
Revolutionizing Financial Portfolio Management: The NSTD Model's Fusion of Macroeconomic Indicators and Sentiment Analysis in a Deep Reinforcement Learning Framework

[Yuchen Liu](#) , [Daniil Mikriukov](#) , [Owen Christopher Tjahyadi](#) , Gangmin Li , Terry R. Payne , [Yong Yue](#) , [Kamran Siddique](#) , [Ka Lok Man](#) *

Posted Date: 12 December 2023

doi: 10.20944/preprints202312.0851.v1

Keywords: Portfolio Management (PM); Deep Reinforcement Learning (DRL); Non-Stationary 13 Transformer; Sequential Processing; Data Heterogeneity; Market Uncertainty; Diverse Knowledge 14 Integration; Multimodal Learning15



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Revolutionizing Financial Portfolio Management: The NSTD Model's Fusion of Macroeconomic Indicators and Sentiment Analysis in a Deep Reinforcement Learning Framework

Yuchen Liu ^{1,2} , Daniil Mikriukov ³ , Owen Christopher Tjahyadi ³ , Gangmin Li ⁴ ,
Terry R. Payne ² , Yong Yue ^{1,2} , Kamran Siddique ⁵  and Ka Lok Man ^{1,2,*} 

¹ Department of Computing, Xi'an Jiaotong-Liverpool University, Suzhou, China

² Department of Computer Science, University of Liverpool, Liverpool, United Kingdom

³ School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University, Suzhou, China

⁴ Faculty of Creative Arts, Technologies and Science, University of Bedfordshire, Luton, United Kingdom

⁵ Department of Computer Science & Engineering, University of Alaska Anchorage, Alaska, United States

* Correspondence: ka.man@xjtlu.edu.cn

Abstract: In the evolving landscape of Portfolio Management (PM), the fusion of advanced machine learning techniques with traditional financial methodologies has opened new avenues for innovation. Our study introduces a cutting-edge model combining Deep Reinforcement Learning (DRL) with a Non-stationary Transformer architecture. This model is designed to decode complex patterns in financial time series data, enhancing portfolio management strategies with deeper insights and robustness. It effectively tackles the challenges of data heterogeneity and market uncertainty, key obstacles in PM. Our approach integrates key macroeconomic indicators and targeted news sentiment analysis into its framework, capturing a comprehensive picture of market dynamics. This amalgamation of varied data types addresses the multifaceted nature of financial markets, enhancing the model's ability to navigate the complexities of asset management. Rigorous testing demonstrates the model's efficacy, highlighting the benefits of blending diverse data sources and sophisticated algorithmic approaches in mastering the nuances of PM.

Keywords: Portfolio Management (PM); Deep Reinforcement Learning (DRL); non-stationary transformer; sequential processing; data heterogeneity; market uncertainty; diverse knowledge integration; multimodal learning

1. Introduction

Portfolio Management (PM) is a multifaceted domain where investors strive to optimise financial assets to achieve long-term goals. Distinguished into passive and active management, each branch adopts unique strategies to fulfil these objectives. Active PM [1,2], characterized by dynamic trading to secure higher profits, contrasts with the conservative nature of passive PM, which aims to mirror market indices for steady, long-term gains. Prevailing practices in active PM often rely on predefined trends, which increasingly fall short against the complexities of volatile financial markets. This limitation has catalyzed a shift toward incorporating advanced machine learning and artificial intelligence technologies [3,4]. Reinforcement Learning (RL), renowned for its dynamic adaptability to market changes, has been applied in various PM models, such as iCNN [5], EIIE [6], SARL [7], AlphaStock [8], and GPM [9]. These models utilize historical asset prices and external information like financial news, extracting features to guide portfolio rebalancing and decision-making. However, the PM still grapples with challenges like extracting non-stationary temporal sequential features, integrating macro and microeconomic knowledge, and diversifying portfolios to minimize risk exposure.

To tackle these challenges, our study introduces a unique algorithmic approach that combines a Non-stationary Transformer with deep reinforcement learning algorithms (NSTD). This innovative

approach is designed to reconstitute non-stationary features post-stationary processing of time-series data, merge multi-modal information from diverse sources, and enrich portfolio diversity to mitigate concentration risks. The proposed model integrates macroeconomic indicators and sentiment scores from financial news into the Non-stationary Transformer framework, enhancing its ability to navigate the complexities of PM with data heterogeneity and environmental uncertainty. The model's enhanced performance, validated through rigorous experimentation, demonstrates the efficacy of combining diverse knowledge sources for PM.

This paper is structured as follows: Section 2 reviews related work in PM, focusing on the evolution of active and passive strategies and the role of machine learning in this domain. Section 3 elaborates on our methodology, detailing the integration of macroeconomic features and sentiment scores from news data into the Non-stationary Transformer and its synergy with DRL. Section 4 and 5 present our experimental setup and findings, offering a comparative analysis of the NSTD model against traditional and modern deep-learning financial strategies. Finally, Section 6 concludes the paper, summarizing our key findings and suggesting avenues for future research in PM.

2. Related Work

Portfolio management (PM) is primarily divided into passive and active strategies. The passive approach focuses on replicating the performance of specific indices. This approach is exemplified in studies like [10], which explore integrating Environmental, Social, and Governance (ESG) factors into passive portfolio strategies. Additionally, [11] discusses dynamic asset allocation strategies for passive management to achieve benchmark tracking objectives. In contrast, [12] emphasizes the evaluation methodologies crucial for aligning passive portfolios with their target indices. In contrast, active portfolio management seeks maximal profit independent of any specific index. Traditional active PM strategies, as categorized by [13], include equal fund allocation, follow-the-winner, follow-the-loser, and pattern-matching approaches. However, these strategies need more adaptability to accommodate real-world financial markets' unpredictable dynamics.

The evolving financial landscape and increasing data availability have highlighted the limitations of traditional PM methods. This realization has spurred a transition towards advanced machine learning technologies. The emergence of Deep Reinforcement Learning (DRL), combining deep learning (DL) and reinforcement learning (RL) [14–16], represents a significant advancement in addressing PM's inherent challenges. DRL synergizes the representational capabilities of DL with RL's strategic decision-making, which is ideal for the volatile finance sector. Early explorations like [17] and [18] showcased neural networks' potential in predicting market behaviour. The subsequent integration of RL, as seen in [5,19], enhanced decision-making capabilities in PM models. Furthermore, the time-series nature of financial data has led to the adoption of transformer mechanisms, as illustrated by [20], to unravel temporal information. This approach has been refined in studies such as [21,22], which aim to reduce the computational complexity of self-attention mechanisms. Further advancements, including Autoformer [23] and Pyraformer [24], have significantly contributed to robust time-series analysis and forecasting.

The Non-stationary Transformer framework, introduced in [25], tackles the challenge of non-stationarity in time-series data, a critical limitation of conventional transformers. This framework, integrating Series Stationarization and De-stationary Attention modules, optimises the processing of non-stationary data, thereby enhancing transformer performance across diverse forecasting scenarios. This novel approach aligns well with reinforcement learning algorithms, offering enhanced capabilities for PM. The framework's ability to adapt to non-stationary data makes it particularly relevant to the challenges and complexities encountered in passive and active PM.

3. Methodology

Integrating Deep Reinforcement Learning (DRL) with macroeconomic modelling has revolutionized portfolio management, leveraging insights from macroeconomic factors and sentiment

analysis from news data. This study further refines this approach by employing a Non-stationary Transformer combined with a Deep Deterministic Policy Gradient (DDPG) architecture [32], as illustrated in Figure 1.

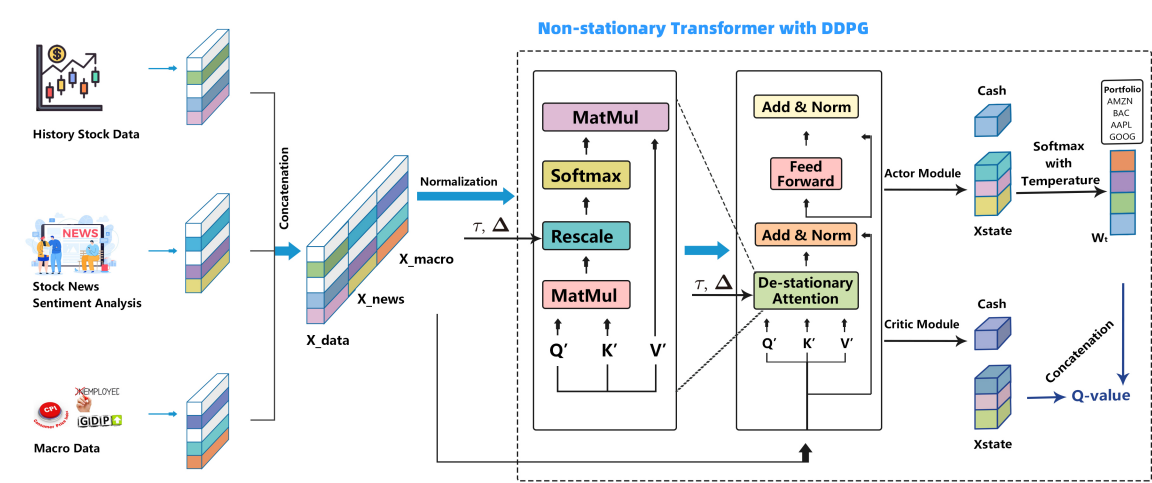


Figure 1. NSTD (Non-stationary Transformer with Deep Reinforcement Learning) Framework.

3.1. Incorporating Macroeconomic Analysis through Composite Indicators

3.1.1. Theoretical Framework for Financial Metrics in Macroeconomics

The interconnection between macroeconomic conditions and financial markets plays a crucial role in shaping investment strategies. Macroeconomic analysis, as a tool, reflects on broader economic states and influences market trajectories, thus informing financial decision-making in a DRL-augmented environment [26,27]. This symbiosis between economic indicators and market dynamics forms the theoretical basis for integrating macroeconomic insights into financial metrics.

3.1.2. Data Collection and Preprocessing

The study utilizes macroeconomic data from 2012 to 2016, with monthly GDP data from BBKI, CPI and Unemployment Rate from the Federal Reserve Economic Data (FRED) database. The stock dataset, operating daily, incorporates these macroeconomic indicators, assigning consistent monthly values from BBKI and FRED to each trading day. This methodology aligns daily stock market activity with broader economic trends, with additional data normalization and alignment steps ensuring consistency across different data sources.

3.1.3. Impact of Macroeconomic Indicators on Market Movements

This section outlines the influence of key macroeconomic indicators on market dynamics:

- **Real GDP:** Real GDP, sourced from the Bureau of Economic Analysis (BEA), represents the total economic output, calculated as $GDP = C + G + I + NX$, where C is consumer spending, G is government expenditures, I is investment, and NX is net exports.
- **Consumer Price Index (CPI):** Provided by the Bureau of Labor Statistics (BLS), the CPI reflects changes in the price level of a consumer goods and services market basket, calculated through a two-stage process involving basic indexes for each item-area combination and aggregate indexes.
- **Unemployment Rate:** Reported by the BLS, the unemployment rate is based on projections from an interindustry model and the National Employment Matrix, assuming full employment conditions.

These indicators, essential in assessing the economy's health, significantly impact forecasting market trends. Incorporating macroeconomic indicators into portfolio management provides a nuanced understanding of market movements, which is essential for informed investment decisions. Portfolio managers can align their strategies with economic trends by analyzing indicators like GDP growth, inflation data, and employment statistics. This process involves using these indicators as predictive tools for market trajectory, aiding in risk mitigation and opportunity identification, especially in the context of major stock indices like the S&P 500.

3.2. Incorporating Sentiment Analysis for Stock Insight from News Data

3.2.1. Framework for Sentiment Analysis

In the realm of stock market analysis, the integration of sentiment analysis has emerged as a pivotal tool for deriving insights from news articles related to individual stocks. This methodology utilises a sentiment analysis framework, which assigns a sentiment score to each news headline. These scores range from -1, denoting a strongly negative sentiment, to 1, indicating a strongly positive sentiment, with 0 representing a neutral stance. This quantification of news sentiment is instrumental in evaluating its potential impact on the microeconomic variables associated with stocks, as highlighted in the works by Devlin et al. [28] and Yang et al. [29].

3.2.2. Data Collection and Preprocessing

The initial phase of this methodology involves procuring a comprehensive dataset of stock-related news articles, as provided by the source [30]. This dataset, comprising a substantial collection of news headlines, is the foundation for the subsequent sentiment analysis. The preprocessing of this dataset includes a thorough cleansing process to eliminate inconsistencies, converting date formats into a standardised form, and organising the data chronologically based on the date and the stock symbol. This meticulous preprocessing ensures that the dataset is primed for accurate sentiment analysis.

3.2.3. Sentiment Analysis and Lexicon Enhancement

The sentiment analysis uses the VADER (Valence Aware Dictionary and sEntiment Reasoner) dictionary [31], a tool recognised for its proficiency in analysing texts from social media and news headlines. Its effectiveness stems from its ability to discern the polarity and the intensity of emotions conveyed in textual data. To refine the sentiment analysis further, the lexicon is enhanced by incorporating the 25 most frequently occurring words in the dataset's headlines. Each word is assigned a specific sentiment value, calibrated on a scale from -1 to 1. This augmentation of the lexicon is aimed at bolstering the precision of the sentiment analysis.

3.2.4. Sentiment Scoring and Data Structuring

Upon completion of the sentiment analysis, each news headline is ascribed a sentiment score based on the augmented VADER lexicon. The outcome of this analysis is a structured dataset, which includes the original news headlines and their corresponding sentiment scores. This dataset is uniquely formatted to encapsulate the date of the news article, the sentiment score, and the associated stock symbol. Such a structured arrangement of data is specifically designed to facilitate its seamless integration into the non-stationary transformer portfolio management model. Including this sentiment-based information is anticipated to enhance the model's ability to make more informed and accurate predictions regarding stock market trends and behaviours.

3.3. Non-stationary Transformer with Deep Deterministic Policy Gradient

The advancement in portfolio management techniques has led us to develop the 'Adaptive Transformer' (Non-stationary Transformer) within the established Deep Deterministic Policy Gradient

(DDPG) framework [32]. This innovative element is tailored to recognise and adapt to the dynamic statistical characteristics of financial time series data, thereby enhancing the model's predictive accuracy. The architecture of our Adaptive Transformer is composed of four key components, each intricately designed to work in harmony, ensuring the model's adeptness in navigating the ever-changing landscape of financial data. We delve into the specifics of these components:

3.3.1. Projector Layer

At the core of the Adaptive Transformer resides the Projector Layer, an intricate multi-layer perceptron (MLP). It recognises and adapts to non-stationary elements within sequential data streams. The procedure initiates with a global average pooling operation to reduce temporal complexity, thereby isolating pivotal attributes:

$$x_{\text{pooled}} = \text{GlobalAvgPooling}(x), \quad (1)$$

After this reduction of data, a sequence of dense layers ensues. These layers are enhanced with ReLU activation functions and L2 regularisation, a strategy that aids in distilling and integrating intricate data features. The process reaches its zenith with the application of a hyperbolic tangent function, culminating in an output that encapsulates the identified non-stationary elements:

$$y = \tanh(\text{Dense}_{\text{hidden}}(x_{\text{pooled}})). \quad (2)$$

3.3.2. Transformer Encoder Layer

The Transformer Encoder Layer, an integral part of the Adaptive Transformer, has a self-attention mechanism. This mechanism is particularly proficient in meticulously examining and interpreting financial data sequences. Embedded within this layer are two pivotal adaptive elements, termed τ_{learner} and Δ_{learner} . These components are instrumental in learning temporal scaling and shifting factors, endowing the model with a refined acuity for adjusting its self-attention outputs to the present non-stationary conditions:

$$\tau = \exp(\text{Projector}_{\tau}(x_{\text{raw}}, \sigma_{\text{enc}})), \quad \Delta = \text{Projector}_{\Delta}(x_{\text{raw}}, \mu_{\text{enc}}), \quad (3)$$

where σ_{enc} and μ_{enc} denote the standard deviation and mean of the input sequences, respectively. The recalibrated attention output is thus articulated as follows:

$$\text{Attn}_{\text{output}} = \text{MHA}(x) \cdot \tau + \Delta. \quad (4)$$

To promote robustness and prevent overfitting, layer normalisation and a dropout strategy are applied after that:

$$x = \text{LayerNorm}(x + \text{Dropout}(\text{Attn}_{\text{output}})). \quad (5)$$

3.3.3. Policy Network with Transformer

The Policy Network with Transformer is meticulously engineered to formulate a strategic policy for decision-making processes. It integrates state representations with the robust architecture of the Transformer, succeeded by batch normalisation and a series of dense layers, thereby creating a refined state-to-action mapping function represented as:

$$\text{Net} = \text{Dense}_{\text{ReLU}}(\text{BatchNorm}(\text{Transformer}(x_{\text{state}}))). \quad (6)$$

Following this, a softmax function, tempered by a specified parameter, delineates the action space. This process is essential in balancing the imperative dichotomy between exploratory and exploitative behaviours:

$$\text{Action} = \text{Softmax}_T(\text{Dense}(\text{Net})) \times \text{action_bound}. \quad (7)$$

This equation signifies the culmination of the network's computation, where the softmax function, adjusted by a temperature factor, contributes to a probabilistic selection of actions constrained within the defined bounds of possible actions.

3.3.4. Q-Value Network with Transformer

Simultaneously, the Q-Value Network with Transformer is tasked with computing the expected returns for distinct state-action combinations. This computation leverages a coeval Transformer encoder framework. The inputs corresponding to states and actions are processed in isolation, normalised, and amalgamated. This precedes their progression through a successive array of dense layers, which culminate in the distillation of the Q-value, a singular measure of the projected reward:

$$Q(s, a) = \text{Dense}(\text{Concat}(\text{BatchNorm}(s), \text{BatchNorm}(a))). \quad (8)$$

The resultant Q-value is then articulated as:

$$Q_{\text{value}} = \text{Dense}(\text{Net}_{\text{combined}}). \quad (9)$$

This expression encapsulates the network's computation, where the Q-value embodies the expected utility of adopting a specific action in a given state, as per the model's learned policy.

3.3.5. Optimisation and Policy Learning within DDPG Framework

The DDPG (Deep Deterministic Policy Gradient) algorithm lies at the heart of our optimisation technique. It is a model-free, off-policy actor-critic algorithm using deep function approximators that can learn policies in high-dimensional, continuous action spaces. Our implementation encapsulates an actor-network designed to map states to actions and a critic network that evaluates the action given the state.

The initialisation of these networks is performed with random weights, which are subsequently refined through training iterations. The actor-network proposes an action given the current state, while the critic network appraises the proposed action by estimating the Q-value. The policy, parameterised by θ , generates actions predicated on the current state, aiming to maximise the cumulative reward:

$$J(\mu_\theta) = \int_S \rho_\mu(s) r(s, \mu_\theta(s)) ds = \mathbb{E}_{s \sim \rho_\mu} [r(s, \mu_\theta(s))]. \quad (10)$$

where μ_θ is the deterministic policy, Q^μ is the action-value function according to policy μ , and ρ^μ is the state distribution under policy μ . Considering the dynamic realm of portfolio management, we recalibrate this function to account for the sequential product of portfolio values:

$$J_T(\mu_\theta) = J_0 \prod_{t=1}^T (1 + r(s_t, \mu_\theta(s_t))), \quad (11)$$

where J_0 represents the initial portfolio valuation.

To ensure stability and promote convergence, we employ target networks for actor and critic, which are slowly updated to track the learned networks. The update for the target networks is governed by a soft update parameter τ , ensuring that the targets change gradually, which helps with the stability of learning:

$$\theta_{\text{target}} \leftarrow \tau \theta + (1 - \tau) \theta_{\text{target}}. \quad (12)$$

The policy is refined through gradient ascent on the expected cumulative reward, while the critic's parameters are adjusted based on the temporal-difference error. To maintain exploration, a noise process \mathcal{N} is added to the actor's policy action:

$$a_t = \mu_\theta(s_t) + \mathcal{N}_t, \quad (13)$$

Where a_t is the chosen action, μ_θ is the current policy, s_t is the current state, and \mathcal{N}_t is the noise term at time t , which decays during training to allow the policy to exploit the learned values as it matures.

This methodical learning process culminates in developing a policy that maximises expected returns, aligning with our objective of enhancing portfolio management strategies.

3.3.6. Action and Reward Formulation

At any discrete time step t , the action a_t is identified with the portfolio's asset allocation vector w_t . The policy endeavours to optimise the allocation for the forthcoming period w_{t+1} , maintaining the constraint of unitary sum across all asset weights. The updated allocation, which adapts to price fluctuations y_t , is formalised as:

$$w_{t+1} = \frac{y_t \odot w_t}{y_t \cdot w_t}, \quad (14)$$

Here, \odot denotes the Hadamard product, and \cdot represents the conventional dot product.

The reward function, a pivotal determinant of the portfolio's efficacy, encompasses transaction costs denoted by β , and is thus defined to capture the realised profit at time t :

$$r_t = w_t \cdot y_t - \beta \sum_{i=1}^n |w_{i,t+1} - w_{i,t}|, \quad (15)$$

Enhancing the practical applicability and differentiability of the reward function, we integrate fixed commission rates c_b for buying and c_s for selling. The reward function is consequently defined as:

$$r_t = \ln(\beta_t \cdot (w_t \cdot y_t)). \quad (16)$$

The ultimate objective is to amplify the cumulative return R across a designated interval T , expressed as:

$$R(s_1, w_1, \dots, s_T, w_T) = \frac{1}{T} \sum_{t=1}^T \ln(\beta_t \cdot (w_t \cdot y_t)). \quad (17)$$

The optimisation conundrum thus revolves around ascertaining the optimal policy μ_θ that maximises this return:

$$\mu_\theta^* = \arg \max_{\mu_\theta} \frac{1}{T} \sum_{t=1}^T \ln(\beta_t \cdot (w_t \cdot y_t)). \quad (18)$$

The policy gradient with respect to the parameters θ is meticulously derived:

$$\nabla_\theta \mu_\theta(\tau) = \frac{1}{T} \sum_{t=1}^T \nabla_\theta \ln(\mu_\theta(w_t, s_t)), \quad (19)$$

and the parameters are then conscientiously updated with a learning rate λ :

$$\theta \leftarrow \theta + \lambda \nabla_\theta J_T(\mu_\theta). \quad (20)$$

The normalisation by T is crucial as it appropriately scales the gradient irrespective of batch sizes, an essential factor for effective mini-batch training in a stochastic environment.

Overall, this structured approach in the methodology section ensures a comprehensive understanding of the sophisticated Non-stationary Transformer model and its integration with the DDPG framework, which is crucial for advanced portfolio management techniques.

4. Experiment Setup

4.1. Dataset

Our analysis focuses on S&P 500 data from 2012 to 2016 (Table 1), chosen for its market stability and diverse economic conditions. We selected the top 100 stocks by trading volume from the 2012 S&P 500 index for efficiency. This selection means our stock portfolio is based on these 100 stocks, allowing for a targeted and effective training approach for our deep reinforcement learning models. To ensure robust model training and validation, we have divided our dataset into distinct subsets:

- **Training Set (2012-2014):** The initial phase of our analysis utilizes data from 2012 to 2014 as the training set. This period is crucial for training our models, tuning hyperparameters, and optimizing algorithms based on historical market trends.
- **Validation Set (2015):** Data from 2015 is used as the validation set. The primary purpose of this phase is to compare different time window hyperparameters and select the optimal ones for application in our models. This step is critical for assessing the performance of our models under different market conditions and ensuring that they can generalize well to new data.
- **Testing Set (2016):** Finally, the 2016 data serves as our testing set. In this phase, we apply the models refined through training and validation to this unseen dataset. This allows us to compare the performance of various models and configurations, providing insights into their effectiveness and robustness.

The rationale behind this structured approach to dataset splitting is to ensure that our models are well-trained on historical data and thoroughly validated before being tested. By comparing the performance of models trained and validated on different subsets, we can more accurately assess their predictive capabilities and adaptability to market changes. This methodical approach significantly enhances the reliability and validity of our findings.

Table 1. Statistics of Train, Validate and Test Datasets.

Period	Date Range	#Examples
Train	2012-01-02 to 2014-12-31	754
Validate	2015-01-04 to 2015-12-31	252
Test	2016-01-02 to 2016-12-30	252

4.2. Compared Models

Our empirical evaluation contextualises the performance of the Non-stationary Transformer with the DDPG (NSTD) model by benchmarking it against a selection of advanced deep learning models and traditional financial strategies. The models in this comparative analysis include the following:

1. Deep Learning Approaches:

- **Transformer [20]:** Employs self-attention mechanisms for long-range dependencies in time series data.
- **LSTM (Long Short-Term Memory) [33]:** Excels in learning from sequences with long temporal dependencies.
- **GRU (Gated Recurrent Unit) [34]:** Designed to capture dependencies of varying time scales in sequential data.
- **RNN (Recurrent Neural Network) [35]:** Processes temporal sequences through a directed graph architecture.

- **EIIE [6] (Ensemble of Identical Independent Evaluators):** Processes time series for each asset independently while maintaining shared network parameters.
- **EI³ [36] (Ensemble of Identical Independent Inception):** Extracts and aggregates multi-scale price movement information, integrated with a recurrent reinforcement learning framework.

2. Traditional Financial Strategies:

- **OLMAR [37] (Online Moving Average Reversion):** Utilizes mean reversion effect based on moving average.
- **UBAH [38] (Uniform Buy and Hold):** Allocates funds uniformly across assets, maintaining them without further trading actions.

4.3. Evaluating Investment Strategies with Financial Metrics

We employ several financial metrics to evaluate the efficiency and risk of investment portfolios:

4.3.1. Accumulative Return (AR) [39]

Accumulative Return (AR) is a crucial metric that quantifies the total return generated by an investment relative to its initial capital. It is calculated as the ratio of the final portfolio value p_T to the initial portfolio value p_0 , directly measuring the overall return on investment. Mathematically, it is expressed as:

$$AR = \frac{p_T}{p_0} \quad (21)$$

where p_T represents the portfolio's value at the end of the investment period, and p_0 is the portfolio's initial value. This metric is particularly useful for assessing an investment's growth or decline over a specified time frame. A higher AR indicates a greater return on the initial capital invested, signifying effective portfolio management and investment strategy. Conversely, an AR of less than one signifies a decline in the value of the initial investment, indicating potential areas for strategic adjustment in portfolio management. The AR is thus an integral component of a comprehensive investment performance analysis, providing a straightforward yet powerful tool for investors to gauge the efficacy of their investment decisions.

4.3.2. Sharpe Ratio [40]

The Sharpe Ratio measures risk-adjusted returns:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \quad (22)$$

where R_p is the return of the portfolio, R_f the risk-free rate, and σ_p the standard deviation of the portfolio's excess return. A Sharpe Ratio greater than one is generally considered good, indicating adequate returns relative to the risk taken.

4.3.3. Maximum Drawdown (MDD) [41]

MDD assesses the largest drop in portfolio value:

$$MDD = \frac{\text{Trough Value} - \text{Peak Value}}{\text{Peak Value}} \quad (23)$$

A lower MDD is preferable, indicating less potential loss.

4.3.4. Sortino Ratio [42]

The Sortino Ratio differentiates harmful volatility from overall volatility:

$$\text{Sortino Ratio} = \frac{R_p - R_f}{\sigma_d} \quad (24)$$

where σ_d is the standard deviation of negative asset returns. A higher Sortino Ratio indicates better performance with respect to downside risk.

These metrics offer valuable insights for strategic portfolio positioning and serve as benchmarks for evaluating post-investment performance. When integrated with DRL strategies, they optimise investment strategies, adapting to changing economic landscapes.

5. Experiment and Analysis

Our experiment used the S&P 500 data from 2012 to 2014 as our training dataset. This selection was instrumental in training our models, providing a comprehensive representation of market activities over this period. To circumvent memory overflow issues commonly encountered in prolonged training sessions, we adopted a strategy of fine-tuning model parameters after every set of 200 episodes. The training process was executed in a phased manner, with each model undergoing training ten times, each consisting of 200 episodes, amounting to a total of 2000 training rounds. The learning rate was subsequently reduced to 90% of its value from the previous phase. This progressive reduction in the learning rate facilitated nuanced model tuning and optimisation. The optimisation was conducted using the Adam optimiser.

In addition to the primary training process, we conducted comprehensive comparative experiments to determine the most effective hyperparameters for our model. These experiments focused on two key hyperparameters: the temperature coefficient and window length. The window length, a crucial hyperparameter in time series analysis, was also rigorously examined. To ascertain the most effective window length for capturing market dynamics, we considered a range of durations, specifically 3, 5, 7, 10, 15, 20, 30, and 50 days. For the temperature coefficient, to control the portfolio diversity, we explored various settings, including 500, 1000, 3000, 5000 and 10000, to observe the impact on model performance. Each model, configured with a distinct set of these hyperparameters, was initially trained on our training dataset (S&P 500 data from 2012 to 2014). Following the training phase, we saved the model parameters and then validated their performance on a separate validation dataset from 2015. This approach allowed us to thoroughly assess the influence of each hyperparameter on the model's efficacy, enabling us to fine-tune our model for optimal performance.

Upon identifying the most effective hyperparameters through our rigorous evaluation process, we applied these optimised settings to a backtest dataset from 2016. This phase of our research involved conducting comparative ablation studies using four distinct dataset compositions, each offering unique insights into the model's performance. The primary dataset encompassed a comprehensive blend of macroeconomic indicators, sentiment scores from news articles, and historical stock trading data, providing a holistic view of market influences. The second dataset was curated by excluding macroeconomic indicators while maintaining sentiment scores and historical trading data, allowing us to assess the impact of macroeconomic factors on model performance. In contrast, the third dataset excluded news sentiment scores but included macroeconomic indicators alongside historical data, enabling us to isolate and evaluate the influence of sentiment analysis. The fourth and final dataset was streamlined to include only historical stock trading data, offering a baseline to gauge the added value of macroeconomic and sentiment data.

Moreover, to contextualise our model's performance within the broader landscape of financial analytics, we conducted comparative analyses against other leading deep learning models and established traditional financial models. This comprehensive approach allowed us to validate our

model’s efficacy and understand its standing relative to the current state-of-the-art methodologies in financial portfolio management.

5.1. Evaluation on Validated Dataset

Upon training our models with the dataset from 2012 to 2014, we evaluated the efficacy of various hyperparameters—specifically, the window length and temperature coefficients—using the 2015 dataset as our validation set. This step was crucial to calibrate the models effectively before final testing.

In Figure 2, the analysis of cumulative returns for the validation dataset across varying window lengths reveals the strategic influence of the chosen time frame. The 10-day window (NSTD-10), marked in red, consistently surpasses other strategies from mid-year to year-end, indicating that a 10-day span effectively leverages historical data to inform current market predictions. Shorter windows like the 3-day (NSTD-3) and 5-day (NSTD-5) show early promise but later falter, suggesting they may be too narrow to grasp the larger cyclical movements of the market that a 10-day window encapsulates. Conversely, the longer window strategies—NSTD-30 and NSTD-50—demonstrate smoother return trajectories but do not achieve the higher returns seen with NSTD-10. This implies that while longer windows provide a broader historical context, potentially capturing more extended market cycles, they may lack the agility to capitalize on immediate market shifts. The 10-day strategy emerges as the most effective, balancing immediate market reactivity with understanding broader market movements, highlighting its efficiency for this period.

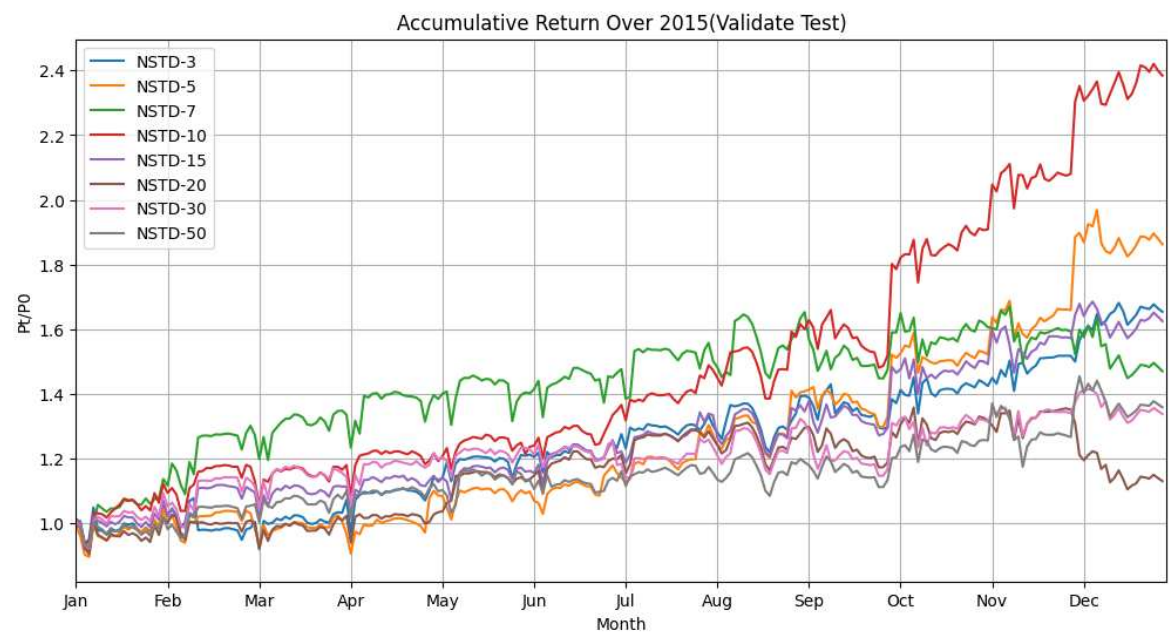


Figure 2. Compared with Different Window Length on 2015 (Validation Set)

In Figure 3, the cumulative returns for the validated dataset are mapped out across different temperature coefficients (T) on portfolio performance. The parameter T influences the dispersion of stock weights within the portfolio, with higher T values promoting diversification and lower values indicating concentration. Portfolios with T=5000 and T=10000, respectively, demonstrate a marked increase in cumulative return, especially noticeable from mid-year onward, with T=10000 displaying a pronounced uptick as the year concludes. This trend suggests that higher temperature settings, while promoting a more balanced weight distribution across the portfolio, may also enable the capture of growth opportunities in a rising market, contributing to higher overall returns. Conversely, the T=500 and T=1000 settings indicate a preference for a more concentrated portfolio. While this strategy may yield higher short-term gains by focusing on a limited number of stocks, it also introduces heightened

risk, as the portfolio is more vulnerable to the volatility of its few constituents. These lower T values produced more modest growth throughout the year, with occasional dips reflecting the increased risk exposure. The T=3000 setting balances the two extremes. Portfolios with this temperature setting deliver steady, consistent growth over the year, suggesting that a moderate T value can effectively balance the benefits of diversification with the potential gains from a more focused investment strategy. Overall, the temperature parameter's calibration is crucial for managing risk and capturing growth within a portfolio.



Figure 3. Compared with Different Temperature Controller on 2015 (Validation Set)

5.2. Ablation Study for Features

After getting optimal hyperparameters, we tested our model based on different features in 2016 (Test Set). The NSTD-10 comprehensive includes macroeconomic indicators, sentiment-derived news scores, and historical stock trading data. The NSTD without Macro excludes macroeconomic indicators while retaining news scores and historical data. The NSTD without News removes news scores featuring macroeconomic indicators and historical data. The NSTD Pure is the most pared-down, solely comprising historical stock trading data.

Figure 4 shows the year-long trajectory of cumulative returns for each NSTD variant. The NSTD-10 model sets a high standard with an impressive performance, indicating that a 10-day time window robustly captures market trends when combined with a full feature set. Meanwhile, the NSTD without the macro variant experiences a moderate decline in return. This suggests that while macroeconomic indicators contribute to performance, their absence does not drastically diminish the model's efficacy. Conversely, the model without news sentiment maintains a competitive AR close to the full model's, suggesting that excluding sentiment indicators might not significantly impact the model's performance in the presence of other data types. Yet, Shown in Table 2, the altered ratios of Sharpe and Sortino indicate that news scores play a role in refining risk-adjusted returns. Most notably, the NSTD Pure model demonstrates a significant decrease in AR and Sharpe Ratio, emphasizing the collective value of integrating both macroeconomic and sentiment indicators into the model. This variant's underperformance highlights the complexity of market dynamics and the importance of a diversified information set for effective market analysis. It is evident that the comprehensive NSTD-10 model, with its full suite of indicators, excels, underscoring the importance of a multifaceted approach to capturing the full spectrum of market signals.

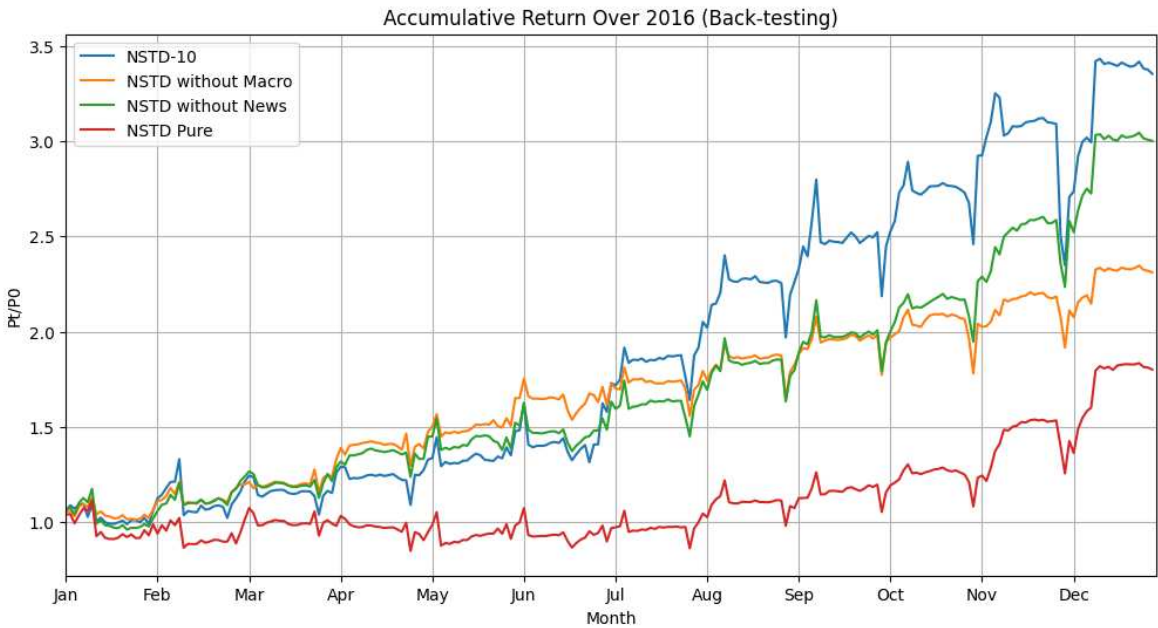


Figure 4. Back-testing Compared with Ablation Study on 2016

Table 2. Different Metrics Performance of Models

Model	AR	MDD	Sharpe Ratio	Sortino Ratio
NSTD 10	3.353	0.277	1.859	282.140
NSTD without News	3.000	0.182	1.273	78.099
NSTD without Macro	2.311	0.158	1.668	0.000
NSTD Pure	1.801	0.237	0.443	3.101
Transformer	1.460	0.231	1.119	7.823
RNN	1.953	0.282	1.303	29.622
LSTM	1.921	0.248	1.824	0.000
GRU	1.622	0.251	1.772	158.717
EI ³	2.196	0.291	1.446	27.188
EIIE	2.174	0.298	1.773	0.000
OLMAR	2.349	0.221	1.517	98.337
UBAH	2.349	0.221	1.517	98.337

5.3. Comparison with Other Deep Learning and Traditional Financial Methods

In demonstrating the superior performance of the Non-stationary Transformer with Deep Deterministic Policy Gradient (NSTD-10) model, we conducted an extensive back-testing analysis over 2016 compared it with deep learning architectures—such as the Transformer, RNN, LSTM, and GRU as well state-of-the-art EIIE, EI³, along with traditional financial algorithms like Online Moving Average Reversion (OLMAR) and Universal Portfolio Algorithm (UBAH). The evaluation hinged on critical financial metrics: Annual Return (AR), Maximum Drawdown (MDD), Sharpe Ratio, and Sortino Ratio. The analysis depicted in Figure 5, when considered alongside the quantitative metrics from Table 2, offers a nuanced assessment of various model performances. The NSTD-10 model emerges as the frontrunner, boasting an Annual Return (AR) of 3.353, indicative of its effective market trend

utilisation for optimal capital increase. Its Sharpe Ratio of 1.859 and a notably high Sortino Ratio of 282.140 reflect its adeptness at securing returns per unit of risk, with a pronounced efficiency in mitigating downside risk. Delving into the NSTD model variations, the omission of news sentiment analysis (NSTD without News) results in a slight decrease in AR to 3.000 and a marked reduction in the Sortino Ratio to 78.099, highlighting the significant contribution of sentiment data to forecasting precision and risk-adjusted returns. Excluding macroeconomic indicators (NSTD without Macro), we see a reduced AR of 2.311. However, it attains the lowest Maximum Drawdown (MDD) at 0.158, signifying macroeconomic insights' protective role against market downturns. The NSTD Pure variant, which solely incorporates historical trading data, reiterates the value of a diversified data approach, as evidenced by its lower metrics across the spectrum. Compared with established deep learning models such as the Transformer, RNN, LSTM, and GRU, these models exhibit commendable results but do not match the NSTD-10's superior standards. Specifically, the LSTM model's high Sharpe Ratio of 1.824 underscores its strength in handling time-series data. Nevertheless, this does not correlate with a high AR or Sortino Ratio, pointing to potential limitations in market dynamics adaptation. The ensemble approaches EP^3 , and EIIE offers more consistent returns. However, their Sharpe and Sortino Ratios fall short of those achieved by NSTD-10, suggesting that while reliable, they may not optimise the balance between risk and return. Traditional financial methods like OLMAR and UBAH maintain stable risk management but lack the NSTD-10's efficacy in capitalising on market conditions. Overall, the NSTD-10 model's unmatched performance in terms of overall returns and risk management sets a new standard in portfolio management. It underscores the strategic benefit of integrating comprehensive market data, from economic indicators to sentiment analysis, within a sophisticated algorithmic framework.

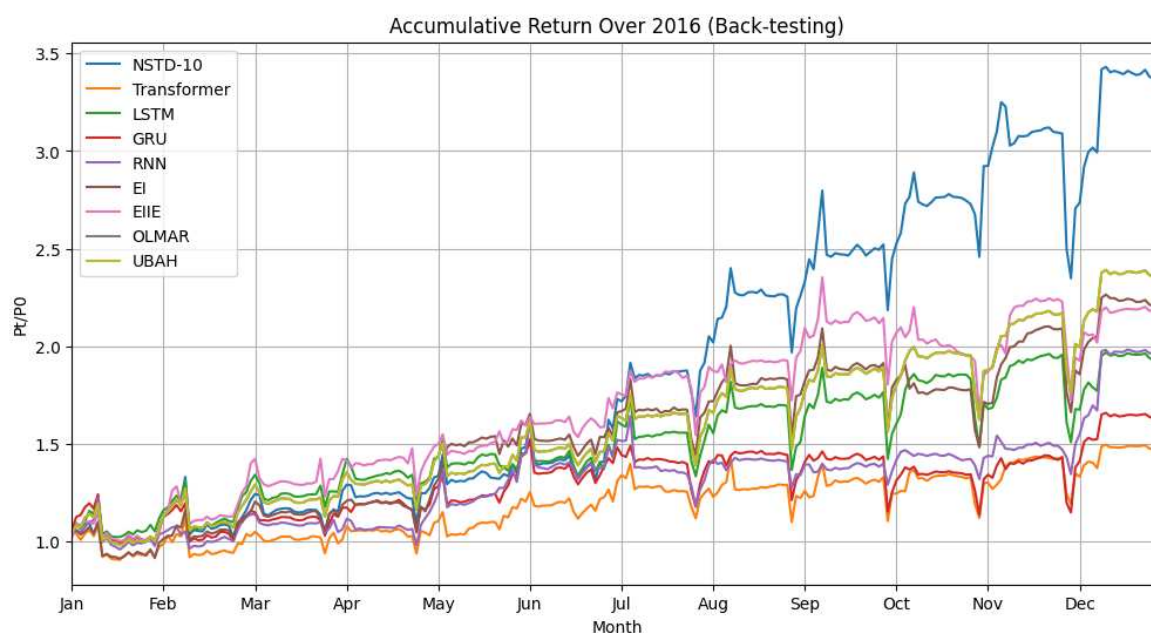


Figure 5. Back-testing Compared with Different Models on 2016

6. Conclusion

This study presents the Non-stationary Transformer with the Deep Deterministic Policy Gradient (NSTD) model, a breakthrough in portfolio management combining macroeconomic insights and sentiment analysis. Our evaluation across various financial strategies demonstrates NSTD's superior performance, particularly using a 10-day window. This model effectively captures market trends, balancing immediate reactivity with long-term insights. Integrating diverse data sources into the NSTD framework significantly enhances investment strategy and decision-making, offering a robust

approach to the complexities of financial markets. The NSTD model marks a substantial advancement in applying machine learning to portfolio management, paving the way for more informed and adaptive investment methodologies.

Author Contributions: Yuchen Liu contributed substantially to the conceptualisation, methodology, software development, validation, formal analysis, and study investigation. He was also heavily involved in writing the original draft, its subsequent review and editing, and data visualisation. Daniil Mikriukov played a vital role in methodology, providing resources, and partaking in the investigation. Owen Christopher Tjahyadi contributed through resource provision, data curation, and investigation. Gangmin Li, Terry R. Payne, Yong Yue, and Ka Lok Man were responsible for supervision and contributed to the review and editing of the writing. Gangmin Li, Yong Yue, and Ka Lok Man also played a crucial role in acquiring funding for the project. Kamran Siddique assisted in the writing-review editing process. All authors have read and agreed to the published version of the manuscript.

Funding: This research is partially funded by the research funding: XJTLU-REF-21-01-002 and the XJTLU Key Program Special Fund (KSF-A-17)

Data Availability Statement: The dataset and algorithm code is available online at <https://github.com/Larry-Liu02/Non-stationary-Transformer-with-DDPG-for-Portfolio-Management>.

Acknowledgments: This work is partially supported by the XJTLU AI University Research Centre and the Jiangsu Province Engineering Research Centre of Data Science and Cognitive Computation at XJTLU. It is also partially funded by the Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (SZS2022004), as well as by the funding: XJTLU-REF-21-01-002 and the XJTLU Key Program Special Fund (KSF-A-17).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. S. Poon, M. Rockinger, J. Tawn. Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications. *RFS* **2004**, [CrossRef].
2. S. Satchell, A. Scowcroft. A demystification of the Black–Litterman model: Managing quantitative and traditional portfolio construction. *JAM* **2000**, [CrossRef].
3. Ruben Lee. *What is an Exchange? The Automation, Management, and Regulation of Financial Markets*, Publisher: Publisher Location, Country, 1998.
4. F. Fabozzi, P. Drake. *The Basics of Finance: An Introduction to Financial Markets, Business Finance, and Portfolio Management*, Publisher: Publisher Location, Country, 2010.
5. Z. Jiang, J. Liang. Cryptocurrency portfolio management with deep reinforcement learning. In *IntelliSys*, 2017, pp. 905–913.
6. Z. Jiang, D. Xu, J. Liang. A deep reinforcement learning framework for the financial portfolio management problem. In *arXiv preprint arXiv:1706.10059*, 2017.
7. Y. Ye, H. Pei, B. Wang, P. Chen, Y. Zhu, J. Xiao, B. Li. Reinforcement-learning based portfolio management with augmented asset movement prediction states. *Proceedings of AAAI* 2020, pp. 1112–1119.
8. J. Wang, Y. Zhang, K. Tang, J. Wu, Z. Xiong. Alphastock: A buying-winners-and-selling-losers investment strategy using interpretable deep reinforcement attention networks. In *KDD*, ACM, 2019, pp. 1900–1908.
9. S. Shi, J. Li, G. Li, P. Pan, Q. Chen, Q. Sun. GPM: A graph convolutional network based reinforcement learning framework for portfolio management. In *Neurocomputing*, 2022, Vol. 498, pp. 14–27.
10. J. Amon, M. Rammerstorfer, K. Weinmayer. Passive ESG portfolio management—the benchmark strategy for socially responsible investors. In *Sustainability*, 2021, Vol. 13, No. 16, pp. 9388.
11. A. Al-Arabi, S. Jaimungal. Outperformance and tracking: Dynamic asset allocation for active and passive portfolio management. In *Applied Mathematical Finance*, 2018, Vol. 25, No. 3, pp. 268–294.
12. K. KF Law, W. K. Li, P. LH Yu. Evaluation methods for portfolio management. In *Applied Stochastic Models in Business and Industry*, 2020, Vol. 36, No. 5, pp. 857–876.
13. B. Li, S. C. H. Hoi. Online portfolio selection: A survey. In *ACM Computing Surveys (CSUR)*, 2014, Vol. 46, No. 3, pp. 1–36.
14. Y. Liu, K. Man, G. Li, T. R. Payne, Y. Yue. Dynamic Pricing Strategies on the Internet. In *Proceedings of International Conference on Digital Contents: AICO (AI, IoT and Contents) Technology*, 2022.
15. Y. Liu, K. Man, G. Li, T. Payne, Y. Yue. Enhancing Sparse Data Performance in E-Commerce Dynamic Pricing with Reinforcement Learning and Pre-Trained Learning. In *2023 International Conference on Platform Technology and Service (PlatCon)*, IEEE, 2023, pp. 39–42.

16. Y. Liu, K. Man, G. Li, T. Payne, Y. Yue. Evaluating and Selecting Deep Reinforcement Learning Models for Optimal Dynamic Pricing: A Systematic Comparison of PPO, DDPG, and SAC. 2023.
17. J. B. Heaton, N. G. Polson, J. H. Witte. Deep learning for finance: deep portfolios. In *Applied Stochastic Models in Business and Industry*, 2017, Vol. 33, No. 1, pp. 3–12.
18. T. H. Nguyen, K. Shirai, J. Velcin. Sentiment analysis on social media for stock movement prediction. In *Expert Systems with Applications*, 2015, Vol. 42, No. 24, pp. 9603–9611.
19. S. Almahdi, S. Y. Yang. An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. In *Expert Systems with Applications*, 2017, Vol. 87, pp. 267–279.
20. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin. Attention is all you need. *Advances in neural information processing systems* **30**, 2017, [CrossRef].
21. H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI conference on artificial intelligence* **35**(12), 2021, 11106–11115, [CrossRef].
22. N. Kitaev, Ł. Kaiser, A. Levskaya. Reformer: The efficient transformer. *arXiv preprint arXiv:2001.04451* **2020**, [CrossRef].
23. H. Wu, J. Xu, J. Wang, M. Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems* **34**, 2021, 22419–22430, [CrossRef].
24. S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A.X. Liu, S. Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. *International conference on learning representations*, 2021, [CrossRef].
25. Y. Liu, H. Wu, J. Wang, M. Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in Neural Information Processing Systems* **35**, 9881–9893, 2022.
26. R.C. Hockett, S.T. Omarova. Public actors in private markets: Toward a developmental finance state. *Wash. UL Rev.* **93**, 2015, 103.
27. B. Hirchoua, B. Ouhbi, B. Frikh. Deep reinforcement learning based trading agents: Risk curiosity driven learning for financial rules-based policy. *Expert Systems with Applications* **170**, 114553, 2021.
28. J. Devlin, M. W. Chang, K. Lee, K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805* **2019**, [CrossRef].
29. Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, Q. V. Le. XLNet: Generalized Autoregressive Pretraining for Language Understanding. *arXiv preprint arXiv:1906.08237* **2019**, [CrossRef].
30. M. Aenlle. Massive Stock News Analysis DB for NLPBacktests. Available online: <https://www.kaggle.com/datasets/miguelaenlle/massive-stock-news-analysis-db-for-nlpbacktests> (accessed on 20 October 2023).
31. C. Hutto, E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, 2014, Vol. 8, No. 1, pp. 216–225.
32. T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.
33. S. Hochreiter, J. Schmidhuber. Long short-term memory. *Neural computation* **9**(8), 1997, 1735–1780.
34. J. Chung, C. Gulcehre, K. Cho, Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
35. K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
36. S. Shi, J. Li, G. Li, P. Pan. A multi-scale temporal feature aggregation convolutional neural network for portfolio management. In *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019, 1613–1622.
37. B. Li, S. CH Hoi. On-line portfolio selection with moving average reversion. *arXiv preprint arXiv:1206.4626*, 2012.
38. E.F. Fama, M.E. Blume. Filter rules and stock-market trading. *The Journal of Business* **39**(1), 1966, 226–241.
39. X.D. Xu, S.X. Zeng, C.M. Tam. Stock market's reaction to disclosure of environmental violations: Evidence from China. *Journal of Business Ethics* **107**, 2012, 227–237.
40. W.F. Sharpe. The Sharpe Ratio. In *Streetwise—the Best of the Journal of Portfolio Management* **3**, 1998, 169–85.
41. M. Magdon-Ismail, A.F. Atiya. Maximum drawdown. *Risk Magazine* **17**(10), 2004, 99–102.

42. T.N. Rollinger, S.T. Hoffman. Sortino: a 'sharper' ratio. *Chicago, Illinois: Red Rock Capital*, 2013.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.