Article

Forecasting Orange Juice Futures: LSTM, ConvLSTM, and Traditional Models Across Trading Horizons

Apostolos Ampountolas [ID]

School of Hospitality Administration, Boston University, Boston, MA 02215, USA; aampount@bu.edu

Abstract: This study evaluated the forecasting accuracy of various models over 5-day and 10-day trading horizons to predict the prices of orange juice futures (OJ = F). The analysis included traditional models like Autoregressive Integrated Moving Average (ARIMA) and advanced neural network models such as Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Backpropagation Neural Network (BPNN), Support Vector Regression (SVR), and Convolutional Long Short-Term Memory (ConvLSTM), incorporating factors like the Commodities Index and the S&P500 Index. We employed loss function metrics and various tests to assess model performance. The results indicated that for the 5-day horizon, the LSTM and ConvLSTM consistently outperformed the other models. LSTM achieved the lowest error rates and demonstrated superior capability in capturing temporal dependencies, especially in single-factor and S&P500 Index predictions. ConvLSTM also performed strongly, effectively modeling spatial and temporal data patterns. In the 10-day horizon, similar trends were observed. LSTM and ConvLSTM models had significantly lower errors and better alignment with actual values. The BPNN model performed well when all factors were included, and the SVR model maintained consistent accuracy, particularly for single-factor predictions. The Diebold–Mariano (DM) test indicated significant differences in forecasting accuracy, favoring advanced neural network models. In addition, incorporating multiple influencing factors further improved predictive performance, enhancing investment outcomes and reducing risk.

Keywords: orange juice futures price; time series forecasting; LSTM; RNN; BPNN; SVR; ConvLSTM; machine learning; commodities

1. Introduction

The study of commodity market dynamics has been a cornerstone of financial research, offering valuable perspectives on price formation, risk management, and market efficiency. The global economy relies heavily on commodities, particularly the oil and natural gas sector, as well as other important commodities such as energy, agriculture, minerals, and metals. Among the range of commodities, orange juice futures (OJ = F) have been a particularly interesting area of study already for several decades (Roll 1984) due to their distinctive market traits and the significant impact of both natural and economic factors. In addition, evaluating orange juice prices is crucial because several interrelated factors significantly impact the global market and consumer behavior (Wang and Wei 2021; Zhang et al. 2018).

Is it possible to reliably predict commodity prices? This question has been the subject of ongoing discussion in the financial and economic literature. For example, the recent surge in orange juice prices can be attributed to extreme weather events and persistent diseases affecting major orange-producing regions, i.e., hurricanes and pest infestations (Durbin and Pollastri 2024). Weather greatly influences orange juice production, unlike other widely produced commodities. Nevertheless, commodity prices are generally considered more unpredictable than stock prices or exchange rates, posing challenges for accurate forecasting. Factors like the interaction of demand and supply, economic expansion, market predictions, government regulations, and unexpected events such as spillover

effects, pandemics, war, and global debt crises all have an impact on commodity futures prices (Zhao et al. 2016). These complex factors are the primary drivers of the significant price fluctuations in the spot market prices of commodities. As a result, the expectation is that predicting commodity price trends, specifically, in this case, the orange juice futures prices, will not only help mitigate volatility and reduce risk in commodity markets but can also support governmental entities in making sound and long-term economic choices. Therefore, the motivation for this research is tied to the unique characteristics of the orange juice futures market, which is heavily influenced by various factors that create challenges for traders and investors. The study highlights the need to develop robust forecasting models to improve investment outcomes and reduce market risks, especially given the significant volatility of orange juice futures prices compared to other commodities like gold or oil.

Given the latest technological advancements, various methods are being utilized to forecast prices in the financial industry (Ampountolas 2023; Gupta and Nigam 2020). Although traditional econometric techniques, for example, the vector autoregressive model (VAR) struggle to accurately predict the non-linear aspects of commodity prices due to their robust linear assumptions (Sun et al. 2022; Wang and Fang 2022), advanced models such as machine learning techniques have gained significant attention due to their ability to observe volatility characteristics, non-linear information, and historical data effectively (Ampountolas 2024; Butler et al. 2021; Zhao et al. 2017) or combining models (Barrow and Crone 2016). Therefore, in the financial and economics literature, we have encountered many authors since the early years, for example, Kroner et al. (1995) who employed machine learning techniques, such as the Support Vector Regression (SVR), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNNs), Multi-Layer Perceptron (MLP), convolutional neural networks (CNNs), gate recurrent units (GRUs), backpropagation (BPNN) models, and many other models to validate the impact of various factors on predicting commodity futures prices.

Limited literature examines the price forecasts of orange juice futures (OJ = F) as an independent asset. Most research focuses on commodities like gold or oil, with many papers examining multiple commodities inclusively. Motivated by this and the enormous price growth during the last two years, we examine various forecasting models—ARIMA, LSTM, RNN, BPNN, SVR, NAR, and ConvLSTM—to predict commodity futures market's prices of orange juice futures prices. We also include other factors such as commodity futures (ES = F) and S&P500 Indexes. Therefore, we employ two forecasting horizons: 5 trading days and 10 trading days. Thus, this study aims to contribute to current research by analyzing and predicting the price trends of orange juice futures in the selected commodity markets. Additionally, we present a comparative analysis of the forecasting models based on loss functions and performance metrics. As such, predicting orange juice futures prices is essential because this market is highly volatile and affected by a range of unpredictable factors that have significant economic impacts. Accurate predictions can help stakeholders mitigate volatility, manage risk, and make more informed decisions in the commodity market. Moreover, given the increasing price volatility in recent years, enhanced forecasting methods can improve profitability for stakeholders involved in futures trading.

Our results revealed that for both the 5-day and 10-day horizons, advanced neural network models, particularly LSTM and ConvLSTM, consistently outperformed the other forecasting models. These models achieved the lowest error rates and demonstrated superior capability in capturing temporal dependencies, with ConvLSTM also effectively modeling spatial and temporal data patterns. The directional accuracy and Diebold and Mariano (1995) test supported the findings. In the 10-day horizon, the LSTM and ConvLSTM models again showed significantly lower errors and better alignment with actual values than ARIMA, which had the highest error rates. The BPNN model performed well when all factors were included, and the SVR model maintained consistent accuracy, especially for single-factor predictions. The DM test indicated significant differences in forecasting accuracy, favoring advanced neural network models.

Section 2 briefly overviews the current literature, and Section 3 discusses the relevant forecasting models and performance assessment metrics and details the data. Section 4 then presents an analysis of the empirical study's findings. Section 5 summarizes the study's conclusions and outlines potential directions for future research.

2. Literature Review

In one of the earliest studies, Roll (1984) confirmed that the weather condition variable impacts the market for frozen concentrated juice. Orange juice prices are impacted by high volatility as a result of concerns about extreme weather events that could affect production. However, he demonstrated that weather accounts for only a small portion of the fluctuations seen in futures prices. In another work, Kroner et al. (1995) utilized time-series approaches to generate long-term predictions of commodity price volatility by integrating investors' anticipated volatility. The authors assessed various forecasts of commodity price volatility, categorizing them into three groups: (1) forecasts based solely on expectations derived from options prices, (2) forecasts relying exclusively on time-series modeling, and (3) forecasts that combine market expectations and time-series techniques. They concluded that the forecasts proposed in category (3) outperformed the other two categories. Brooks et al. (2013) analyzed whether there is consistency in the evidence supporting two theories on commodity future pricing over time. The authors explored if the ability of commodity futures to predict prices is related to their seasonal fluctuations, and they also examined if there are changes in the pricing relationships at different times. They found more compelling evidence of seasonal patterns in the basis, which aligns with the storage theory. The findings reveal that structural changes mainly involve adjustments in the starting points rather than the trends, indicating that the predictive power of the basis remains consistent across various economic conditions. The study by Black et al. (2014) investigates how stock and commodity prices are related and whether this connection can be utilized to predict stock returns. Since both prices are associated with anticipated future economic performance, they are expected to have a lasting relationship, while shifts in sentiment toward commodity investments may impact how the response to imbalances occurs. The findings indicated that there is a long-term relationship between stock and commodity prices, and further tests identify disruptions in the predictive regression. The paper by Atsalakis et al. (2016) introduces an innovative method for predicting the price direction of 25 commodities on the global market using a neuro-fuzzy controller. The prediction system utilizes two adaptive neural fuzzy inference systems (ANFISs) to create an inverse controller for each commodity. The findings demonstrate a 68.33% hit rate with a significant improvement in return on equity compared to the buy-and-hold strategy.

In addition to traditional econometric approaches, various machine learning techniques are used to uncover the inherent complexity of commodity prices. The most common machine learning methods include neural networks (NN) and support vector machines (SVM), which are favored for their ability to model intricate characteristics like nonlinearity and volatility. Hybrid models have also demonstrated superior forecasting accuracy compared to their machine learning models. Drachal and Pawłowski (2021) briefly overviews how genetic algorithms (GA) are used to predict commodity prices. The authors concentrated on a hybrid method (i.e., combining genetic algorithms with other approaches) used in situations like determining if a complete forecasting technique can be split into two or more distinct parts, with one part being based on a GA and the other parts based on different methods. Another study by Jiang et al. (2022) utilized various machine learning techniques to confirm the influence of investor sentiment on estimating the price of crude oil futures. The authors included several forecasting models, such as the MLP, LSTM, SVR, RNN, and GRU models. The results indicated that the Long Short-Term Memory model yielded the best results when combined with the composite sentiment index. This was attributed to a reduced rate of accuracy errors and improved directional accuracy when forecasting next-day-ahead prices for time-series analysis. In a similar study, Guo et al. (2023) utilized machine learning to analyze historical data, volatility, and

non-linear characteristics. They assessed the predictive capabilities of neural network models such as the GRU, MLP, LSTM, RNN, CNN, SVR, and BPNN models on crude oil futures. The set of assessment tests illustrated that the GRU model surpassed other models in terms of accuracy and performance when forecasting crude oil futures prices. Moreover, the incorporation of relevant factors resulted in enhanced forecast accuracy for the proposed models. A recent study by Zheng et al. (2024) reported the effectiveness of hybrid models in enhancing the accuracy of crude oil price forecasts when compared to single models. Their research introduces an innovative interval-based approach. Initially, they apply variational mode decomposition (VMD) to split the original training series into low- and high-frequency components. The low-frequency component is considered an inseparable random set. It is forecasted using a newly developed autoregressive conditional interval (ACI) model, while the high-frequency component is predicted using interval Long Short-Term Memory (iLSTM) networks. The final interval-valued prediction is obtained by combining the forecasts of both components. Additionally, the study designs and implements a daily trading strategy based on interval-valued data.

Ren et al. (2024) introduced an innovative imaging technique to predict the daily price data of crude oil futures. Utilizing convolutional neural networks (CNNs), they achieved higher accuracy in predicting future price trends than other standard forecasting methods. The findings indicate that images can capture more nonlinear information, which is advantageous for energy price prediction, particularly during significant fluctuations in crude oil prices. In a different study, Ampountolas (2024) studied GARCH models and the Support Vector Regression (SVR) model to understand better how volatility changes in commodity returns, like gold and cocoa, as well as the financial market index S&P500. The evaluation showed that Support Vector Regression (SVR) performs better than traditional GARCH models for short-term forecasting, suggesting it could be a valuable alternative for predicting financial market trends. These results highlight the importance of choosing the right modeling techniques for specific types of assets and forecasting time frames.

In conclusion, an extensive body of literature discusses predicting volatility in commodity futures markets, mainly for energy, crude oil, or metals. Throughout the years, forecasting techniques have progressed from traditional econometric approaches to innovative machine learning methods. Consequently, the accuracy of forecast models is gradually increasing, and at the same time, it has been demonstrated that the variables influencing the prediction of commodity futures prices are varied.

3. Data and Methodology

3.1. Data

This study's data set contained daily historical time series data for three financial assets: orange juice futures (OJ = F), S&P500 futures (ES = F), and the S&P500 Index (GSPC). The dependent variable for this study is the price of orange juice futures. We aim to accurately forecast the price of orange juice futures in the USA market. The dataset covers the period from July 2022 to June 2024 and has 504 observations. The data were obtained from Yahoo Finance. In addition, we utilized the S&P500 futures index and the stock market, i.e., the S&P500 Index, as impact factors in the orange juice futures price estimation model.

3.2. Descriptive Statistics

3.2.1. Dataset Trend

Figure 1 illustrates a noticeable upward trend in the price of orange juice futures (OJ = F) over the two years, characterized by substantial volatility and periodic corrections. Similarly, the trend index confirms the bullish trajectory, with the futures price increasing by 145.17% from the beginning date for the dataset. In this context, the analysis highlights the potential for significant returns while emphasizing the volatility of commodity futures markets. Such insights are vital for investors and market analysts to make informed trading and investment decisions.
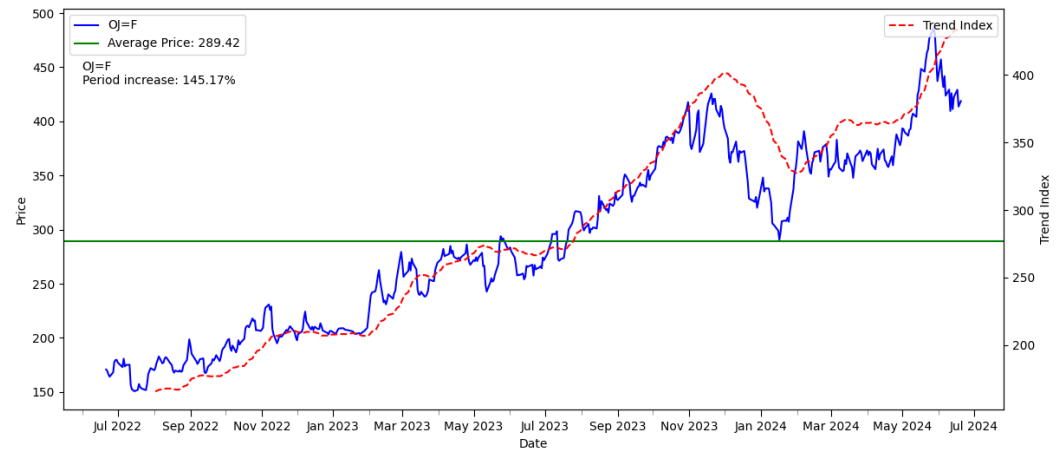
Figure 1. Orange juice futures price and trend.

3.2.2. Summary Statistics

Moreover, Table 1 reports a comprehensive overview of the study's summary statistics of the three financial assets: OJ = F, ES = F, and GSPC. OJ = F has a standard deviation of 82.74, which, relative to its mean, suggests significant variability in prices. This is also supported by the OJ = F price range, which shows a broad price range that aligns with its high standard deviation. In contrast, ES = F and GSPC have higher absolute standard deviations of 483.97 and 478.46, respectively, but these are small relative to their higher mean values and thus have relatively low volatility compared to OJ = F. Finally, OJ = F has a kurtosis of −1.1256, suggesting less frequent extreme deviations from the mean. ES = F and GSPC have kurtosis values of −0.7725 and −0.7436, respectively, indicating a similar but slightly less pronounced platykurtic distribution. At the same time, OJ = F shows a slight positive skewness of 0.1262, suggesting a marginally longer right tail. The financial indices, ES = F and GSPC, have higher positive skewness values of 0.5726 and 0.5850, respectively, indicating a more noticeable asymmetry with a longer right tail.

Table 1. Dataset summary statistics.

Asset	Obs	Mean	Std. Dev	Min	Max	Kurtosis	Skewness
OJ = F	504	289.4238	82.7351	150.6500	487.2000	−1.1256	0.1262
Financial Indices							
ES = F	504	4394.4995	483.9666	3588.5000	5491.0000	−0.7725	0.5726
GSPC	503	4376.1165	478.4562	3577.0300	5487.0298	−0.7436	0.5850

3.3. Forecasting Models

3.3.1. Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) model is a prominent statistical forecasting technique within the ARMA linear model class. According to Hyndman and Athanasopoulos (2018), the development of exponential smoothing models hinges on identifying trends and seasonality in the data. In contrast, ARIMA models are adept at handling stationary, non-stationary, and seasonal processes of order (p, d, q). The general form of the ARIMA model is represented as

$$(1 − \phi_1 B)\,(1 − \Phi_1 B^4)(1 − B)(1 − B^4)y_t = (1 + \theta_1 B)\,(1 + \Theta_1 B^4)\varepsilon_t \tag{1}$$

In this equation, y_t denotes the observed value at time t, and ε_t represents the error term, assumed to be white noise with a Gaussian distribution, having a mean of zero and a constant variance σ^2. The ARIMA model is denoted by ARIMA(p, d, q), where selecting the appropriate order (p, d, q) is a critical aspect of the ARIMA modeling procedure.

ARIMA models can be applied to both seasonal and non-seasonal data. Seasonal ARIMA requires a more intricate specification of the model components. Prior to estimating the time series models, it is essential to perform the augmented Dickey–Fuller (ADF) test Dickey and Fuller (1979) to determine the stationarity of the dataset. If the series is found to be non-stationary, data transformation is necessary. The ADF test is defined as follows:

$$\Delta x_t = \alpha_0 + b_0 x_{t-1} + \sum_{i=1}^{k} c_0 \Delta x_{t-1} + w_t \tag{2}$$

Here, Δ denotes the difference operator; α_0, b_0, and c_0 are coefficients to be estimated; x is the variable under examination; and w is the white noise error term. The null hypothesis ($b_0 = 0$) indicates that the series is non-stationary, while the alternative hypothesis ($b_0 < 0$) suggests that the series is stationary.

3.3.2. Recurrent Neural Network (RNN)

The RNN is structured with input, hidden, and output layers, allowing it to handle and retain new data simultaneously, thus enabling information transfer to subsequent periods (Henrique et al. 2018). Due to its feedback mechanism, the RNN incorporates historical data in its predictions. However, it struggles with retaining long-term data and may suffer from gradient explosion issues (Jiang et al. 2022). The RNN calculations are as follows:

$$h_t = f_h(u_t x_t + W_{t-1} h_{t-1}) \tag{3}$$

$$y_{t+T} = f_y(v_t h_t + b_y) \tag{4}$$

where h_t represents the hidden layer vector, x_t is the input layer vector, y_{t+T} is the output layer, u_t is the input-to-hidden weight at time t, v_t is the hidden-to-output weight at time t, and W_{t-1} is the weight from the output state at time $t-1$ to the hidden state at time t.

3.3.3. Long Short-Term Memory (LSTM)

LSTM is an advanced version of RNN featuring forget, input, and output gates. It leverages RNN's strengths while mitigating its weaknesses, making it suitable for time series prediction. Based on Jiang et al. (2022), the transfer process is detailed as follows:

$$F_t = \rho(W_{fx} x_t + W_{fh} h_{t-1} + b_f) \tag{5}$$

$$I_t = \rho(W_{ix} x_t + W_{ih} h_{t-1} + b_i) \tag{6}$$

$$O_t = \rho(W_{ox} x_t + W_{oh} h_{t-1} + b_o) \tag{7}$$

$$C_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_{cx} x_t + W_{ch} h_{t-1} + b_c) \tag{8}$$

$$h_t = o_t \circ \tanh(c_t) \tag{9}$$

where F_t is the forget gate at time t, W is the weight matrix, x_t is the input vector at time t, b the bias parameter, h_t is the hidden state vector at time t, I_t is the input gate at time t, O_t is the output gate at time t, ρ and tanh are the activation functions, and C_t is the candidate set.

3.3.4. Convolutional Long Short-Term Memory (ConvLSTM)

The Convolutional Long Short-Term Memory (ConvLSTM) model represents an advanced neural network architecture specifically designed to handle spatiotemporal data by integrating convolutional operations within the LSTM framework. The traditional fully connected LSTM (FC-LSTM) is powerful for sequence modeling but lacks the capability to effectively capture spatial correlations, as it uses fully connected layers that disregard spatial information. ConvLSTM addresses this limitation by incorporating convolutional structures in both the input-to-state and state-to-state transitions, allowing it to capture local spatial dependencies better.

The fundamental equations governing ConvLSTM are as follows:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \tag{10}$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \tag{11}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \tag{12}$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \tag{13}$$

$$H_t = o_t \circ \tanh(C_t) \tag{14}$$

Here, i_t, f_t, and o_t represent the input, forget, and output gates, respectively. The symbols X_t and H_t denote the input and hidden state at time t, while C_t is the cell state. The convolution operator is represented by *, and the Hadamard product is represented by $\circ$.

The ConvLSTM model thus maintains the advantages of traditional LSTM in handling long-term dependencies while enhancing its ability to process data with spatial structures. This makes ConvLSTM particularly suitable for applications like precipitation nowcasting, where capturing both spatial and temporal patterns is crucial. By stacking multiple ConvLSTM layers and forming an encoding–forecasting structure, the model achieves robust performance predicting future states from historical data, significantly outperforming traditional FC-LSTM models in spatiotemporal sequence forecasting tasks.

3.3.5. Backpropagation Neural Network (BPNN)

The Backpropagation Neural Network (BPNN) is among the most popular and extensively used models in artificial neural networks, renowned for its robustness and simplicity. BPNN employs a Multi-Layer Perceptron structure, typically consisting of an input layer, one or more hidden layers, and an output layer. The core principle of BPNN is the backpropagation algorithm, which adjusts the network weights to minimize the error between the predicted outputs and the actual targets. This is achieved through an iterative process of forward and backward passes.

During the forward pass, input data are propagated through the network, generating an output. The error is then calculated using a loss function, such as the mean squared error (MSE):

$$E = \frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{15}$$

where y_i is the actual target value and $\hat{y}_i$ is the value predicted by the network.

This error is propagated backward through the network to update the weights in the backward pass. The learning algorithm performs a gradient descent optimization on the weights linking the nodes in each layer. The weight update rule is derived from the gradient descent method, where the weights are adjusted in the direction that reduces the error. The update for a weight w_{ij} from neuron i to neuron j is given by

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \tag{16}$$

where η is the learning rate, controlling the step size of the weight update. The partial derivative $\frac{\partial E}{\partial w_{ij}}$ is computed using the chain rule, which involves calculating the gradient of the error concerning the weights.

Despite its advantages, BPNN has shortcomings, such as long training times and potential overtraining. However, its robustness and generally good performance across a wide range of applications make it a valuable tool in neural network modeling. Due to its effectiveness and ease of use, BPNN is often considered a benchmark for comparing the performance of other neural network models. This iterative weight adjustment process continues until the network converges to a state where the error is minimized, thereby

improving the model's accuracy. BPNN's ability to fine-tune weights through gradient descent is highly effective for various applications, including pattern recognition, time series forecasting, and complex function approximation.

3.3.6. Support Vector Regression (SVR)

Support Vector Regression (SVR) is a non-linear regression technique based on Support Vector Machine (SVM) principles. SVR excels in approximating functions and works by identifying a regression hyperplane in a high-dimensional feature space with minimal risk. According to Kazem et al. (2013), the formulation of SVR can be expressed as follows:

$$f(x) = w^T \phi(x) + b \tag{17}$$

Minimize

$$\frac{1}{2}||w||^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{18}$$

Subject to

$$y_i - (w \cdot \phi(x_i) + b) \leq \epsilon + \xi_i \tag{19}$$

$$(w \cdot \phi(x_i) + b) - y_i \leq \epsilon + \xi_i^* \tag{20}$$

$$\xi_i, \xi_i^* \geq 0 \tag{21}$$

In this formulation, $x_i \in \mathbb{R}^k$ for $i = 1, 2, ..., n$ and $y_i \in \mathbb{R}$. Here, y_i represents the target value of x_i, w is the weight vector, $\phi(x)$ denotes a non-linear mapping function and b is a bias term. The variables ξ_i and ξ_i^* are slack variables that account for deviations from the margin of tolerance ϵ.

SVR aims to determine the optimal hyperplane that approximates the data with a minimal margin of error and maintains the model's generalization ability by managing the trade-off between the hyperplane's flatness and the error tolerance.

3.3.7. Non-Linear Autoregressive (NAR) Neural Network

A neural network is a computational model designed for data processing that can capture relationships within data. One of the significant advantages of artificial neural networks (ANNs) over other forecasting and modeling approaches is their ability to approximate complex functions with high precision and identify nonlinear patterns in input data without preset assumptions. Dynamic neural networks, particularly the NAR model, are extensively utilized for modeling and forecasting time series data, such as financial time series.

The NAR model addresses nonlinear time series problems by utilizing a single time series and predicting its future values based solely on its past values. Mathematically, the future value of a time series Y_t is forecasted using its previous values $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-d}$, where f represents the mapping function performed by the neural network:

$$Y_t = f(Y_{t-1}, \ldots, Y_{t-d}) \tag{22}$$

This model aims to learn the optimal weights for the neurons to minimize the error between the network's output and the actual values. A crucial aspect of neural-network-based forecasting is the network's architecture, which defines the number of neurons in each layer and the connections between them. A feed-forward network with a hidden layer is commonly employed for time series modeling and forecasting. The NAR neural network typically features a feed-forward structure with a tansigmoid transfer function in the hidden layer and a linear transfer function in the output layer.

Determining the number of hidden neurons and the number of delays in observations (denoted by d) is essential because these parameters significantly influence the autocorrelation structure of the time series. Researchers often rely on trial-and-error experiments to choose these parameters due to the lack of a theoretical method for their

determination. In one-step-ahead forecasting tasks, the number of neurons in the output layer is usually set to one.

3.4. Assessment Indicators

3.4.1. Loss Functions

The study presents a comprehensive analysis of the forecasting accuracy of various loss functions, including the commonly used Mean Absolute Percentage Error (MAPE), as well as the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

$$\text{Mean Absolute Error: MAE} = \frac{1}{h} \sum_{j=1}^{h} \left| y_{t+j} - \hat{y}_{t+j} \right|, \tag{23}$$

$$\text{Root Mean Square Error: RMSE} = \sqrt{\frac{1}{h} \sum_{j=1}^{h} \left(y_{t+j} - \hat{y}_{t+j} \right)^2}, \tag{24}$$

$$\text{Mean Absolute Percentage Error: MAPE} = \frac{1}{h} \sum_{j=1}^{h} \frac{y_{t+j} - \hat{y}_{t+j}}{y_{t+j}}, \tag{25}$$

where $\hat{y}_{t+j}$ indicates the model's forecast at time t. y_{t+j} refers to the dataset's actual values, h refers to the forecasting horizon, and finally, j indicates the number of historical observations. A lower value obtained from these evaluation indicators signifies a smaller error, indicating that the predictive model effectively converges toward accurate results.

3.4.2. Forecasting Performance Metrics

In addition, we utilize the directional accuracy (DA) and accuracy improvement (AI) metrics and the Diebold and Mariano (DM) test to evaluate the performance of forecasting models.

The directional accuracy (DA) is a metric used to assess forecasting models by measuring their ability to predict the direction of changes in observed values. This is especially valuable in financial forecasting, where accurately predicting whether prices will increase or decrease is often more important than predicting the exact value. A higher DA indicates better forecasting model performance in predicting the direction of changes.

$$DA = \frac{100}{T} \sum_{t=1}^{T} d_t \tag{26}$$

where d_t is defined as

$$d_t = \begin{cases} 1 & \text{if } (Y(t) - Y(t-1))(\hat{Y}(t) - \hat{Y}(t-1)) \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

where $Y(t)$ and $\hat{Y}(t)$ are the actual and predicted values at time t, respectively, and T is the sample size. The indicator function d_t checks whether the predicted change in the value (from $t-1$ to t) matches the actual change in the value. If both the actual and predicted changes are in the same direction (both up or both down), d_t equals 1, indicating a correct prediction. If the directions do not match, d_t equals 0, indicating an incorrect prediction.

We have also employed the accuracy improvement (AI) and the Diebold and Mariano (DM) tests to compare the forecasting models more accurately.

The accuracy improvement (AI) is designed to compare two forecasting models. The accuracy improvement is defined as

$$\text{AI} = \frac{S - S_p}{S} \times 100\% \tag{28}$$

where the sum of the absolute errors for a specified model is denoted as S, and the sum of the absolute errors for the proposed model is denoted as S_p. If AI > 0, it indicates that the proposed forecasting model performs better, whereas if AI < 0, it implies that the proposed model has not overcome the specified model's drawback. The index AI provides a more intuitive way to compare precision.

Finally, we used the predictive accuracy test suggested by Diebold and Mariano (1995) to assess the statistical significance of enhancements in forecast accuracy. This test is commonly utilized to compare the predictive capabilities of various models and ascertain if the differences in accuracy are statistically meaningful.

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2 / T}} \tag{29}$$

where $\bar{d}$ refers to the mean of the loss differential series d_t and d_t represents the difference between the loss from the first model and the loss from the second model. T represents the number of observations and $\hat{\sigma}_d^2$ indicates an estimate of the variance of d_t.

4. Estimation Results

This study's estimation results present forecasts with a horizon of 5 trading days and 10 trading days. According to the whole evaluation forecasting model, we initially conducted a forecast without influencing factors (single factor—OJ = F). Then, considering the influencing variables, we added the ES = F factor along with the OJ = F. Afterward, we introduced the OJ = F and the S&P500 Index factor, conducting a new estimation, and finally, we performed estimations including all factors in the forecasting process.

4.1. Forecast Results in the 5-Trading Day Horizon

Table 2 presents the forecast accuracy results for the study's various models and the evaluation indicators across different financial indices (Single-factor (OJ = F), Commodities Index, S&P500 Index, and a combined category of all factors) in a 5-trading-day horizon (steps). Compared to advanced models, the traditional ARIMA model shows the highest error rates across all metrics, indicating its limited capacity to handle complex time series data. LSTM stands out with the lowest error rates, particularly excelling in single-factor and S&P500 Index predictions, showcasing its strength in modeling temporal dependencies with high accuracy (e.g., MAE of 12.4155 and 9.4766 and MAPE of 3.1107% and 2.4010%, respectively). RNN and BPNN also perform well, though RNN shows higher errors in the Commodities Index, indicating variability in performance across different data types. SVR exhibits consistent but moderate accuracy, with relatively low errors but less effectiveness than neural networks. BPNN shows low errors, specifically when introducing the Commodities Index and when we combine all factors to forecast the daily price of orange juice futures. While improving over ARIMA, NAR still presents higher errors, especially in the combined category of all factors. ConvLSTM demonstrates robust performance with low errors across most categories, second only to LSTM, highlighting its efficacy in capturing spatial and temporal data patterns. Overall, the results emphasize the superiority of advanced neural network models, particularly BPNN, LSTM, and ConvLSTM, in achieving accurate forecasts in financial time series data.

Table 3 compares the performance of various models in terms of directional accuracy and average improvement over a 5-day prediction period. The ARIMA model, serving as a baseline, shows a directional accuracy of 50.53%. The ConvLSTM model outperforms all others, with the highest directional accuracy of 62.11% and an average improvement of 65.33%. SVR also demonstrates strong performance with a 58.95% directional accuracy and 56.13% average improvement. The LSTM and BPNN models provide moderate enhancements, with directional accuracies of 55.79% and 51.58%, respectively, and average improvements of 39.11% and 51.82%. Interestingly, despite its poor directional accuracy of 46.32%, the NAR model shows the highest average improvement at 66.86%, suggesting it

may excel in other prediction aspects. RNN, with a directional accuracy of 47.37%, shows a significant average improvement of 53.59%. Overall, ConvLSTM is the most reliable model for directional predictions, followed by SVR, while NAR and RNN might enhance different prediction metrics beyond directional accuracy.

Table 2. Accuracy of models' forecasting results in 5-trading-day steps.

Models	Metrics	Single Factor	Commodities Index	S&P500 Index	All
ARIMA	MAE	75.3148	77.0628	52.4859	52.3981
	MSE	6904.3332	7212.5565	3565.8827	3546.0604
	RMSE	83.0923	84.9268	59.7150	59.5488
	MAPE	18.8555	19.2984	13.0741	13.0551
LSTM	MAE	12.4155	19.1915	9.4766	14.4386
	MSE	254.8753	522.2959	158.0147	341.6509
	RMSE	15.9648	22.8538	12.5704	18.4838
	MAPE	3.1107	4.8310	2.4010	3.5943
RNN	MAE	13.7396	26.1952	14.7374	14.3007
	MSE	316.8579	970.6563	401.1044	329.8499
	RMSE	17.8005	31.1554	20.0276	18.1618
	MAPE	3.4733	6.6225	3.6936	3.6512
BPNN	MAE	8.4186	42.2337	10.2134	9.9090
	MSE	137.5138	1943.8305	155.2525	159.2351
	RMSE	11.7266	44.0889	12.4600	12.6188
	MAPE	2.1074	10.9688	2.6142	2.5316
SVR	MAE	8.6390	10.7829	10.8827	10.1447
	MSE	142.4235	165.7649	166.9276	156.7200
	RMSE	11.9341	12.8750	12.9200	12.5188
	MAPE	2.1728	2.8067	2.8366	2.6253
NAR	MAE	14.1338	13.7667	13.6553	19.1757
	MSE	289.0329	288.4339	288.7256	501.5138
	RMSE	17.0010	16.9833	16.9919	22.3945
	MAPE	3.6222	3.5077	3.4741	4.9971
ConvLSTM	MAE	11.4094	11.4070	10.4159	14.5008
	MSE	185.1033	182.2130	162.6494	272.5280
	RMSE	13.6053	13.4986	12.7534	16.5084
	MAPE	2.9474	2.9387	2.6706	3.7759

Table 3. Directional accuracy and accuracy improvement—5-day steps (all factors).

	Models	Directional Accuracy (%)	Average Improvement (%)
5-day steps	ARIMA	50.53	–
	LSTM	55.79	39.11
	RNN	47.37	53.59
	BPNN	51.58	51.82
	SVR	58.95	56.13
	NAR	46.32	66.86
	ConvLSTM	62.11	65.33

Next, in Figure 2, we compare each prediction model's estimated results and actual values in the 5-trading day horizon. We observe that the ARIMA model shows poor performance, as its forecasted values are flat and do not capture the trend and volatility of the actual data. LSTM and ConvLSTM models, on the other hand, performed well, closely aligning with the actual data and accurately capturing both the upward trend

and fluctuations. The RNN model shows moderate accuracy, better than ARIMA, but still missed some key variations in the data. The BPNN also performs reasonably well, capturing the overall trend with some deviations. However, the SVR displayed substantial divergence from the actual data, indicating its inadequacy in this forecasting context. The NAR model improves upon ARIMA and SVR, capturing the general trend but still missing several peaks and troughs. We conclude that advanced neural network models, particularly LSTM and ConvLSTM, demonstrate superior forecasting capabilities and effectively handle the complexities and nonlinearities present in the time series data.
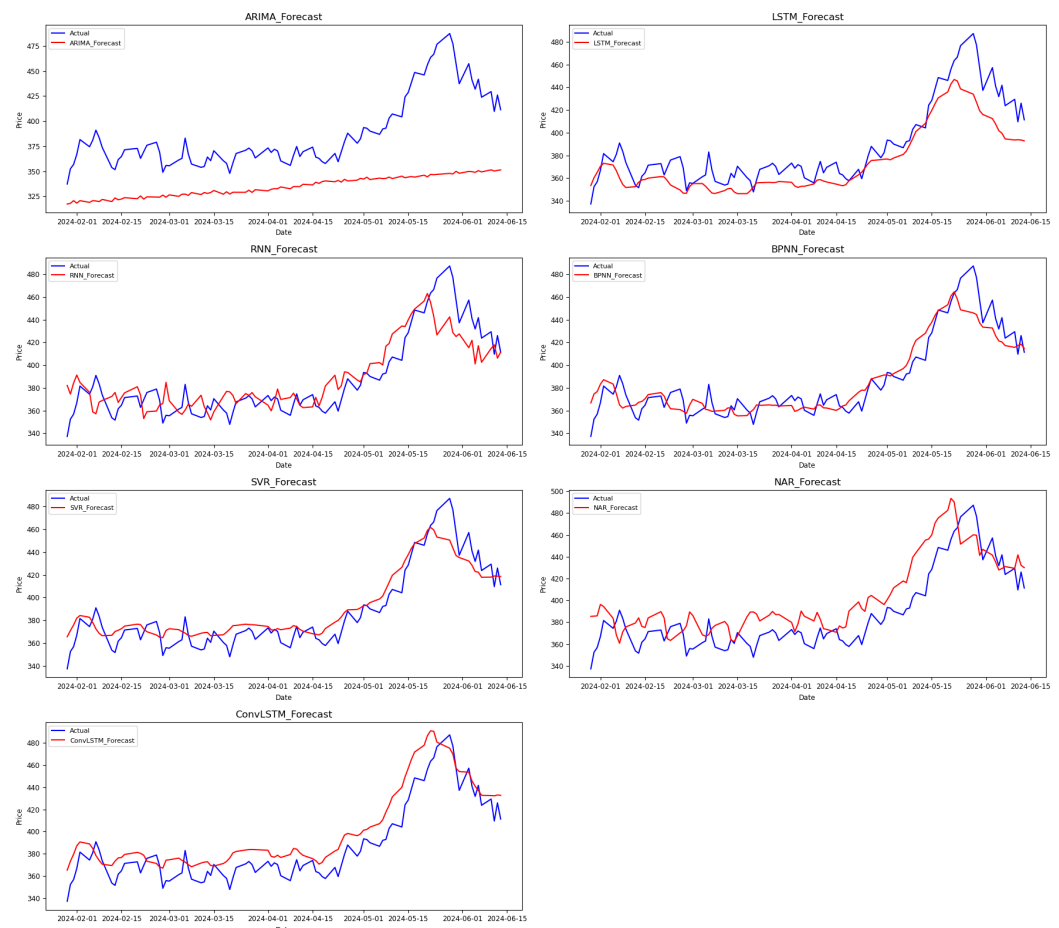

Figure 2. Comparison of forecasting models' actual and predicted values—5-day steps.

Finally, in Table 4, we present the results of the Diebold–Mariano test. The findings reveal significant differences in forecasting accuracy among various models, with ARIMA showing highly significant differences ($p < 0.01$) compared to all other models, suggesting its distinct performance characteristics. Notably, LSTM consistently outperforms other models, as indicated by significant positive DM statistics across all comparisons ($p < 0.01$). In contrast, RNN exhibits a notable negative DM value when compared with the BPNN, indicating inferior performance while showing better performance against SVR, NAR, and ConvLSTM. The BPNN model shows significant differences with SVR, NAR, and ConvLSTM, highlighting its unique predictive capabilities. The SVR and NAR comparisons also indicate significant differences, suggesting varied forecasting strengths. Furthermore, the comparison between ConvLSTM and NAR shows no significant difference, implying similar performance. As such, we observe that the DM test results underscore the variability in forecasting accuracy among the models.

Table 4. Diebold–Mariano (DM) test results among forecasting models in 5-day steps.

Models	Benchmark					
	LSTM	RNN	BPNN	SVR	NAR	ConvLSTM
ARIMA	17.8053 ***	24.3258 ***	20.7271 ***	25.5028 ***	24.4659 ***	22.6057 ***
LSTM		16.8237 ***	26.5703 ***	50.8174 ***	26.8106 ***	30.5143 ***
RNN			−1.9965 **	3.0291 *	18.2003 ***	9.1926 ***
BPNN				9.5119 ***	18.0418 ***	19.7829 ***
SVR					11.7014 ***	12.5263 ***
NAR						−1.6399

Note: * Significance at 10% level, ** at the 5% level, *** at the 1% level.

4.2. Forecast Results in 10-Trading Day Horizon

Table 5 presents the accuracy of the forecasting models evaluated over 10-day (horizon—trading day) steps across different datasets, similar to 5-day steps. The ARIMA model consistently shows the highest error rates across all metrics and datasets, with an MAE ranging from 51.4101 to 75.0684 and a MAPE from 12.8525% to 18.8773%, indicating its limited efficacy in forecasting complex time series data. In contrast, the LSTM model demonstrates significantly lower errors, with an MAE between 17.7515 and 19.3754 and a MAPE around 4.4288% to 4.7611%, highlighting its superior ability to capture temporal dependencies. The RNN model also performs well, particularly for the Commodities Index, but shows higher variability with an MAE from 18.9532 to 40.0245 and a MAPE from 4.6796% to 10.0940%. The BPNN model exhibits robust performance with the lowest errors among the neural networks in the category of all factors, achieving an MAE of 11.7737 and a MAPE of 2.9399%. The SVR outperforms the other models and maintains consistent, higher accuracy across datasets with an MAE from 12.4496 to 13.9587 and a MAPE of around 3.0931% to 3.4438%. It is noticeable that the SVR is the most accurate model for a single factor, including the Commodities Index and the S&P500 Index. While improving over ARIMA, the NAR model still shows higher errors than LSTM and BPNN, with an MAE between 18.7106 and 19.4742 and a MAPE from 4.7287% to 4.8448%. Lastly, ConvLSTM displays strong forecasting capability, excelling in the Commodities Index with the lowest MSE of 407.6756 and an RMSE of 20.1910 and maintaining a competitive performance across other datasets. These empirical results indicate that the SVR model and, in one case, the BPNN model are the most accurate models for the forecasting of orange juice futures prices, even if additional influencing factors are included in the prediction process. These findings are, to an extent, similar to the estimation results in the 5-trading day horizon, although in the 5-day step forecasts, the LSTM and the ConvLSTM demonstrated superior forecasting accuracy in some cases. Additionally, concurrently incorporating extra relevant factors can enhance all predictive models' performance. Thus, it has been shown once more that integrating influencing factors can decrease the forecasting model's prediction error and boost the accuracy of forecasting orange juice (OJ = F) futures prices.

Table 6 presents the results of the directional accuracy (DA) and accuracy improvement (AI) criterion for various forecasting models over 10-day steps. Directional accuracy measures the percentage of the correctly predicted direction of changes, while AI reflects the percentage improvement over a baseline model, in this study the ARIMA model. ARIMA shows a directional accuracy of 48.89%, serving as the AI baseline. Among the models, BPNN achieves the highest directional accuracy at 54.44%, indicating superior predictive capability in capturing the direction of changes. ConvLSTM, despite having a directional accuracy of 50.00%, shows the most substantial average improvement (63.75%) over ARIMA, highlighting its efficacy in enhancing forecasting accuracy. NAR and LSTM exhibit notable average improvements despite lower directional accuracies (45.56% and 44.44%). RNN and SVR demonstrate moderate directional accuracies (51.11% and 52.22%) but differ in accuracy improvement, with RNN showing a significant improvement (33.30%)

compared to SVR (8.20%). Therefore, the results suggest that while directional accuracy varies across models, advanced neural networks like ConvLSTM and NAR substantially improve forecasting accuracy.

Table 5. Models forecast results accuracy in 10-trading day horizon.

Models	Metrics	Single Factor	Commodities Index	S&P500 Index	All
ARIMA	MAE	67.8574	75.0684	51.4101	51.4660
	MSE	5836.0380	6898.8519	3465.5084	3473.0407
	RMSE	76.3940	83.0593	58.8686	58.9325
	MAPE	17.0000	18.8773	12.8525	12.8667
LSTM	MAE	18.8399	19.2947	17.7515	19.3754
	MSE	640.2273	668.2661	586.0565	707.8805
	RMSE	25.3027	25.8508	24.2086	26.6060
	MAPE	4.6487	4.7541	4.4288	4.7611
RNN	MAE	40.0245	20.8836	18.9532	21.0377
	MSE	2206.0296	812.4532	647.3866	900.3059
	RMSE	46.9684	28.5036	25.4438	30.0051
	MAPE	10.0940	5.1418	4.6796	5.1378
BPNN	MAE	14.5993	17.1953	14.3158	11.7737
	MSE	377.5428	426.3074	344.9291	256.9264
	RMSE	19.4305	20.6472	18.5723	16.0289
	MAPE	3.6595	4.4805	3.5939	2.9399
SVR	MAE	12.9888	12.4810	12.4496	13.9587
	MSE	345.7689	285.5075	284.0731	374.0397
	RMSE	18.5949	16.8970	16.8545	19.3401
	MAPE	3.1915	3.1000	3.0931	3.4438
NAR	MAE	18.7106	18.9656	18.9775	19.4742
	MSE	622.8136	621.6618	633.2322	649.2775
	RMSE	24.9562	24.9331	25.1641	25.4809
	MAPE	4.7287	4.7720	4.7523	4.8448
ConvLSTM	MAE	19.2422	15.5235	15.8455	17.5845
	MSE	621.6677	407.6756	408.7932	495.4774
	RMSE	24.9333	20.1910	20.2186	22.2593
	MAPE	4.9771	3.9664	4.0803	4.5623

Table 6. Directional accuracy and accuracy improvement—10-day steps (all factors).

	Models	Directional Accuracy (%)	Average Improvement (%)
10-day steps	ARIMA	48.89	–
	LSTM	44.44	40.33
	RNN	51.11	33.30
	BPNN	54.44	42.88
	SVR	52.22	8.20
	NAR	45.56	47.96
	ConvLSTM	50.00	63.75

Then, we evaluated the forecasted values from the estimation models, which take into account all the relevant factors with the actual value, as shown in Figure 3. We can observe that the ARIMA model significantly underperforms, failing to capture the upward trend and volatility, indicating its limitations in forecasting complex, nonlinear patterns. On the other hand, the LSTM and ConvLSTM models demonstrate a closer alignment with actual values, particularly in capturing the general upward trend and peak levels, highlighting

their superior ability to handle time series data with temporal dependencies. The RNN and BPNN models exhibit moderate performance, capturing some trends but with notable deviations and missed volatility. The SVR model shows substantial divergence from actual values, at least in this context, reflecting its inadequacy. Finally, while better than ARIMA, the NAR model still shows significant discrepancies, particularly in capturing peak values. Therefore, by comparing the predicted and actual values, we find that the LSTM and ConvLSTM models stand out for their enhanced forecasting capabilities, which effectively model the underlying patterns and trends in the data.
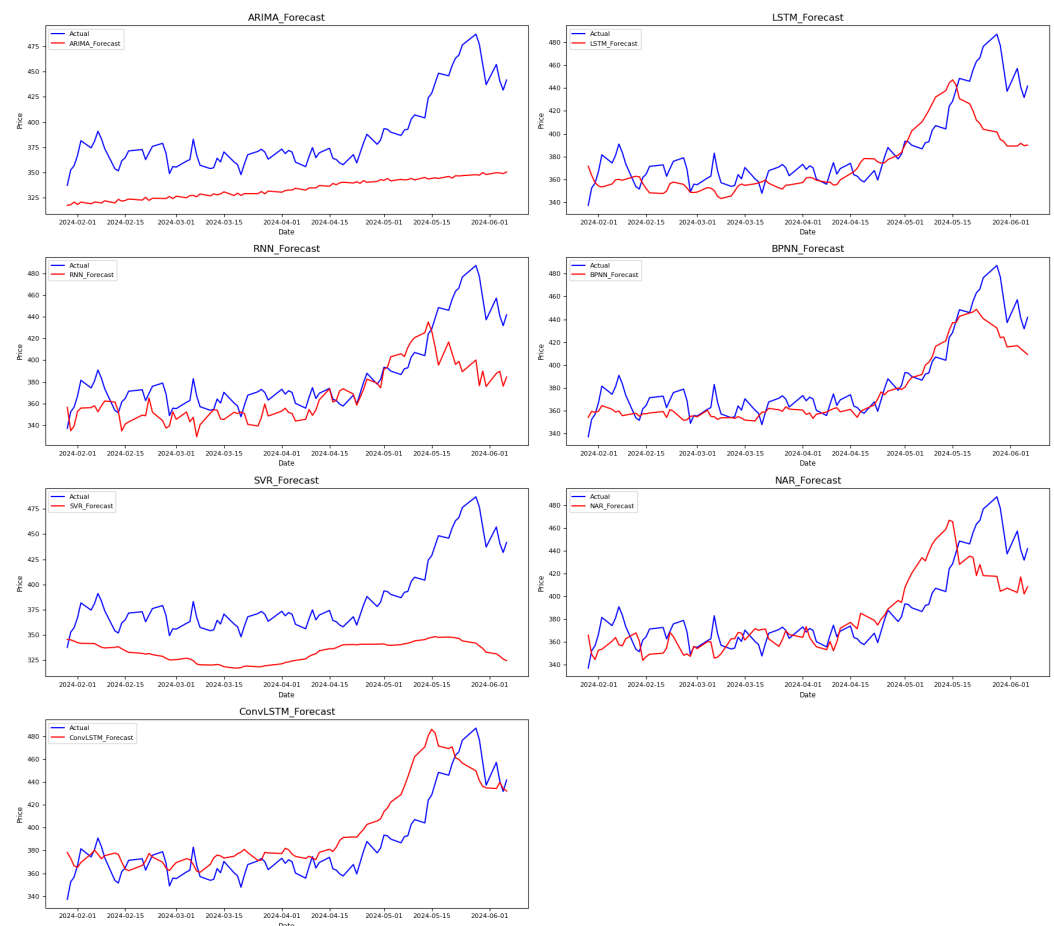

Figure 3. Forecasts model comparison of actual and predicted values—10-day steps.

Finally, we employed the Diebold–Mariano (DM) test to examine the forecasting accuracy of the various models over 10-day steps. Table 7 presents the test results. More specifically, the ARIMA model shows highly significant differences ($p < 0.01$) with all models except SVR, where it slightly underperforms (-0.6466 ***), suggesting ARIMA's generally distinct predictive behavior. LSTM exhibits significantly better performance compared to RNN (-8.3953 ***) and SVR (-17.9334 ***) but slightly outperforms BPNN (1.8406 ***) and demonstrates superior accuracy against NAR and ConvLSTM. RNN's performance is significantly worse than SVR (-16.0307 ***) but better than BPNN (5.8633 ***) and significantly improved over NAR and ConvLSTM. BPNN shows similar trends, underperforming against SVR (-16.2270 ***) but outperforming NAR and ConvLSTM. SVR demonstrates significant superiority over NAR (16.5782 ***) and ConvLSTM (19.7874 ***). Lastly, NAR's performance is significantly outperformed by ConvLSTM (15.6224 ***). In these findings, we observe that the variable forecasting capabilities across models with advanced neural networks like LSTM and ConvLSTM often exhibit superior performance compared to traditional models such as ARIMA, particularly in handling complex time series data.

Table 7. Diebold–Mariano (DM) test results among forecasting models in 10-day steps.

Models	Benchmark					
	LSTM	**RNN**	**BPNN**	**SVR**	**NAR**	**ConvLSTM**
ARIMA	18.5606 ***	17.0005 ***	17.7258 ***	−0.6466 ***	18.1408 ***	21.2885 ***
LSTM		−8.3953 ***	1.8406 ***	−17.9334 ***	8.5297 ***	19.5791 ***
RNN			5.8633 ***	−16.0307 ***	14.6996 ***	19.0879 ***
BPNN				−16.2270 ***	3.1734 ***	19.2480 ***
SVR					16.5782 ***	19.7874 ***
NAR						15.6224 ***

Note: *** Significance at the 1% level.

5. Conclusions

This study evaluated the forecasting accuracy of various models with different configurations over 5-day and 10-day trading horizons to forecast orange juice futures (OJ = F) prices. We have employed a dataset from July 2022 to June 2024. Our analysis included traditional models like ARIMA and advanced neural network models such as LSTM, RNN, BPNN, SVR, and ConvLSTM, with varying influencing factors like the Commodities Index and the S&P500 Index. In addition, we have adopted a set of loss function metrics to evaluate the accuracy of each model and various tests to assess the performance of each forecasting model.

For the 5-trading day forecasting horizon, the advanced neural network models, particularly LSTM and ConvLSTM, consistently outperformed traditional models like ARIMA. LSTM achieved the lowest error rates and demonstrated superior capability in capturing temporal dependencies, especially in single-factor and S&P500 Index predictions. ConvLSTM also exhibited strong performance, highlighting its effectiveness in modeling spatial and temporal data patterns. The directional accuracy and Diebold–Mariano test further supported the superiority of LSTM and ConvLSTM over other models.

In the 10-trading day forecasting period, we observed similar trends. While ARIMA displayed the highest error rates, the LSTM and ConvLSTM models showed significantly lower errors and better alignment with actual values. The BPNN model also performed well, mainly when we incorporated all factors. The SVR model maintained consistent accuracy across datasets, especially for single-factor predictions. The Diebold–Mariano test results indicated significant differences in forecasting accuracy, with advanced neural network models generally outperforming traditional models.

The findings of this study demonstrate that advanced models such as LSTM and ConvLSTM outperform traditional methods like ARIMA in forecasting orange juice futures prices. Specifically, LSTM achieved the lowest error rates across various factors, including the Commodities Index and S&P500 Index. This differs from previous research on commodities such as crude oil and gold, which favored machine learning techniques (e.g., LSTM, GRU) while emphasizing different influencing factors such as investor sentiment and macroeconomic indicators. For example, research by Guo et al. (2023) highlighted the superior performance of GRU in crude oil forecasting, particularly when considering relevant factors like volatility and historical data. Furthermore, while previous studies applied hybrid models to energy commodities, this study demonstrates the advantage of neural network models for commodity markets, emphasizing the need to customize forecasting tools to the distinctive characteristics of each market.

Our empirical results also have practical implications. Therefore, investors and analysts can promptly analyze market trends and identify potential risks based on the forecasting model results. As we have observed, the findings emphasize the superiority of advanced neural network models, particularly LSTM and ConvLSTM, in forecasting complex time series data. These models effectively capture underlying patterns and trends, offering enhanced forecasting capabilities compared to traditional models like ARIMA. Incorporating influencing factors further improves the predictive performance of these models, underscoring the importance of considering multiple variables in the forecasting

of financial assets. This optimization enhances investors' investment performance and reduces risk. Therefore, the DM test in both periods supports the above findings by indicating that models like LSTM and ConvLSTM not only provide statistically better predictions but also can offer traders and investors more reliable forecasts for decision-making. This could lead to improved returns and reduced risks, especially in volatile markets such as orange juice futures.

Limitations and Further Research

Despite the promising results, this study, like any other, has limitations that warrant further research. First, the dataset was limited to specific financial indices and assets. As such, future research could explore a broader range of variables and datasets to enhance the generalizability of the findings. In addition, we observed only two forecasting horizons (5-trading day and 10-trading day steps). Examining shorter- or longer-term forecasting estimations could provide more insights into the robustness and reliability of these models.

Second, while advanced neural network models showed superior performance, model optimization is also very important. Future studies should explore different optimization methods to enhance forecasting accuracy or incorporate additional forecasting models, such as hybrid models and parameters. At the same time, although neural networks like LSTM and ConvLSTM can effectively model nonlinear relationships, they require extensive data for training to avoid overfitting, especially in highly volatile markets like orange juice futures. Additionally, neural networks are computationally intensive, requiring significant time and resources for both training and fine-tuning, particularly as the complexity of the network increases. Finally, another practical consideration is the interpretability of these models. Neural networks are often seen as black boxes, making it difficult for users to understand how predictions are derived.

Finally, the study primarily focused on point forecasts. Introducing probabilistic forecasting methods could offer a more comprehensive evaluation of model performance by considering uncertainty and confidence intervals in predictions. Furthermore, the economic implications of these forecasts were not analyzed. Future research should assess the practical applications and financial benefits of employing advanced neural network models for trading and investment strategies.

Nevertheless, the study demonstrated the potential of the models utilized in financial forecasting, and further research could lead to even more robust and practical forecasting solutions.

Data Availability Statement: Data are publicly available.

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

References

Ampountolas, Apostolos. 2023. Comparative analysis of machine learning, hybrid, and deep learning forecasting models: Evidence from european financial markets and bitcoins. *Forecasting* 5: 472–486. [CrossRef]

Ampountolas, Apostolos. 2024. Enhancing forecasting accuracy in commodity and financial markets: Insights from garch and svr models. *International Journal of Financial Studies* 12: 59. [CrossRef]

Atsalakis, George, Dimitrios Frantzis, and Constantin Zopounidis. 2016. Commodities' price trend forecasting by a neuro-fuzzy controller. *Energy Systems* 7: 73–102. [CrossRef]

Barrow, Devon K., and Sven F. Crone. 2016. Cross-validation aggregation for combining autoregressive neural network forecasts. *International Journal of Forecasting* 32: 1120–1137. [CrossRef]

Black, Angela J., Olga Klinkowska, David G. McMillan, and Fiona J. McMillan. 2014. Forecasting stock returns: Do commodity prices help? *Journal of Forecasting* 33: 627–639. [CrossRef]

Brooks, Chris, Marcel Prokopczuk, and Yingying Wu. 2013. Commodity futures prices: More evidence on forecast power, risk premia and the theory of storage. *The Quarterly Review of Economics and Finance* 53: 73–85. [CrossRef]

Butler, Sunil, Piotr Kokoszka, Hong Miao, and Han Lin Shang. 2021. Neural network prediction of crude oil futures using b-splines. *Energy Economics* 94: 105080. [CrossRef]

Dickey, David A., and Wayne A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427–431.

Diebold, X. Francis, and S. Roberto Mariano. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13: 134–144.

Drachal, Krzysztof, and Michał Pawłowski. 2021. A review of the applications of genetic algorithms to forecasting prices of commodities. *Economies* 9: 6. [CrossRef]

Durbin, Dee-Ann, and Tatianna Pollastri. 2024. High Orange Juice Prices May Be on the Table for a While Due to Disease and Extreme Weather—finance.yahoo.com. Available online: https://finance.yahoo.com/news/high-orange-juice-prices-may-151316322.html (accessed on 18 July 2024).

Guo, Lili, Xinya Huang, Yanjiao Li, and Houjian Li. 2023. Forecasting crude oil futures price using machine learning methods: Evidence from china. *Energy Economics* 127: 107089. [CrossRef]

Gupta, Nalini, and Shobhit Nigam. 2020. Crude oil price prediction using artificial neural network. *Procedia Computer Science* 170: 642–647. [CrossRef]

Henrique, Bruno Miranda, Vinicius Amorim Sobreiro, and Herbert Kimura. 2018. Stock price prediction using Support Vector Regression on daily and up to the minute prices. *The Journal of Finance and Data Science* 4: 183–201. [CrossRef]

Hyndman, Rob J., and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. Melbourne: OTexts. Available online: http://OTexts.com/fpp2/ (accessed on 10 July 2024).

Jiang, Zhe, Lin Zhang, Lingling Zhang, and Bo Wen. 2022. Investor sentiment and machine learning: Predicting the price of china's crude oil futures market. *Energy* 247: 123471. [CrossRef]

Kazem, Ahmad, Ebrahim Sharifi, Farookh Khadeer Hussain, Morteza Saberi, and Omar Khadeer Hussain. 2013. Support Vector Regression with chaos-based firefly algorithm for stock market price forecasting. *Applied Soft Computing* 13: 947–958. [CrossRef]

Kroner, Kenneth F., Kevin P. Kneafsey, and Stijn Claessens. 1995. Forecasting volatility in commodity markets. *Journal of Forecasting* 14: 77–95. [CrossRef]

Ren, Xiaohang, Wenting Jiang, Qiang Ji, and Pengxiang Zhai. 2024. Seeing is believing: Forecasting crude oil price trend from the perspective of images. *Journal of Forecasting* 43: 2809–2821. [CrossRef]

Roll, Richard. 1984. Orange juice and weather. *The American Economic Review* 74: 861–880.

Sun, Yongxuan, Bowen Zhang, Zhizhong Ding, Momiao Zhou, Mingxi Geng, Xi Wu, Jie Li, and Wei Sun. 2022. Environment-aware vehicle lane change prediction using a cumulative probability mapping model. *International Journal of Sensor Networks* 40: 1–9. [CrossRef]

Wang, Donghua, and Tianhui Fang. 2022. Forecasting crude oil prices with a wt-fnn model. *Energies* 15: 1955. [CrossRef]

Wang, Wenting, and Longbao Wei. 2021. Impacts of agricultural price support policy on price variability and welfare: Evidence from china's soybean market. *Agricultural Economics* 52: 3–17. [CrossRef]

Zhang, Dongqing, Guangming Zang, Jing Li, Kaiping Ma, and Huan Liu. 2018. Prediction of soybean price in china using qr-rbf neural network model. *Computers and Electronics in Agriculture* 154: 10–17. [CrossRef]

Zhao, Lin, Xun Zhang, Shouyang Wang, and Shanying Xu. 2016. The effects of oil price shocks on output and inflation in china. *Energy Economics* 53: 101–110. [CrossRef]

Zhao, Yang, Jianping Li, and Lean Yu. 2017. A deep learning ensemble approach for crude oil price forecasting. *Energy Economics* 66: 9–16. [CrossRef]

Zheng, Li, Yuying Sun, and Shouyang Wang. 2024. A novel interval-based hybrid framework for crude oil price forecasting and trading. *Energy Economics* 130: 107266. [CrossRef]