Journal of Economic Development and Global Markets

GLINT OPEN ACCESS

Volume 1, Issue 1

Research Article

Date of Submission: 29 September 2025
Date of Acceptance: 27 October 2025
Date of Publication: 31 October 2025

Tactical Asset Allocation Using Deep Reinforcement Learning and Latent Macroeconomic Conditions

Tauseef Kazi[1], Ken Abbott[2], and Joe Wayne Byers[3]*

[1]MScFE Program, WorldQuant University, USA

[2]WorldQuant University, USA

[3]WorldQuant University, Oklahoma State University, USA

***Corresponding Author:** Joe Wayne Byers, WorldQuant University, Oklahoma State University, USA.

Citation: Kazi, T., Abbott, K., Byers, J, W. (2025). Tactical Asset Allocation Using Deep Reinforcement Learning and Latent Macroeconomic Conditions. *Econ Dev Glob Mark*, 1(1), 01-32.

Abstract

Traditional asset allocation techniques fail to adapt to sudden and severe economic downturns and lead to a loss of opportunities for investors. This paper seeks to address this problem by automating the Tactical Asset Allocation (TAA) framework that allocates the asset weights based on the latent macroeconomic conditions and market regimes. This project employs Deep Reinforcement Learning (DRL) to allocate funds to 36 Exchange Traded funds (ETFs) representing multiple geographies and asset classes. This study demonstrates that a macro, temporal, and spatially aware transformer agent guided by the LSTM-based macroeconomic regime shift encoder can outperform the traditional benchmarks, while taking the transaction cost into consideration.

Keywords: Tactical Asset Allocation, Hidden Markov Model, Deep Reinforcement Learning, Macroeconomics.

Introduction

Traditional strategic asset allocation techniques, which rely on the static weight adjustment method, such as mean-variance optimization, fail to adapt to sudden macroeconomic and regime shifts. These extreme events expose the investors to either extreme drawdowns or missed opportunities because of either being greedy or fearful. Tactical Asset Allocation (TAA) helps solve this problem by allocating funds across different assets based on changing market conditions. Their performance is frequently impacted due to the high transaction costs, model uncertainty, and subjective interpretation of the macroeconomic conditions by the fund manager.

Recent advances in the field of reinforcement learning, Deep reinforcement learning (DRL), provide a stable platform to address these problems. By deploying these agents to automate the asset allocation problem, they can interact with the market as an environment and thus reduce human biases and adapt to latent economic regimes. Existing DRL-based research does not take into account the macroeconomic fundamentals and market regime dynamics, which limits their ability to generalize well across changing economic cycles. This gap is more profound in portfolios spanning across multiple geographical regions and asset classes because the idiosyncratic risks and regional economic divergences complicate the asset allocation problem.

This paper proposes a novel TAA framework that combines the DRL with latent macroeconomic conditions and the price regime of the asset. The model follows a three-stage approach. First, it employs the Hidden Markov (HMM) model to detect regime shifts and calculates technical indicators to identify momentum and trends in ETF prices. Second, the feature extractor inside the DRL agent calculates the latent macroeconomic regime shift using the Long Short-Term Memory (LSTM) for 13 regions (12 Countries and the European Union) by consuming the macroeconomic data such as GDP growth, inflation, and unemployment for each area. Third, the attention mechanism inside the DRL agent uses the macroeconomic regime shift signal from the feature extractor along with the price features from the first stage to adapt the weight allocation to the cyclical and structural market changes.

The contribution of this work is fourfold. First, it introduces a systematic method to integrate macroeconomic data, which is usually sparse, asynchronous, and of lower frequency than the price data, into agentic asset allocation decisions through latent macroeconomic regime shift. Second, it introduces a DRL environment that penalizes the agent for the higher transaction cost and rewards them for higher returns or a better Sharpe ratio. Third, the framework is thoroughly tested against benchmarks such as a buy-and-hold portfolio, an equally-weighted portfolio, SPY, and DJIA, with the focus on checking the robustness of the model during a crisis like COVID-19 in 2020. Fourth, it introduces a systematic approach to include the assets that are not yet listed or have been delisted during the training or testing of the DRL agent.

Literature Review

Understanding the underlying macroeconomic state of the economy for investment into diverse assets has been the key objective of many investment firms. Bridgewater Associates is an example of how factor models are used to extract a score for an economic regime, utilizing "economic temperature" to make informed investment decisions. DFM has been traditionally used in modelling high-dimensional macro data into a few latent factors. However, they are linear and may not be able to capture all the complexities in the economic data. Park et al. demonstrated the use of the LSTM to make macroeconomic state interpretable for intelligent investment decisions [1]. Staffini makes use of bidirectional LSTM to forecast the macroeconomic state [2].

The Hidden Markov model (HMM) is a probabilistic model that is memoryless and models a time series as a Markov chain [3]. Markov chains are a sequence of discrete finite states. Kim demonstrated that the use of HMM to represent the market regimes had significantly increased the performance of the portfolio across a variety of asset classes [4]. Kritzman et al., instead of using price action as a regime classifier, used factors that drive these prices, such as economic growth, inflation, etc, for regime classification [5]. In a similar approach, Nguyen and Nguyen used macroeconomic variables like the inflation rate and industry production index for regime classification with HMMs [6]. According to Wang et al., concerning efficient market hypothesis, the price and volatility being action data would be the most immediate representation of the current market condition and thus would be ideal for market regime classification using HMMs [7].

Researchers have recently applied DRL algorithms to train agents to allocate funds during the portfolio optimization task efficiently [8]. These applications of DRL algorithms can be classified into two main types, viz, policy-based and value-based algorithms. Policy-based application, train the agents to learn a policy to provide action for portfolio optimization as the output. Value-based application trains the agents to directly rank the effective portfolio management actions in the action space. The DRL framework proposed by has two units, one for asset-scoring and the other for market-scoring [9]. The market-scoring unit used market sentiments as an indicator, whereas the asset-scoring unit used portfolio asset-price rise to score future allocations. The method proposed by Sun et al. used a hybrid approach that combines the traditional Black-Litterman model with the DRL agent consisting of a transformer layer [10]. The model learns dynamic correlation between portfolio assets and uses it to run a long/short strategy.

The model proposed by Katongo et al. uses DRL agents to predict the weights of individual Dow Jones Industrial Average (DJIA) constituents based on the technical indicators [11]. In this approach, technical indicators are calculated on the price and volume data of the individual assets, and then Autoencoders are applied for feature reduction. This approach performs better than the DJIA benchmark; however, it is more volatile due to a smaller number of assets in the portfolio and is less diversified. Another approach proposed by Li et al. uses an Ensemble of DRL agents to predict the weights of the assets in the portfolio [12]. This work uses the attention layer to extract the cross-sectional and temporal relationships between the tokens to efficiently predict the portfolio weights. Precomputed directional changes (DC) of the asset price are used as input features to the Attention mechanism. A customized framework with an ensemble of DRL agents consisting of attention layers for spatial and temporal representation is used; the whole network is then optimized using the policy gradient method.

Methodology

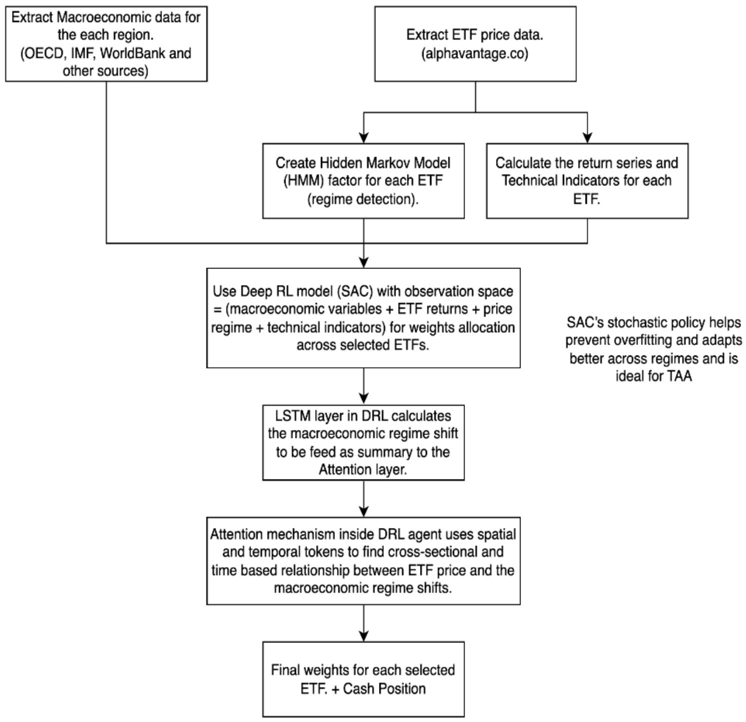

Figure 1: High-level solution with the dataflow pipeline of the proposed model

This section outlines all the tools and techniques used during research and formulation of the complete data pipeline that was used in the current experiment (Figure 1).

Data Collection

The following regions of interest were identified for the experiment: United States of America (USA), Canada, United Kingdom (UK), European Union (EU), China, Japan, South Korea, Indonesia, India, South Africa, Australia, New Zealand, and Brazil. There are ETFs available on the US stock exchanges from multiple asset management firms to invest in these countries and asset classes. Out of multiple ETFs available, 36 were selected (Table 2) to represent each of these regions. For the USA region, instead of using a single SPY ETF, underlying sector-based ETFs were chosen so that each sector's cyclical and seasonal trends could be exploited while tactically allocating funds to the current booming sector. Daily prices, dividends, and stock splits time series of these assets were downloaded from Alphavantage using the publicly available APIs. The adjusted close price was calculated using the following formula from the corporate actions, such as dividends and the stock split, representing the backward adjustment method.

$$A_t = P_t \times \prod_{k<t} \left(\frac{1}{S_k}\right) \times \prod_{k<t} \left(1 - \frac{Dk}{P_k}\right)$$

Where, At is the calculated adjusted closing price at time t, P_t the raw closing price at time t, Sk is the stock split ratio at time $k < t$, and Dk is the dividend paid at time $k < t$

Macroeconomic data was primarily sourced from OECD using their APIs, and among many variables available for each region of interest, only the most important variables were selected. The selected macroeconomic variables are listed in Table 3 along with the reasons for choosing each of the variables. The selected macroeconomic variables were reindexed to the month-end, and missing values were forward-filled. Since macroeconomic data from the OECD has a different starting point, and not all countries maintain data for each macroeconomic variable from 1996. The missing data points were imputed from other sources as listed in Table 4. Missing values from 1996 were backfilled with the first available value, and an additional masking column was added for each macroeconomic variable, with zero representing the backfilled value and one representing the non-backfilled section of the timeseries. The masking column would help improve the performance of machine learning algorithms by regularizing the missing part of the macroeconomic time series [13]. The final macroeconomic dataset has 28 features, with 14 representing the original variables as described in Table 3, and the remaining 14 have masking information related to the original variables for all 13 regions.

Exploratory Data Analysis (EDA)

Data for most of the ETFs mentioned in Table 2 of the appendix have been available since the beginning of the 21st Century, except for XLC, XLRE, VKG, MCHI, EIDO, INDA, ENZL, SLV, and REET. The latter were added late to the stock exchanges and marked as inactive for the initial phase. There is a special feature, "is tradable," added to the price series with zero for the non-active phase of the asset, for example, the asset was delisted or not yet listed on the exchange, and one for the active phase of the asset. This feature can be used by the agent to identify the inactive asset when it is penalized by the environment for assigning the weights to inactive assets.

The returns distribution of all of the ETFs, as seen in Figure 6. a to d in the Appendix, have thick tails indicating the presence of kurtosis. ETFs with the left-skew have the risk of large sudden price drops, whereas ETFs with right-skewed distributions can occasionally have large gains. IYR, EWU, EZU, EWJ, and EZA show signs of being susceptible to extreme outliers because they have extremely high kurtosis. The return distribution of AGG, EWG, and EWJ shows strong negative skewness and can face sudden drops rather than rallies. DRL-based approach is a natural choice for non-normal asset return distribution as it can adapt to the skewed and fat-tailed environments as compared to Gaussian-based approaches like mean-variance optimization. The Augmented Dickey-Fuller (ADF) test suggests that all ETF return series are stationary; hence, applying further differencing or other transformation techniques is not required.

The Correlation Heatmap of various geographies (Figure 7 in the Appendix) shows a negative correlation between the Unemployment rate and the Gross Domestic Product (GDP), Retail sales, and Industrial production. Consumer Price Index (CPI) are negatively correlated with interest rates, reflecting the impact of monetary tightening by the Central Banks. The LSTM is used to extract latent macroeconomic regime shifts for each region would help the cross-sectional attention mechanism inside the DRL agent to capture the underlying economic structure of each region.

The presence of heteroskedasticity in the return prices of ETFs is indicated by rolling volatility plots in Figure 8. a to f under the Appendix, which show persistent periods of high volatility followed by low volatility. For example, during the 2020 COVID-19 crisis, IWM, EWY, and INDA show spikes in volatility during the same time, confirming the presence of macro-driven risk. The Ljung-Box test performed on these assets further confirms the presence of volatility clustering and autocorrelated volatility for most of the assets. EWU, EZU, EWJ, and EZA show exceptional behaviour as they exhibit lower autocorrelation, which can be due to less persistent volatility or their underlying structure, such as low liquidity or pricing methods, might mask volatility persistence. As volatility clustering is observed in the price returns of these ETFs, DRL agents could adapt to this time-varying volatility by exploring the recent volatility regimes and exploiting this information to adjust the portfolio weights.

Hidden Markov Model and Technical Indicators

A 2-state HMM was precomputed with a 3-year rolling window and stored as a variable for computation efficiency to avoid look-ahead bias. Smoothed regime probabilities for states 0 and 1 are stored as separate features in the price series. The Statsmodel library was used for HMM calculations and is based on the method described by Jurafsky et al. [14]. The following technical indicators were precomputed and stored as features to further improve the attention of the DRL agent:

- Moving Average Convergence Divergence (MACD)
- Upper and Lower Bollinger Bands
- Relative Strength Index (RSI) - 30 period
- Commodity Channel Index (CCI) - 30 period
- Directional Movement Index (DX) - 30 period
- Average True Range (ATR) - 14 period
- Money Flow Index (MFI) - 14 period
- 30-day Exponential Moving Average (EMA) of Close Price
- 60-day Simple Moving Average (SMA) of Close Price
- 90-day SMA of Close Price
- 150-day SMA of Close Price
- 200-day SMA of Close Price

Environment

Reinforcement learning consists of an environment and an agent (Figure 2), such that the agent takes an action on the environment and the environment returns the current state and the rewards for the current action taken by the agent. The agent iteratively learns from its actions and improves over time. The state of the environment would represent the features related to the tradable universe until the current time step. A separate set of features representing the macroeconomic variables of 13 regions was passed to the agent. Shape ratio and log returns were used as two types of reward mechanisms in the current experiment. The agent would pass portfolio weights as actions in each time step to the environment. The environment would calculate the returns after applying fees and pass back the reward calculated based on this action. The transaction fees in this experiment were set to 2 percent. An extra cash position is also maintained, so the action space of the agent has 37 items. The sum of the weights assigned to these 37 assets in action space sums up to one to avoid short or leveraged positions. The observation space of the environment consists of two tensors of dimensions sums up to one to avoid short or leveraged positions. The observation space of the environment consists of two tensors of dimension (f_p, n, t_p) for price and (f_m, r, t_m) for macroeconomic time series, where f_p is the number of features in the price series, f_m is the number of features in the macroeconomic series, n the number of assets, r the number of regions, t_p represents the time window in days of backdated information to include for price-related time series and t_m represents the time window in months of backdated information to include for macro-related time series. The time window for the price time series was kept constant t_p = 30, and the time window for macroeconomic time series was experimented with values t_m = 60 and t_m = 96. Suppose the agent sets the weights on an inactive asset in a particular time step. The environment adds a penalty to the reward returned to the agent and adds the weight assigned to the inactive asset by the agent to the cash component. This penalty mechanism and the "is tradable" feature supplied in the price series would enable agents not to pick the inactive asset.

Soft-Actor Critic (SAC)

SAC is an off-policy actor-critic method that makes use of reinforcement learning principles along with an entropy-regularization framework. It works in a continuous action space, thus making it ideal for selecting weights of assets in a portfolio. It is an off-policy method that maintains past experiences in a replay buffer and reuses them to make action selection instead of just relying on the recent actions, as in the on-policy method. This approach makes it sample-efficient than the other on-policy methods. The agent being trained maintains a policy network, which consists of two critic networks and an actor network. Each critic network is accompanied by a slow-moving target network. The actor is trained to not just maximize the expected reward but also to maximize the entropy of its action distribution. It learns a stochastic policy π^θ and produces a Gaussian probability distribution over actions, where the mean and variance of this distribution are given by a neural network. A *tanh* squashing function is applied to the sampled Gaussian actions so that they are bounded to a finite range. SAC maintains two critic networks that estimate separate Q-functions $Q_{\varphi 1}$ and $Q_{\varphi 2}$. This technique is called double Q-learning and is used to mitigate the overestimation bias by using the smaller of the two estimates for stable learning [15]. A target network is maintained for each critic such that the copy of critic weights is slowly updated into the respective target network to reduce instability during the training process. The Polyak averaging technique is used to soft update the weights in the target network. SAC also maintains an entropy regularization coefficient α that can be constant or varying throughout training. This coefficient controls the explore-exploit tradeoff, with a higher value α enabling the agent to explore more, whereas a lower α value makes the agent exploit more the information it has learned. The α was set to automatic tuning mode with an initial value of 2.0 in the current work.

SAC is trained by minimizing three loss functions: actor loss, critic loss, and entropy temperature loss separately. However, they work together to form a complete policy network. The following equation represents the actor loss function [16].

$$J_\pi(\theta) = \mathbb{E}_{s_t \sim \mathcal{D}}\left[\mathbb{E}_{a_t \sim \pi_\theta(\cdot \mid s_t)}\left[\alpha \log \pi_\theta\left(a_t \mid s_t\right) - Q\left(s_t, a_t\right)\right]\right]$$

Where $\pi^{\theta}(a_t \mid s_t)$ is the actor's policy, and $Q(s_t, a_t)$ denote the current Q-value estimate for (s_t, a_t) which is taken from the minimum of the two critic networks, i.e. $Q_{\varphi 1}'$ or $Q_{\varphi 2}'$. The term $\alpha \log \pi$ penalizes actions with low entropy, and the overall loss function drives the policy to choose the actions that maximize the expected return. The policy update is done via gradient ascent on the objective. With the reparameterization trick, the gradient can be estimated through backpropagation using the sampled actions $a_t \sim \pi^{\theta}$.

Similarly, the critic loss function is given by the following equation [15].

$$J_Q(\phi_i) = \mathbb{E}_{(s_t,a_t,r_t,s_{t+1})\sim\mathcal{D}}\left[\left(Q_{\phi_i}(s_t, a_t) - y_t\right)^2\right]$$

SAC updates the two critics $Q_{\phi 1} Q_{\phi 2}$ by minimizing the mean squared error (MSE) against a common target y_t, which is given by the following equation.

$$y_t = r_t + \gamma(1 - d_t)[\min_{j=1,2} Q_{\phi_j}(s_{t+1}, a'_{t+1}) - \alpha \log \pi_\theta(a'_{t+1} \mid s_{t+1})]$$

Where r_t is the reward, γ is the discount factor, and d_t is the terminal indicator. Finally, entropy temperature loss is given by the following equation [16].

$$J(\alpha) = \mathbb{E}_{s_t \sim \mathcal{D}, a_t \sim \pi_\theta(\cdot \mid s_t)}\left[-\alpha\left(\log \pi_\theta(a_t \mid s_t) + \mathcal{H}_{target}\right)\right]$$

Where $\mathcal{H}_{target}$ is the target entropy. By minimizing the above loss function, the α increases if the entropy is lower than the target, enabling the agent to explore more; the α decreases if the entropy is higher than the target, which enables the agent to exploit more. This dynamic adjustment of the SAC algorithm allows the agent to balance between exploration and exploitation during the training.

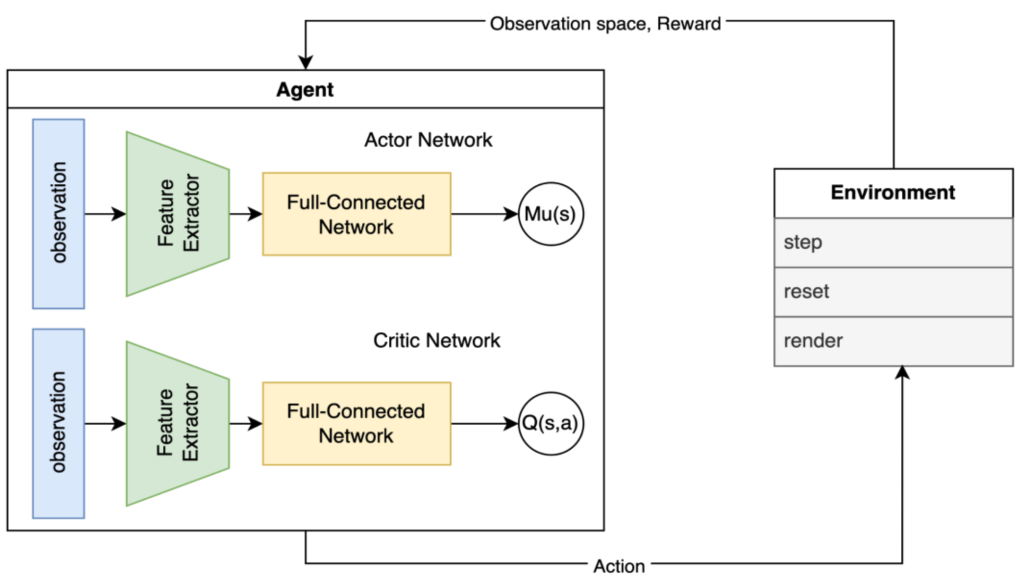

Figure 2: Agent, Environment, and the Policy Network in DRL

Feature Extractor

The policy network in the Stable Baselines3 (SB3) implementation of SAC is further divided into two parts: a feature extractor and a fully-connected network. The role of the feature extractor is to extract the features from the high-dimensional space, and the fully-connected network maps the features to actions/value (Figure 2). The feature extractors were configured to be shared between the actor and critic networks to improve the performance of the agents. The input feature observation space consists of two tensors. The first tensor is a 30-day rolling financial time series of individual assets in the portfolio trading universe, along with other precomputed indicators as mentioned in section 3.3. The second tensor is a 60-month or 96-month rolling macroeconomic time series of 13 regions with a total of 28 features as described in section 3.1. To improve the asset weight selection capabilities of the agent, the input tensors of dimension (f_p, n, t_p) and (f_m, r, t_m) are encoded into a latent space using different configurations of the attention mechanism in the current work. Each of these feature extractors will be used in separate agents and backtested against the benchmarks to find the best feature extraction technique. The following three custom implementations of feature extractors have been used: No Macro Multi-Head Attention (NM), Single-direction LSTM with Transformer Encoder Layer (LT), and Bidirectional LSTM with Transformer Encoder Layer (BT).

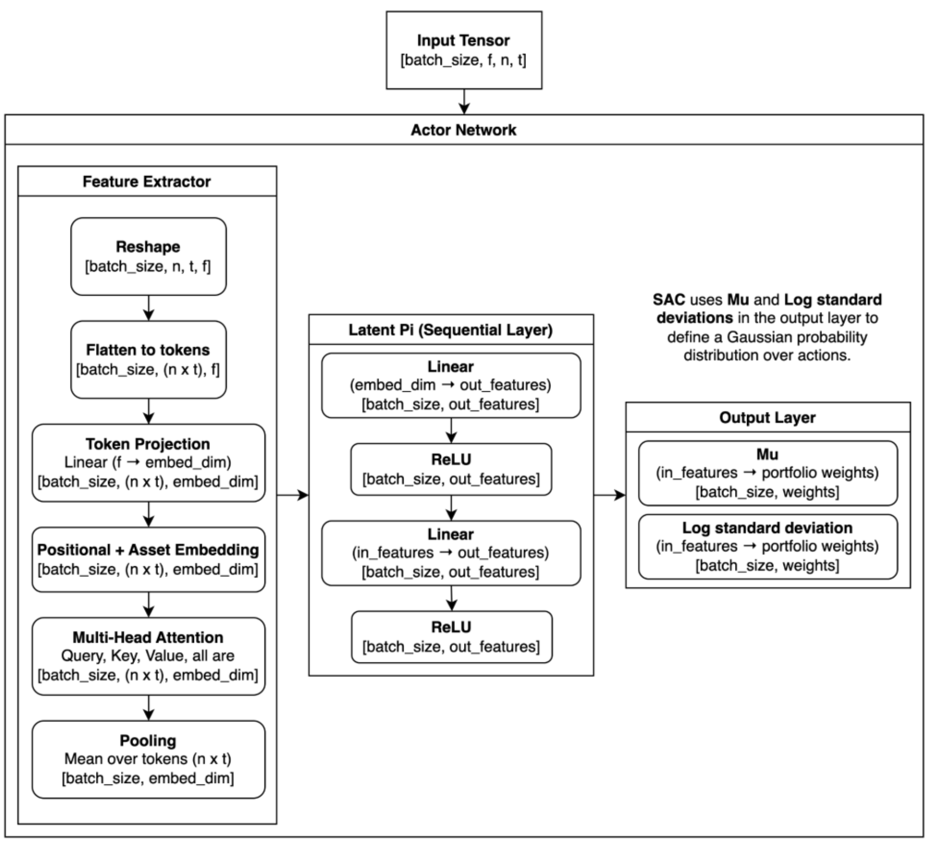

Figure 3: Neural network layers inside the DRL agent (Actor Network) with NM feature extractor.

The NM feature extractor only uses the price tensor received from the observation space and discards the macro tensor. It is designed to capture the cross-sectional and temporal information at the same time using the transformer-based self-attention mechanism by considering each asset at each time step as a token. These tokens are then fed into a multi-head self-attention module to obtain a contextual representation across assets and time steps. The output feature token is then aggregated using an average pooling layer.

The LT feature extractor ingests both price and macro tensor from the observation space. Each asset time slice $x_{a,t} \in \mathbb{R}f_p$ for the asset index a is linearly projected to a d-dimensional token. The price tokens of size $n \times t_p$ are created by adding element-wise an asset ID embedding and a temporal embedding to the projected price tokens. The macro matrix for every region is fed through a unidirectional LSTM to create a summary vector at the region level, serving as the final hidden state. A learnable region ID embedding is then added element-wise to create r macro tokens. A learnable Classification (CLS) token is prepended to a combined sequence of the price and macro tokens. The complete sequence is fed into a Transformer Encoder to generate the CLS output. This CLS output now jointly represents the cross-sectional, temporal, and macro information and becomes the feature extractor vector that would be finally shared by both the Actor and Critic networks. Single-direction LSTM runs forward only, and every macro summary generated by it encodes the past information relative to each time step.

Similar to the LT feature extractor BT feature extractor ingests both price and macro tensor from the observation space. The Bidirectional LSTM passes through each regional macro sequence both forward and backward, resulting in two hidden states. These hidden states are then concatenated and projected back to d-dimension. This doubles the contextual capacity and allows the token to capture both leading and lagging macro patterns. The rest of the pipeline for the BT remains the same as LT, which includes CLS token, concatenation, and pooling.

Figures 3 and 4 show the design of the neural network inside the Actor and Critic networks, along with the separation between feature extractors and other components of respective fully-connected layers in the Actor and Critic networks.

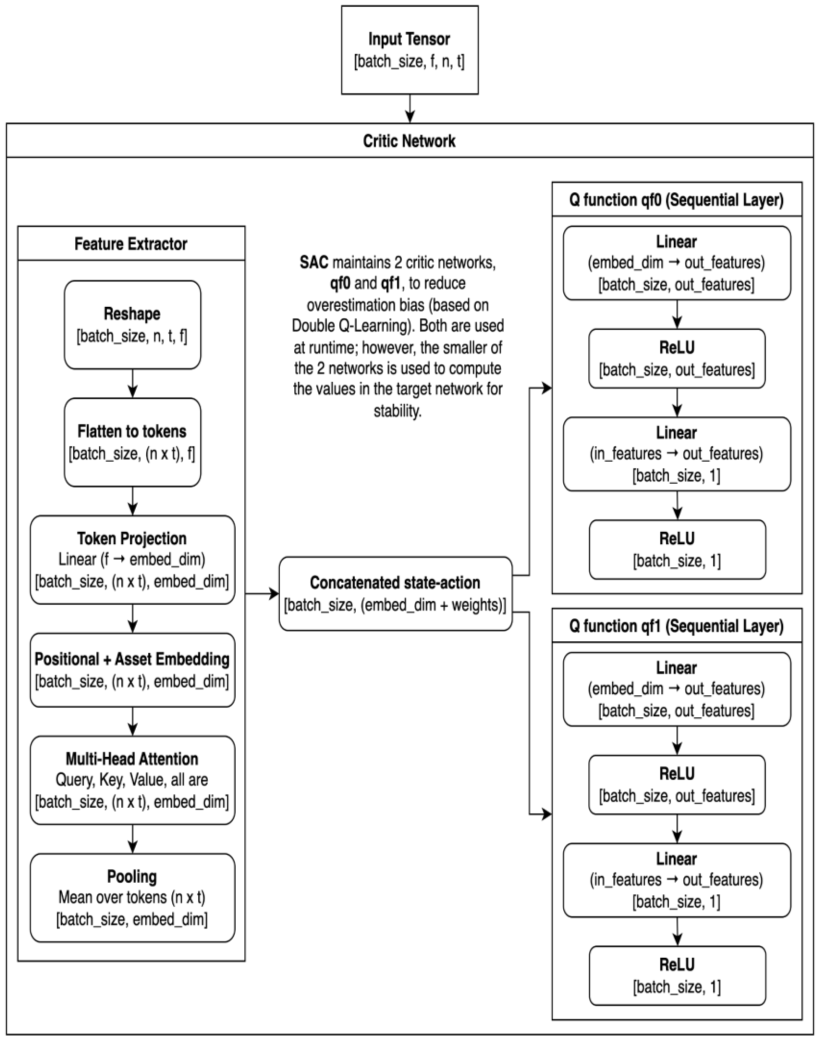

Figure 4: Neural network layers inside the DRL agent (Critic Network) with NM feature extractor.

Model Training and Assessment

The selected models are assessed against the benchmark portfolios that have historically performed well, like buy-and-hold portfolios, equally weighted portfolios, and SPY. One of the major criticisms about the TAA is that it underperforms in the long run due to high transaction costs associated with portfolio rebalancing and underperforms against these benchmark portfolios. A comparison of the proposed agents against the benchmark portfolios is performed using the following metrics:

- CAGR
- Final Wealth
- Max Drawdown
- Average Drawdown
- Value-at-Risk
- Alpha
- Beta
- Sortino Ratio
- Sharpe Ratio
- Information Ratio

All the selected models are trained on the data between 10th July 2003 and 1st November 2016. Then, the trained models are used to evaluate the portfolio weights between 2nd November 2016 and 11th April 2025 in the deterministic mode, utilizing only the information agents have learned during the training phase. Separate environments were created for the Train and Test phases to represent the underlying observation space. To keep the prediction stable and consistent, all the features were normalized using the maximum absolute scaler method. Instead of training the agents by executing a fixed number of simulation steps, a different approach of training by convergence was used. The training is executed for an infinite number of simulation steps, with evaluation checks being performed after every 500 steps against a separate evaluation environment. The agent is considered to have converged if performance in terms of cumulative rewards did not improve in the last 1500 steps, and the training is stopped. This validation is not performed for the initial 2500 steps so that the training process does not stop prematurely. A callback mechanism frequently checks and saves the current best-performing model and returns the last best model on convergence during the training process. All the agents are registered to log the performance metrics during training with Tensorboard to track the training progress in a real-time dashboard. Finally, the output weights of trained agents in the test are combined using the Rolling Kelly portfolio as an ensemble technique to further stabilize the results. This ensemble technique can be thought of as a process of turning volatile opinions from different agents into a disciplined investment meta-strategy analogous to opinions of different portfolio managers into a consolidated portfolio. The Environment assumes the market to be liquid and executes the order immediately at the close price; thus, slippage is not considered in the current environment configuration.

Results

The designed agents were trained in the Google Colab environment using an A100 GPU with 40 GB of dedicated GPU memory and 40 GB of system memory. The speed was 4 times slower when executed using a lower Colab environment with the L4 GPU. On average, the agents with bidirectional LSTM took more time to converge than the other agents. Similarly, agents trained with the reward as log returns took comparatively more time to converge than their equivalent agents with the reward as the Sharpe ratio. The following is the description of 15 portfolios as in Table 1.a. and Table 1.b. used in this experiment, of which 4 are benchmark portfolios, 11 represent the portfolio constructed using the weights received from different types of agents used in the current experiment, and 1 portfolio constructed using the rolling Kelly ensemble technique:

- **SPY:** ETF representing the S&P index and is used as a benchmark.
- **DIA:** ETF representing the Dow Jones Industrial Average (DJIA) and is used as a benchmark.
- **BAH:** Buy-and-Hold portfolio containing the 36 ETFs that represent the trading universe. It is used as a benchmark.
- **EQW:** An equal-weighted portfolio containing 36 ETFs representing the trading universe is rebalanced daily to maintain equal weights. The same transaction fees are applied as with the other backtested portfolio to maintain consistency. This portfolio is used as a benchmark.
- **NML:** Agent with NM feature extractor. Log return was used as the reward mechanism to train the agent, and the macroeconomic data was not used to train the agent.
- **NMS:** Agent with NM feature extractor. Sharpe ratio was used as the reward mechanism to train the agent, and the macroeconomic data was not used to train the agent.
- **LL5Y:** Agent with LT feature extractor. Log return was used as the reward mechanism, and the model was trained using a rolling 5-year window $tm = 60$.
- **LS5Y:** Agent with LT feature extractor. Sharpe ratio was used as the reward mechanism, and the model was trained using a rolling 5-year window $t_m = 60$.
- **BL5Y:** Agent with BT feature extractor. Log return was used as the reward mechanism, and the model was trained using a rolling 5-year window $t_m = 60$.
- **BS5Y:** Agent with BT feature extractor. Sharpe ratio was used as the reward mechanism, and the model was trained using a rolling 5-year window $t_m = 60$.
- **LL8Y:** Agent with LT feature extractor. Log return was used as the reward mechanism, and the model was trained using a rolling 8-year window $t_m = 96$.
- **LS8Y:** Agent with LT feature extractor. Shape ratio was used as the reward mechanism, and the model was trained using a rolling 8-year window $t_m = 96$.
- **BL8Y:** Agent with BT feature extractor. Log return was used as the reward mechanism, and the model was trained using a rolling 8-year window $t_m = 96$.
- **BS8Y:** Agent with BT feature extractor. Shape ratio was used as the reward mechanism, and the model was trained using a rolling 8-year window $t_m = 96$.
- **KELE:** Portfolio generated by the 40-day rolling Kelly ensembling technique, combining the portfolio weights of NMS, LS5Y, and BS8Y agents.

It was observed that agents trained with the Sharpe ratio as the reward mechanism in general converged faster and provided the highest performance in each segment compared to the log returns. Single-direction LSTM produced better results on the rolling 5-year window $t_m = 60$, and Bidirectional LSTM on the rolling 8-year window $t_m = 96$. NMS outperforms all the other portfolios with a Cumulative return of 5955.16% and a CAGR of 41.47%, while maintaining the same Sharpe ratio of 0.35 as the equally weighted portfolio. Overall, all the agents had beta lower than 1.0; as a result, they have lower market sensitivity while maintaining a positive and higher alpha. These results reinforce that the selected agents generated higher returns beyond market exposure. However, all the agents have higher tail risk, which is measured in terms of Daily Value-at-Risk, than most of the benchmarks, suggesting that there is further room to improve the risk management capability of these agents during the extreme events. KELE, which was generated using the rolling Kelly ensembling technique and combining the 3 best agents in their respective segments, was able to produce similar results as the top-performing NMS but with reduced risk exposure.

Table 1.a. Backtest results comparing portfolios and the benchmark using the performance metric (Part 1)

Metric	SPY	DIA	BAH	EQW	NML	NMS	LL5Y	LS5Y	BL5Y	BS5Y
Cumulative Return (%)	160.83	130.39	91.54	5947.89	5947.35	5955.16	5812.92	5790.22	5720.13	5825.14
Final Wealth	260825.16	230387.33	191541.6	6047893.96	6047345.53	6055164.01	5912922.76	5890224.29	5820129.91	5952142.99
CAGR (%)	8.44	7.31	5.65	41.45	41.45	41.47	41.18	41.14	40.99	41.21
Sharpe Ratio	0.46	0.38	0.29	0.35	0.35	0.35	0.35	0.35	0.35	0.35
Sortino Ratio	0.64	0.53	0.39	107.07	100.24	105.67	105.96	106.52	100.64	106.92
Information Ratio	0.00	-0.02	-0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Calmar Ratio	0.25	0.20	0.20	1.75	1.78	1.89	1.81	1.77	1.86	1.81
Skew	-0.31	-0.52	-0.68	45.31	45.31	45.31	45.31	45.31	45.31	45.31
Kurtosis	14.27	20.64	17.65	2052.92	2052.91	2052.91	2052.91	2052.91	2052.91	2052.91
Expected Daily (%)	0.05	0.04	0.03	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Expected Monthly (%)	0.97	0.85	0.66	4.23	4.23	4.23	4.21	4.02	4.19	4.21
Expected Yearly (%)	11.24	9.72	7.49	57.75	57.74	57.77	57.35	57.28	57.07	57.39
Daily Value-at-Risk (%)	-1.91	-1.88	-1.46	-216.28	-209.79	-212.52	-209.60	-209.69	-201.97	-210.56
Volatility (ann.)	0.1900	0.1856	0.1441	21.1568	20.5221	20.7894	20.5329	20.5124	19.7575	20.9435
Max Drawdown (%)	-3369.74	-3668.50	-2843.17	-2369.66	-2329.15	-2193.71	-2280.37	-2323.07	-2209.86	-2280.59
Longest DD Days	708	693	778	1018	975	968	1016	1018	980	1018
Avg. Drawdown (%)	-175.61	-208.36	-156.18	-96.71	-97.21	-94.43	-98.28	-90.29	-108.92	-88.32

Avg. Drawdown Days	16.64	23.37	20.45	40.00	35.14	37.25	42.00	38.93	40.70	38.14
Beta	1.0000	0.9273	0.7296	-0.6946	-0.6311	-0.6732	-0.6685	-0.6694	-0.611	-0.6726
Alpha	0.0000	-0.0062	-0.0089	7.5093	7.2823	7.3799	7.2878	7.2804	7.0112	7.3109

Table 1.b. Backtest results comparing portfolios and the benchmark using the performance metric (Part 2)

Metric	SPY	DIA	BAH	EQW	LL8Y	LS8Y	BL8Y	BS8Y	KELE
Cumulative Return (%)	160.83	130.39	91.54	5947.89	5497.42	5630.73	5824.13	5924.49	5954.65
Final Wealth	260825.16	230387.33	191541.6	6047893.96	6047345.53	6055164.01	5912922.76	5890224.29	5820129.91
CAGR (%)	8.44	7.31	5.65	41.45	40.53	40.81	41.20	41.41	41.47
Sharpe Ratio	0.46	0.38	0.29	0.35	0.35	0.35	0.35	0.35	0.35
Sortino Ratio	0.64	0.53	0.39	107.07	98.45	100.60	106.97	115.43	108.94
Information Ratio	0.00	-0.02	-0.04	0.02	0.02	0.02	0.02	0.02	0.02
Calmar Ratio	0.25	0.20	0.20	1.75	1.73	1.82	1.77	1.93	1.95
Skew	-0.31	-0.52	-0.68	45.31	45.31	45.31	45.31	45.31	45.31
Kurtosis	14.27	20.64	17.65	2052.92	2052.90	2052.91	2052.91	2052.92	2052.91
Expected Daily (%)	0.05	0.04	0.03	0.20	0.20	0.20	0.20	0.20	0.20
Expected Monthly (%)	0.97	0.85	0.66	4.23	4.15	4.17	4.21	4.23	4.23
Expected Yearly (%)	11.24	9.72	7.49	57.75	56.39	56.80	57.38	57.68	57.76
Daily Value-at-Risk (%)	-1.91	-1.88	-1.46	-216.28	-200.22	-197.40	-212.07	-214.10	-209.79
Volatility (ann.)	0.1900	0.1856	0.1441	21.1568	20.5843	19.3104	20.7454	20.9425	20.5403
Max Drawdown (%)	-3369.74	-3668.50	-2843.17	-2369.66	-2347.82	-2245.99	-2332.44	-2144.02	-2123.17
Longest DD Days	708	693	778	1018	1050	980	1040	1018	974
Avg. Drawdown (%)	-175.61	-208.36	-156.18	-96.71	-97.43	-94.43	-82.39	-92.42	-91.62
Avg. Drawdown Days	16.64	23.37	20.45	40.00	39.58	35.71	34.86	42.02	36.70
Beta	1.0000	0.9273	0.7296	-0.6946	-0.5955	-0.5970	-0.6703	-0.7127	-0.6730
Alpha	0.0000	-0.0062	-0.0089	7.5093	6.9455	6.8533	7.3615	7.4371	7.2937

Figure 5: Cumulative returns of backtested portfolios versus benchmark portfolios.

Discussion

Agents that are training using the Sharpe ratio as the reward mechanism were able to converge faster and produce higher risk-adjusted returns compared to the agents training using the log returns as the reward mechanism. Sharpe ratio normalises reward by volatility, which rescales the gradient signal across market regimes. Thus stabilizing the policy updates and reducing the exploration time. It amplifies large single-period gains and losses when an agent is trained using the log return as the reward mechanism. This results in the generation of higher-variance gradients, in turn leading to longer training time. This drawback with the log return as the reward mechanism also impacts the out-of-sample performance of the trained agent.

Empirical results show that single-direction LSTM produced better results on the shorter 5-year rolling window of macroeconomic data, whereas the Bidirectional LSTM on the longer 8-year rolling window. The single-directional LSTM is a forward-looking model and uses past information to predict the present state. A shorter sequence of macroeconomic data helps this model because the risk of vanishing gradients is reduced in a single direction. The bidirectional LSTM captures both the leading and lagging relationships between the macroeconomic variables; thus, it generalizes better with longer sequences of the data. A longer time horizon allows the bidirectional LSTM to identify structural shifts and cyclical patterns within an economy.

Every training cycle produces a slightly different version of the agent, which might show differences in performance. Thus, training different versions of agents as in the current approach, and then combining their results using an ensemble technique, would give more reliable results. This ensemble technique can be thought of as a process of turning volatile opinions from different agents into a disciplined investment meta-strategy analogous to opinions of different portfolio managers into a consolidated portfolio. An inexpensive ensemble technique, such as the rolling Kelly, was able to retain the performance of the best single agent while reducing the tail risk of individual policy errors.

A higher daily Value-at-Risk signifies that the portfolio is vulnerable to a single-day price shock despite higher positive alpha and lower market beta against SPY, especially during portfolio rebalancing activity. Thus, there is a need to add an intra-day risk measure or fractional-trade execution. Finally, each agent takes around 3-4 hours of training time on a high-performance GPU, which raises questions about the cost-benefit tradeoff for production deployment of this strategy.

Conclusion

This study demonstrates that a macro, temporal, and spatially aware transformer agent guided by the LSTM-based macroeconomic regime shift encoder can outperform the traditional benchmarks, while including the transaction cost. The study also shows that a globally diversified portfolio outperforms the trusted ETFs like SPY by a huge margin. The agent-based TAA technique proposed in the current research further improves the performance of this globally diversified portfolio even during the COVID-19 crisis period. The approach to include the assets that are not yet listed or have been delisted does not impact the performance of the trained DRL agents, as XLC was introduced on 19th June 2018 during the testing phase. The novel approach introduced in the current study includes the macroeconomic data in addition to the price data in the observation space of the DRL agent, which has shown promising results and can be exploited in future studies. This approach can be considered analogous to passing the meta information in addition to the price data, which would help the DRL agents to make more informed decisions by systematically combining it with the price information related to the assets and the frequency of this meta information does not need to be the same as that of the price data [17-21].

Disclaimer

This paper was created as part of a WorldQuant University degree program towards an MSc in Financial Engineering. This paper is reproduced with the consent and permission of WorldQuant University.

Appendix

All the extended information used in this project, including tables and images, is presented under this section. The **reference code**, trained agents, tensorboard logs and raw data, and engineered data are maintained at the following GitHub repository (link).

Table 2. Asset selection per region of interest

<table>
<tr><th>Region</th><th>ETFs</th></tr>
<tr><td>USA</td><td>iShares Russell 1000 Growth ETF (IWF): Large-cap growth stocks.
iShares Russell 1000 Value ETF (IWD): Large-cap value stocks.
iShares Russell 2000 ETF (IWM): Small-Cap stocks.
iShares U.S. Real Estate ETF (IYR): Broad U.S. real estate sector.

SPDR Select Sector ETFs:
XLY: Consumer Discretionary Select Sector SPDR Fund.
XLP: Consumer Staples Select Sector SPDR Fund.
XLU: Utilities Select Sector SPDR Fund.
XLE: Energy Select Sector SPDR Fund.
XLC: Services Select Sector SPDR Fund.
XLF: Financial Select Sector SPDR Fund.
XLV: Health Care Select Sector SPDR Fund.
XLI: Industrial Select Sector SPDR Fund.
XLB: Materials Select Sector SPDR Fund.
XLK: Technology Select Sector SPDR Fund.
XLRE: Real Estate Select Sector SPDR Fund.

Bond-Based ETFs:
iShares Core U.S. Aggregate Bond ETF (AGG): Broad U.S. bond exposure.
iShares 20+ Year Treasury Bond ETF (TLT): Long-term U.S. Treasuries.
iShares 1-3 Year Treasury Bond ETF (SHY): Short-term U.S. Treasuries.
iShares iBoxx $ Investment Grade Corp Bond ETF (LQD): Corporate bonds.

ETFs for alternative assets:
SPDR Gold Shares (GLD): Most liquid gold ETF.
iShares Silver Trust (SLV): The most popular silver ETF.
iShares Global REIT ETF (REET): Global real estate exposure (U.S. + intl).</td></tr>
<tr><td>Canada</td><td>iShares MSCI Canada ETF (EWC): Offers exposure to large and mid-cap Canadian equities.</td></tr>
<tr><td>UK</td><td>iShares MSCI United Kingdom ETF (EWU): Tracks the performance of the MSCI United Kingdom Index, representing large and mid-cap segments of the UK market.</td></tr>
</table>

European Union (EU)	**Vanguard FTSE Europe ETF (VGK):** Provides exposure to large and mid-cap companies in developed European markets. **iShares MSCI EMU ETF (EZU):** Focuses on companies within the European Economic and Monetary Union.
China	**iShares MSCI China ETF (MCHI):** Tracks the performance of the MSCI China Index, covering large and mid-cap Chinese equities. **iShares MSCI Hong Kong ETF (EWH):** Large and mid-cap companies listed in Hong Kong.
Japan	**iShares MSCI Japan ETF (EWJ):** Focuses on Large and mid-cap Japanese companies.
South Korea	**iShares MSCI South Korea ETF (EWY):** Tracks large and mid-sized South Korean companies.
Indonesia	**iShares MSCI Indonesia ETF (EIDO):** Provides access to a broad range of Indonesian companies.
India	**iShares MSCI India ETF (INDA):** Offers exposure to large and mid-cap Indian equities.
South Africa	**iShares MSCI South Africa ETF (EZA):** Focuses on Large and mid-cap South African companies.
Australia	**iShares MSCI Australia ETF (EWA):** Targets large and mid-cap Australian companies.
New Zealand	**iShares MSCI New Zealand ETF (ENZL):** Provides access to publicly traded companies in New Zealand.
Brazil	**iShares MSCI Brazil ETF (EWZ):** Focuses on Large-cap Brazilian stocks.

Table 3. Macroeconomic variables used for the experiment, along with the category and reason for using them for the analysis

Variable	Reason	Category
GDP (real)	The primary measure of economic activity.	Growth & Output Indicators
Industrial Production	High-frequency proxy for GDP.	Growth & Output Indicators
Retail Sales	Consumer demand trends.	Growth & Output Indicators
Unemployment Rate	Classic lagging recession indicator.	Labor Market Indicators
Participation Rate	Labor force trends.	Labor Market Indicators

Consumer Price Index (CPI)	Headline inflation measure.	Inflation Indicators
3-Month Government Bond Yield	Captures short-term expectations.	Monetary Policy / Rates
10-year Government Bond Yield	Captures long-term expectations.	Monetary Policy / Rates
Yield Curve (10Y - 3M spread)	Strong recession predictor.	Monetary Policy / Rates
Trade Balance	Net external position.	Trade & External Sector
Current Account Balance	Broader external health indicator.	Trade & External Sector
Consumer Confidence Index (CCI)	Tells about future consumer spending.	Sentiment & Leading Indices
Business Confidence / Surveys (BCI)	Early signals of a slowdown or a boom.	Sentiment & Leading Indices
Leading Economic Index (CLI)	Composite leading indicator.	Sentiment & Leading Indices

Table 4. Additional sources for Macroeconomic data of countries that were missing from the OECD API

Region	**Variable**	**Source**
South Korea	Trade Balance	**IMF Dataset:** Balance of Payments (BOP), Quarterly, Net (credits less debits) **Indicator:** Goods and services
South Korea	Current Account Balance	**IMF Dataset:** Balance of Payments (BOP), Quarterly, Net (credits less debits) **Indicator:** Current account balance (credit less debit)
South Korea	Participation Rate	**IMF Dataset:** Labor Statistics (LS), Monthly & Yearly **Indicator:** Labour Force **Type of transformations:** Persons
Canada	Consumer Confidence Index	Bank of Canada
Australia	Retail Sales	Australian Bureau of Statistics

New Zealand	GDP	**GDP_QOQ:** Real q/q% s.a. under Expenditure-based GDP (Reserve Bank of New Zealand) **GDP_YOY:** Real y/y% under Expenditure-based GDP (Reserve Bank of New Zealand)
New Zealand	Retail Sales	Reserve Bank of New Zealand
New Zealand	Consumer Price Index	CPI (y/y%) (Reserve Bank of New Zealand)
New Zealand	Leading Economic Index (LEI)	Survey of expectations (M14) - Annual CPI growth - 1 year out (Reserve Bank of New Zealand)
New Zealand	Unemployment Rate	Labour market (M9) - Quarterly (Reserve Bank of New Zealand)
		Indicator: Unemployment rate Type of transformations: Percent
South Africa	Participation Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Labour Force Type of transformations: Persons
Brazil	Unemployment Rate	IMF Dataset: Labor Statistics (LS), Monthly, Quarterly & Yearly Indicator: Unemployment rate Type of transformations: Percent
Brazil	Participation Rate	IMF Dataset: Labor Statistics (LS), Monthly & Yearly Indicator: Labour Force Type of transformations: Persons
Brazil	3-Month Government Bond Yield	FRED
Brazil	Trade Balance	IMF Dataset: Balance of Payments (BOP), Quarterly, Net (credits less debits) Indicator: Goods and services
Brazil	Current Account Balance	IMF Dataset: Balance of Payments (BOP), Quarterly, Net (credits less debits) Indicator: Current account balance (credit less debit)
India	Retail Sales	IMF Dataset: National Economic Accounts (NEA), Quarterly Indicator: Final consumption expenditure

		Timeseries: IND.P3.Q.SA.XDC.Q
India	Unemployment Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Unemployment rate Type of transformations: Percent
India	Participation Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Labour Force Type of transformations: Persons
India	3-Month Government Bond Yield	3MINY.B (stooq.com)
India	10-year Government Bond Yield	10YINY.B (stooq.com)
Indonesia	Retail Sales	IMF Dataset: National Economic Accounts (NEA), Quarterly Indicator: Final consumption expenditure Timeseries: IDN.P3.Q.SA.XDC.Q
		Indicator: Unemployment rate Type of transformations: Percent
South Africa	Participation Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Labour Force Type of transformations: Persons
Brazil	Unemployment Rate	IMF Dataset: Labor Statistics (LS), Monthly, Quarterly & Yearly Indicator: Unemployment rate Type of transformations: Percent
Brazil	Participation Rate	IMF Dataset: Labor Statistics (LS), Monthly & Yearly Indicator: Labour Force Type of transformations: Persons
Brazil	3-Month Government Bond Yield	FRED
Brazil	Trade Balance	IMF Dataset: Balance of Payments (BOP), Quarterly, Net (credits less debits) Indicator: Goods and services

Brazil	Current Account Balance	IMF Dataset: Balance of Payments (BOP), Quarterly, Net (credits less debits) Indicator: Current account balance (credit less debit)
India	Retail Sales	IMF Dataset: National Economic Accounts (NEA), Quarterly Indicator: Final consumption expenditure Timeseries: IND.P3.Q.SA.XDC.Q
India	Unemployment Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Unemployment rate Type of transformations: Percent
India	Participation Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Labour Force Type of transformations: Persons
India	3-Month Government Bond Yield	3MINY.B (stooq.com)
India	10-year Government Bond Yield	10YINY.B (stooq.com)
Indonesia	Retail Sales	IMF Dataset: National Economic Accounts (NEA), Quarterly Indicator: Final consumption expenditure Timeseries: IDN.P3.Q.SA.XDC.Q
Indonesia	Unemployment Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Unemployment rate Type of transformations: Percent
Indonesia	Participation Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Labour Force Type of transformations: Persons
Indonesia	Trade Balance	IMF Dataset: Balance of Payments (BOP), Quarterly, Net (credits less debits) Indicator: Goods and services
Indonesia	Current Account Balance	IMF Dataset: Balance of Payments (BOP), Quarterly, Net (credits less debits) Indicator: Current account balance (credit less debit)

Indonesia	10-year Government Bond Yield	10YIDY.B (stooq.com)
China	GDP	Nominal Quarterly GDP - Not Seasonally adjusted (National Bureau of Statistics of China)
China	Retail Sales	Total Retail Sales of Consumer Goods, Current Period - 100 million yuan (National Bureau of Statistics of China)
China	Unemployment Rate	IMF Dataset: Labor Statistics (LS), Quarterly Indicator: Unemployment rate Type of transformations: Percent Country: Hong Kong
China	Participation Rate	IMF Dataset: Labor Statistics (LS), Quarterly & Yearly Indicator: Labour Force Type of transformations: Persons Country: China, Hong Kong
China	Trade Balance	IMF Dataset: Balance of Payments (BOP), Quarterly & Yearly, Net (credits less debits) Indicator: Goods and services Country: China
China	Current Account Balance	IMF Dataset: Balance of Payments (BOP), Quarterly & Yearly, Net (credits less debits) Indicator: Current account balance (credit less debit) Country: China
China	10-year Government Bond Yield	10YCNY.B (stooq.com)
Global	Working age population	World Bank
Global	Total Unemployment	World Bank

Figure 6. a: ETF Return Histograms (Part 1)

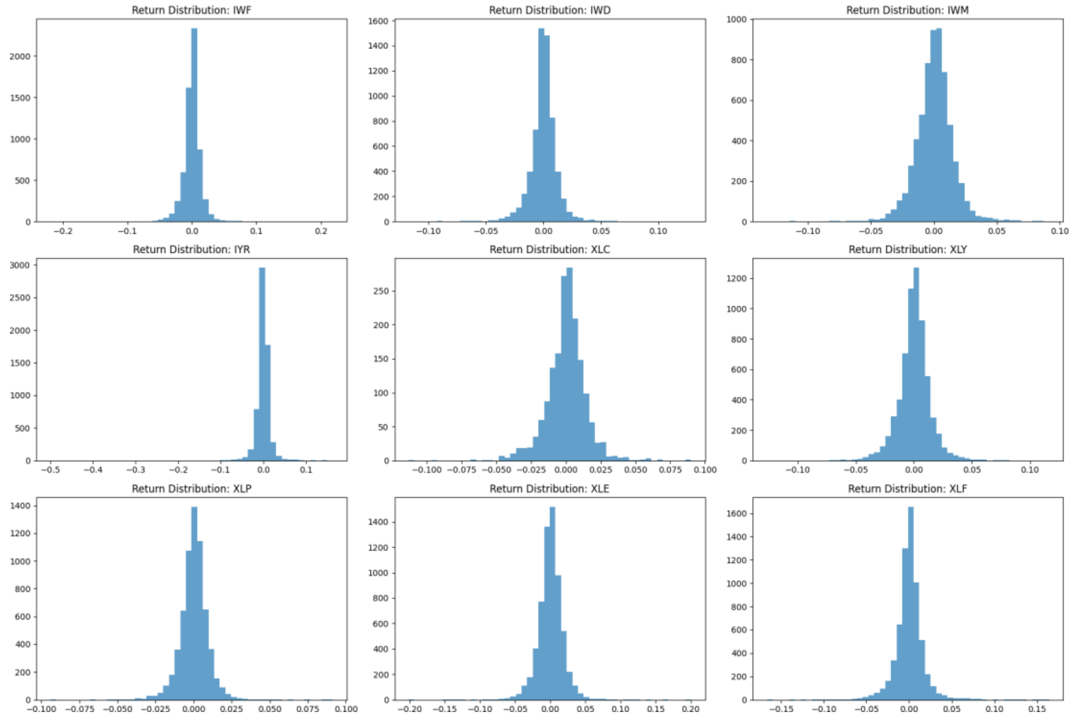

Figure 6. a: ETF Return Histograms (Part 1)

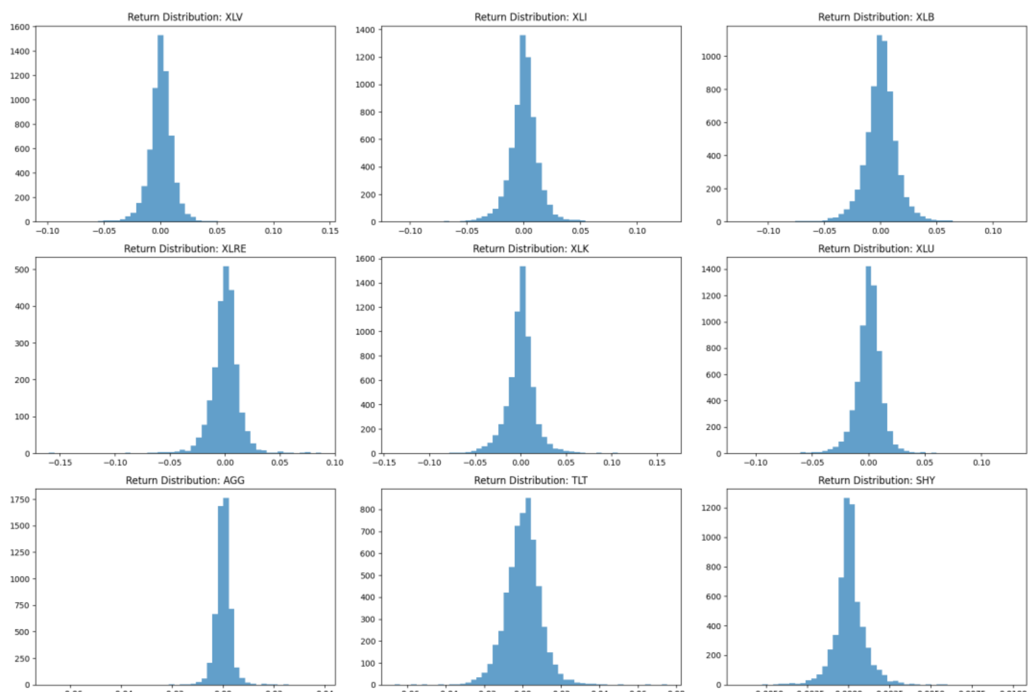

Figure 6. b: ETF Return Histograms (Part 2)

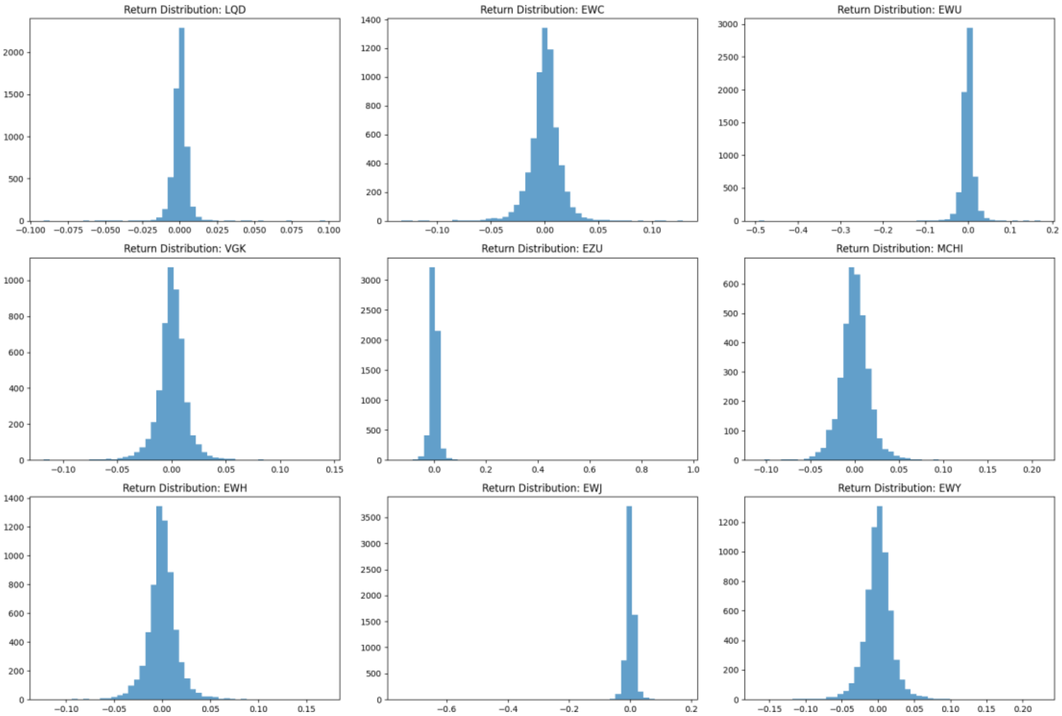

Figure 6. c: ETF Return Histograms (Part 3)

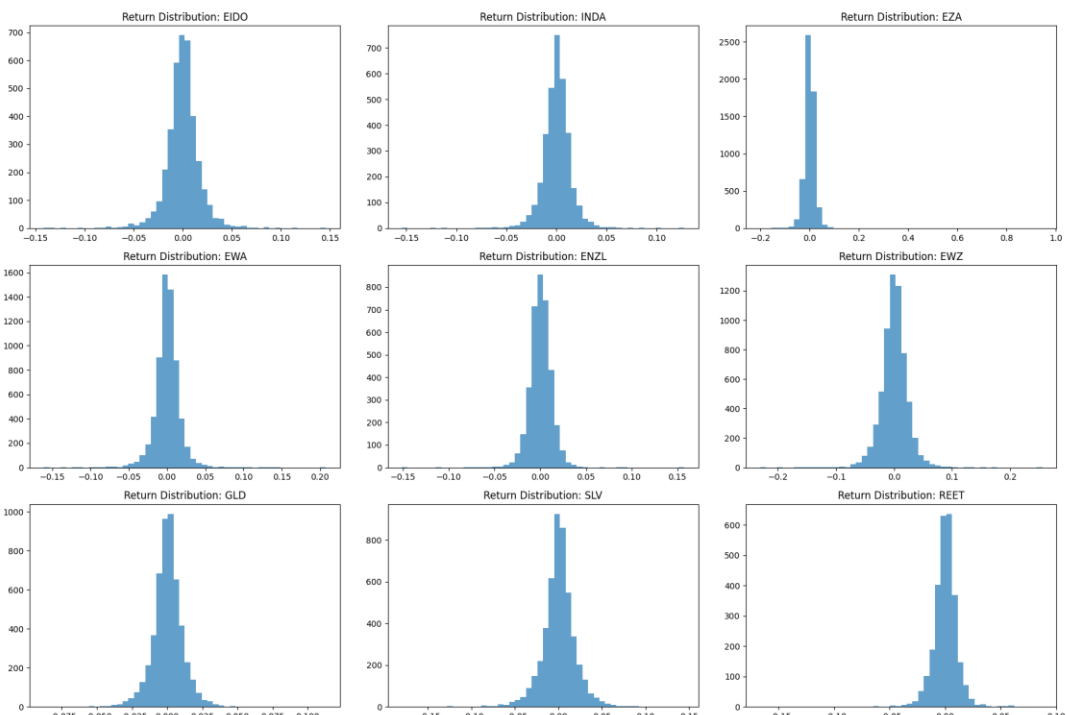

Figure 6. d: ETF Return Histograms (Part 4)

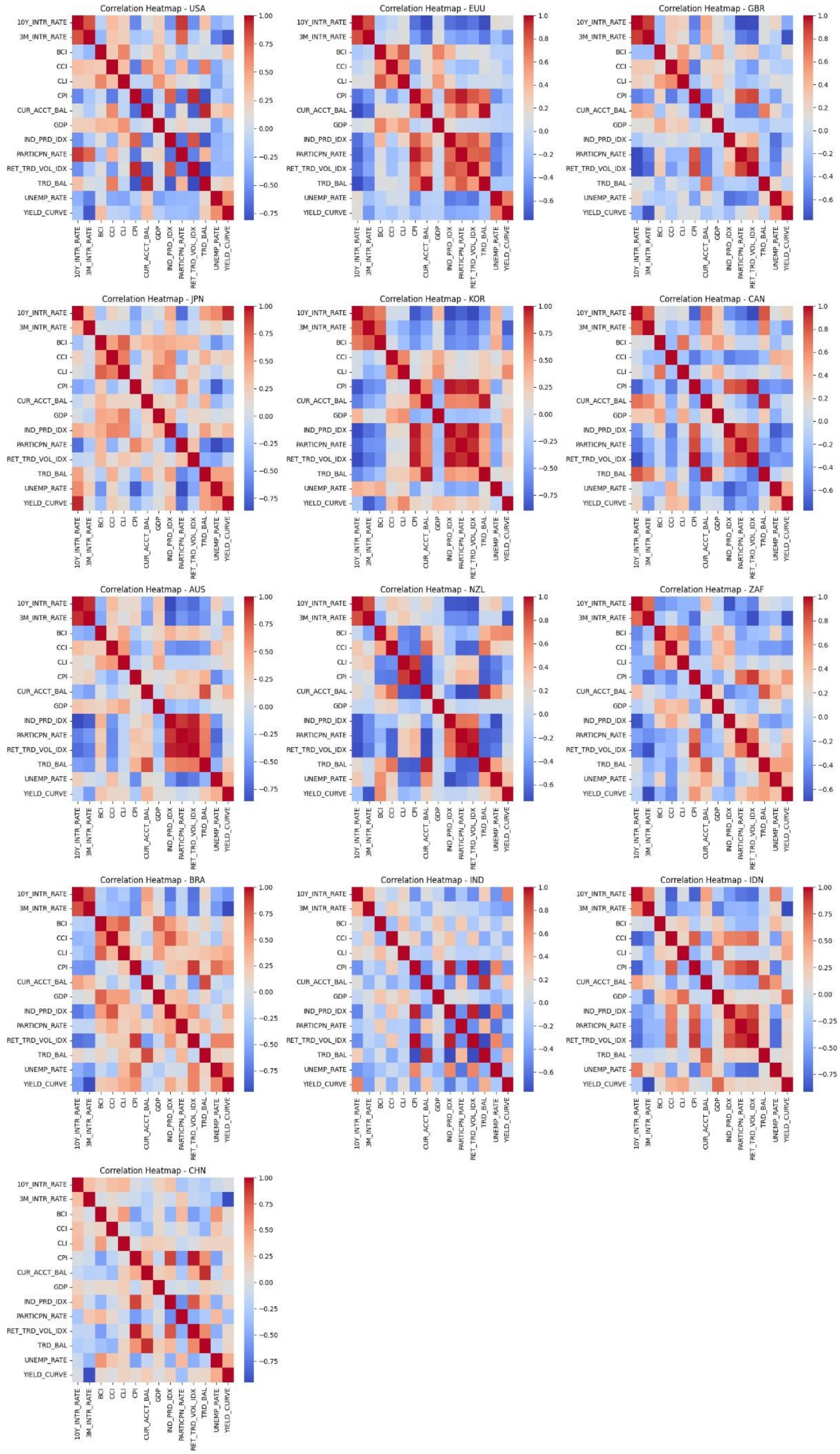

Figure 7: Correlation Heatmap of Macro Variables

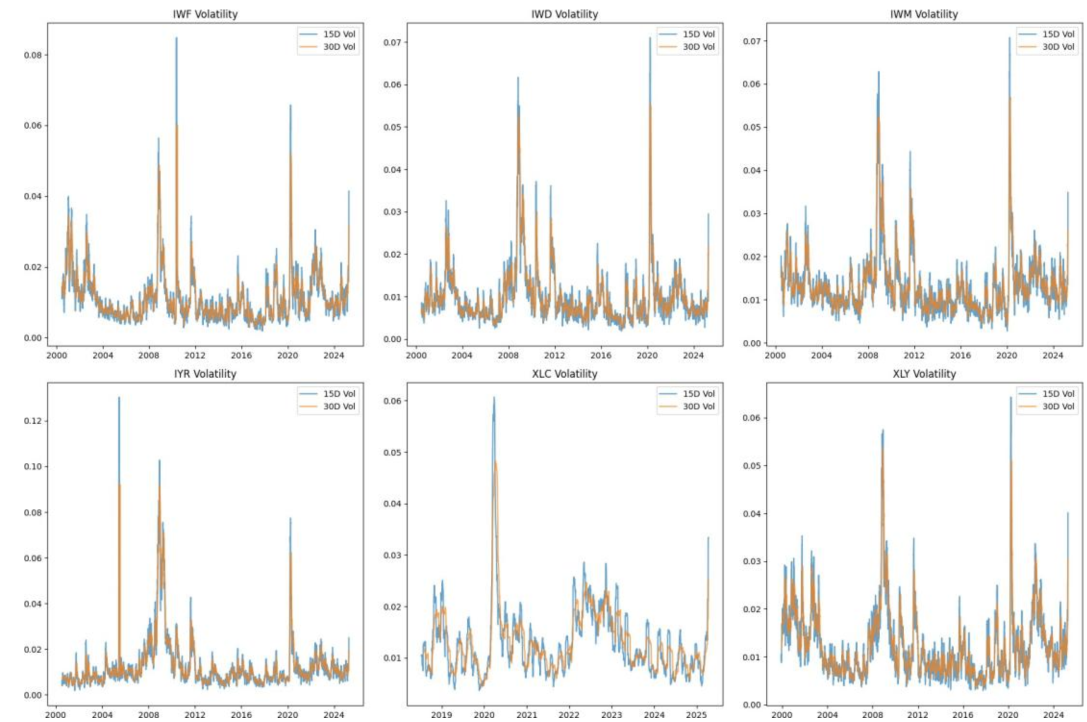

Figure 8. a: Rolling Volatility on ETF Returns (Part 1)

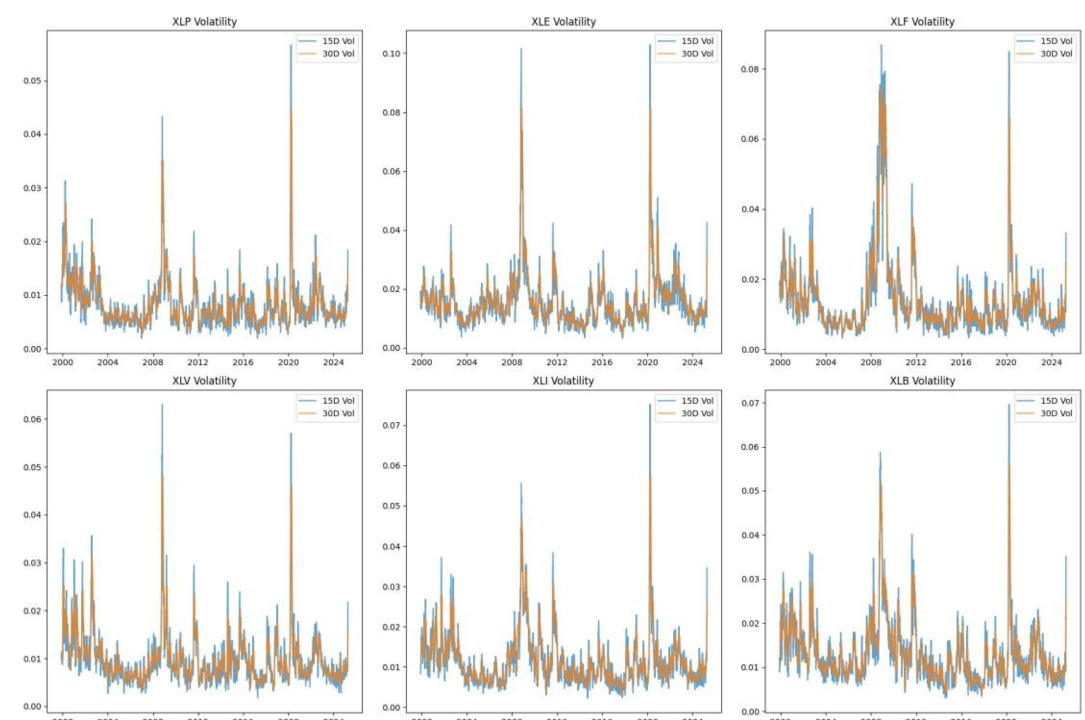

Figure 8. b: Rolling Volatility on ETF Returns (Part 2)

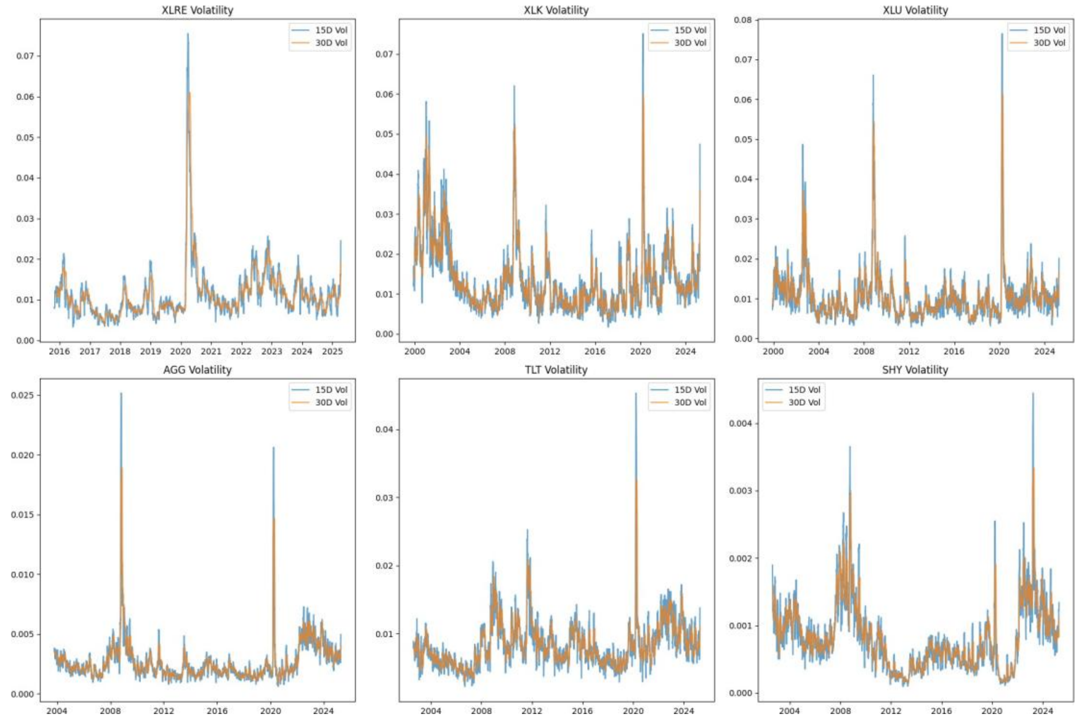

Figure 8. c: Rolling Volatility on ETF Returns (Part 3)

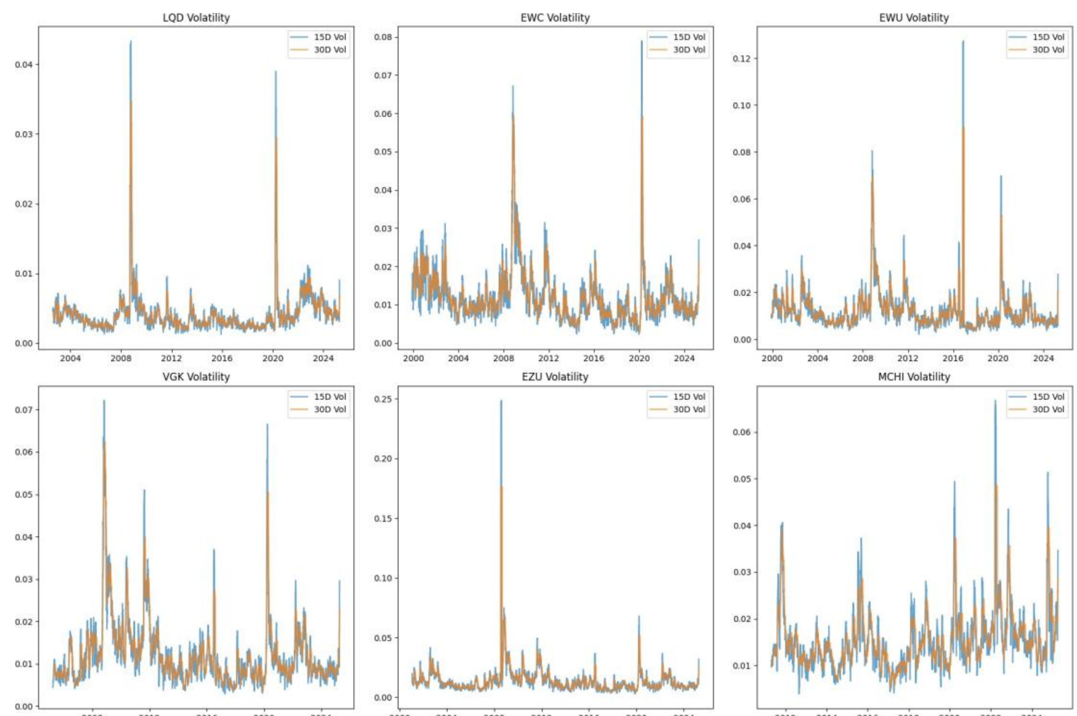

Figure 8. d: Rolling Volatility on ETF Returns (Part 4)

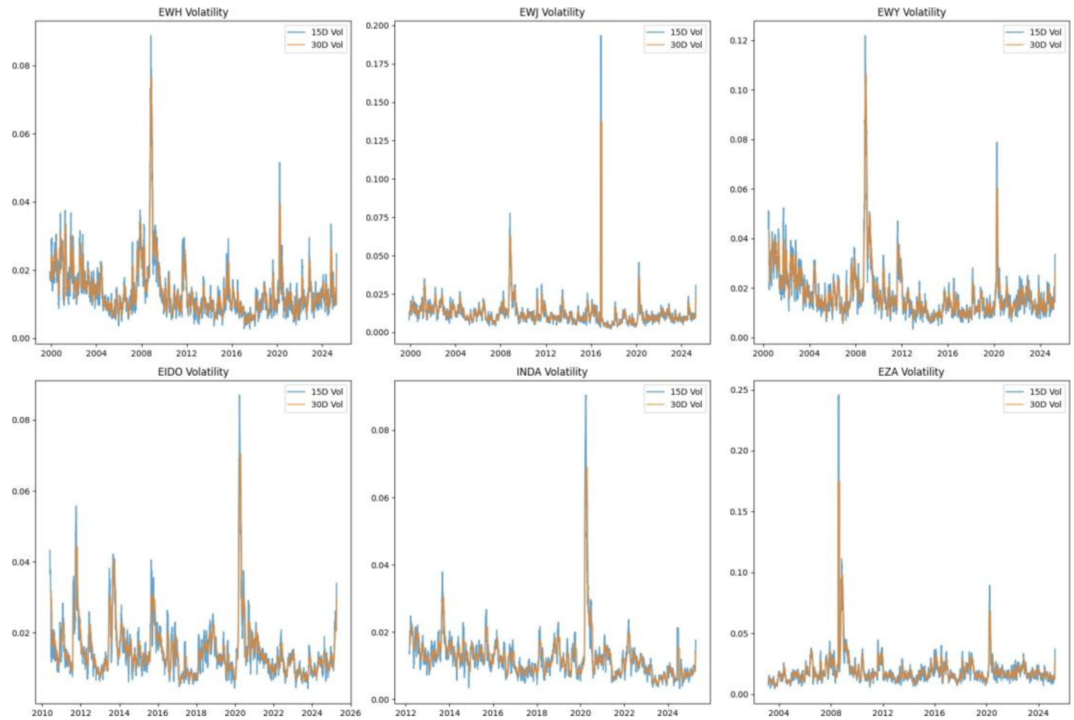

Figure 8. e: Rolling Volatility on ETF Returns (Part 5)

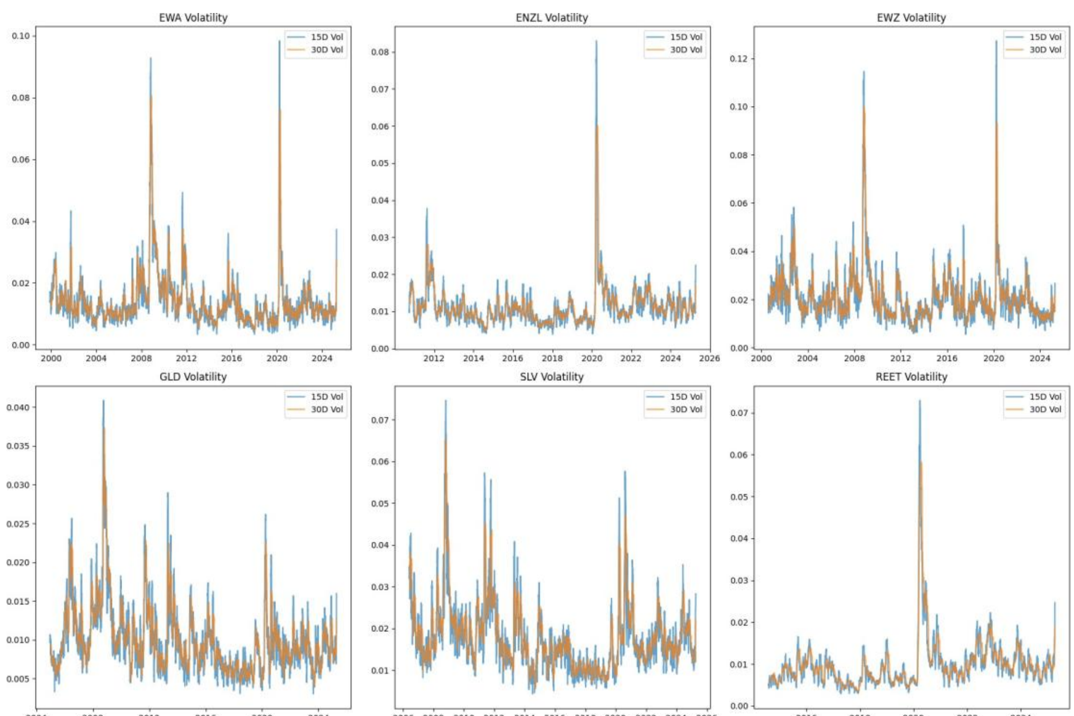

Figure 8. f: Rolling Volatility on ETF Returns (Part 6)

References

1. Park, Sangjin, and Jae-Suk Yang. “Interpretable Deep Learning LSTM Model for Intelligent Economic Decision-Making.” Knowledge-Based Systems, vol. 248, July 2022, p. 108907. Elsevier.
2. Staffini, Alessio. “A CNN–BiLSTM Architecture for Macroeconomic Time Series Forecasting.” Engineering Proceedings, vol. 39, no. 1, 2023, p. 33. MDPI.
3. Ramage, Daniel. Hidden Markov Models Fundamentals. CS229 Lecture Notes, Stanford University, 2007.
4. Kim, Eun-Chong, Han-Wook Jeong, and Nak-Young Lee. "Global Asset Allocation Strategy Using a Hidden Markov Model." Journal of Risk and Financial Management, vol. 12, no. 4, 2019, p. 168. MDPI,
5. Kritzman, Mark, David Turkington, and Page, Sébastien. "Identifying and Explaining the Number of Regimes Driving Asset Returns." The Journal of Portfolio Management, vol. 38, no. 2, 2012, pp. 44–53.
6. Nguyen, Nguyet, and Dung Nguyen. "Hidden Markov Model for Stock Selection." Risks, vol. 3, no. 4, 2015, pp. 455–472. MDPI.
7. Wang, Matthew, Yi-Hong Lin, and Ilya Mikhelson. "Regime-Switching Factor Investing with Hidden Markov Models." Journal of Risk and Financial Management, vol. 13, no. 12, 2020, p. 311. MDPI
8. Hambly, Ben, Ruyu Xu, and Haoyang Yang. Recent Advances in Reinforcement Learning in Finance. arXiv, 2021,.
9. Wang, Matthew, Yi-Hong Lin, and Ilya Mikhelson. "DeepTrader: A Deep Reinforcement Learning Approach for Risk-Return Balanced Portfolio Management with Market Conditions Embedding." Journal of Risk and Financial Management, vol. 13, no. 12, 2020, p. 311. MDPI.
10. Sun, Ruoyu, Angelos Stefanidis, Zhengyong Jiang, and Jionglong Su. Combining Transformer Based Deep Reinforcement Learning with Black-Litterman Model for Portfolio Optimization. arXiv, 23 Feb. 2024.
11. Katongo, Musonda, Ritabrata Bhattacharyya. The Use of Deep Reinforcement Learning in Tactical Asset Allocation. SSRN, 2021. SSRN ID: 3812609.
12. Li, Zhenglong, and Vincent Tam. Developing an Attention-Based Ensemble Learning Framework for Financial Portfolio Optimisation. arXiv, 13 Apr. 2024, arXiv:2404.08935.
13. Che, Zhengping, et al. “Recurrent Neural Networks for Multivariate Time Series with Missing Values.” Scientific Reports, vol. 8, no. 1, 2018, pp. 1–12.
14. Jurafsky, Daniel, and James H. Martin. “Hidden Markov Models.” Speech and Language Processing, draft 3rd ed., 12 Jan. 2025, web.stanford.edu/~jurafsky/slp3/A.pdf. Accessed 11 June 2025.
15. Achiam, Joshua. Spinning Up in Deep Reinforcement Learning. OpenAI, 2018.
16. Haarnoja, Tuomas, et al. Soft Actor-Critic Algorithms and Applications. arXiv preprint arXiv:1812.05905v2, 29 Jan. 2019.

17. Liu, Xiao-Yang, et al. FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance. arXiv:2011.09607v2 [q-fin.TR], 2 Mar. 2022. arXiv.
18. Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen. "Low-Risk Investing Without Industry Bets." SSRN, 10 May 2013.
19. Clare, Andrew, James Seaton, Peter N. Smith, and Stephen Thomas. "The Trend is Our Friend: Risk Parity, Momentum and Trend Following in Global Asset Allocation." SSRN, 31 July 2015.
20. Bekaert, Geert, and Campbell R. Harvey. "Emerging Markets Finance." Journal of Empirical Finance, vol. 10, no. 1-2, 2003, pp. 3–55. SSRN.
21. Bai, Jushan, and Serena Ng. "Determining the Number of Factors in Approximate Factor Models." Econometrica, vol. 70, no. 1, Jan. 2002, pp. 191–221.