OPTIVOL
OPTION TRADING VIA PREDICTED VOLATILITY

Haoran Liu, Leo Zeng, Rainie Yan, Yang Liu, Miranda Lu
Duke University

ABSTRACT

This project focuses on implementing volatility trading strategies in the options market by forecasting future volatility. Machine learning and deep learning models, including LSTM, Bi-LSTM, and CNN-Bi-LSTM, are employed for predicting future volatility. The volatility trading strategies encompass three main approaches: short straddle options, VIX strategy, and IV-HV strategy. The selection of underlying assets for these trading strategies involves the technology sector, specifically AMD, MU, and NVDA, chosen based on historical annual trading volumes and volatility.

1 INTRODUCTION

Options are financial instruments that provide investors with the right, but not the obligation, to buy or sell an underlying asset at a predetermined price within a specified period. This flexibility allows traders to hedge risk, speculating on price movements, and creating various sophisticated strategies based on market expectations. Options are categorized into call and put. Call options allow the purchase of the asset at a predetermined price, while put options enable selling at a specified price, providing versatile strategies for risk management and speculation.

Option strategies, which are combinations of various calls and puts, enable traders to construct positions that align with their market predictions and risk appetite. Options strategies usually have a higher probability of winning more in certain scenarios. Derived from a combination of call and put options, the straddle strategy involves simultaneously purchasing both with the same strike price and expiration date. This dynamic approach capitalizes on anticipating substantial price movement, making it ideal when market participants expect heightened volatility but are uncertain about the direction of the underlying asset's price.

The financial markets of the past decade have experienced unprecedented volatility, emphasizing the need to understand the underlying drivers of these movements. Financial markets, characterized by their complexity, non-linearity, and low signal-to-noise ratio, pose significant challenges in predicting short-term asset returns. Studies such as Gu et al. (2020) and Chen et al. (2019) highlight the inherent difficulties in forecasting asset returns. However, certain characteristics, like the slow decay of auto-correlation in absolute returns, suggest that volatility, a critical factor in asset pricing, portfolio allocation, and risk management, can be predicted to some extent.

Historically, the analysis of financial market volatility traces back to foundational works like Engle (1982), Bollerslev (1986), and Taylor (1982). These authors developed discrete-time models like GARCH and stochastic volatility processes to model auto-regressive conditional heteroskedasticity. A comprehensive review by Lunde & Hansen (2005) offers broader insights into these models. Corsi (2009), however, pointed out limitations in standard GARCH and stochastic volatility models, leading to the introduction of the Heterogeneous auto-regressive (HAR) model. This model, combining volatility components at different frequencies, aimed to reflect the heterogeneous nature of market participants.

Despite its parsimony, the HAR model, along with its extensions (e.g., Andersen et al. (2007); Corsi & Renò (2012); Bollerslev et al. (2016)), primarily rely on past return histories, overlooking the significant impact of news announcements on volatility documented in research (e.g., Schwert (1989); Engle et al. (2013); Bollerslev et al. (2018)).

Recent advancements in computational power and machine learning have introduced novel approaches to volatility forecasting. Techniques like artificial neural networks and natural language processing are being increasingly employed to analyze the influence of macroeconomic indicators and news sentiment on market volatility. These methods, evolving beyond traditional models, aim to capture the complex dynamics of financial markets more effectively. However, they also introduce new challenges, such as overfitting and the need for large, high-quality datasets.

Furthermore, the integration of alternative data sources, like social media sentiment and geopolitical events, into volatility models, underscores the shifting paradigm in financial analysis. This evolution reflects a broader recognition of the multifaceted nature of market dynamics, where psychological and social factors play a significant role. As the financial industry continues to grapple with these changes, the importance of interdisciplinary approaches, combining finance, economics, psychology, and data science, becomes increasingly evident. These developments highlight the ongoing need for robust, adaptable models capable of understanding and forecasting market volatility in an ever-changing financial landscape.

In this research, we used GARCH, FCN, LSTM, CNN-Bi-LSTM, and ResNet-LSTM to predict the next day's intra-day volatility for stocks in the technical sector. Then, we used the prediction results to implement straddle-shorting strategies and Volatility Index (VIX) strategies. We first tested the strategies separately and then combined them. Our backtesting results show that by using appropriate models and timing techniques, our strategy reaches a total return of 46.6 percent and a Sharpe ratio of 0.614. Our research makes contributions to the literature on using machine learning models for sequential data, especially for financial asset volatility, as well as options trading strategies.

The remainder of this paper is organized as follows: Section 2 looks back at the relevant literature for machine learning models and options trading; Section 3 describes the data preparation for prediction, including data source, exploratory data analysis (EDA), and feature engineering; Section 4 introduce the methodology; Section 5 presents the empirical results of our models and performance of our strategies; and Section 6 finally concludes the research and suggests limitations and future expansions.

2 Literature Review

2.1 Option and Volatility Trading

The volatility trading on options starts from the theory of Black & Scholes (1973). The Black-Scholes model provides a formula for the price of a European call option. The model assumes that the price of the underlying stock follows a geometric Brownian motion with constant volatility and interest rate. The derivation starts by constructing a risk-free portfolio consisting of the option and the underlying stock.

Let:

- $C(S, t)$ be the price of the call option as a function of the stock price S and time t.
- S be the current stock price.
- r be the risk-free interest rate.
- σ be the volatility of the stock price.
- t be the current time with T being the expiration date of the option.

The Black-Scholes partial differential equation (PDE) is given by:

$$\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} - rC = 0 \tag{1}$$

Derivation:

The price of the option can be modeled by the stochastic differential equation (SDE):

$$dC = \left(\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2}\right) dt + \frac{\partial C}{\partial S} dS \tag{2}$$

By Ito's lemma, the differential of the call option price dC is composed of a deterministic part and a stochastic part.

The risk-free portfolio Π is constructed as follows:

$$\Pi = C - \Delta S \tag{3}$$

where Δ is the number of shares held, which is chosen to make the portfolio risk-free, i.e., $\Delta = \frac{\partial C}{\partial S}$.

The return on this portfolio should be equal to the return on a risk-free asset:

$$d\Pi = r\Pi dt = r(C - \Delta S)dt \tag{4}$$

Substituting the SDE of C and Δ into the equation above and equating coefficients yields the Black-Scholes PDE.

Solution to the Black-Scholes PDE:

The Black-Scholes formula for a European call option, which is the solution to the Black-Scholes PDE, is:

$$C(S,t) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \tag{5}$$

where:

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \tag{6}$$

$$d_2 = d_1 - \sigma\sqrt{T-t} \tag{7}$$

and Φ is the cumulative distribution function of the standard normal distribution, K is the strike price of the option.

Implied Volatility

Historical volatility, realistic volatility, and implied volatility are the three main types. Historical volatility is calculated using historical data, while implied volatility is deduced from existing option quotes through option pricing models, reflecting the market's expectations for actual volatility.

Changes in implied volatility and underlying prices exhibit certain regularities. Statistical analyses reveal a positive correlation, particularly in commodities, between implied volatility and underlying prices, especially in trending markets. However, this correlation is less evident in ETF options. During asset declines, there is an increased market demand for bearish hedging, often accompanied by a rapid rise in volatility.

The relationship between implied volatility and the strike price demonstrates that at-the-money options have relatively low implied volatility. Additionally, as the option's real or intrinsic value deepens, the corresponding implied volatility gradually expands, exhibiting the implied volatility smile property.

The implied volatility surface graphically represents the relationship between option exercise prices and implied volatility. However, the actual market volatility surface is not uniformly distributed and often shows skewness.

Apart from the relationship with the strike price, implied volatility also shows a dependence on time. Short-term volatility being low often leads to the market expecting future volatility to rise, while high short-term volatility tends to result in a decreasing function of time for volatility.

Volatility Trading Strategies

Trading strategies based on volatility characteristics encompass four types:

1. Mean Reversion-Based Strategy: Capitalizing on mean reversion characteristics, investors can go short or long on volatility after it exceeds or falls below the normal range. For instance, shorting volatility by selling straddle options when it's higher than the 90th percentile and has fallen for consecutive days.
2. Volatility and Price Rules Strategy: Leveraging the positive relationship between the price and volatility of certain varieties, investors use the trend of volatility as a reference for underlying asset movements. The concept of IV-HV, the difference between implied and historical volatility, is introduced for this strategy.
3. Volatility Smile-Based Strategy: Exploiting the patterns in the implied volatility smile, investors can identify opportunities to return the pattern to a standard smile. Buying options with a strike price of 5000 and selling options with a strike price of 4950 can be a strategy when the implied volatility of the former is expected to rise relative to the latter.

These trading strategies showcase the diverse approaches investors can take to navigate the dynamic landscape of financial derivatives and capitalize on volatility trends.

2.2 Machine Learning models

Recent years have witnessed a rise in machine learning techniques, especially for high-dimensional problems. Even though compared to other areas such as genetics and image recognition, the dimension of the financial area is not so high, researchers have made significant achievements in this area. Traditional financial and economics research in Econometric usually uses simpler linear models. Because linear models offer excellent interpretability, they are highly effective in describing the impact relationships in past economic contexts. However, those models do not perform well in prediction. More financial institutions such as investment firms care more about predictions of the financial asset price than past economic impact. So academic researchers are also trying to use complex models for prediction in the financial area.

Financial data is one kind of time series data essentially. Machine learning models for time series data usually have sequential inputs and outputs. So models for NLP such as Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gate Recurrent Unit (GRU), and Transformers are also suitable for time series data. LSTM and GRU are two variants of RNNs that incorporate gate mechanisms into the RNN architecture. They are designed to address issues related to long-term memory and vanishing gradient problems. With the rise of LLMs, Transformer has become very popular these days. Transformer (Vaswani et al. (2017)) uses self-attention and positional encoding to detect sequential inputs, so it's also applicable to time series data. Literature on applications of machine learning models on time series data is in multiple fields such as influenza prevalence (Wu et al. (2020)), energy consumption (Weeraddana et al. (2021)), weather prediction (Sønderby et al. (2020)), and financial asset price and volatility prediction.

Volatility prediction is important for both risk management and investment. Literature on financial volatility prediction usually uses GARCH as a benchmark and combines GARCH with RNN, LSTM (Kim & Won

(2018)), Transformers (Ramos-Pérez et al. (2021); Liu et al. (2023); Bilokon & Qiu (2023)), and a mixture of them. Lim et al. (2021) created Temporal Fusion Transformers (TFT) structure designed for volatility prediction that combines LSTM, GRN, and Self-Attention.

3 Data Preparation

3.1 Data Source

The dataset for this study is sourced exclusively from QuantConnect, encompassing both fundamental data of technology companies and stock price bar data.

3.2 Date range and Focus

Date Range: The dataset spans from January 1, 2017, to January 1, 2021. Specific Stocks: The primary focus is on the stocks of AMD, MU, NVDA, and SPY, offering a concentrated view of their market behavior during this period.

3.3 Frequency

The data is recorded at a minute-level frequency, providing a granular view of stock movements.

3.4 Objective

Intraday Volatility Analysis: The core objective is to assess the intraday volatility of each stock. This is achieved by calculating the standard deviation of returns measured at one-minute intervals.

3.5 Features

Total Features: A comprehensive set of 76 features is utilized.

- Price Data: This includes variables such as bid size, ask size, and high/low prices for both bids and asks.
- Fundamental Data: Financial ratios pertinent to the companies, including PE (Price-to-Earnings) ratio, PB (Price-to-Book) ratio, EPS (Earnings Per Share), and ROE (Return on Equity).

3.6 Data Preprocessing and Feature Engineering

- Extensive Feature Engineering: The price bar data undergoes significant transformation to extract meaningful features.
- Global Features: These capture overarching market trends and characteristics.
- Imbalance Features: Metrics focusing on the disparities between bid and ask sizes/prices.
- Distribution Features: These features analyze the distribution characteristics of prices and sizes for each stock.

3.7 Feature Details

For specific feature descriptions, refer to the provided tables.

median size	the sum of the median values of bid size and ask size for each trading day
std size	the sum of the standard deviations of both bid size and ask size for each day
ptp size	the peak-to-peak size for bid size on each day
median price	the sum of the median values of bid price and ask price for each day
std price	the sum of the standard deviations of both bid price and ask price for each day
ptp price	the difference between the maximum bid price and the minimum ask price for each date

Table 1: Global Features

Liquidity Imbalance	the difference between bid size and ask size divided by their sum
Size Imbalance	the ratio of bid size to ask size
Mid Price Movement	the direction of movement in the mid-price
Price Spread	the difference between the ask price and the bid price
Spread Intensity	the price spread from one period to the next
Market Urgency	product of the price spread and the liquidity imbalance
Spread Depth Ratio	the price spread divided by the sum of bid size and ask size
Weighted Average Price (WAP)	the price that reflects both the bid and ask sides of the market
Relative Spread	the price spread divided by the weighted average price

Table 2: imbalanced features

mean	the mean of prices and sizes
std	the standard deviation of prices and sizes
skew	the skew of prices and sizes
kurt	the kurtosis of prices and sizes

Table 3: distribution features

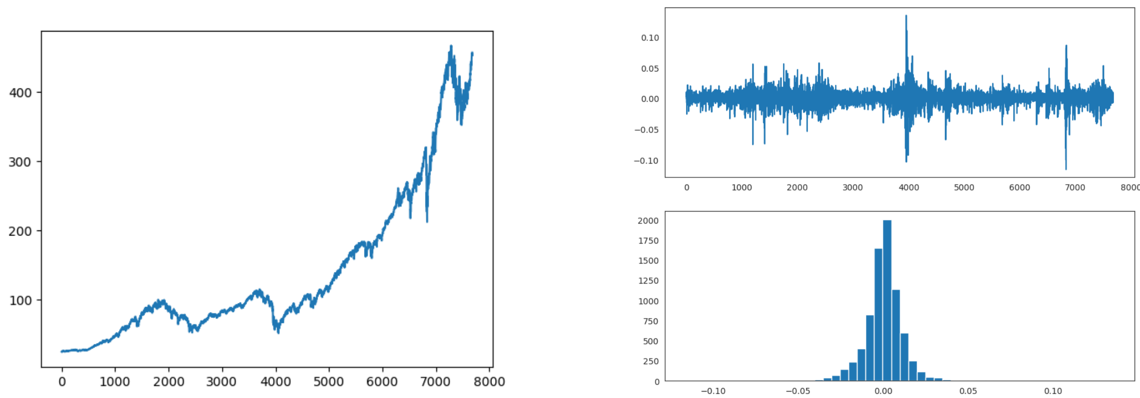
4 Methodology

4.1 Volatility Prediction

In our investigation, we delve into a diverse array of models for time series prediction, as outlined in the introductory section of our study. This exploration is centered around the application of these models to a straightforward SPY dataset, with the primary objective of identifying the most effective model. Our approach involves rigorous testing and evaluation of each model's performance on this dataset. The intent is not only to ascertain the best-performing model but also to extrapolate its structure as a foundational blueprint for future modeling efforts. This methodology ensures that the insights gained from the current analysis can be effectively applied to subsequent predictive models, potentially enhancing their accuracy and efficiency in forecasting time series data.

4.1.1 EDA

To gain an initial understanding of our model selection process, it's essential to examine the test data utilized in our study. We have chosen a comprehensive set of SPY closing price data spanning a lengthy period. This approach is critical in ensuring the robustness and generalizability of our results. The SPY, with its extensive historical record and relatively stable performance, offers an ideal dataset compared to most individual stocks. By using this extensive dataset, we aim to mitigate the volatility and idiosyncrasies often encountered in the stock market, thereby providing a more reliable foundation for our analysis. This choice of data not only enhances the validity of our model evaluation but also ensures that the findings are applicable to a broad spectrum of similar financial time series scenarios.



(a) SPY Close Plot

(b) SPY Log Return Plot

Figure 1: SPY Close and Log Return Plots

Building upon the foundation of SPY close data, our next analytical step involves calculating the logarithmic returns, a crucial measure in financial analysis. The log return, derived from consecutive closing prices, offers a more normalized and continuous metric for assessing financial returns, particularly beneficial for time series analysis.

After calculating the volatility from the log returns, our next step is to examine its distribution. While classic financial theory suggests that returns should follow a normal distribution, the distribution of volatility is less straightforward and more complex. This analysis is vital as it helps us understand the true nature of market volatility, which often deviates from standard distribution models, and has significant implications for risk management and financial modeling.

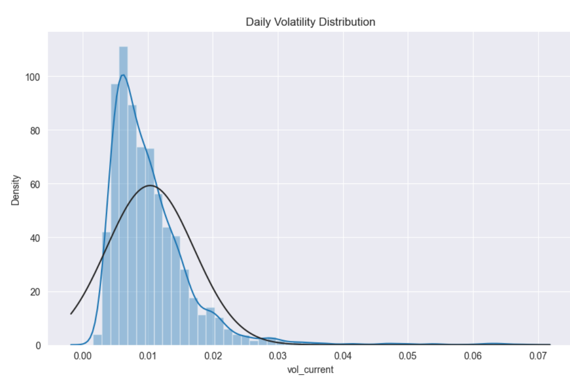


Figure 2: SPY Volatility Distribution

After processing the data, we split the SPY dataset into training and testing sets, applying a scaler trained on the training set to both. This approach ensures model validation and unbiased performance assessment. Here's a brief look at the datasets for training and testing our model.

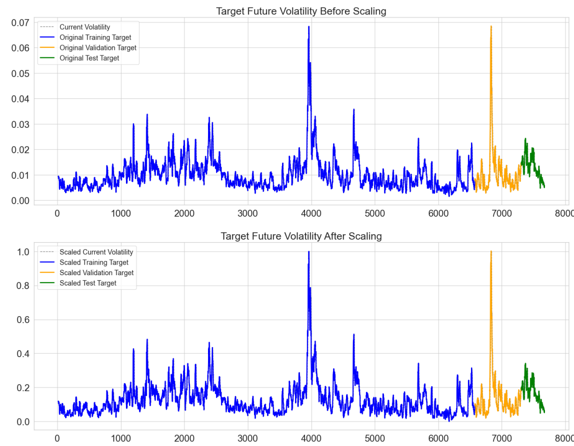


Figure 3: train-test data slot

4.1.2 Previous predict Future

The first approach we employ to predict volatility involves using previous data under the assumption that volatility follows a random-walk pattern. This method is straightforward but operates on the premise that volatility is unpredictable and merely continues its most recent trend. While this assumption is simplistic and may not fully capture the complexities of market behavior, it provides us with a baseline for comparison.

To evaluate the effectiveness of this and other models, we calculate the Root Mean Square Error (RMSE) and the Root Mean Square Percentage Error (RMSPE). These metrics will serve as benchmarks, allowing us to quantitatively assess each model's accuracy in predicting future volatility. By comparing these values, we can determine how well each model performs relative to the simple random-walk hypothesis and identify more sophisticated approaches that offer improved predictive capabilities. The result is shown in Figure 17 in the appendix.

4.1.3 GARCH

Volatility prediction is a critical aspect in financial econometrics, impacting risk management, portfolio optimization, and derivative pricing. Among various models, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model stands out for its robustness and effectiveness.

Introduced by Tim Bollerslev in 1986, the GARCH model extends the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle (1982). The ARCH model was a significant step in modeling time-varying volatility, but faced limitations in handling long data lags. GARCH addresses this by incorporating lagged conditional variances.

The GARCH(p, q) model is mathematically expressed as:

- **Return Equation**:

$$R_t = \mu + \epsilon_t \tag{8}$$

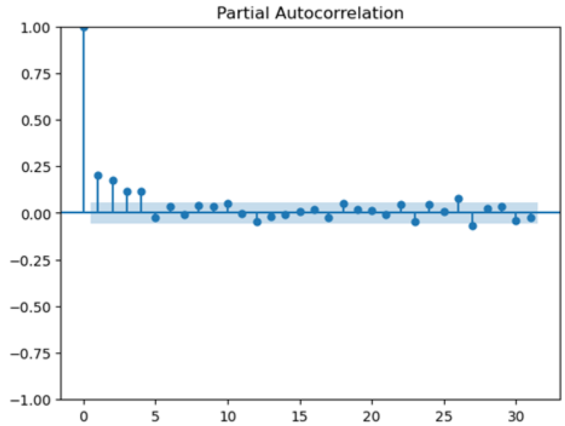


Figure 4: GARCH Partial Autocorrelation

where R_t is the asset return at time t, μ is the mean return, and ϵ_t is the error term.

- **Conditional Variance Equation**:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{9}$$

Here, σ_t^2 is the conditional variance at time t, α_0 is a constant term, α_i are coefficients for the lagged error terms, and β_j are coefficients for the lagged conditional variances. p and q represent the order of the GARCH model.

The GARCH model assumes ϵ_t follows a normal distribution with mean zero and conditional variance σ_t^2, i.e., $\epsilon_t \sim N(0, \sigma_t^2)$. This model captures the "volatility clustering" phenomenon, where high volatility periods tend to be followed by high volatility and vice versa.

Parameters α_i and β_j in the GARCH model are typically estimated using Maximum Likelihood Estimation (MLE), maximizing the likelihood of observing the given return series.

In summary, the GARCH model represents a significant advancement in volatility modeling, extensively used in financial econometrics for applications ranging from risk management to derivative pricing. Our focus is on the GARCH(1,1) model, which, based on our evaluations, demonstrates superior performance compared to other parameter configurations. This preference for GARCH(1,1) is reinforced by our tests on various settings derived from the Partial Autocorrelation Function (PACF) analysis. These tests, aimed at optimizing the model parameters, did not yield significantly better p-values, suggesting that the simpler GARCH(1,1) model is sufficiently robust for our purposes. Figure 5 shows the result of PACF, which indicates that we should choose 1 parameter for GARCH.

Hence, we proceed with the GARCH(1,1) model for our analysis. This choice is guided by its proven efficacy in historical performance and the aim to maintain a balance between model complexity and predictive accuracy. Our analysis with GARCH will provide a deeper understanding of volatility dynamics and serve as a crucial part of our comprehensive evaluation of various volatility prediction models. The result is shown in Figure 16 in the appendix.

4.1.4 FCN

We proceed to explore machine learning methods, starting with the Fully-Connected Network (FCN), which essentially represents a linear regression approach in this context. In the realm of neural networks, Fully Connected Networks (FCNs) are foundational, serving as the bedrock for many complex architectures in deep learning. FCNs, characterized by their structure where each neuron is connected to every neuron in the preceding and subsequent layers, play a vital role in understanding the dynamics of neural networks.

The structure of FCNs can be concisely represented through their layers. Each layer in an FCN is composed of units or neurons, and the output of each neuron in one layer is connected to the input of every neuron in the next layer.

The mathematical formulation of a single layer in an FCN can be expressed as:

$$\mathbf{y} = f(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{10}$$

Here, $\mathbf{x}$ represents the input vector to the layer, $\mathbf{W}$ is the weight matrix associated with the layer, $\mathbf{b}$ is the bias vector, and f is the activation function applied element-wise. The function f introduces non-linear properties to the model, allowing the network to learn complex patterns.

In an FCN, the depth of the network, i.e., the number of hidden layers, and the width, i.e., the number of neurons in each layer, are crucial parameters. These parameters significantly influence the network's learning capacity and its ability to approximate complex functions.

FCNs are widely used in various applications such as regression, classification, and even in more complex structures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The universal approximation theorem states that a sufficiently large FCN can approximate any continuous function, highlighting the significance and versatility of FCNs in the field of neural networks.

Despite their simplicity, FCNs often face challenges such as overfitting, especially in scenarios with limited data or overly complex network structures. Regularization techniques, dropout, and proper selection of network architecture are critical in mitigating these issues.

In conclusion, Fully Connected Networks form an essential component in the landscape of neural network architectures. Their ability to model complex relationships and their foundational role in the development of more specialized network structures underscore their continued relevance in the field of deep learning.

As you can see below, the initial results are unsatisfactory, indicating that a linear relationship may not adequately capture the time-series dynamics of volatility. This outcome is attributed to the limited features available in the test dataset, suggesting that more sophisticated modeling techniques are required to uncover the complex temporal patterns within the volatility data. The result is shown in Figure 18 in the Appendix.

4.1.5 LSTM

To capture non-linear and time-series relationships, we employ the LSTM (Long Short-Term Memory) model. LSTM networks, a type of Recurrent Neural Network (RNN), have revolutionized the field of sequence data analysis. Introduced by Hochreiter & Schmidhuber (1997), LSTMs are specifically designed to overcome the challenges of learning long-term dependencies, a limitation prevalent in traditional RNNs.

The core idea of LSTMs is to maintain a memory cell that can preserve information over extended periods, making them exceptionally suited for tasks involving sequential data such as time series prediction, natural language processing, and speech recognition.

An LSTM unit is composed of four main components: the cell state, the input gate, the output gate, and the forget gate. These components work together to regulate the flow of information into and out of the cell, and to decide what to retain or discard from the cell state.

The mathematical formulation of an LSTM cell can be described as follows:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \tag{11}$$

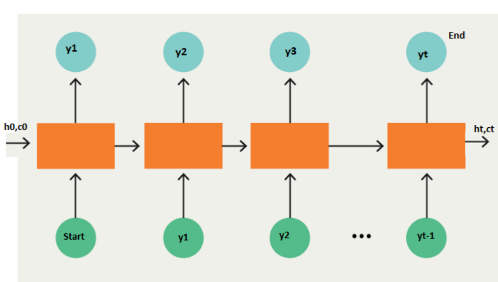
In these equations, x_t is the input at time step t, h_{t-1} is the output of the previous LSTM unit, and C_{t-1} is the cell state from the previous unit. σ represents the sigmoid function, and $\tanh$ is the hyperbolic tangent function. W and b are the weights and biases associated with different gates: the forget gate (f_t), the input gate (i_t), the output gate (o_t), and the cell state updates ($\tilde{C}_t$).

LSTMs have been instrumental in advancing the performance of various complex tasks in sequential data analysis. Their ability to remember information for long durations and to avoid the vanishing gradient problem, common in standard RNNs, makes them a powerful tool in deep learning.

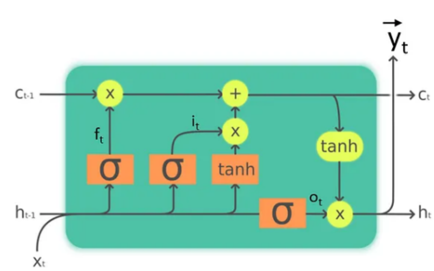
However, LSTMs also come with their set of challenges, such as being computationally intensive and requiring careful tuning of parameters. Despite these challenges, the versatility and effectiveness of LSTMs make them a staple in the toolkit of modern neural network architectures.

In summary, Long Short-Term Memory networks represent a significant advancement in the field of sequential data analysis. Their unique structure and the ability to capture long-term dependencies enable them to tackle complex tasks that were previously challenging for traditional neural network architectures.

The provided graph illustrates a basic LSTM (Long Short-Term Memory) model for time series forecasting. While the internal mechanisms, including the gate formulas, are represented as black boxes, the core concept is clear. Sequential input data, representing previous volatility, is fed into the model, resulting in a sequence of output data, with the final prediction being based on the last element of this sequence. This one-layer LSTM model processes hidden variables (h) in a unidirectional manner, capturing time-series relationships for volatility prediction.



(a) LSTM Architecture



(b) LSTM Unit

Figure 5: Detailed LSTM Architectures

The performance of the trained LSTM model on the validation and test sets is summarized in Figure 20, Figure 21 and Figure 19 in the Appendix.

The loss plot for the LSTM model shows its superior performance compared to all previously tested models, including GARCH(1,1). This indicates that the LSTM model has effectively captured the underlying patterns in the data, resulting in improved predictive accuracy and better loss outcomes.

4.1.6 Bi-LSTM

Expanding upon the foundation of Long Short-Term Memory (LSTM) networks, Bidirectional LSTMs (BiLSTMs) introduce an advanced approach to further enhance the capabilities of sequence data analysis. BiLSTMs, as the name suggests, process data in both forward and backward directions, providing a more comprehensive understanding of the sequence context.

The key innovation of BiLSTMs lies in their dual structure. While a traditional LSTM processes data in a forward sequence, capturing dependencies from past to present, a BiLSTM adds a parallel layer that processes data in reverse, from future to past. This dual mechanism allows the network to have information from both past and future states simultaneously, leading to a richer and more nuanced understanding of the sequence.

The structure of a BiLSTM can be represented as follows:

$$
\begin{aligned}
\overrightarrow{h_t} &= \text{LSTM}(\overrightarrow{h_{t-1}}, x_t) \\
\overleftarrow{h_t} &= \text{LSTM}(\overleftarrow{h_{t+1}}, x_t) \\
H_t &= [\overrightarrow{h_t}; \overleftarrow{h_t}]
\end{aligned}
\tag{12}
$$

Here, $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ represent the hidden states of the forward and backward LSTMs at time step t, respectively. x_t is the input at time step t, and H_t is the concatenated hidden state which combines the information from both directions.

In conclusion, Bidirectional Long Short-Term Memory networks represent a significant leap in the field of sequence analysis. By synthesizing information from both past and future states, BiLSTMs offer a more comprehensive framework for understanding and predicting sequence data, solidifying their position as a powerful tool in the arsenal of deep learning architectures. Bi-LSTM also shows a better understanding of sequence in natural language understanding tasks, which might a better understanding of financial data sequence. This Bi-LSTM model is poised to become the foundational network for future applications due to its superior results. The loss plot for Bi-LSTM demonstrates lower loss values(see Figure 22, Figure 23 and Figure 24 in the Appendix), reaffirming its effectiveness in capturing complex temporal dependencies in the data.

4.1.7 CNN-Bi-LSTM

In the realm of sequence data analysis, particularly in financial markets, One-Dimensional Convolutional Neural Networks (1D CNNs) have emerged as a powerful tool. Unlike their more commonly known two-dimensional counterparts used in image processing, 1D CNNs are specifically designed for univariate or multivariate time series data, like financial sequences, where the temporal dimension is key.

1D CNNs function by applying convolutional operations along the time axis of the data. This approach allows them to effectively identify patterns and trends in time series data, making them highly suitable for tasks such as trend analysis, anomaly detection, and forecasting in financial markets.

The architecture of a 1D CNN involves:

- **Convolutional Layer**: Applies one-dimensional filters across the temporal sequence.

$$F_i = \text{ReLU}(W * X_i + b) \tag{13}$$

  Here, F_i represents the feature at the i-th position in the sequence, W is the filter weight, X_i is the input segment at the i-th position, b is the bias, and $*$ denotes the convolution operation.

- **Pooling Layer** (optional): Similar to 2D CNNs, reduces the length of the output, thus decreasing the number of parameters and computation in the model.

- **Fully Connected Layer**: Processes the output of the convolutional and pooling layers to make final predictions or classifications.

1D CNNs are particularly advantageous in financial data analysis due to their ability to efficiently process long sequences of data and capture temporal dependencies. They are less computationally intensive than RNNs, including LSTMs, and often provide faster training and inference times, which is crucial in financial applications where speed and efficiency are essential.

Incorporating 1D CNNs into a CNN-BiLSTM architecture offers a compelling approach for handling financial time series data. The 1D CNN layers first extract temporal features from the sequence data, capturing trends and patterns over time. These extracted features are then fed into BiLSTM layers, which process the information in both forward and backward directions, capturing complex long-term dependencies in the data.

This integrated CNN-BiLSTM model, with 1D CNNs for temporal feature extraction and BiLSTMs for sequential dependency modeling, is particularly adept at analyzing financial sequences. It can discern subtle patterns and dependencies in market data, making it highly effective for forecasting, risk management, and algorithmic trading strategies.

The combination of 1D CNN and BiLSTM layers not only leverages the strengths of both models in spatial feature extraction and temporal sequence processing but also addresses the challenges of analyzing high-dimensional financial data. This makes the CNN-BiLSTM framework a state-of-the-art approach for sophisticated financial data analysis and prediction.

In the proposed model, CNN layers precede Bi-LSTM, with the convolutional results being input to the Bi-LSTM layer, ultimately leading to predictions. The results for the CNN-Bi-LSTM model(see Figure 25 and Figure 26 in the Appendix) indicate improved performance, underscoring the utility of this hybrid approach in capturing intricate temporal dependencies.

No.	Model	Validation RMSPE	Validation RMSE
0	Mean Baseline	1.010889	0.144979
1	Random Walk Naive Forecasting	0.450441	0.076689
2	GARCH(1,1), Constant Mean, Normal Dist	0.530965	0.185607
3	Simple LR Fully Connected NN, n_past=14	0.638177	0.234520
4	LSTM 1 layer 20 units, n_past=14	0.643199	0.052535
5	2 layers Bidirect LSTM (32/16 units), n_past=14	0.512388	0.034322
6	1 Conv1D 2 Bi-LSTM layers (32/16), n_past=14	0.480372	0.037696

Table 4: Comparison of Models Based on Validation RMPSE and RMSE

Final Model Comparison We also summarize and compare the results of the models here. In a nutshell, the CNN-Bi-LSTM model outperforms all others and will be applied in QuantConnect for volatility prediction.

4.1.8 RESNET-LSTM

Residual Networks (ResNets), introduced by He et al. (2016), mark a significant advancement in deep neural networks. ResNets address the vanishing gradient problem in deep networks through skip connections, allowing the flow of gradients directly across layers. These connections enable the network to learn an identity function, preventing degradation in deeper layers.

A typical ResNet block is mathematically expressed as:

$$\mathbf{y} = F(\mathbf{x}, \{W_i\}) + \mathbf{x} \tag{14}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the input and output of the layers within the block, F represents the residual mapping, and $\{W_i\}$ are the block's weights. The element-wise operation $F + \mathbf{x}$ forms the shortcut connection.

Future Direction and Preliminary Exploration In our preliminary exploration, we experimented with integrating ResNet with LSTM, forming a ResNet-LSTM model. This integration aimed to leverage ResNet's robust feature extraction capabilities in combination with LSTM's proficiency in sequential data processing. However, this approach did not yield superior results and occasionally led to unintended outcomes. The ResNet in this context acts as a more complex CNN layer placed before the LSTM.

$$\text{Output} = \text{LSTM}(\text{ResNet}(\text{Input})) \tag{15}$$

Despite the challenges encountered, this exploration underlines the potential for future enhancements to the CNN-Bi-LSTM model. By refining the integration of ResNet's deep feature extraction with the BiLSTM's temporal analysis capabilities, there's a possibility to create a more powerful hybrid model. This direction holds promise for complex applications, including advanced time series analysis and sophisticated natural language processing tasks, and warrants further investigation and optimization.

4.2 STRATEGIES

4.2.1 UNIVERSE SELECTION

In the pursuit of crafting a strategy for volatility option trading, the selection of the Technology sector as our focal point is grounded in a nuanced understanding of market dynamics and the distinctive characteristics inherent to this domain. The technology sector, marked by its inherent dynamism and rapid technological advancements, engenders a heightened level of volatility, thereby presenting a fertile ground for exploiting volatility-driven trading strategies. The sector's susceptibility to macroeconomic trends, coupled with the intricate interplay of innovation cycles, renders it uniquely responsive to market sentiment and external influences, amplifying potential trading opportunities.

The universe selection of underlyings within the technology sector for volatility trading involves a multifaceted approach grounded in both quantitative and qualitative analyses. Initially, our screening process focused on the top ten stocks with the highest trading volumes, including 'IBM,' 'AMD,' 'STX,' 'MU,' 'MSFT,' 'NVDA,' 'AAPL,' 'CSCO,' 'QCOM,' and 'INTC.' This selection criteria was pivotal in ensuring sufficient liquidity and market participation, essential for effective implementation of volatility-based strategies.

Subsequently, a quantitative assessment spanning the period from 2017 to 2021 was conducted to ascertain the historical annualized volatility of each underlying asset. It resulted in a list of 8 underlying that exhibited notable historical volatility characteristics. The 8 chosen stocks, 'IBM,' 'AMD,' 'STX,' 'MU,' 'MSFT,' 'NVDA,' 'AAPL,' and 'CSCO,' emerged as particularly compelling candidates for inclusion in our volatility trading portfolio. Their historical annualized volatility not only showcased responsiveness to market fluctuations but also demonstrated a resilience that aligns with our risk-adjusted return objectives.

Upon careful verification of the compatibility of each stock with our predefined volatility trading strategy, 'AMD,' 'MU,' and 'NVDA' emerged as the most strategically aligned candidates. This selection was informed by a nuanced evaluation of factors such as historical price movements, responsiveness to market dynamics, and the inherent characteristics that make these stocks well-suited for effective implementation within the defined volatility-driven framework.

In summary, the comprehensive selection process involving both quantitative metrics and sector-specific considerations has resulted in a tailored portfolio of technology stocks. This approach positions us strategically to navigate the complexities of the market, leveraging historical volatility patterns to optimize risk-adjusted returns in the context of our volatility trading objectives.

4.2.2 Straddle Strategy

Our primary strategy is articulated to employ a short straddle position, which is contingent on the probability of profit extrapolated from a normal distribution framework. We are using 0DTE options so the PnL here is linear. As delineated in Figure 6, the strategy is predicated on the premise that a state of profitability is achieved if the price of the underlying stock at maturity nestles within the bracket demarcated by [k-p, k+p], where k symbolizes the strike price of both the call and put options, and p epitomizes the sum of two options' price.

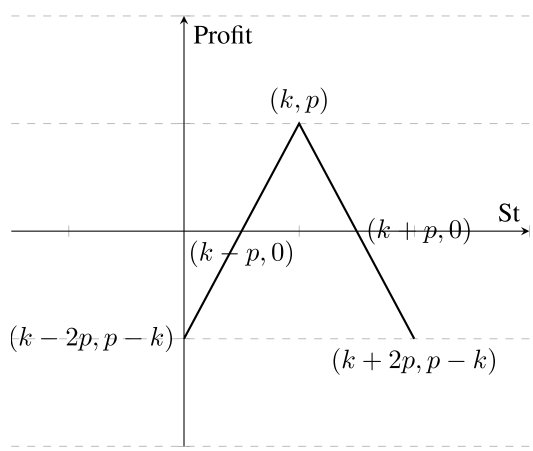


Figure 6: Short Straddle

To model stock price $S(t)$, we assume its log returns $\log(S(t))$ follows a Brownian motion, that is, setting $X(t) = \log(S(t))$ in equation (16) below, hence the term Geometric Brownian Motion, or Lognormal Process. A salient characteristic of this paradigm is the absence of 'memory' in the stock price movements, which implies that alterations from one day to the next are entirely discrete and exhibit no serial dependence.

$$X(T) - X(t) = \xi(t,T) \sim \mathcal{N}(0, \sigma\sqrt{T-t}) \tag{16}$$

The strategic inclination to establish a short straddle is predicated on a probability exceeding a meticulously defined threshold, suggesting that the price of the stock will reside within the interval [k-p, k+p]. That probability is our trading signal, which is calculated as follows:

$$r_t = \frac{S_t - 1}{S_{t-1}} \sim N(0, \sigma^2)$$

$$S_{t-1} \approx k$$

$$\sigma^2 = \text{pred(vol)}$$

$$P(S_t \in [k-p, k+p]) = P\left(\left|\frac{r_t}{k}\right| \leq \frac{p}{k}\right) = \text{cdf}\left(\frac{p}{k * \sigma^2}\right) - \text{cdf}\left(-\frac{p}{k * \sigma^2}\right)$$

In which k is the strike price for both call and put options, and p is the summation of the premium for call and put.

For the quantification of this probability, we have employed the predicted return volatility of the underlying stocks, which serves as the standard deviation in the normal distribution of our returns. This volatility is extrapolated from historical pricing data, thereby providing an empirical basis for the probabilistic assessment of our strategic positions.

The strategy harnesses sophisticated financial models to predict the probability of market movements and to place trades accordingly, illustrating a confluence of statistical theory and market acumen. This meticulous approach to option trading seeks to capitalize on the inherent volatility of the market by implementing a probabilistic framework that is grounded in the principles of stochastic processes and the empirical behavior of asset prices.

4.2.3 VIX Strategy

Upon the implementation and subsequent backtesting of the primary volatility strategy, it was observed that the algorithm did not furnish trading signals with daily regularity. This irregularity in signal generation necessitates the adoption of an ancillary strategic approach that ensures the effective deployment of capital over temporal intervals. The secondary strategy serves to mitigate the temporal inefficiency of capital deployment inherent to the primary strategy's intermittent signal output. It is predicated on the understanding that in the absence of actionable signals from the primary strategy, the temporal value of capital should not remain idle. To this end, the supportive framework is structured to actively engage in market positions, thereby leveraging the latent potential of the invested capital.

This supplementary strategy is structured to capitalize on discrepancies within the volatility pricing market. At the heart of this strategy lies the trading of Volatility Index (VIX) derivatives alongside VIX-correlated Exchange Traded Funds (ETFs), notably VIXY and SVXY. VIXY represents the Long VIX Short-Term Futures ETF, while SVXY denotes the Short VIX Short-Term Futures ETF.

The core computational element of this strategy is the 'basis,' which is quantitatively expressed as the proportional difference between the price of VIX futures and the spot price of VIX, as shown in equation (17).

$$basis = \frac{\text{VIX future price}}{\text{VIX spot price}} - 1 \tag{17}$$

The directional impetus of trading decisions is predicated upon the sign of the VIX basis. In instances where the VIX basis is affirmative, it is indicative of a contango market condition, prompting the strategy to divest any holdings in VIXY in favor of establishing a long position in SVXY. Conversely, a negative VIX basis is indicative of a backwardation in the market, catalyzing the liquidation of SVXY holdings, if present, and the subsequent procurement of a long position in VIXY.

This dual-faceted approach to strategy formulation allows for implicit risk mitigation, achieved through a diversified portfolio of volatility-centric assets. This is complemented by a dynamic adjustment mechanism responsive to the ongoing fluctuations within market conditions. By continuously recalibrating the asset mix and positioning, the strategy aims to maintain a balanced exposure, thereby harnessing volatility for potential financial gain while striving to minimize undue risk.

4.2.4 IV-PV Strategy

The tertiary strategy developed in this project exploits the mean-reverting nature of the divergence between realized volatility and implied volatility. It postulates that a sustained elevation of implied volatility above historical volatility, particularly when it persists at heightened levels, may be indicative of a systematic overvaluation of prospective asset volatility within the options market. Under such circumstances, there is an inclination to posit that the market, in its collective wisdom, is erring on the side of caution, potentially due to risk aversion or anticipated future events that are presumed to engender increased volatility.

To operationalize this insight, the strategy posits a threshold-based mechanism where options are sold when the discrepancy between the market's implied volatility and the volatility forecasted by our model surpasses a predetermined upper limit. This approach is premised on the assumption that the market's heightened volatility expectations will, over time, revert to more historically normative levels, thereby yielding profits from the initial option sale.

In contrast, a persistent minimal gap between the forecasted and market-implied volatilities may signal an underestimation of future volatility by the market. The strategy capitalizes on such scenarios by purchasing options, predicated on the expectation that an eventual correction or convergence to our model's predicted volatility will materialize, resulting in a profitable outcome.

4.3 PART 3: Risk Management

4.3.1 Stop-loss

In implementing the short volatility strategy for NVDA based on historical volatility, we have incorporated a robust risk management mechanism using a stop-loss condition. Specifically, if the strategy incurs losses for three consecutive days, we initiate a full liquidation of the portfolio. Subsequently, a mandatory 14-hour freeze period is enforced during which no trading activities are permitted. This proactive approach serves to mitigate potential further losses and allows for a reassessment of market conditions before resuming trading operations.

Expanding our strategy to encompass NVDA, AMD, and MU within the short volatility framework, we maintain a similar stop-loss condition as part of our risk management protocol. In this case, if the cumulative losses on any given day surpass 8%, an immediate liquidation of the entire portfolio is triggered. This precautionary measure is designed to contain losses and protect capital, reflecting a commitment to prudent risk management practices.

These stop-loss conditions not only act as safeguards against adverse market movements but also embody our commitment to disciplined risk management. By incorporating historical volatility analysis and setting predefined stop-loss thresholds, we aim to strike a balance between capital preservation and the pursuit of returns, fostering a resilient and adaptive approach to navigating the complexities of the financial markets.

4.3.2 VALUE AT RISK(VAR)

In the initial strategy of VIX ETFs, the drawdown is higher than 70%, so we tried several risk management tools, including the value at risk. The methodological approach to implementing VaR commenced with the compilation of a comprehensive list of hourly portfolio valuations derived from the VIX ETFs trading strategy. These valuations were systematically documented in a sequential list for subsequent analysis. Utilizing this dataset, the subsequent phase involved the calculation of hourly returns, which were then methodically arranged in ascending order to facilitate the identification of potential loss thresholds. Armed with the ordered return data, the VaR was computed by selecting a predetermined confidence level, reflective of the risk tolerance of the strategy. This confidence level served as the basis for identifying the precise index within the ordered return list, corresponding to the maximum acceptable loss. The return value at this index, invariably negative due to its position in the lower tail of the distribution, represented the VaR metric for the strategy. Upon determination, the VaR served as a critical threshold, signaling the necessity for strategic adjustments to the portfolio's positions.

The utilization of VaR in this context provided a quantifiable benchmark for initiating preemptive risk mitigation actions. By anchoring position adjustments to the VaR, the strategy aimed to proactively manage the exposure to extreme market movements, thus preserving the integrity of the investment portfolio and aligning with the overarching objective of risk-adjusted return optimization.

5 RESULTS & PERFORMANCE

5.1 PART 1: PREDICTION (CNN-BI-LSTM)

In the QuantConnect environment, we leverage a multi-input approach with the CNN-Bi-LSTM model. The data, as previously described, is processed with a window size of 10, generating sequences of features. These feature sequences are then fed into the CNN-Bi-LSTM network for training. While specific training results are not provided here, this methodology enhances our model's ability to capture complex temporal relationships and is instrumental in achieving improved predictive performance for volatility in financial markets.

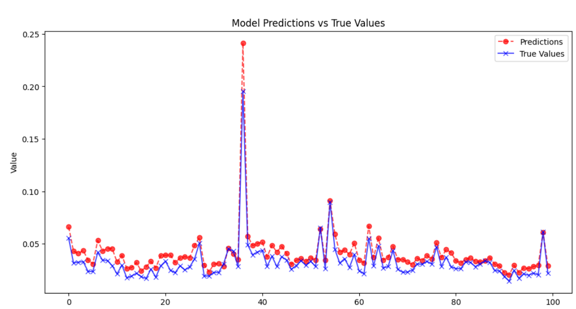
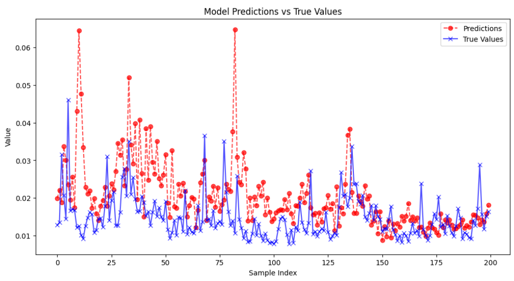
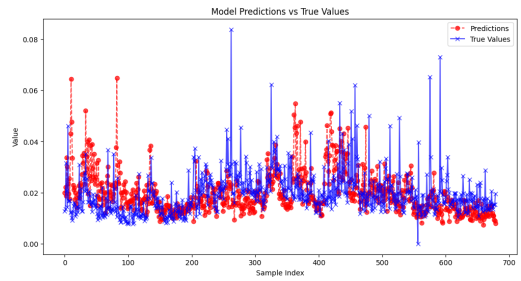


Figure 7: CNN-Bi-LSTM result on Train Set in QC

The training results indicate that the model has learned the training data effectively, nearly predicting all the ground truth. In the test set, while the predictions are not perfect, they show promise as valuable indicators for future volatility trading strategies. The first figure provides a snapshot of the test set results, while the second figure illustrates the overall performance of the test set, demonstrating the model's potential utility in volatility forecasting.



(a) CNN-Bi-LSTM result on part of Test Set in QC



(b) CNN-Bi-LSTM whole result of Test Set in QC

Figure 8: CNN-Bi-LSTM results on QC Test Set

5.2 PART 2: STRATEGY (WITH HISTORICAL VOL)

5.2.1 VOLATILITY-STRATEGY

a.AMD

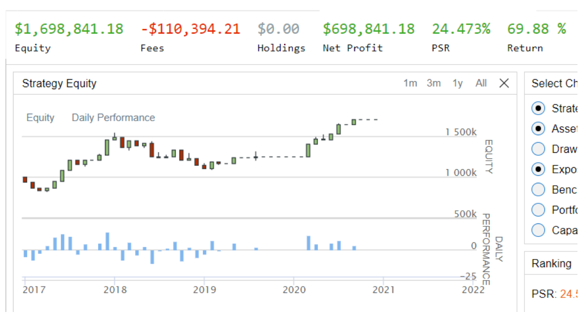


Figure 9: Return of volatility strategy with Historical Volatility - AMD 2017/01/01 - 2021/01/01

Period	Sharpe Ratio	Total Return	Max Drawdown
2017/01/01-2021/01/01	0.617	69.88%	28.7%
2022/01/01-2022/11/01	0	0%	0%
2016/01/01-2017/01/01	-1.293	-35.91%	41.4%
2023/03/10-2023/10/10	-0.732	-3.06%	15.5%

Table 5: Results of volatility strategy with Historical Volatility - AMD

b.MU

Figure 10: Return of volatility strategy with Historical Volatility - MU 2017/01/01 - 2021/01/01

Period	Sharpe Ratio	Total Return	Max Drawdown
2017/01/01-2021/01/01	0.493	33.4%	11.1%
2022/01/01-2022/11/01	0.618	10.51%	13.1%
2016/01/01-2017/01/01	-0.094	-2.29%	17.5%
2023/03/10-2023/10/10	0.575	10.13%	5.9%

Table 6: Results of volatility strategy with Historical Volatility - MU

c.NVDA

Figure 11: Return of volatility strategy with Historical Volatility - NVDA 2017/01/01 - 2021/01/01

Period	Sharpe Ratio	Total Return	Max Drawdown
2017/01/01-2021/01/01	0.242	27.01%	22.4%
2022/01/01-2022/11/01	0	0%	0%
2016/01/01-2017/01/01	0.825	14.95%	13.2%
2023/03/10-2023/10/10	0	0%	0%

Table 7: Results of volatility strategy with Historical Volatility - NVDA

d.AMD + MU + NVDA

Figure 12: Return of volatility strategy with Historical Volatility - AMD + MU + NVDA 2017/01/01 - 2021/01/01

Figure 13: Return of VIX strategy with Historical Volatility 2017/01/01 - 2021/01/01

Period	Sharpe Ratio	Total Return	Max Drawdown
2017/01/01-2021/06/30	0.53	61.07%	21.7%
2022/01/01-2022/11/01	0	0%	0%
2016/01/01-2017/01/01	-1.913	-51.8%	53.7%
2023/03/10-2023/10/10	0.888	14.77%	9%

Table 8: Results of volatility strategy with Historical Volatility - AMD + MU + NVDA

5.2.2 VIX-Strategy

Period	Sharpe Ratio	Total Return	Max Drawdown
2017/01/01-2021/01/01	0.568	130.99%	68.4%
2022/01/01-2022/11/01	-1.203	-67.3%	68.2%
2016/01/01-2017/01/01	3.129	240.97%	41.2%
‘2023/03/10-2023/10/10	-0.248	1.87%	31%

Table 9: Results of VIX strategy with Historical Volatility

5.3 Part 3: Backtesting Performance

5.3.1 Performance Assessment

This is our project link which contains our research and strategy code: QC Project Link

5.3.2 In-Sample Period: 2017/01/01-2021/01/01

The date range of In-Sample Period is

Period	Sharpe Ratio	Total Return	Max Drawdown
2017/01/01-2021/01/01	0.787	60.691%	15.200%

Table 10: In-Sample Results

2017-2021 IS QC Backtest Link

5.3.3 Out-of-Sample Period A: 2022/01/01-2022/11/01

Period	Sharpe Ratio	Total Return	Max Drawdown
2022/01/01-2022/11/01	-1.356	-11.46%	16.400%

Table 11: Out-of-Sample A Results

2022 OOS QC Backtest Link

5.3.4 Out-of-Sample Period B: 2016/01/01-2017/01/01

Period	Sharpe Ratio	Total Return	Max Drawdown
2016/01/01-2017/01/01	0.147	1.82%	25.3%

Table 12: Out-of-Sample B Results

2016 OOS QC Backtest Link

5.3.5 Out-of-Sample Period C: 2023/03/10-2023/10/10

Period	Sharpe Ratio	Total Return	Max Drawdown
2023/03/10-2023/10/10	1.325	23.170%	6.91%

Table 13: Out-of-Sample C Results

2023 OOS QC Backtest Link

5.4 PART 4: Live Trading Attempt

We've also tried to deploy our strategy on QuantConnect's and interactive broker's live trading platform. Due to the lack of data for options and futures, live trading does not have any results.

Figure 14: Live Trading Result

```
: Algorithm starting warm up...
: Warning: cfe Future QuoteBar data not supported. Please consider reviewing the data providers selection.
: Warning: cfe Future TradeBar data not supported. Please consider reviewing the data providers selection.
: Warning: cfe Future OpenInterest data not supported. Please consider reviewing the data providers selection.
: Warning: usa IndexOption QuoteBar data not supported. Please consider reviewing the data providers selection.
: Warning: usa IndexOption TradeBar data not supported. Please consider reviewing the data providers selection.
: Warning: usa IndexOption OpenInterest data not supported. Please consider reviewing the data providers selection.
: Processing algorithm warm-up request 2%...
```

Figure 15: Live Trading Error Message

6 Conclusion

6.1 Analysis of Limitations and Future Improvements

6.1.1 Limitations in Data Sources and Model Input

1. Restricted Data Access: Currently, our model inputs are confined to data available from QuantConnect (QC), as access to external APIs is limited. This restriction impacts the diversity and comprehensiveness of our data.
   - Potential Enhancement: Incorporating additional data sources, such as Tiingo for news, could significantly enrich our dataset and improve pattern recognition accuracy.
2. Substitute for 0DTE Options: QC lacks comprehensive data on options with zero days to expiry (0DTE). As a workaround, we utilize weekly options data, focusing on options with 0 to 1 day until expiry. This limitation confines our trading frequency to once a week, potentially underutilizing our prediction capabilities.

6.1.2 Strategy Diversification and Computational Constraints

- Limited Strategy Diversification: Due to constrained computational resources and node capacity, our strategy lacks diversity. This limitation not only slows down the model and strategy computations but also restricts our trading scope.
- Current Scope: We are currently trading options for only three technology sector stocks.
- Future Expansion: Expanding to options of underlying assets from varied sectors can diversify our strategy and potentially enhance performance.

6.1.3 Model Tuning and Strategy Deployment

- Over-Tuning Concerns: Our model and strategy are heavily tailored for the period from 2016 to 2023. While they fit this specific timeframe well, there's a risk that they may not adapt effectively to different market conditions.
- Testing on Unseen Data: To validate our strategy further, deploying it on a live trading platform with an unseen dataset is crucial. This step is fundamental in assessing the real-world applicability of our strategy.
- Deployment Challenges: Currently, both QC and Interactive Brokerage do not support the data necessary for options and futures trading in our context.
- Future Opportunities: Exploring other platforms or data sources that offer the required data for live trading is a potential avenue to overcome this hurdle.

6.1.4 Competition with Institutional Investors

- Challenges in Competing with Institutions: The financial market is highly competitive, with institutional investors often having the upper hand due to their vast resources and access to extensive data.
- Data vs. Models: In such a market, having unique data or innovative models is crucial, as common patterns are quickly discovered and exploited by others.
- Frequent Updates Required: To remain competitive, it's essential to continuously update our data and models to capture new market patterns.
- Exploring Unconventional Approaches: Utilizing less popular models or signals to uncover unrecognized patterns could offer an edge in this dynamic environment.

6.2 Conclusion

Assessing predictive models, specifically CNN-Bi-LSTM and LSTM, reveals that each exhibits distinct strengths when considering both RMSE and RMSPE as comprehensive prediction measurements. Notably, both models outperform other counterparts significantly, underscoring their efficacy in forecasting future volatility.

From the perspective of volatility trading strategies, the short straddle options strategy stands out for its relatively minimal drawdown and stability. The qualified orders are limited, contributing to a more stable performance. Conversely, the VIX strategy presents a trade-off between risk and return, showcasing substantial gains alongside significant drawdowns, rendering it less stable. The IV-HV (PV) strategy mirrors the characteristics of the short straddle options, emphasizing stability and demonstrating similarities in performance.

In summary, the volatility trading strategies vary in their risk-return profiles and stability. In the mean time, predicted volality is not always similar to the historical volatility, which also contribute to the difference. However, we still capture some insights about volatility trading.

Firstly, further research and development efforts can be directed towards refining and fine-tuning predictive models, such as CNN-Bi-LSTM and LSTM, aiming to harness the strengths of each model for specific market conditions. Continuous advancements in machine learning and deep learning techniques offer the potential to enhance predictive accuracy, providing traders and investors with more reliable forecasts for future volatility.

In the realm of volatility trading strategies, the identified strengths and weaknesses of each approach present opportunities for the development of hybrid strategies that combine the stability of the short straddle options strategy with the potential returns offered by the VIX strategy. This synthesis could lead to the creation of more robust and adaptive trading strategies that capitalize on market opportunities while mitigating risks.

Additionally, we figure out the importance of incorporating additional trading signals to complement predictive models and volatility trading strategies. Future research could explore the integration of macroeconomic indicators, sentiment analysis, or other relevant factors to further enhance decision-making processes and improve overall strategy performance.

Moreover, as financial markets continue to evolve, the integration of real-time data and adaptive learning mechanisms into predictive models and trading strategies will be essential. This dynamic approach ensures that algorithms can adapt swiftly to changing market conditions, enhancing their responsiveness and effectiveness.

In conclusion, the future holds promising prospects for the intersection of predictive modeling and volatility trading strategies. Continued advancements in technology, combined with a comprehensive understanding of the strengths and weaknesses of existing approaches, will contribute to the development of more sophisticated, adaptive, and successful strategies in the realm of algorithmic trading.

REFERENCES

Torben G. Andersen, Tim Bollerslev, and Francis X. Diebold. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720, 2007.

Paul Bilokon and Yitao Qiu. Transformers versus lstms for electronic trading. *arXiv preprint arXiv:2309.11400*, 2023.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3): 307–327, 1986. doi: 10.1016/0304-4076(86)90063-1.

Tim Bollerslev, Andrew Patton, and Rogier Quaedvlieg. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18, 2016. URL `https://EconPapers.repec.org/RePEc:eee:econom:v:192:y:2016:i:1:p:1-18`.

Tim Bollerslev, Jia Li, and Yuan Xue. Volume, volatility, and public news announcements. *Review of Economic Studies*, 85(4):2005–2041, 2018. doi: 10.1093/restud/rdy003.

Luyang Chen, Markus Pelger, and Jason Zhu. Deep learning in asset pricing. *Available at SSRN*, 2019. URL `https://ssrn.com/abstract=3350138`.

Fulvio Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009. doi: NotIdentified.

Fulvio Corsi and Roberto Renò. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30(3):368–380, 2012. doi: 10.1080/07350015.2012.663261.

Robert Engle, Eric Ghysels, and Bumjean Sohn. Stock market volatility and macroeconomic fundamentals. *The Review of Economics and Statistics*, 95(3):776–797, 2013.

Robert F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. doi: 10.2307/1912773.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020. doi: 10.1093/rfs/hhaa009.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778. IEEE, 2016.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Ha Young Kim and Chang Hyun Won. Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models. *Expert Systems with Applications*, 103:25–37, 2018.

Bryan Lim, Sercan Ark, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

Wenting Liu, Zhaozhong Gui, Guilin Jiang, Lihua Tang, Lichun Zhou, Wan Leng, Xulong Zhang, and Yujiang Liu. Stock volatility prediction based on transformer model using mixed-frequency data. *arXiv preprint arXiv:2309.16196*, 2023.

Asger Lunde and Peter Reinhard Hansen. A forecast comparison of volatility models: does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20(7):873–889, 2005. doi: NotIdentified.

Eduardo Ramos-Pérez, Pablo J Alonso-González, and José Javier Núñez-Velázquez. Multi-transformer: A new neural network-based architecture for forecasting s&p volatility. *Mathematics*, 9(15):1794, 2021.

G. William Schwert. Why does stock market volatility change over time? *The Journal of Finance*, 44(5): 1115–1153, 1989.

Casper Kaae Sønderby, Lasse Espeholt, Jonathan Heek, Mostafa Dehghani, Avital Oliver, Tim Salimans, Shreya Agrawal, Jason Hickey, and Nal Kalchbrenner. Metnet: A neural weather model for precipitation forecasting. *arXiv preprint arXiv:2003.12140*, 2020.

S.J. Taylor. Financial returns modelled by the product of two stochastic processes—a study of daily sugar prices 1961-79. In O.D. Anderson (ed.), *Time Series Analysis: Theory and Practice 1*, pp. 203–226, Amsterdam, 1982. North-Holland.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Dilusha Weeraddana, Nguyen Lu Dang Khoa, Lachlan O'Neil, Weihong Wang, and Chen Cai. Energy consumption forecasting using a stacked nonparametric bayesian approach. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part V*, pp. 19–35. Springer, 2021.

Neo Wu, Bradley Green, Xue Ben, and Shawn O'Banion. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317*, 2020.

A Appendix

Figure 16: GARCH(1,1) result on train set

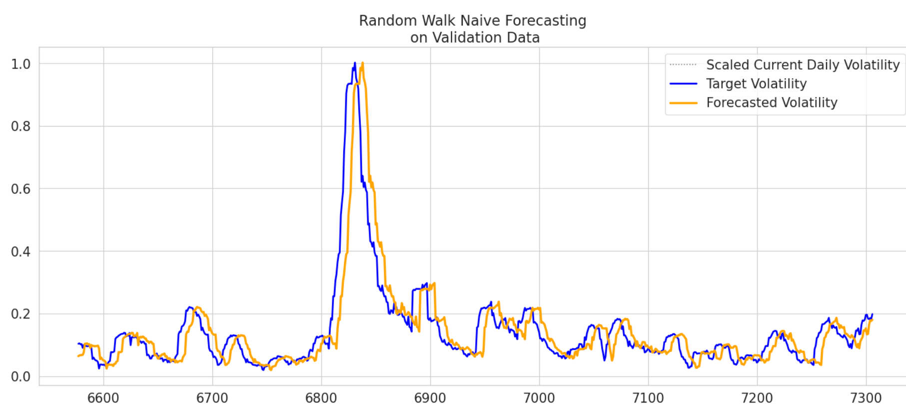


Figure 17: The result of random walk

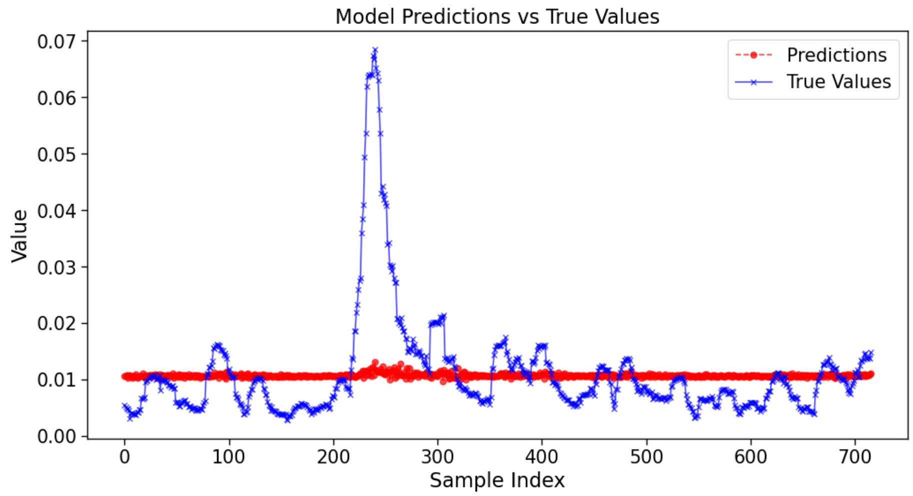


Figure 18: The result of FCN

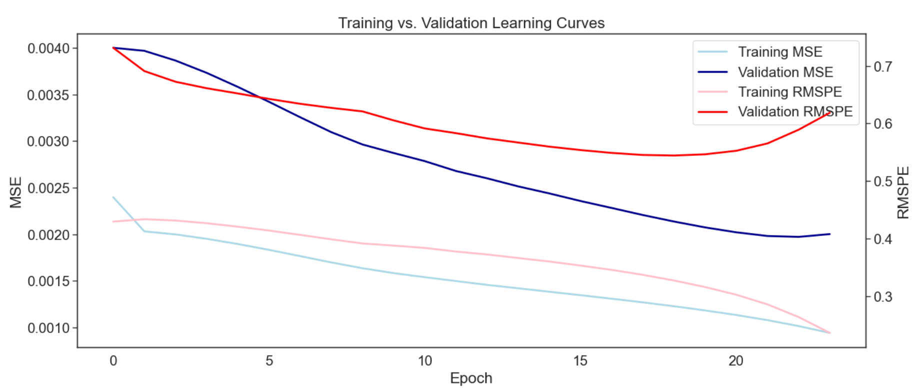


Figure 19: LSTM Loss Plot

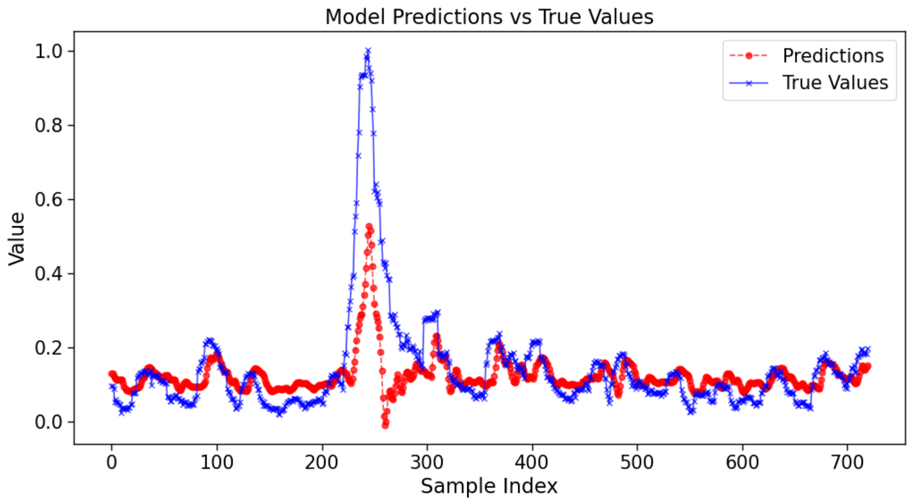


Figure 20: LSTM result on Validation set

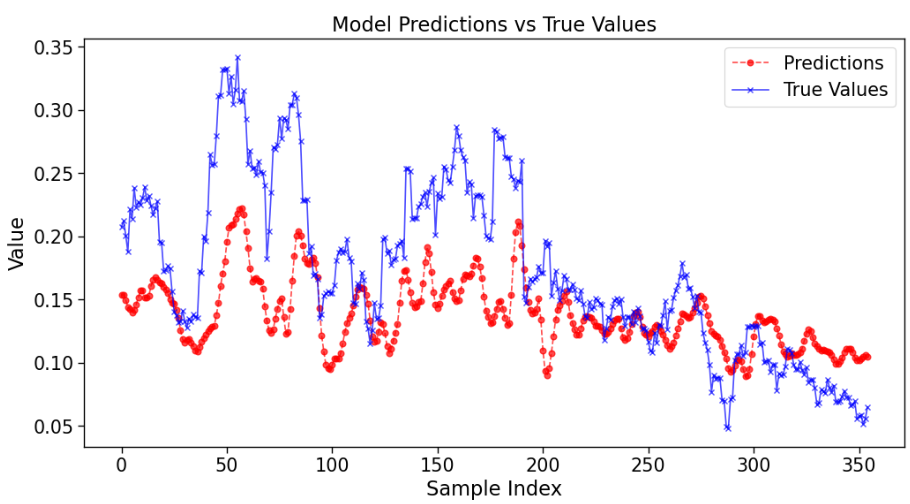


Figure 21: LSTM result on Test set

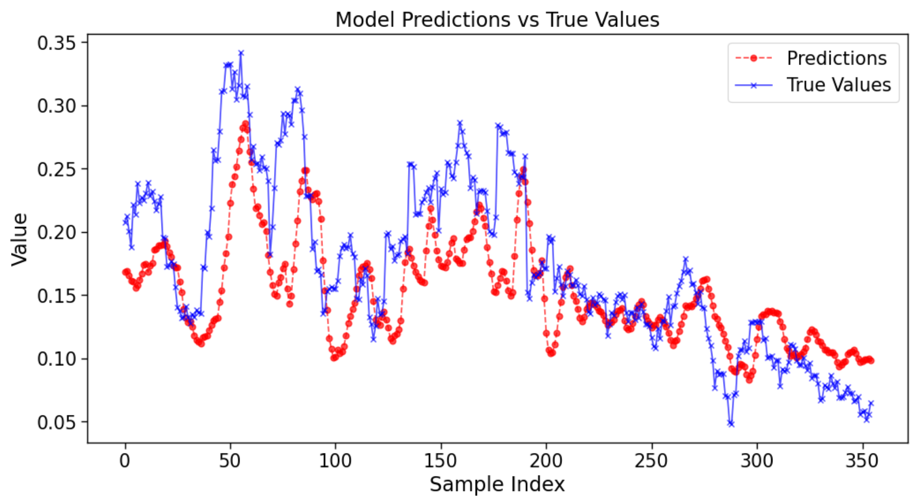


Figure 22: Bi-LSTM result on Validation set

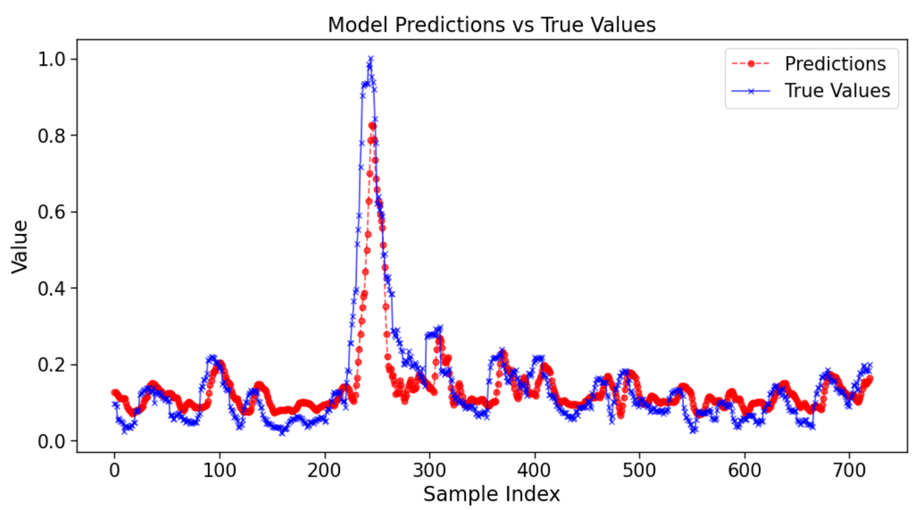


Figure 23: Bi-LSTM result on Test set

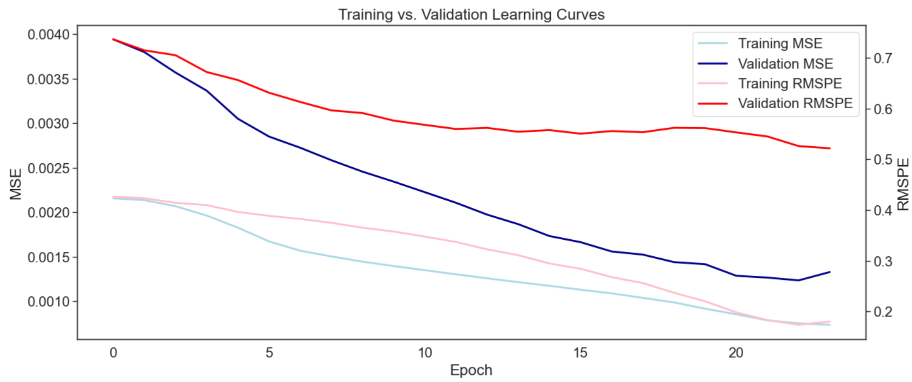


Figure 24: Bi-LSTM Loss Plot

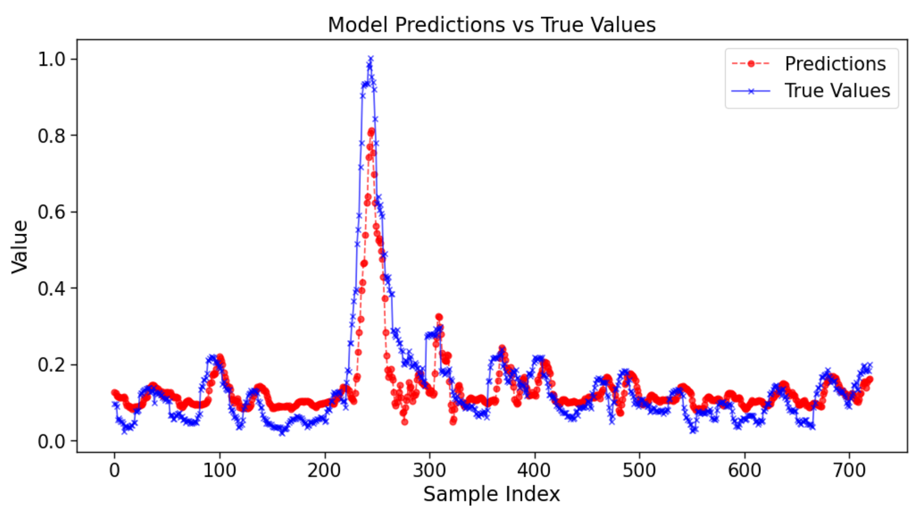


Figure 25: CNN-Bi-LSTM result on Test set

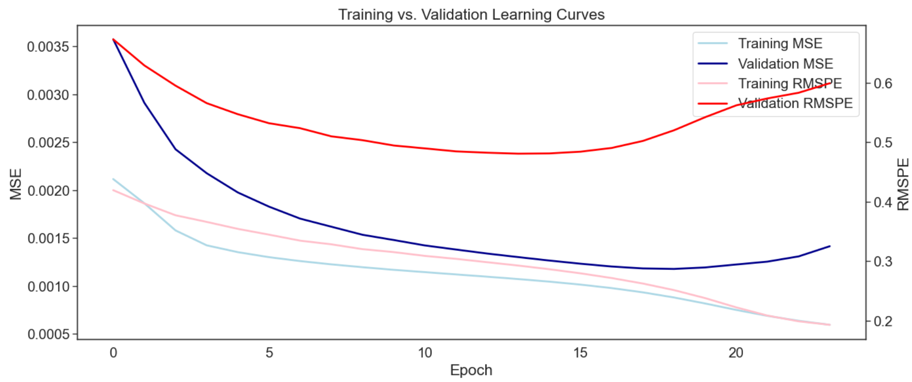


Figure 26: CNN-Bi-LSTM Loss Plot