffectiveness of metrics. Each crypto in CryptoRLPM is reallocated by a dedicated and decentralized module (CM) of Portfolio Agent Unit (PAU), which enables CrytoRLPM to be a scalable PM system. By this setting, once a CM is trained, it becomes reusable and can be combined with other CMs for the weighted reallocation of any given portfolio. The backtesting results of the three portfolios indicate that CryptoRLPM robustly outperforms the baselines. Also, we prove that CryptoRLPM stands unbiased by using FAIB proposed in Chapter 9, which is aligned with the goal of the thesis to build a scalable, robust and unbiased RL-based PM system. Thus, by achieving the research goals in studies covered in Chapter 8, Chapter 9 and Chapter 10), the **Q4** and **Q5** in Section 1.2 are practically answered.

11.4 INSIGHTS INTO FUTURE STUDIES

As displayed by the roadmap (Fig 11.1), the studies covered by this thesis aim to establish structured and organic connections among reinforcement learning, financial portfolio management, behavioral finance, and blockchain technologies.

In this thesis, we have investigated and revealed if two RL-based PM systems are unbiased compared to human investors. This is exceptionally critical because sentiment-based factors may be expected to be constructed or selected by RL-based methods for investing in future studies [6, 17]. Particularly, as assets like cryptocurrency are considerably propelled by investors' sentiments, resulting in the dynamics of the interaction between the market and its participants (investors), it is important to consider this interaction and its consequences in the system / environment / reward function designs. On the other hand, the incorporation of such sentiment-based factors into an RL-based PM system should also be expected. The investment biases themselves may be considered when designing reward functions as well.

Furthermore, since topics of heterogeneous data are investigated and discussed in this thesis, and as the aforementioned information, such as social sentiments, is accessible by implementing various methods, it is foreseeable that RL-based methods for on-chain arbitrage based on information like on-chain metrics and sentiments will emerge.

As a general framework for investing investment biases in PM systems (FAIB) being proposed in Chapter 8, it will be more interesting if FAIB can be extended to reveal the internal mechanism of how the system overcomes more investment biases in future studies.

Due to the increasing demand, openness, and transparency of blockchains, developers, individual or institutional investors, and organizations wish to build their crypto portfolios on blockchains, which are managed by autonomous algorithms or smart contracts. Several existing products seek to fulfill this demand, e.g., Enzyme Finance [28]. Nevertheless, the existing tools are hard to implement by calling API and lack flexibility. More importantly, they do not provide SDKs supporting multiple programming languages. It will be more convenient if a tool is designed for and provided to people who want to build, allocate, manage, and report their cryptos or any decentralized financial assets through an API-enabled service.

BIBLIOGRAPHY

[1] Bashar Yaser Al-Mansour. "Cryptocurrency Market: Behavioral Finance Perspective." In: *The Journal of Asian Finance, Economics and Business* (2020).

[2] Dogu Araci. *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models*. 2019. arXiv: 1908.10063 [cs.CL].

[3] Warren Bailey, Alok Kumar, and David Ng. "Behavioral biases of mutual fund investors." In: *Journal of Financial Economics* 102.1 (2011), pp. 1–27. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2011.05.002. URL: https://www.sciencedirect.com/science/article/pii/S0304405X11001140.

[4] The World Bank. *Market capitalization of listed domestic companies*. 2022. URL: https://data.worldbank.org/indicator/CM.MKT.LCAP.CD/.

[5] Dirk G Baur, Kihoon Hong, and Adrian D Lee. "Bitcoin: Medium of exchange or speculative assets?" In: *Journal of International Financial Markets, Institutions and Money* 54 (2018), pp. 177–189.

[6] Jennifer Bender, Remy Briand, Dimitris Melas, and Raman Aylur Subramanian. "Foundations of factor investing." In: *Available at SSRN 2543990* (2013).

[7] Dimitri Bertsekas and John N Tsitsiklis. *Neuro-dynamic programming*. Athena Scientific, 1996.

[8] Binance. *Market Data Endpoints*. 2022. URL: https://binance-docs.github.io/apidocs/spot/en/.

[9] Thomas J. Brennan and Andrew W. Lo. "The Origin of Behavior." In: *The Quarterly Journal of Finance* 01.01 (2011), pp. 55–108. DOI: 10.1142/S201013921100002X. eprint: https://doi.org/10.1142/S201013921100002X. URL: https://doi.org/10.1142/S201013921100002X.

[10] Vitalik Buterin et al. "A next-generation smart contract and decentralized application platform." In: *NA* 3.37 (2014), pp. 2–1.

[11] Jonathan Chiu and Thorsten V Koeppl. "The economics of cryptocurrencies–bitcoin and beyond." In: *Available at SSRN 3048124* (2017).

[12] CoinAPI.io. *API Documentation*. 2022. URL: https://docs.coinapi.io.

[13] CoinGecko.io. *API Documentation*. 2022. URL: https://www.coingecko.com/en/api/documentation.

[14] CoinMarketCap.com. *Cryptocurrencies Ranking*. 2022. URL: https://coinmarketcap.com.

[15] CoinMetrics.com. *Introducing Realized Capitalization*. 2018. URL: https://coinmetrics.io/realized-capitalization/.

[16] Coingecko. *CoinGecko 2021 Report*. 2021. URL: https://www.coingecko.com/en/publications/reports.

[17] Guillaume Coqueret and Eric André. "Factor investing with reinforcement learning." In: *Available at SSRN 4103045* (2022).

[18] Henrik Cronqvist and Stephan Siegel. "The genetics of investment biases." In: *Journal of Financial Economics* 113.2 (2014), pp. 215–234. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2014.04.004. URL: https://www.sciencedirect.com/science/article/pii/S0304405X14000889.

[19] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" In: *The Review of Financial Studies* 22.5 (Dec. 2007), pp. 1915–1953. ISSN: 0893-9454. DOI: 10.1093/rfs/hhm075. eprint: https://academic.oup.com/rfs/article-pdf/22/5/1915/24429471/hhm075.pdf.

[20] Paul Delfabbro, Daniel L King, and Jennifer Williams. "The psychology of cryptocurrency trading: Risk and protective factors." In: *Journal of behavioral addictions* 10.2 (2021), pp. 201–207.

[21] Paul Delfabbro, Daniel L. King, and Jennifer Williams. "The psychology of cryptocurrency trading: Risk and protective factors." In: *Journal of Behavioral Addictions* 10.2 (2021), pp. 201 –207. DOI: https://doi.org/10.1556/2006.2021.00037. URL: https://akjournals.com/view/journals/2006/10/2/article-p201.xml.

[22] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

[23] Etherscan.com. *DEX Tracker*. 2022. URL: https://etherscan.io/dex.

[24] Jonathan St B. T. Evans. "Heuristic and analytic processes in reasoning*." In: *British Journal of Psychology* 75.4 (1984), pp. 451–468. DOI: https://doi.org/10.1111/j.2044-8295.1984.tb01915.x. eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2044-8295.1984.tb01915.x. URL:

https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8295.1984.tb01915.x.

[25] MICHELLE FAVERIO and NAVID MASSARAT. *46% of Americans who have invested in cryptocurrency say it's done worse than expected*. 2022. URL: https://www.pewresearch.org/fact-tank/2022/08/23/46-of-americans-who-have-invested-in-cryptocurrency-say-its-done-worse-than-expected/ (visited on 08/23/2022).

[26] Eugene F. Fama. "Efficient Capital Markets: A Review of Theory and Empirical Work." In: *The Journal of Finance* 25.2 (1970), pp. 383–417. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2325486.

[27] Eugene F. Fama. "Efficient capital markets: A review of theory and empirical work." In: *Journal of Finance* 25 (1970), pp. 383–417.

[28] Avantgarde Finance. *Enzyme Finance: On-chain Asset Management*. 2022. URL: https://enzyme.finance.

[29] Cary Frydman and Colin F. Camerer. "The Psychology and Neuroscience of Financial Decision Making." In: *Trends in Cognitive Sciences* 20.9 (2016), pp. 661–675. ISSN: 1364-6613. DOI: https://doi.org/10.1016/j.tics.2016.07.003. URL: https://www.sciencedirect.com/science/article/pii/S1364661316300997.

[30] Sally M Gainsbury and Alex Blaszczynski. "How blockchain and cryptocurrency technology could revolutionize online gambling." In: *Gaming Law Review* 21.7 (2017), pp. 482–492. DOI: http://doi.org/10.1089/glr2.2017.2174.

[31] Xiu Xiao Gao and Lai-Wan Chan. "An Algorithm for Trading and Portfolio Management Using Q-learning and Sharpe Ratio Maximization." In: 2000.

[32] Fernando García-Monleón, Ignacio Danvila-del Valle, and Francisco J Lara. "Intrinsic value in crypto currencies." In: *Technological Forecasting and Social Change* 162 (2021), p. 120393.

[33] Glassnode.com. *Glassnode Studio Ethereum: Number of Active Addresses*. 2022. URL: https://studio.glassnode.com/metrics?a=ETH&m=addresses.ActiveCount.

[34] Cleotilde Gonzalez. "249Decision-Making: A Cognitive Science Perspective." In: *The Oxford Handbook of Cognitive Science*. Oxford University Press, Oct. 2017. ISBN: 9780199842193. DOI: 10.1093/oxfordhb/9780199842193.013.6. eprint: https://academic.oup.com/book/0/chapter/295176048/chapter-ag-pdf/44511413/book_34641_section_295176048.ag.pdf. URL: https://doi.org/10.1093/oxfordhb/9780199842193.013.6.

[35] Ivo Grondman. "Online model learning algorithms for actor-critic control." In: (2015).

[36] Hado van Hasselt, Arthur Guez, and David Silver. "Deep Reinforcement Learning with Double Q-Learning." In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI'16. Phoenix, Arizona: AAAI Press, 2016, pp. 2094–2100.

[37] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie. "ENCASE: An ENsemble ClASsifiEr for ECG classification using expert features and deep neural networks." In: *2017 Computing in Cardiology (CinC)*. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.178-245.

[38] Zhenhan Huang and Fumihide Tanaka. "Behavioral Biases of Cryptocurrency Investors." In: *Available at SSRN 4280610* (2022).

[39] Zhenhan Huang and Fumihide Tanaka. "Investment Biases in Reinforcement Learning-based Financial Portfolio Management." In: *2022 61st Annual Conference of the Society of Instrument and Control Engineers (SICE)*. 2022, pp. 494–501. DOI: 10.23919/SICE56594.2022.9905789.

[40] Zhenhan Huang and Fumihide Tanaka. "MSPM: A modularized and scalable multi-agent reinforcement learning-based system for financial portfolio management." In: *PLOS ONE* 17.2 (Feb. 2022), pp. 1–24. DOI: 10.1371/journal.pone.0263689. URL: https://doi.org/10.1371/journal.pone.0263689.

[41] Hudson and Thames Quantitative Research. *Machine Learning Financial Laboratory (MlFinLab)*. https://github.com/hudson-and-thames/mlfinlab. 2021.

[42] John Manoogian III and Buster Benson. *The Cognitive Bias Codex - 180+ biases, designed by John Manoogian III and Buster Benson — Wikimedia Commons, the free media repository*. [Online; accessed 25-December-2022]. 2022. URL: https://commons.wikimedia.org/w/index.php?title=File:The_Cognitive_Bias_Codex_-_180%2B_biases,_designed_by_John_Manoogian_III_(jm3).png&oldid=718710619.

[43] InfoTrie. *FinSentS Web News Sentiment*. https://data.nasdaq.com/databases/NS1/data. 2021.

[44] Nishant Jagannath, Tudor Barbulescu, Karam M. Sallam, Ibrahim Elgendi, Braden Mcgrath, Abbas Jamalipour, Mohamed Abdel-Basset, and Kumudu Munasinghe. "An On-Chain Analysis-Based Approach to Predict Ethereum Prices." In: *IEEE Access* 9 (2021), pp. 167972–167989. DOI: 10.1109/ACCESS.2021.3135620.

[45] Huisu Jang and Jaewook Lee. "An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information." In: *Ieee Access* 6 (2017), pp. 5427–5437.

[46] Patel Jay, Vasu Kalariya, Pushpendra Parmar, Sudeep Tanwar, Neeraj Kumar, and Mamoun Alazab. "Stochastic neural networks for cryptocurrency price prediction." In: *Ieee access* 8 (2020), pp. 82804–82818.

[47] Zhengyao Jiang, Dixing Xu, and Jinjun Liang. *A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem*. arXiv:1706.10059. 2017. arXiv: 1706.10059 [q-fin.CP]. URL: https://arxiv.org/abs/1706.10059.

[48] Daniel Kahneman. "Maps of Bounded Rationality: Psychology for Behavioral Economics." In: *The American Economic Review* 93.5 (2003), pp. 1449–1475. ISSN: 00028282. URL: http://www.jstor.org/stable/3132137.

[49] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011, p. 98.

[50] Daniel Kahneman and Dan Lovallo. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." In: *Management Science* 39.1 (1993), pp. 17–31. DOI: 10.1287/mnsc.39.1.17. eprint: https://doi.org/10.1287/mnsc.39.1.17. URL: https://doi.org/10.1287/mnsc.39.1.17.

[51] Daniel Kahneman and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk." In: *Econometrica* 47.2 (1979), pp. 263–291. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1914185.

[52] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. DOI: 10.48550/ARXIV.1412.6980. URL: https://arxiv.org/abs/1412.6980.

[53] David Kirk. "NVIDIA Cuda Software and Gpu Parallel Computing Architecture." In: *Proceedings of the 6th International Symposium on Memory Management*. ISMM '07. New York, NY, USA: Association for Computing Machinery, 2007, pp. 103–104. ISBN: 9781595938930. DOI: 10.1145/1296907.1296909. URL: https://doi.org/10.1145/1296907.1296909.

[54] Robert C. Klemkosky and John D. Martin. "The Effect of Market Risk on Portfolio Diversification." In: *The Journal of Finance* 30.1 (1975), pp. 147–154. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2978437 (visited on 01/10/2023).

[55] Olivier Kraaijeveld and Johannes De Smedt. "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices." In: *Journal of International Financial Markets, Institutions and Money* 65 (2020), p. 101188. ISSN: 1042-4431. DOI:

https://doi.org/10.1016/j.intfin.2020.101188. URL: https://www.sciencedirect.com/science/article/pii/S104244312030072X.

[56] Alok Kumar. "Hard-to-Value Stocks, Behavioral Biases, and Informed Trading." In: *Journal of Financial and Quantitative Analysis* 44.6 (2009), pp. 1375–1401. DOI: 10.1017/S0022109009990342.

[57] Alok Kumar. "Who Gambles in the Stock Market?" In: *The Journal of Finance* 64.4 (2009), pp. 1889–1933. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/27735154.

[58] Alok Kumar and Sonya Seongyeon Lim. "How Do Decision Frames Influence the Stock Investment Choices of Individual Investors?" In: *Management Science* 54.6 (2008), pp. 1052–1064. ISSN: 00251909, 15265501. URL: http://www.jstor.org/stable/20122454.

[59] Rajesh Kumar. "1 - Perspectives on strategic finance." In: *Strategic Financial Management Casebook*. Ed. by Rajesh Kumar. Academic Press, 2017, pp. 1–29. ISBN: 978-0-12-805475-8. DOI: https://doi.org/10.1016/B978-0-12-805475-8.00001-X. URL: https://www.sciencedirect.com/science/article/pii/B978012805475800001X.

[60] Dominique Marcel Lammer, Tobin Hanspal, and Andreas Hackethal. *Who are the Bitcoin investors? Evidence from indirect cryptocurrency investments*. eng. SAFE Working Paper 277. urn:nbn:de:hebis:30:3-534536. Frankfurt a. M., 2020. DOI: 10.2139/ssrn.3501549. URL: http://hdl.handle.net/10419/218737.

[61] Maxim Lapan. *Deep Reinforcement Learning Hands-On: Apply Modern RL Methods, with Deep Q-Networks, Value Iteration, Policy Gradients, TRPO, AlphaGo Zero and More*. Packt Publishing, 2018, pp. 259–282. ISBN: 1788834240.

[62] Jonathan Law and John Smullen. *A Dictionary of Finance and Banking*. Oxford University Press, 2008. ISBN: 9780191726668. DOI: 10.1093/acref/9780199229741.001.0001. URL: https://www.oxfordreference.com/view/10.1093/acref/9780199229741.001.0001/acref-9780199229741.

[63] Jinho Lee, Raehyun Kim, Seok-Won Yi, and Jaewoo Kang. "MAPS: Multi-Agent reinforcement learning-based Portfolio management System." In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence* (2020). DOI: 10.24963/ijcai.2020/623.

[64] Bin Li and Steven C. H. Hoi. "Online Portfolio Selection: A Survey." In: *ACM Comput. Surv.* 46.3 (Jan. 2014). ISSN: 0360-0300. DOI: 10.1145/2512962. URL: https://doi.org/10.1145/2512962.

[65] Zhipeng Liang, Hao Chen, Junhao Zhu, Kangkang Jiang, and Yanran Li. *Adversarial Deep Reinforcement Learning in Portfolio Management*. arXiv:1808.09940. 2018. arXiv: 1808.09940 [q-fin.PM]. URL: https://arxiv.org/abs/1808.09940.

[66] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. *Continuous control with deep reinforcement learning*. 2019. arXiv: 1509.02971 [cs.LG]. URL: https://arxiv.org/abs/1509.02971.

[67] Long-Ji Lin. *Reinforcement learning for robots using neural networks*. Carnegie Mellon University, 1992.

[68] Yang Liu, Qi Liu, Hongke Zhao, Zhen Pan, and Chuanren Liu. "Adaptive Quantitative Trading: An Imitative Deep Reinforcement Learning Approach." In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.02 (2020), pp. 2128–2135. DOI: 10.1609/aaai.v34i02.5587. URL: https://ojs.aaai.org/index.php/AAAI/article/view/5587.

[69] Vivian F. López, Noel Alonso, Luis Alonso, and María N. Moreno. "A Multiagent System for Efficient Portfolio Management." In: *Trends in Practical Applications of Agents and Multiagent Systems*. Ed. by Yves Demazeau, Frank Dignum, Juan M. Corchado, Javier Bajo, Rafael Corchuelo, Emilio Corchado, Florentino Fernández-Riverola, Vicente J. Julián, Pawel Pawlewski, and Andrew Campbell. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 53–60. ISBN: 978-3-642-12433-4.

[70] Oscar Lundström and Måns Pettersson Spångäng. *Cryptocurrencies and Investor Disparities : A research paper about demographic factors' effect on investment purpose and herd behavior among Swedish cryptocurrency investors*. 2022.

[71] William J Luther. "Is bitcoin intrinsically worthless?" In: *AIER Sound Money Project Working Paper* 2018-07 (2018).

[72] H. B. Mann and D. R. Whitney. "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other." In: *The Annals of Mathematical Statistics* 18.1 (1947), pp. 50–60. ISSN: 00034851. URL: http://www.jstor.org/stable/2236101.

[73] Harry Markowitz. "Portfolio Selection." In: *The Journal of Finance* 7.1 (1952), pp. 77–91.

[74] Sanford I Millar. "Cryptocurrency expands online gambling." In: *Gaming Law Review* 22.3 (2018), pp. 174–174. DOI: https://doi.org/10.1089/glr2.2018.2232.

[75] Devin J. Mills and Lia Nower. "Preliminary findings on cryptocurrency trading among regular gamblers: A new risk for problem gambling?" In: *Addictive Behaviors* 92 (2019), pp. 136–140. ISSN: 0306-4603. DOI: https://doi.org/10.1016/j.addbeh.2019.01.005. URL: https://www.sciencedirect.com/science/article/pii/S0306460318311900.

[76] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. "Playing Atari with Deep Reinforcement Learning." In: (2013). cite arxiv:1312.5602Comment: NIPS Deep Learning Workshop 2013.

[77] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. "Human-level control through deep reinforcement learning." In: *nature* 518.7540 (2015), pp. 529–533.

[78] J. Moody and M. Saffell. "Learning to trade via direct reinforcement." In: *IEEE Transactions on Neural Networks* 12.4 (2001), pp. 875–889. DOI: 10.1109/72.935097.

[79] John Moody, Lizhong Wu, Yuansong Liao, and Matthew Saffell. "Performance functions and reinforcement learning for trading systems and portfolios." In: *Journal of Forecasting* 17.5-6 (1998), pp. 441–470. DOI: https://doi.org/10.1002/(SICI)1099-131X(1998090)17:5/6<441::AID-FOR707>3.0.CO;2-\#.

[80] John J. Murphy. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York Institute of Finance, 1999, pp. 24–31. ISBN: 0262039249.

[81] Vinod Nair and Geoffrey E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines." In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Madison, WI, USA: Omnipress, 2010, pp. 807–814. ISBN: 9781605589077.

[82] Satoshi Nakamoto. "Bitcoin: A peer-to-peer electronic cash system." In: *Decentralized Business Review* (2008), p. 21260.

[83] Terrance Odean. "Are Investors Reluctant to Realize Their Losses?" In: *The Journal of Finance* 53.5 (1998), pp. 1775–1798. DOI: https://doi.org/10.1111/0022-1082.00072.

[84] Terrance Odean. "Do Investors Trade Too Much?" In: *American Economic Review* 89.5 (1999), pp. 1279–1298.

[85] Ingram Olkin and Harold Hotelling. *Contributions to probability and statistics: Essays in honor of Harold Hotelling*. Stanford University Press, 1960, pp. 278–292.

[86] Mihály Ormos and András Urbán. "Performance analysis of log-optimal portfolio strategies with transaction costs." In: *Quantitative Finance* 13.10 (2013), pp. 1587–1597. DOI: 10.1080/14697688.2011.570368.

[87] Quandl. *End-Of-Day Data*. 2020. URL: https://www.quandl.com/data/EOD-End-of-Day-US-Stock-Prices (visited on 12/07/2020).

[88] QuoteMedia. *End-Of-Day Data*. https://data.nasdaq.com/data/EOD-end-of-day-us-stock-prices. 2020.

[89] Luis Rayo and Gary S. Becker. "Evolutionary Efficiency and Happiness." In: *Journal of Political Economy* 115.2 (2007), pp. 302–337. DOI: 10.1086/516737. eprint: https://doi.org/10.1086/516737. URL: https://doi.org/10.1086/516737.

[90] The Block Research and GSR. *The Block Research 2022 Digital Asset Outlook*. 2021. URL: https://www.tbstat.com/wp/uploads/2021/12/The-Block-Research-2022-Digital-Asset-Outlook.v2.pdf.

[91] Lior Rokach. "Ensemble-based classifiers." In: *Artif. Intell. Rev.* 33 (Feb. 2010), pp. 1–39. DOI: 10.1007/s10462-009-9124-7.

[92] Olga Romanchenko, Olga Shemetkova, Victoria Piatanova, and Denis Kornienko. "Approach of estimation of the fair value of assets on a cryptocurrency market." In: *The 2018 International Conference on Digital Science*. Springer. 2018, pp. 245–253.

[93] S. S. SHAPIRO and M. B. WILK. "An analysis of variance test for normality (complete samples)†." In: *Biometrika* 52.3-4 (Dec. 1965), pp. 591–611. ISSN: 0006-3444. DOI: 10.1093/biomet/52.3-4.591. eprint: https://academic.oup.com/biomet/article-pdf/52/3-4/591/962907/52-3-4-591.pdf. URL: https://doi.org/10.1093/biomet/52.3-4.591.

[94] Muhammad Saad, Jinchun Choi, DaeHun Nyang, Joongheon Kim, and Aziz Mohaisen. "Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions." In: *IEEE Systems Journal* 14.1 (2019), pp. 321–332.

[95] Santiment. *On-chain, Social and Financial API*. 2022. URL: https://api.santiment.net.

[96] Santiment. *daily-active-addresses*. 2022. URL: https://academy.santiment.net/metrics/daily-active-addresses/.

[97] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. *Proximal Policy Optimization Algorithms*. arXiv:1707.06347. 2017. arXiv: 1707.06347 [cs.LG]. URL: https://arxiv.org/abs/1707.06347.

[98] Esther-Mirjam Sent. "Rationality and bounded rationality: you can't have one without the other." In: *The European Journal of the History of Economic Thought* 25.6 (2018), pp. 1370–1386. DOI: 10.1080/09672567.2018.1523206. eprint: https://doi.org/10.1080/09672567.2018.1523206. URL: https://doi.org/10.1080/09672567.2018.1523206.

[99] Hersh Shefrin and Meir Statman. "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence." In: *The Journal of Finance* 40.3 (1985), pp. 777–790. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2327802.

[100] Aamna Al Shehhi, Mayada Oudah, and Zeyar Aung. "Investigating factors behind choosing a cryptocurrency." In: *2014 IEEE International Conference on Industrial Engineering and Engineering Management*. 2014, pp. 1443–1447. DOI: 10.1109/IEEM.2014.7058877.

[101] David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. "Deterministic Policy Gradient Algorithms." In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. ICML'14. Beijing, China: JMLR.org, 2014, pp. I–387–I–395.

[102] Herbert A. Simon. "A Behavioral Model of Rational Choice." In: *The Quarterly Journal of Economics* 69.1 (1955), pp. 99–118. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/1884852.

[103] Frank A. Sortino and Lee N. Price. "Performance Measurement in a Downside Risk Framework." In: *The Journal of Investing* 3.3 (1994), pp. 59–64. ISSN: 1068-0896. DOI: 10.3905/joi.3.3.59. URL: https://joi.pm-research.com/content/3/3/59.

[104] Thomas Spooner. *Algorithmic Trading and Reinforcement Learning: Robust methodologies for AI in finance*. The University of Liverpool (United Kingdom), 2021.

[105] Keith E. Stanovich and Richard F. West. "Individual differences in reasoning: Implications for the rationality debate?" In: *Behavioral and Brain Sciences* 23.5 (2000), pp. 645–665. DOI: 10.1017/S0140525X00003435.

[106] Statista. *Largest stock exchange operators worldwide as of October 2022, by market capitalization of listed companies*. 2022. URL: https://www.statista.com/statistics/270126/largest-stock-exchange-operators-by-market-capitalization-of-listed-companies/.

[107] Shuo Sun, Rundong Wang, and Bo An. "Reinforcement learning for quantitative trading." In: *arXiv preprint arXiv:2109.13851* (2021).

[108] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018. ISBN: 0262039249.

[109] Katia Sycara, K. Decker, and Dajun Zeng. "Designing a Multi-Agent Portfolio Management System." In: *Proceedings of the AAAI Workshop on Internet Information Systems*. 1995.

[110] Tõnn Talpsepp. "Reverse Disposition Effect of Foreign Investors." In: *Journal of Behavioral Finance* 12.4 (2011), pp. 183–200. DOI: 10.1080/15427560.2011.606387. URL: https://doi.org/10.1080/15427560.2011.606387.

[111] Tether. *What are Tether tokens and how do they work?* 2022. URL: https://tether.to/en/how-it-works.

[112] "The intraday dynamics of bitcoin." In: *Research in International Business and Finance* 49 (2019), pp. 71–81. ISSN: 0275-5319. DOI: https://doi.org/10.1016/j.ribaf.2019.01.008. URL: https://www.sciencedirect.com/science/article/pii/S027553191830552X.

[113] Horst Treiblmaier. "Do cryptocurrencies really have (no) intrinsic value?" In: *Electronic Markets* 32.3 (2022), pp. 1749–1758.

[114] Arianna Trozze, Josh Kamps, Eray Arda Akartuna, Florian J Hetzel, Bennett Kleinberg, Toby Davies, and Shane D Johnson. "Cryptocurrencies and future financial crime." In: *Crime Science* 11.1 (2022), pp. 1–35.

[115] Amos Tversky and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases." In: *Science* 185.4157 (1974), pp. 1124–1131. DOI: 10.1126/science.185.4157.1124. eprint: https://www.science.org/doi/pdf/10.1126/science.185.4157.1124. URL: https://www.science.org/doi/abs/10.1126/science.185.4157.1124.

[116] Amos Tversky and Daniel Kahneman. "Judgments of and by representativeness." In: *Judgment under Uncertainty: Heuristics and Biases*. Ed. by Daniel Kahneman, Paul Slovic, and AmosEditors Tversky. Cambridge University Press, 1982, pp. 84–98. DOI: 10.1017/CBO9780511809477.007.

[117] Amos Tversky and Daniel Kahneman. "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." In: *Psychological review* 90.4 (1983), p. 293.

[118] OpenAI Spinning Up. *Proximal Policy Optimization*. 2018. URL: https://spinningup.openai.com/en/latest/algorithms/ppo.html.

[119] Sha Wang and Jean-Philippe Vergne. "Buzz Factor or Innovation Potential: What Explains Cryptocurrencies' Returns?" In: *PLOS ONE* 12.1 (Jan. 2017), pp. 1–17. DOI: 10.1371/journal.pone.0169556. URL: https://doi.org/10.1371/journal.pone.0169556.

[120] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado van Hasselt, Marc Lanctot, and Nando de Freitas. *Dueling Network Architectures for Deep Reinforcement Learning*. 2016. arXiv: 1511.06581 [cs.LG].

[121] Yunan Ye, Hengzhi Pei, Boxin Wang, Pin-Yu Chen, Yada Zhu, Ju Xiao, and Bo Li. "Reinforcement-Learning Based Portfolio Management with Augmented Asset Movement Prediction States." In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.01 (2020), pp. 1112–1119. DOI: 10.1609/aaai.v34i01.5462.

[122] HowMuch.net, a financial literacy website. *Visualizing Bitcoin's Wild Ride in the Last Decade*. March 2020. URL: https://howmuch.net/articles/timeline-bitcoin-major-events.