JOURNAL OF ECONOMIC DEVELOPMENT
Volume 50, Number 3, September 2025

REINFORCEMENT LEARNING BASED DYNAMIC ASSET ALLOCATION WITH TECHNICAL AND MACRO-ECONOMIC ANALYSIS

JONG HA JEON [a], ZOONKY LEE [a] AND DOJOON PARK [b]

[a] *Yonsei University, South Korea*
[b] *Kongju National University, South Korea*

This study proposes a reinforcement learning based dynamic asset allocation framework that incorporates both market-based technical indicators and macroeconomic variables. The model aims to learn an optimal investment policy that maximizes cumulative portfolio returns while adapting to changing market regimes and economic conditions. We construct a multi-asset portfolio including equities, government bonds, corporate bonds, and cash equivalents, and train the reinforcement learning agent using Proximal Policy Optimization algorithms within a Markov Decision Process framework. Extensive ablation studies reveal that the inclusion of select macroeconomic variables enhances both portfolio returns and downside risk control. Regime-specific analyses confirm that macro-informed models outperform baseline strategies in market downturns while maintaining competitive performance during bull and bear markets. This study bridges the gap between the literature in the field of computer science and financial economics, offering an empirically validated, end-to-end decision-making tool for regime-aware, multi-asset portfolio management.

Keywords: Reinforcement Learning, Asset Allocation, Technical Indicators, Macroeconomic Variables, Multi-asset Portfolio
JEL Classification: C51, C52, G11, E44

1. INTRODUCTION

The main objective of this study is to develop a reinforcement learning based asset allocation model that optimizes multi-asset portfolios by incorporating variations in asset prices and shifts in the macroeconomic environment. For sophisticated investors, portfolio decisions can be guided by expectations of future returns, influenced by asset prices and macroeconomic conditions. Previous studies have addressed the portfolio optimization problem, ranging from mean-variance optimization to dynamic asset allocation strategies that incorporate regime switching and changes in the market condition and business cycle. However, these methods typically rely on static inputs and raise concerns regarding potential estimation errors (Chopra and Ziemba, 1993). Recent developments in reinforcement learning have elucidated this field by enabling dynamic data-driven learning frameworks that can effectively mitigate these limitations.

Despite the advantages of reinforcement learning based asset allocation models, research in this area remains in its nascent stages. In particular, few studies have focused on integrating insights from widely studied topics in the finance and economic literature, such as business cycles and regime changes, into reinforcement learning based frameworks (Kelly and Xiu, 2023). Existing approaches that incorporate regime-switching dynamics often rely solely on the price information of individual assets to detect regime shifts, thereby underutilizing broader macroeconomic variables. This is a significant gap, considering that macroeconomic regime analysis using variables such as inflation, employment, and industrial activity is not only prevalent in academic research, but is also widely adopted by investors or portfolio managers in the financial industry to guide portfolio allocation decisions.

This study proposes an autonomous asset allocation model that leverages key variables emphasized in both financial literature and practice, combined with reinforcement learning. The model is designed to learn an optimal investment policy that maximizes cumulative returns, enabling it to make asset allocation decisions independently. Specifically, the model is trained using three categories of information: (1) asset prices and trading volume, (2) technical indicators derived from prices and volume data, and (3) macroeconomic variables. Through continuous interaction with the changes in the market environment, the model dynamically adjusts its investment policy in response to the market and macroeconomic conditions.

This study employs exchange-traded funds (ETFs) listed in the U.S. market that represent four primary asset classes: equities, government bonds, corporate bonds, and cash equivalents. Using ETFs enhances the practical feasibility of implementing the models' proposed strategy. The reinforcement learning agent is trained using asset price data, technical indicators, and macroeconomic variables as state inputs. The objective is to learn an optimal investment policy that maximizes cumulative portfolio returns. Based on this policy, the agent autonomously reallocates portfolio weights on a weekly basis. We propose an end-to-end model in which the agent learns continuously from both changes in market conditions and portfolio performance. Hence, the agent can make dynamic and informed asset allocation decisions.

This paper examines whether reinforcement learning based portfolio strategies deliver superior investment performance. Following are the two main contributions. First, we augment the commonly used asset price data with macroeconomic variables that reflect the changes in market condition and economic cycle in constructing the reinforcement learning models. While reinforcement learning has been applied to various field in finance, its application to portfolio optimization has been limited and paid little attention to macroeconomic considerations. We address this gap by aligning reinforcement learning frameworks more closely with financial economic theory. By incorporating macroeconomic data, we design models that learn efficiently from a carefully selected set of economically informative state variables, thereby enhancing the performance of reinforcement learning based portfolio optimization. Second, we expand the investment universe to include equities, government bonds, corporate bonds, and cash equivalents, thereby aligning the optimization problem with the strategic asset allocation decisions faced by sophisticated investors. The sophisticated investors usually begin with an asset allocation decision across broad asset classes before selecting individual stocks. Whereas most prior studies focus on stock selection, our focus on multi-asset allocation offers insights that are both theoretically grounded and practically relevant for institutional and individual investors alike. We propose an adaptive asset allocation framework that is applicable to multi-asset portfolios and robust across varying macroeconomic regimes.

The remainder of this study is organized as follows. Section 2 reviews the related literature. Section 3 introduces the reinforcement learning model used in this study, including the architecture, investment universe, and training data sets. Section 4 outlines the analysis methodology. Section 5 presents the results, and Section 6 concludes the paper with a summary of the contributions and suggestions for future research.

2. LITERATURE REVIEW

2.1. Portfolio Optimization

The theory of portfolio optimization explores how investors can construct and manage portfolios comprising multiple assets to achieve an optimal balance between return and risk. The foundational work in this field was established by Markowitz (1952), who introduced the mean-variance optimization framework. Through appropriate weighting of diverse assets, they aim to construct efficient portfolios that deliver the highest possible return for a given level of risk.

Building on this foundation, Merton (1969) extended portfolio theory from the static framework to an intertemporal formulation. Additionally, academic attention has increasingly turned toward developing asset allocation strategies that incorporate macroeconomic conditions and their transitions over time. Hamilton (1996) demonstrated that the volatility of equity returns varies depending on the phase of the business cycle, with recessions exerting particularly strong effects. He proposed a regime-switching model that uses economic variables, such as industrial production, to predict the changes in macroeconomic conditions. Recently, Vliet et al. (2011) uses economic variables to forecast business cycles as defined by the National Bureau of

Economic Research. Their approach dynamically adjusts asset allocations based on the predicted economic cycle, thereby improving responsiveness to macroeconomic fluctuations.

Portfolio optimization research has been based on historical and expected returns as well as the covariances between assets. However, this approach presents several notable limitations. First, the theory uses estimated or expected covariances. In reality, asset volatilities and correlations are highly dynamic and often change significantly under turbulent market conditions. During periods of heightened volatility, asset correlations tend to increase, thereby reducing the benefits of diversification and leading to instability in mean-variance optimized portfolios (López de Prado, 2016). Second, portfolio optimization is highly sensitive to estimation errors in variables. Expected returns and volatilities are typically derived from historical data, and any inaccuracies in these estimates can significantly distort optimization outcomes. Given the inherent uncertainty in forecasting future returns, the reliability of such estimates is often questionable. Third, mean-variance optimization is based on a single-period framework, which limits its applicability in dynamic, multi-period investment environments. Finally, conventional statistical approaches have difficulty capturing the complex non-linear relationships often observed in financial data (Sugadev et al., 2023).

To address these issues, researchers have explored regime-switching frameworks that dynamically incorporate changing market conditions into asset allocation strategies (Collin-Dufresne et al., 2022). Methods for identifying regime shifts -changes that significantly affect asset returns, volatilities, and covariances- typically rely either on macroeconomic variables or inferences drawn from asset price dynamics. Incorporating regime-switching into asset allocation represents an important advancement beyond static optimization models, offering improved adaptability to evolving market environments. However, conventional methods still struggle to capture the complex non-linear relationships and dynamics prevalent in financial data. Typically, these models first classify the market into discrete regimes and then proceed with optimization procedures based on regime-specific assumptions about expected returns and covariances.

Reinforcement learning models interact with the environment and learn optimal policies through a process of trial and error, guided by feedback in the form of rewards. Statistical methods can be affected by regime misclassification and inaccurate estimation of regime-specific moments. However, reinforcement learning does not rely on such predictive estimates, which helps mitigate the impact of model misspecification and estimation errors. In addition, reinforcement learning models may demonstrate superior performance in capturing the non-linear relationships between explanatory variables and asset prices. Rather than forecasting asset returns and volatilities for the purpose of optimization, reinforcement learning models aim to characterize dynamic market environments and derive asset allocation strategies based on the available information set. Despite its advantages, reinforcement learning has not yet been widely adopted in the finance literature. However, a growing body of research has begun to explore the application of reinforcement learning to portfolio optimization, as detailed in the next subsection.

2.2. Portfolio Optimization Using Reinforcement Learning

Recent advancements in reinforcement learning have led to significant performance improvements across a wide range of research domains, including finance. In the context of portfolio optimization, reinforcement learning algorithms are typically categorized based on whether or not they require an explicit model of the environment. This results in two primary classes: model-based and model-free approaches. Model-free methods can be further divided into value-based and policy-based approaches.

Value-based agents learn only a value function while maintaining an implicit policy. Notable algorithms in this category include State-Action-Reward-State-Action (SARSA), Q-learning, and Deep Q-Networks. One advantage of this approach is that it typically makes more efficient use of available data. In contrast, policy-based agents learn a policy directly without estimating a value function. Examples include algorithms based on policy gradients. This approach offers the benefit of directly optimizing the desired objective, often resulting in more stable learning outcomes. Some agents combine both value and policy learning, and these are known as Actor-Critic models. They leverage the strengths of both value-based and policy-based methods (Tang, 2018).

Notable studies using value-based approaches include the work of Yang et al. (2018), who applied both Q-learning and Recurrent Reinforcement Learning (RRL) to evaluate trading performance between risky and risk-free assets. The study tested multiple value functions, including internal profit, the Sharpe ratio, and variants of the Sharpe ratio. Among these, models utilizing the variants of the Sharpe ratio and RRL demonstrated more stable and superior performance outcomes. Pendharkar et al. (2018) addressed the retirement portfolio optimization problem using value-based reinforcement learning, specifically employing on-policy SARSA and off-policy Q-learning algorithms. The portfolio consisted of two asset classes: the S&P 500 Index and a broad bond market index. The objective was to maximize either total return or the Sharpe ratio. The authors defined market state variables by segmenting asset returns into four distinct regimes, enabling the model to learn customized representations of the market environment.

Park et al. (2020) explored a portfolio trading strategy using deep Q-learning, extending previous research from single-asset settings to multi-asset portfolios. The constructed portfolio included cash, large-cap, mid-cap, and small-cap equity indices, allowing the model to learn style-based allocation among risky assets. The state variables consisted of five technical indicators, such as closing return, intraday highs and lows, to inform decision-making. Das et al. (2024) proposed a Q-learning-based portfolio optimization framework to address the goal-based wealth management problem. The study demonstrated that reinforcement learning provides a flexible solution in environments characterized by path-dependency and large state spaces -key challenges in goal-based wealth management applications.

Research on reinforcement learning based portfolio optimization has explored a wide array of algorithmic frameworks and parameter configurations. The types of assets used in these studies vary considerably, ranging from individual stocks to equity-bond indices and style-based indices. Most of the models rely heavily on asset price data and technical indicators as input features. However, studies that incorporate macroeconomic variables as state inputs remain relatively scarce. The limited incorporation of macroeconomic variables constrains the models' ability to adapt to broader economic regime changes.

2.3. Regime-Switching and Reinforcement Learning

Research on regime-switching in the context of reinforcement learning dates back to Moody and Wu (1997), which introduced RRL to directly optimizing trading strategies and portfolios. Building on this, Maringer and Ramtohul (2012) proposed a method to identify market regimes by analyzing stock price data using RRL techniques. Bauman et al. (2024) introduced a hybrid approach that integrates regime identification via Hidden Markov Models into reinforcement learning environments. Specifically, they classified market conditions into two regimes based on CPI-adjusted returns, and incorporated the regime-switching signals into the reinforcement learning framework to construct a portfolio optimization model.

Although the integration of regime-switching into reinforcement learning-based portfolio optimization is still in its early stages, some studies have made initial at the portfolio theory. These approaches are notable for introducing regime information as an additional feature. However, they tend to rely primarily on asset price data, thereby capturing momentum-based signals rather than broader macroeconomic dynamics. Such methods implicitly assume that past asset returns are predictive of future returns. However, this assumption overlooks the broader context of financial market shifts, which cannot be fully understood through asset price movements alone.

In the financial literature, regime analysis often extends beyond asset prices to include macroeconomic variables that capture changes in the broader business cycle. Similarly, in the asset management industry, regime-based asset allocation strategies that rely on macroeconomic indicators are widely implemented in practice. Despite this, reinforcement learning-based portfolio optimization research has rarely incorporated macroeconomic variables as part of the state representation within the learning environment (Fischer, 2018). To address this gap, this study develops a reinforcement learning based model that explicitly includes macroeconomic variables previously underexplored in the reinforcement learning literature as part of its input features. This enables the model to learn and respond to changes in the economic environment, adapting its asset allocation decisions accordingly.

3. REINFORCEMENT LEARNING MODEL

3.1. Problem Description

This study formulates an asset allocation task as a reinforcement learning problem, wherein the agent dynamically adjusts the weights of a multi-asset portfolio to maximize the cumulative portfolio return. Because reinforcement learning models learn through interactions with an environment based on the observed states and received rewards, the selection of input data is a critical design consideration. Financial markets are highly complex systems in which asset prices, macroeconomic conditions, and investor sentiment interact dynamically. To address this complexity, the reinforcement learning environment is carefully structured to allow the agent to learn from both evolving price patterns and macroeconomic regimes.

The agent is trained on a comprehensive state space that includes (1) raw asset price data, (2) technical indicators used to measure momentum, trading volume dynamics, and price volatility, (3) macroeconomic indicators that signal shifts in the economic condition and business cycle, (4) covariances between asset returns, and (5) reward feedback based on portfolio performance. To ensure that the agent can effectively learn from this diverse set of features, dimension reduction techniques are employed to manage the complexity of the input space.

The investment universe is composed of four major asset classes, and the initial portfolio is constructed using equal weighting. The reinforcement learning agent rebalances the portfolio on a weekly basis throughout the training and testing periods. We propose an end-to-end reinforcement learning framework in which the agent autonomously learns an optimal asset allocation policy by continuously interacting with portfolio data, asset-level signals, and market conditions.

3.2. Reinforcement Learning Model

The asset allocation problem is formulated as a Markov decision process, which provides a framework for sequential decision-making in dynamic environments. In the context of portfolio optimization, the Markov decision process enables agents to make investment decisions that maximize expected rewards while accounting for the inherent uncertainty and temporal variability of financial markets. In the Markov decision process framework, the agent learns from past states and rewards to determine optimal actions at each time step. This process allows the agent to maximize investment performance while managing risk. By leveraging reinforcement learning algorithms, the agent can learn directly from data and dynamically adapt its asset allocation policy over time.

The Markov decision process is defined by five core components: the state space S, action space A, state transition probabilities Pa, reward function Ra, and discount factor γ gamma. At each time step t, the agent observes a state S_t, selects an action A_t, receives a reward R_t, and transitions to a new state S_{t+1}. The objective of reinforcement learning is to learn a policy that maximizes the expected cumulative reward over time by iteratively interacting with the environment.

The agent begins with an initial portfolio balance at time $t=0$ and takes actions at each subsequent time step. While time intervals can be configured on a daily, weekly, monthly, or annual basis depending on the research objective, this study adopts a weekly frequency to account for asset liquidity and transaction costs. In real-world trading, transactions typically incur costs that vary by asset type and counterparty. For simplicity, this study assumes a transaction cost of 0.1% of the trade value for both buy and sell operations. Market impact is ignored, and it is assumed that trades can be executed continuously without delay. The state represents the agent's perception of the market environment at each time step. In this model, the state vector includes information such as account balance, the number of shares held for each asset, portfolio performance metrics, technical indicators, macroeconomic variables, and the covariance matrix between asset returns. Given the weekly rebalancing structure of the portfolio, the time step for the reinforcement learning process is set to a weekly frequency.

In this study, ten technical indicators are incorporated into the state space to capture information on volatility, trading volume, and momentum. These indicators help detect trends, momentum shifts, and buying or selling pressures -insights that cannot be easily derived from raw price data alone. Macroeconomic indicators included in the state space offer insights into broader economic conditions. Since asset prices and cross-asset correlations are sensitive to fluctuations in inflation and economic growth, the integration of macroeconomic variables enables the reinforcement learning agent to make more informed and context-aware allocation decisions, ultimately enhancing portfolio performance.

The simulation assumes an initial balance of USD on million. At each time step, the portfolio's net asset value (NAV) is computed based on the asset-specific returns observed within the corresponding state. The reward function in this study is defined as the portfolio return. The agent receives rewards based on realized portfolio performance and learns a policy aimed at maximizing cumulative returns over time. Both buy and sell transactions are assumed to incur a fee of 0.1% of the trade value. At each time step, this transaction cost is deducted from portfolio returns and incorporated into the reward calculation, ensuring that the agent learns under realistic cost constraints.

3.3. Deep Reinforcement Learning Agent and Investment Universe

The training algorithms used Proximal Policy Optimization (PPO). The PPO is an on-policy method that addresses some of the stability issues associated with traditional policy gradient algorithms. By restricting the size of policy updates using a clipping mechanism, PPO enhances training stability and has been shown to perform robustly in complex environments, such as financial markets. The PPO incorporates a clipping mechanism within its objective function to constrain the magnitude of policy updates. This helps prevent drastic policy changes, thereby promoting more stable and controlled learning.

This study constructs a portfolio comprising equities, U.S. Treasury bonds, corporate bonds, and Treasury bills to reflect the distinct risk-return characteristics of core asset classes, as detailed in Table 1. These assets exhibit heterogeneous risk profiles and varying correlations across market regimes: equities represent high-risk assets, treasury bonds involve interest rate risk without credit risk, corporate bonds entail both credit and interest rate risk, and treasury bills serve as near-risk-free instruments. This framework facilitates the analysis of cross-asset diversification and hedging effects and may be extended in future research to include alternative assets, such as gold or emerging market debt.

Table 1. The List of Investable Assets

Ticker	Asset class	Security Name	Structure
SPY	Equity	SPDR S&P 500	ETF
TLT	Treasury Bonds	iShares 20+ Year Treasury Bond	ETF
LQD	Corporate Bonds	iShares iBoxx $ Investment Grade Corporate Bond	ETF
BIL	Treasury Bill	SPDR Bloomberg 1-3 Month T-Bill	ETF

Note: This table presents the list of investable assets used in this study to construct the portfolio. Each asset is represented by its ticker symbol and asset class. All instruments are exchange-traded funds (ETFs) and listed primarily on the NYSE or NASDAQ. All returns are computed from adjusted close prices with dividends reinvested.

3.4. Data Sets

The model utilizes a comprehensive set of features, including five basic features for each ETF, ten technical indicators, and six macroeconomic indicators. Table 2 presents the key technical indicators employed to build the model. The five basic features of each asset are the closing price, highest price, lowest price, opening price, and trading volume. These metrics capture essential information on asset price movements and liquidity conditions. The model was trained using ten technical indicators, including Volatility Average True Range, Volatility Bollinger Band Width, and Momentum Relative Strength Index, among others. These indicators are designed to capture key aspects of market behavior such as price momentum, volatility, and overbought or oversold conditions -features that are not readily apparent from raw price data alone.

Table 3 presents macroeconomic variables. Macroeconomic data comprise five key indicators that reflect the health and direction of the U.S. economy. All macroeconomic indicators are released on a monthly basis by their respective agencies. The following five variables are widely recognized in both academic research and industry practice as effective tools for distinguishing different phases of the economic cycle.

Table 2: Technical Indicators

Indicators	Key measurements
Volatility Average True Range	Volatility of a price movement
Volatility Bollinger Band Width	Volatility of a Bollinger Band
Volume On-balance Volume	Trading volume
Volume Chaikin Money Flow	Trading volume
Trend Moving Average Convergence Divergence	Momentum / Trend
Trend Average Directional Index	Momentum / Trend
Trend Fast Simple Moving Average	Momentum / Trend
Trend Fast Exponential Moving Average	Momentum / Trend
Trend Commodity Channel Index	Momentum / Trend
Momentum Relative Strength Index	Momentum / Trend

Note: The table outlines the key technical indicators employed to build the model. Indicators span volatility, trading volume, and momentum or trend.

Table 3. Macroeconomic Variables

Macro variables	Descriptions	Reporting Agency
Consumer Price Index (CPI)	Month-over-month change in Consumer Price Index, a measure of inflation	Bureau of Labor Statistics
Non-farm Payrolls	Employment growth in the non-agricultural sector	Bureau of Labor Statistics
Unemployment Rate	Reflects overall economic health and recessionary pressures	Bureau of Labor Statistics
Leading Economic Index (LEI)	Composite indicator used to forecast future economic activity	The Conference Board
ISM Manufacturing New Orders Index (ISM)	Measures new orders in the manufacturing sector	Institute for Supply Management

Note: The table presents macroeconomic variables. These variables are released on a monthly basis by their respective agencies and mapped to the weekly reinforcement learning environment by carrying the latest released value forward until the next release. The data for each indicator were obtained from the Federal Reserve Bank of St. Louis FRED database.

- US Consumer Price Index (CPI), MoM: CPI is a key macroeconomic indicator that measures the average change over time in the prices of a fixed basket of goods and services typically consumed by urban households
- US Non-Farm Payrolls (NFP): This metric captures monthly employment growth in the non-agricultural sector and serves as a timely gauge of labor market strength.
- US Unemployment Rate: Calculated by the U.S. Bureau of Labor Statistics (BLS) through the monthly Current Population Survey (CPS), it serves as a coincident indicator of economic activity, capturing contemporaneous shifts in labor demand.

- US Leading Economic Index (LEI): Published by the Conference Board, the LEI aggregates several forward-looking variables to project the direction of future economic activity.
- ISM Manufacturing New Orders Index (ISM): As a component of the broader ISM manufacturing index, this indicator reflects demand conditions in the manufacturing sector and is closely watched for early signs of economic expansion or contraction.

Kim (2023) identifies economic growth and inflation indicators as key measures of the business cycle, noting that asset returns, volatility, and correlations tend to shift with changes in these macro trends. Based on this, CPI and growth-related indicators were included as core macro variables. Inflation is a key determinant that significantly influences asset returns. (Marshall, 1992). Regarding labor market indicators, Goldberg and Grisse (2013) confirm that payrolls announcements exert a significant influence across the U.S. yield curve. Hornstein (2016) highlights the unemployment rate as a robust indicator of recession risk and notes its significant impact on asset performance. Boyd et al. (2005) examines the reaction of equity markets to unemployment rate announcements. The authors demonstrate that the market's response is conditional on the state of the business cycle: stock prices tend to rise in response to higher unemployment during economic expansions, but decline during contractions.

Long et al. (2022) provide robust evidence that leading indicators possess significant predictive power for equity returns and can serve as effective tools for enhancing investment strategies. Specifically, the LEI in the United States is a composite measure that combines key economic and financial indicators -such as stock prices, building permits, and interest rate spreads- to anticipate future economic activity, particularly turning points in the business cycle. McGuckin (2004) demonstrated that the LEI significantly enhances the predictive power for the Composite Index of Coincident Indicators, a broad monthly measure of current economic conditions. These macroeconomic variables were incorporated into the model to allow the agent to learn and respond to changes in the macroeconomic regime during the investment decision-making process.

The training dataset spans from May 7, 2012, to June 28, 2018, a period characterized by relatively stable post-crisis recovery in global financial markets following the aftermath of the global financial crisis of 2008–2009 and the European sovereign debt crisis. This training window provides the model with exposure to standard cyclical fluctuations under moderate monetary policy conditions. In contrast, the testing period extends from July 4, 2018, to December 19, 2022, and includes several distinct macroeconomic regimes, most notably the unprecedented volatility observed during the COVID-19 pandemic. This out-of-sample window captures a wide range of market dynamics, including sharp contractions in economic activity, aggressive monetary and fiscal interventions, inflationary pressures, and shifts in investor sentiment. By incorporating both relatively normal and highly turbulent conditions, the dataset ensures a robust evaluation of the model's ability to generalize across diverse economic environments and stress scenarios.

4. ANALYSIS METHODOLOGY

4.1. Ablation Studies

Ablation study has emerged as a critical methodological tool in machine learning research, enabling researchers to systematically evaluate the marginal contribution of individual model components or input variables to overall model performance (Souto and Louzada, 2024). The core idea of an ablation study is to iteratively remove or modify specific features, data categories, or modules and observe the resulting impact on key evaluation metrics such as Sharpe ratio, cumulative returns, and prediction accuracy.

Table 4. Models and Corresponding Data Sets

Models	Descriptions / Trained data sets
UW	Uniform Weight
T	Basic and Technical indicators
M	Basic and Macroeconomic variables
T+M	Basic, Technical, and Macroeconomic variables
T+1a	Basic and Technical indicators, CPI
T+1b	Basic and Technical indicators, ISM
T+1c	Basic and Technical indicators, Labor
T+1d	Basic and Technical indicators, LEI
T+2a	Basic and Technical indicators, CPI, ISM
T+2b	Basic and Technical indicators, CPI, Labor
T+2c	Basic and Technical indicators, Labor, ISM
T+2d	Basic and Technical indicators, Labor, LEI
T+2e	Basic and Technical indicators, ISM, LEI
T+3	Basic and Technical indicators, CPI, ISM, Labor

Note: This table presents the configuration of the trained models and corresponding input feature sets. Variable definitions are provided in Table 3. Basic denotes the five features of closing price, highest price, lowest price, opening price, and trading volume. Labor comprises the U.S. non-farm payrolls and the U.S. unemployment rate. Model acronyms: UW (uniform 25% each), T (technical only), M (macro only), T+M (technical + macro), and T+1/2/3 variants (selected macro subsets).

In this study, we employ an ablation framework to evaluate the impact of different groups of input variables on the performance of a reinforcement learning-based asset allocation model. Table 4 shows the configuration of the trained models and corresponding input feature sets. These input groups include technical indicators, macroeconomic indicators such as CPI, ISM, LEI, and labor market variables, as well as their various combinations. Specifically, we compare the performance of four models: a uniform weight (UW) strategy with equal allocation, a technical-only (T) model using only technical indicators, a macro-only (M) model using only macroeconomic variables, and a combined (T+M) model incorporating both feature types. As shown in Table 4, we further analyzed different combinations of macroeconomic variables to evaluate the marginal contribution of each indicator and identify the key variables that drive superior performance. Through this process, we aim to identify which variable combinations yield statistically and economically superior allocation policies and provide guidance for future model designs and variable selection strategies in data-driven investment frameworks.

Portfolio performance is evaluated using multiple metrics, including the cumulative return, Sharpe ratio, volatility of returns, maximum drawdown, Sortino ratio, and the volatility of negative excess returns, which are calculated based on returns that fall below risk-free rates. These indicators enable a comprehensive assessment of both the return and risk characteristics.

4.2. Performance Analysis Across Market Regimes

To evaluate the model's adaptability to market dynamics, the testing period was divided into six regimes -three bull markets and three corrections- based on the S&P 500 trends. This classification captures both bull and bear markets over a four-and-a-half-year period. This segmentation captures the shifts in financial market asset correlations. To quantify the evolving inter-asset relationships across regimes, we calculate the beta of government and corporate bonds relative to equities during both bull and correction phases. This enables us to observe how these sensitivities shift depending on broader market conditions. As presented in Table 5, in regimes 1 through 3, government bonds exhibited a negative beta to equities, consistent with conventional market behavior. In contrast, regimes 4 through 6 showed a breakdown in this inverse relationship, with both asset classes declining simultaneously. These observations highlight the importance of dynamically adjusting asset allocation policies to reflect changing inter-asset dependencies.

Table 5. Betas of Each Asset Against SPY

Phase	Duration (weeks)	BIL	LQD	SPY	TLT
1 – 6	233	0.00	0.39	1.00	-0.03
1 – 3	92	0.00	0.35	1.00	-0.28
4 – 6	141	0.00	0.45	1.00	0.26

Note: This table presents the regime-specific beta coefficients of BIL, LQD, and TLT relative to SPY. The betas are computed using weekly returns over each regime period. The data spans the period from July 2018 to December 2022.

This study evaluates the adaptability of each reinforcement learning model by analyzing how asset allocations were adjusted across market regimes. The excess returns generated by each model are benchmarked against the uniform weight strategy to assess relative performance under varying market conditions. In addition, portfolio performance is evaluated using multiple metrics, including cumulative return, average return, Sharpe ratio, Sortino ratio, maximum drawdown, and the volatility of negative excess returns. These indicators enable a comprehensive assessment of both return and risk characteristics.

5. RESULTS

5.1. Performance Evaluation

In this section, we evaluate the trained models. Table 6 reports the cumulative return, annualized return, volatility, and the volatility of negative excess returns relative to the risk-free rate over the test period. The UW model, which allocates 25% to each of the four asset classes, achieved a cumulative return of 13.89% and a volatility of 8.94%. These values serve as a baseline for evaluating the performance of alternative models. The T model, trained exclusively using technical indicators, achieved a cumulative return of 17.02%, outperforming the UW model by 3.13 percentage points. Both the overall volatility and the volatility of negative returns -defined as the standard deviation of returns falling below the risk-free rate- increased relative to UW.

Next, we examine the performance of models that incorporate macroeconomic variables. The M model generated higher returns than the UW model, but underperformed relative to the T model. The T+M model, which combines technical and macroeconomic features, achieved the highest cumulative return, improving on UW by 3.74 percentage points. Although the portfolio's volatility increased to 9.18%, the volatility of negative returns decreased slightly, from 7.21% to 7.18%. This suggests that the T+M model achieved superior returns and improved downside risk management. The performance and volatility outcomes can be further elucidated by comparing risk-adjusted return metrics.

The risk-adjusted performances of the trained models are examined using the Sharpe ratio, the Sortino ratio and the maximum drawdown (MDD). Table 7 shows the results. Although the volatility of the T model increased slightly, its Sharpe ratio improved from 0.19 to 0.24, reflecting enhanced risk-adjusted performance. However, the model's maximum drawdown marginally widened to -11.98%. This increase in both volatility and MDD is likely due to the reinforcement learning agent’s reliance on technical indicators, particularly momentum signals, suggesting that the model adopts a trend-following allocation strategy which performs well in sustained upward markets.

On the other hand, the M model produced higher returns and a superior Sharpe ratio compared to the UW model but still lagged behind the T model. Its maximum drawdown reached -12.62%, indicating a deterioration in downside protection compared to -11.54%

for the UW model. The T+M model demonstrated notable improvements in both return and risk-adjusted performance. Its Sharpe ratio increased significantly to 0.27, and its maximum drawdown fell to -10.06%, demonstrating effective mitigation of downside risk. This is further evidenced by an increase in the Sortino ratio to 0.34, representing an improvement of 0.11 over the UW model.

To better understand the behavior of each model, we analyzed the portfolio performance and asset allocation patterns across market regimes, specifically during bull and bear markets, throughout the test period. Table 8 illustrates the relative return performance of each model across different market regimes. In bull markets, the T model generated average weekly excess returns of 1.82 basis points, compared to 0.45 basis points in bear markets. This outcome is consistent with the design of the model, which uses technical indicators to identify trends in price and volume.

By contrast, the M model performed better in bear markets, achieving an average weekly excess return of 1.70 basis, which was 1.50 basis points higher than its performance in bull markets. The T+M model exhibited the most robust performance under adverse market conditions, generating 3.70 basis points of weekly outperformance in bear markets and surpassing the M model's performance. This aligns with earlier findings that the T+M model demonstrates improved measures of downside risk, such as lower negative return volatility and a higher Sortino ratio. It also suggests that models incorporating macroeconomic variables are more effective at mitigating downside risk.

Table 6. Returns and Volatilities

Model	**Cumulative Return**	**Rel. Perf.**	**Annualized Return**	**Rel. Perf.**	**Vol**	**Rel. Perf.**	**Vol^{NegR}**	**Rel. Perf.**
UW	13.89	-	2.97	-	8.94	-	7.21	-
T	17.02	3.12	3.60	0.63	9.74	0.79	7.66	0.45
M	15.68	1.79	3.33	0.36	9.34	0.39	7.41	0.19
T+M	17.63	3.74	3.72	0.75	9.18	0.23	7.18	-0.03
T+1a	19.31	5.42	4.05	1.08	8.82	-0.12	6.06	-1.16
T+1b	13.20	-0.70	2.83	-0.14	9.54	0.60	7.78	0.57
T+1c	17.61	3.72	3.71	0.75	8.94	0.00	6.87	-0.34
T+1d	17.38	3.48	3.67	0.70	9.50	0.56	7.44	0.22
T+2a	27.22	13.32	5.56	2.59	9.32	0.38	6.75	-0.46
T+2b	26.00	12.11	5.33	2.36	9.41	0.47	7.46	0.24
T+2c	24.20	10.30	4.99	2.02	8.54	-0.41	6.16	-1.05
T+2d	16.24	2.34	3.44	0.47	8.96	0.02	6.33	-0.89
T+2e	6.08	-7.81	1.34	-1.63	9.00	0.06	6.88	-0.33
T+3	15.62	1.72	3.32	0.35	8.49	-0.45	6.67	-0.54

Note: This table reports the cumulative return, annualized return, volatility (Vol), and the volatility of negative excess returns relative to the risk-free rate (Vol^{NegR}) over the test period for each model, expressed in percentage terms. Models are defined in Table 4. The 3-month U.S. Treasury bill rate is used for the risk-free interest rate. Relative performance to the uniform weight (Rel. Pef.) strategy is reported in percentage points. Volatility is calculated as the standard deviation of weekly returns. All data are sampled from July 2018 to December 2022.

Table 7. Performance Metrics

Portfolio	Sharp ratio	Rel. Perf.	Sortino ratio	Rel. Perf.	MDD (Monthly)	Rel. Perf.
UW	0.19	-	0.24	-	-11.54	-
T	0.24	0.05	0.31	0.07	-11.98	-0.44
M	0.22	0.03	0.28	0.04	-12.62	-1.08
T+M	0.27	0.08	0.34	0.11	-10.06	1.48
T+1a	0.32	0.13	0.46	0.22	-9.28	2.26
T+1b	0.17	-0.03	0.20	-0.04	-12.77	-1.23
T+1c	0.28	0.08	0.36	0.12	-8.41	3.13
T+1d	0.25	0.06	0.32	0.09	-13.50	-1.96
T+2a	0.46	0.27	0.64	0.40	-9.39	2.15
T+2b	0.43	0.24	0.55	0.31	-10.87	0.67
T+2c	0.44	0.25	0.61	0.37	-6.89	4.65
T+2d	0.24	0.05	0.35	0.11	-9.70	1.84
T+2e	0.01	-0.18	0.01	-0.23	-11.02	0.52
T+3	0.24	0.05	0.31	0.07	-8.38	3.16

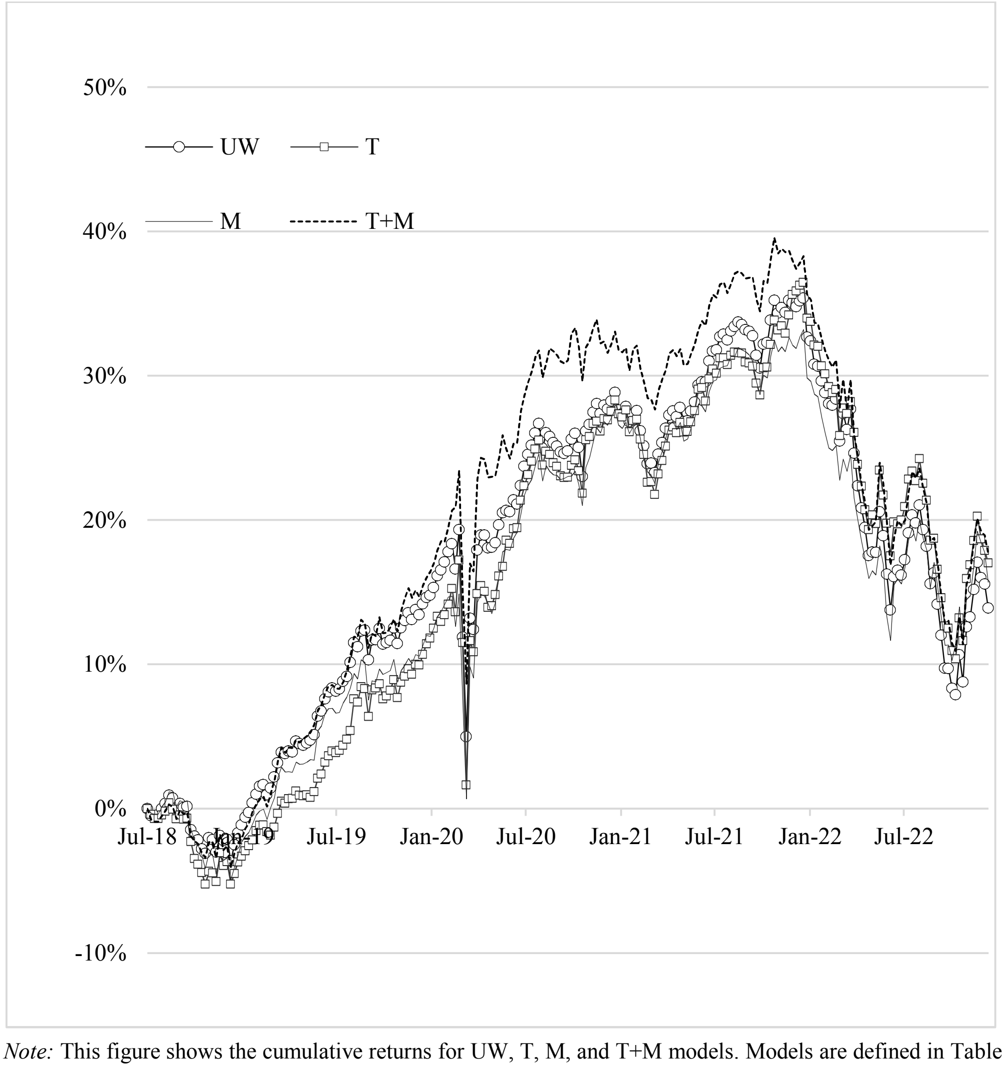
Note: Sharp ratio and Sortino ratio are measures of risk-adjusted performance. Models are defined in Table 4. The Sharp ratio is calculated using the mean and standard deviation of weekly returns, while the Sortino ratio uses Vol^{NegR}, defined as the standard deviation of weekly returns that fall below the risk-free rate, proxied by the 3-month U.S. Treasury bill return. Maximum drawdown (MDD) is computed based on monthly returns. Each variable's relative performance to uniform weight represents the difference in outcome compared to the uniform weight (Rel. Pef.) model, allowing for more effective comparison of risk-adjusted performance across strategies. All data are sampled from July 2018 to December 2022.

Table 8. Relative Returns to Uniform Weight Model returns by Periods

Period	1	2	3	4	5	6	1/3/5	2/4/6	1~6
Regimes	Bear	Bull	Bear	Bull	Bear	Bull	Bear	Bull	Bull/Bear
T	-7.65	-2.07	5.11	3.82	3.76	3.25	0.45	1.82	1.33
M	-5.98	-1.77	-2.73	1.65	8.30	-2.26	1.70	0.25	0.76
T+M	-2.45	3.63	18.57	-2.09	1.80	3.80	3.70	0.21	1.46
T+1a	-7.05	-3.69	14.49	4.61	3.14	5.72	2.14	1.91	1.99
T+1b	3.80	1.65	22.97	0.46	8.24	10.85	9.69	1.55	4.45
T+1c	6.22	2.84	4.99	-3.52	5.05	5.35	5.41	-0.81	1.40
T+1d	-3.66	6.17	10.34	-1.93	1.93	4.58	1.80	1.20	1.42
T+2a	5.49	1.33	22.83	-0.12	12.44	5.44	12.27	0.73	4.84
T+2b	3.80	1.65	22.97	0.46	8.24	10.85	9.69	1.55	4.45
T+2c	4.09	2.07	21.15	2.01	-3.41	26.68	3.68	3.68	3.68
T+2d	-2.26	5.16	6.23	1.66	-6.59	3.23	-2.76	2.93	0.90
T+2e	-4.30	-1.30	6.22	0.66	-12.79	-16.05	-6.47	-1.10	-3.01
T+3	5.98	2.64	-2.04	0.72	-3.64	-3.15	-0.32	1.10	0.60

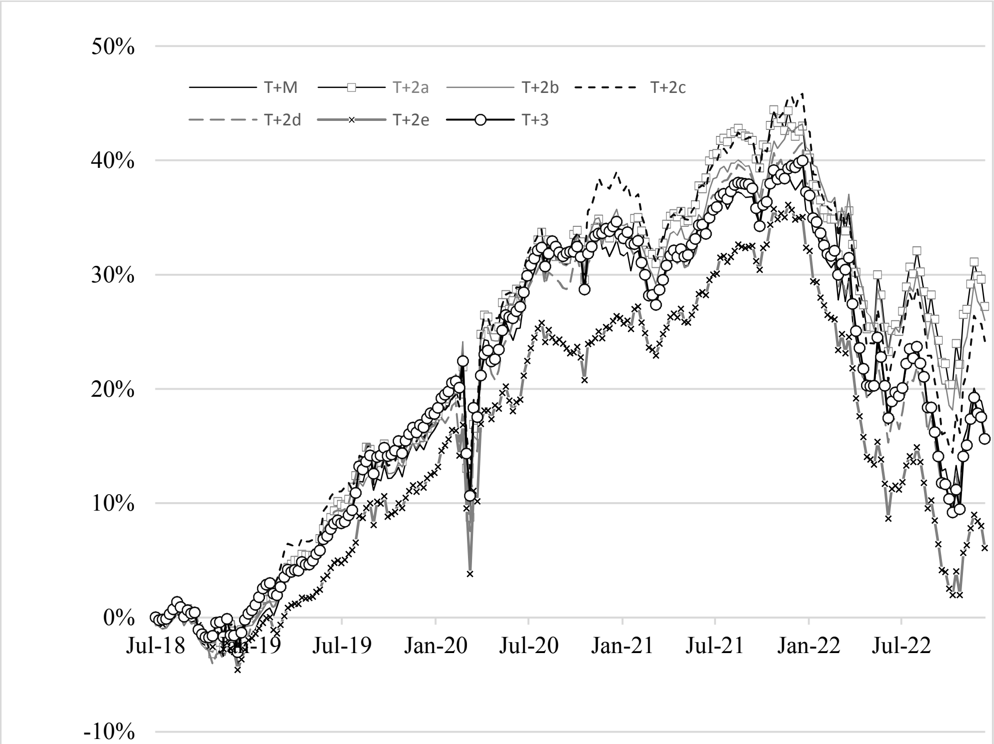
Note: This table presents the relative return performance of each strategy across different market periods. Models are defined in Table 4. All values represent weekly excess returns relative to the uniform weight model and are expressed in basis points. The data spans the period from July 2018 to December 2022.

Figure 1 and Figure 2 presents the cumulative returns of the selected models. Overall, the comparative performance of the models highlights an important point. Although technical indicators are effective at capturing price momentum, they neglect the significant influence of macroeconomic factors on asset prices. Incorporating features related to economic growth, labor market conditions, and inflation enables the model to make more informed allocation decisions and enhances its capacity to manage downside risk in volatile markets.



Note: This figure shows the cumulative returns for UW, T, M, and T+M models. Models are defined in Table 4. The data spans the period from July 2018 to December 2022.

Figure 1. Cumulative Returns of UW, T, M, and T+M



Note: This figure presents the cumulative returns of models trained on different combinations of macroeconomic and technical variables. Models are defined in Table 4. The data spans the period from July 2018 to December 2022.

Figure 2. Cumulative Returns of Models with Macroeconomic Variables

5.2. Contribution of Individual Macroeconomic Variables

Previous analyses have demonstrated that incorporating technical and macroeconomic indicators enhances portfolio performance. To investigate the specific impact of each macro-variable, we conducted a series of experiments using different combinations of these macro-indicators. The unemployment rate and nonfarm payrolls were grouped as labor market indicators, while the CPI, ISM, and LEI were treated as separate variables.

When each macroeconomic variable was added individually to the technical indicators, the CPI produced the most substantial performance gains, whereas the ISM underperformed the UW benchmark. Among all two-variable combinations, the CPI + ISM pairing (T+2a) achieved the highest cumulative return and lowest downside risk. This combination, which reflects both inflation dynamics and business cycle phases, is consistent with the widely adopted regime-based asset allocation approach in investment practice. Section 5.1. shows that the resulting T+2a model, trained on CPI and ISM along with technical indicators, notably achieved a 13.32 percentage point improvement in cumulative returns over the UW benchmark, with substantial reductions in both maximum drawdown and downside deviation. The T+2a model achieved 12.27bp excess return during bear markets and 0.73bp during bull markets, demonstrating exceptional resilience in adverse conditions. These results demonstrate the ability of reinforcement learning models to adaptively implement regime-aware asset allocation strategies when macroeconomic information is incorporated into the learning process.

Other combinations such as CPI + Labor (T+2b) and Labor + ISM (T+2c) also demonstrated strong performance. Overall, several selective combinations outperformed the model trained on the full macroeconomic set (T+M), highlighting the importance of identifying and focusing on impactful variables. These findings suggest that the targeted selection of key macroeconomic variables is more effective for performance enhancement than simply increasing the number of input features.

5.3. Regime-sensitive Allocation Decision

The effectiveness of a model's asset allocation strategy is closely tied to its ability to adapt to shifting market regimes. The T model, agent-adjusted asset weights are based solely on technical indicators, often fails to adequately adjust asset exposures in response to macroeconomic or structural changes in the market environment. In contrast, models that incorporate macroeconomic information can dynamically rebalance portfolios to align with prevailing economic conditions, thereby enhancing performance and risk management. This section investigates how reinforcement learning-based models, with different inputs, vary their asset weights across different market regimes and evaluates their effectiveness in capturing regime-specific risks and opportunities.

Table 9 shows the average weekly allocation of each model to different assets across various market conditions. The T model does not exhibit significant variation in allocation between bear and bull markets, maintaining an average distribution of close to 25% per asset across the regimes. By contrast, the T+M model, which incorporates macroeconomic learning, reduced its equity allocation to 23.12% in the 3rd period, a bear market, and increased it to above 25.74% in the following period, a bull market, demonstrating regime-sensitive flexibility in managing equity exposure. During market downturns, the agent strategically decreases its equity holdings and increases allocations to government and corporate bonds -assets with a lower equity beta, effectively enhancing downside protection.

The T+2a model increased its average allocation to government bonds to 27.38% during bear markets, while expanding its positions in equities and corporate bonds during bull markets to capture upside potential. For instance, in Regimes 1 and 3, where equities and government bonds have a negative correlation, the model allocated 28.09% and 30.31%, respectively, to bonds. In regime 5, where this correlation is positive, the bond allocation is reduced to 25.79%. By dynamically adjusting asset weights in response to changing cross-asset betas, the T+2 model exhibited a high degree of adaptability to evolving market conditions and achieved the best overall performance during the test period. In summary, T+2a, T+2b, and T+2c model achieve excess returns in both the bull and bear markets, providing empirical support for the effectiveness of the dynamic asset allocation strategies enabled by reinforcement learning.

Table 9. Asset Allocation Weights by Market Regime

Markets	BIL	LQD	SPY	TLT
Model: T				
Bear	25.22%	24.59%	24.53%	25.66%
Bull	25.17%	25.21%	24.74%	24.88%
Period #1 (Bear)	25.55%	23.64%	23.97%	26.83%
Period #2 (Bull)	26.99%	26.40%	23.36%	23.24%
Period #3 (Bear)	21.01%	24.71%	27.01%	27.27%
Period #4 (Bull)	23.89%	24.81%	25.40%	25.90%
Period #5 (Bear)	26.66%	25.14%	23.92%	24.28%
Period #6 (Bull)	27.56%	22.79%	25.76%	23.88%
Model: T+M				
Bear	24.41%	26.80%	22.99%	25.81%
Bull	24.40%	24.64%	24.68%	26.29%
Period #1 (Bear)	26.27%	24.80%	24.43%	24.49%
Period #2 (Bull)	24.79%	24.84%	23.29%	27.08%
Period #3 (Bear)	20.88%	27.40%	23.12%	28.60%
Period #4 (Bull)	24.23%	24.35%	25.74%	25.68%
Period #5 (Bear)	24.60%	27.84%	22.01%	25.55%
Period #6 (Bull)	23.97%	26.14%	22.10%	27.79%
Model: T+2a				
Bear	24.63%	23.71%	24.27%	27.38%
Bull	24.39%	25.32%	25.46%	24.83%
Period #1 (Bear)	23.16%	26.12%	22.63%	28.09%
Period #2 (Bull)	25.86%	24.18%	24.48%	25.48%
Period #3 (Bear)	25.09%	20.68%	23.92%	30.31%
Period #4 (Bull)	23.41%	26.26%	25.79%	24.54%
Period #5 (Bear)	25.39%	23.37%	25.45%	25.79%
Period #6 (Bull)	25.99%	22.55%	27.35%	24.11%

Note: This table shows the average weekly allocation of each portfolio to different assets across various market conditions (bull vs. bear). Models are defined in Table 4. The values show how each strategy adjusts asset weights in response to changes in market conditions. The data spans the period from July 2018 to December 2022.

6. CONCLUSION

This study makes several key contributions to the existing literature on reinforcement learning asset allocation and portfolio optimization. First, we developed an end-to-end reinforcement learning framework in which the agent autonomously learns an optimal investment policy to maximize cumulative returns. The model incorporates a rich set of state variables -fundamental, technical, and macroeconomic- and reallocates portfolio weights on a weekly basis, eliminating reliance on traditional parameter estimations. This structure enhances portfolio performance while reduces downside risk by enabling the agent to respond adaptively to changing market and economic conditions.

Second, we identify the macroeconomic variables that significantly enhance portfolio performance. Specifically, the combination of the CPI and the ISM leads to increase the portfolio excess returns, while the inclusion of other variables, such as labor market indicators, does not yield further improvements. This finding underscores the importance of feature or variable selection and highlights the tradeoff between input complexity and learning efficiency.

Third, we conduct extensive ablation studies to isolate the effects of macroeconomic inputs and evaluate the performance across different market regimes. By stratifying the results into bull and bear markets, we provide a detailed understanding of the behavior of the model under changes in economic conditions. In addition to conventional performance metrics, such as cumulative return and the Sharpe ratio, we incorporate downside risk measures, including the volatility of negative excess returns and the Sortino ratio. These measures provide a more comprehensive assessment of performance, especially during unfavorable market conditions.

Finally, this study makes an academic and practical contribution by expanding the applications of reinforcement learning beyond equity selection to multi-asset allocation across equities, government bonds, corporate bonds, and cash equivalents. This structure aligns with the decision-making processes typically employed by sophisticated investors. The proposed framework incorporates macroeconomic dynamics and enables cross-asset diversification. This provides a scalable, regime-sensitive approach with meaningful implications for academic research and applied portfolio management.

This study presents a reinforcement learning-based asset allocation framework that integrates both technical and macroeconomic variables to capture market dynamics. The model demonstrates the ability to adapt portfolio weights in response to shifts in asset prices and market conditions, achieving strong performance in terms of return enhancement and risk reduction, particularly in mitigating downside risks. These findings provide empirical evidence of the possible application of reinforcement learning in multi-asset portfolio management. Future research can enhance the practicality and scalability of reinforcement learning based asset allocation models by extending the investment universe or incorporating global macroeconomic variables.

REFERENCES

Bauman, T., S. Goluža, B. Gasperov, Z. Kostanjcaret (2024), "Deep Reinforcement Learning for Goal-Based Investing under Regime-Switching," Proceedings of the 5th Northern Lights Deep Learning Conference (NLDL), PMLR 233:13-19.

Boyd, J.H., J. Hu and R. Jagannathan (2005), "The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks," *Journal of Finance*, 60 (2): 649-672.

Chopra, V.K., and W.T. Ziemba (1993), "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice," *Journal of Portfolio Management*, 19(2), 6-11.

Collin-Dufresne, P., K. Daniel and M. Sağlam (2022), "Liquidity Regimes and Optimal Dynamic Asset Allocation," *Journal of Financial Economics*, 146 (2), 526-550.

Das, S.R., D. Ostrov; S. Mittal; A. Radhakrishnan; D.R. Srivastav; H. Wang (2024), "Reinforcement Learning for Multiple Goals in Goals-Based Wealth Management," *2024 Artificial Intelligence for Business (AIxB),* Laguna Hills, CA, USA, 2024, pp. 1-8.

Fischer, T.G (2018), "Reinforcement learning in financial markets-a survey," FAU Discussion Papers in Economics, No. 12/2018, University of Erlangen-Nuremberg, Institute for Economics.

Goldberg, L.S. and C. Grisse (2013), "Time Variation in Asset Price Responses to Macro Announcements," National Bureau of Economic Research Working Paper No.w19523.

Hamilton, J.D. and G. Lin. (1996), "Stock Market Volatility and the Business Cycle," *Journal of Applied Econometrics*, 11(5), 573-593.

Hornstein, A. (2016), "Unemployment Changes as Recession Indicators," Federal Reserve Bank of Richmond Economic Brief No. 23-13

Kelly, B. and D. Xiu (2023), "Financial Machine Learning," *Foundations and Trends in Finance*, 13 (3–4), 205-363.

Kim, M.J. and D. Kwon (2023), "Dynamic Asset Allocation Strategy: An Economic Regime Approach," *Journal of Asset Management*, 24(2), 136-147.

López de Prado, M. (2016), "Building Diversified Portfolios That Outperform out of Sample," *Journal of Portfolio Management*, 42(4), 59-69.

Long, H., A. Zaremba, W. Zhou, and E. Bouri (2022), "Macroeconomics Matter: Leading Economic Indicators and the Cross-Section of Global Stock Returns," *Journal of Financial Markets*, 61, 100736.

Maringer, D. and T. Ramtohul (2012), "Regime-Switching Recurrent Reinforcement Learning for Investment Decision Making," *Computational Management Science*, 9(1), 89-107.

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

Marshall, D.A (1992), "Inflation and Asset Returns in a Monetary Economy," *Journal of Finance*, 47(4), 1315-1342.

McGuckin, R.H., A. Ozyildirim and V. Zarnowitz (2004), "Real-Time Tests of the Leading Economic Index: Do Changes in the Index Composition Matter?" *Journal of Business Cycle Measurement and Analysis*, OECD Publishing, Centre for International Research on Economic Tendency Surveys, 2, 171-191.

Merton, R.C (1969), "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *Review of Economics and Statistics*, 51(3), 247-257.

Moody, J. and L. Wu. (1997), "Optimization of Trading Systems and Portfolios," Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr), IEEE.

Park, H.J., M.K. Sim and D.G. Choi (2020), "An Intelligent Financial Portfolio Trading Strategy Using Deep Q-Learning," *Expert Systems with Applications*, 158, 113573.

Pendharkar, P.C. and P. Cusatis (2018), "Trading Financial Indices with Reinforcement Learning Agents," *Expert Systems with Applications*, 103, 1-13.

Souto, H.G. and F. Louzada (2024), "Ablation Studies for Novel Treatment Effect Estimation Models," Statistics: Methodology Paper No.arXiv-2410. 15560v2, Cornell University

Sugadev, T., N.S. Hameed; S. Vijayakumar; P. Tamilarasan; Md. Sahidul Islam (2023), "Portfolio Optimization Using Machine Learning Techniques," 2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 1-7.

Tang, L. (2018), "An Actor-Critic-Based Portfolio Investment Method Inspired by Benefit-Risk Optimization," *Journal of Algorithms and Computational Technology*, 12(4), 351-360.

Van Vliet, P. and D. Blitz (2011), "Dynamic Strategic Asset Allocation: Risk and Return across the Business Cycle," *Journal of Asset Management*, 12, 360-375.

Yang, H.Y., X. Liu, and Q. Wu (2018), "A Practical Machine Learning Approach for Dynamic Stock Recommendation," 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing And Communications/ 12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE), New York, NY, USA, 1693-1697.

Mailing Address: Zoonky Lee, Yonsei University, Graduate School of Information, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, South Korea, E-mail: zlee@yonsei.ac.kr.

Received June 28, 2025, Accepted September 18, 2025.