Forecasting Carbon Prices Using Machine Learning: A Comparative Study of ARIMA, Random Forest, and LSTM Models

Erica Zhao

Preprint submitted to Elsevier *2024 Spring*

1. Introduction

Accurate carbon price forecasting within the European Union Emissions Trading System (EU ETS) is essential for informing policy and investment decisions in the transition to a low-carbon economy. However, the EU ETS presents a unique challenge for forecasting due to the dynamic and volatile nature of carbon prices, influenced by regulatory changes, economic indicators, and energy prices. Traditional time series models, like ARIMA, have been widely applied in carbon price forecasting, valued for their ability to capture linear trends and autocorrelations under stable market conditions (Zhu et al. 2017). However, as carbon markets become more complex and volatile, especially in recent EU ETS phases, ARIMA's limitations in handling nonlinearity have become a potential issue (Huang et al. 2021).

Machine learning models such as Random Forest (RF) have emerged as alternatives due to their capability to manage nonlinear dependencies and interactions among predictors. For example, Xu et al. (2024) demonstrated that RF could effectively incorporate external predictors, such as energy prices and economic indicators, to improve forecast accuracy in carbon price prediction compared to traditional econometric models. Moreover, deep learning models like Long Short-Term Memory (LSTM) networks offer a complementary approach by leveraging sequential dependencies in time series data, enabling them to capture long-term temporal patterns and fluctuations typical of carbon markets (Zhang et al. 2023).

In this study, we conduct a comparative analysis of ARIMA, RF, and LSTM models under identical market conditions to forecast carbon prices within the EU ETS. Specifically, we assess each model's ability to capture the dynamics of carbon prices by applying consistent predictor variables, including economic indicators, energy prices, weather conditions and regulatory factors, across all models. The ARIMA model was chosen for its interpretability and capacity to model stationary data. RF was selected for its robustness in handling nonlinear relationships, and LSTM was chosen due to its capacity to learn long-term dependencies in time series data. Each model was trained on the same dataset with time-series cross-validation to preserve temporal order, and a rolling forecast approach was used to evaluate predictive performance.

The contributions of this study are twofold. First, it provides a rigorous comparison of ARIMA, RF, and LSTM models for carbon price forecasting under the EU ETS, with insights into each models performance in handling volatility and nonlinearities. Second, this study illustrates the specific market conditions under which each model performs optimally. We show that while ARIMA can be effective for relatively stable periods, RF and LSTM provide more accurate predictions in periods of heightened market volatility. This evaluation offers both theoretical insights

and practical guidance for model selection in carbon price forecasting, ultimately supporting policymakers and investors in navigating the complexities of carbon markets.

2. Literature Review

Research in carbon price forecasting has evolved substantially, beginning with traditional econometric models and advancing through machine learning and hybrid approaches. Initial studies predominantly relied on linear models such as ARIMA and GARCH, which have been widely valued for their simplicity and statistical rigor. Zhu and Wei (2017) demonstrated ARIMA's effectiveness in capturing linear trends in stable periods of the carbon market, providing a baseline for predictive models in early carbon market phases. However, as carbon markets have matured and volatility increased, especially after 2017, these linear models have faced limitations in addressing non-linear patterns, a shortcoming noted by several studies (Huang et al. 2021). This limitation has driven researchers to explore alternative methods better suited to the complex dynamics of carbon pricing.

To overcome these challenges, machine learning techniques, especially ensemble models like Random Forest (RF), have gained prominence due to their ability to capture complex, nonlinear relationships. Xu et al. (2024) illustrated RF's advantages over traditional econometric models by incorporating dynamic variables from energy markets, leading to significant improvements in predictive accuracy. Similarly, Wang et al. (2021) developed an advanced RF-based framework that improved forecasting by integrating feature extraction techniques, which enhanced the models adaptability across varying market conditions.

The rise of deep learning models, particularly Long Short-Term Memory (LSTM) networks, has further advanced carbon price forecasting. LSTM models are well-suited for time series data due to their architecture, which captures long-term dependenciesa critical advantage when modeling sequential patterns in volatile carbon markets. For example, Yang et al. (2023) found that an LSTM-based model significantly outperformed traditional approaches by accounting for non-linear dependencies and temporal dynamics in carbon prices. Building on this, Zhang et al. (2023) utilized a CNN-BILSTM hybrid model, combining convolutional layers with bidirectional LSTM units to capture multiscale temporal patterns, thereby achieving higher accuracy in both stable and volatile phases of the market.

In recent years, hybrid models have emerged as a powerful approach, combining the strengths of traditional econometric and advanced machine learning techniques to address the limitations of standalone models. For instance, the hybridization of

ARIMA and machine learning methods, such as Least Squares Support Vector Machines (LSSVM), leverages ARIMAs linear modeling with the non-linear flexibility of LSSVM, resulting in enhanced forecasting accuracy in volatile markets (Zhu et al. 2017). Similarly, Wang et al. (2021) proposed a hybrid model using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose carbon price series into different frequencies before feeding them into LSTM and RF models for multiscale predictions, a method that successfully captured complex price fluctuations and improved robustness across different market conditions. Kamdem et al. (2023) further demonstrated that integrating time-frequency decomposition with machine learning models can enhance predictive performance by isolating specific frequency patterns within the EU ETS market.

3. Methodology

This study employs three distinct modelsARIMA, Random Forest, and LSTMto forecast carbon prices, implementing a consistent set of extraneous variables, cross-validation techniques, and a rolling forecast methodology.

Each model leverages a comprehensive set of extraneous variables to capture the effect of external influences on carbon prices. These variables include macroeconomic indicators (such as Fit_for_55, Law, and the Economic Policy Uncertainty (EPU) index), environmental factors (temperature, precipitation, solar energy, and wind speed), as well as energy prices (Crude Oil, Coal, and Natural Gas) and major stock indices (DAX, STOXX, FTMIB, CAC40, AEX, and SPGTCLNT).

For data preparation across all models, the dataset was split into an 80% training set and a 20% testing set. Time-series cross-validation was applied to ensure that the temporal order of the data was maintained, thus reflecting realistic forecasting conditions. Furthermore, we employed a rolling forecast strategy, whereby each model is updated iteratively with newly observed data.

3.1. ARIMA Model

The ARIMA model is frequently applied in carbon and energy price forecasting due to its capability to capture linear trends and seasonality in time series data. Previous studies have shown ARIMAs effectiveness in emissions and financial markets, particularly in relatively stable conditions (Contreras et al. 2003; Adebiyi et al. 2014). Our study utilizes the ARIMA(p, d, q) model, where p represents the autoregressive order, d the differencing order, and q the moving average order. The model

is expressed as:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{k=1}^{m} \beta_k X_{t-k} + \varepsilon_t$$

where y_t represents the forecasted carbon price at time t, c is a constant, and ϕ_i, θ_j, and β_k are the coefficients for the AR terms, MA terms, and extraneous variables, respectively, while ε_t denotes the error term.

We began by conducting a preliminary analysis to assess trends and autocorrelations in the carbon price data, using diagnostic tools such as the autocorrelation function (ACF) and partial autocorrelation function (PACF) to inform our choice of parameters p, d, and q. And we then uses maximum likelihood estimation (MLE) to determine the parameters. This process iteratively adjusts parameters to maximize the log-likelihood function:

$$\text{LogL} = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$$

where $\hat{y}_t$ represents fitted values, and σ^2 denotes residual variance.

To address multicollinearity, we applied Principal Component Analysis (PCA) to two feature groupsstock indices (DAX, STOXX, FTMIB, CAC40, AEX, SPGTCLNT) and energy prices (Crude Oil, Coal, and Natural Gas). This dimensionality reduction allowed us to condense these variables into two principal components, PCA_{Stock_Index} and PCA_{Energy_Prices}, which we then included as exogenous variables in the ARIMA model. Before conducting PCA, we standardized the data using a StandardScaler to maintain uniform scaling and prevent any single feature from disproportionately influencing the resulting components.

3.2. Random Forest

Random Forest (RF) is an ensemble learning approach that is well-suited for capturing complex, non-linear relationships in time series forecasting. Studies by Khaidem et al. (2016) and Lahouar and Slama 2015) demonstrated RFs success in stock and energy load forecasting. In this study, RF operates through Bootstrap Aggregating (Bagging), where the original dataset is sampled with replacement to generate multiple datasets for training. This sampling technique, involving random selection with replacement, generates subsets where individual instances may repeat or be excluded. The RF algorithm constructs a sequence of decision trees $(T_1, T_2, \ldots, T_N)$ during the training process. For each regression task, a tree T_i is

trained on a bootstrap sample D_i, drawn with replacement from the complete training dataset D. During tree construction, a random subset of m features is selected at each node for splitting. Given a new input x, each tree T_i yields a prediction y_i, and the RF's final prediction $\hat{y}$ for input x is calculated as the average of these predictions:

$$\hat{y}(x) = \frac{1}{N} \sum_{i=1}^{N} y_i(x)$$

Within each tree, the learning objective is to minimize prediction error, quantified by mean squared error (MSE), as shown by:

$$\mathrm{Err}(t) = \frac{1}{n_t} \sum_{D_t} (y_i - k_t)^2$$

where n_t represents the number of instances in node t, and k_t is the average of instances within each node.

We applied two variations of the Random Forest model in this study: one using the original undifferenced carbon price data and another using a differenced series to address non-stationarity. Differencing was performed by subtracting the previous value of the time series from the current value:

$$y'_t = y_t - y_{t-1}$$

where y'_t represents the differenced carbon price at time t, and y_t and y_{t-1} are the current and previous carbon prices, respectively. This transformation removed trends and other non-stationary components, preparing the series for machine learning. The variables used as extraneous variables were not differenced.

To implement the Random Forest model effectively, we optimize several hyperparameters, including the number of trees (`n_estimators`), maximum tree depth (`max_depth`), minimum samples required to split a node (`min_samples_split`), minimum samples required at a leaf node (`min_samples_leaf`), and the number of features considered at each split (`max_features`). Hyperparameter tuning is conducted via grid search based on minimizing Mean Squared Error (MSE), with time-series cross-validation applied to capture the temporal structure of the data and minimize overfitting. Each decision tree is trained on a bootstrap sample, and the models final predictions were obtained by averaging outputs across individual trees.

3.3. LSTM Model

Long Short-Term Memory (LSTM) networks, known for their ability to model sequential dependencies, are increasingly applied in volatile financial markets, includ-

ing carbon price forecasting. LSTM has proven effective in prior studies for capturing temporal dynamics and non-linear relationships across various market conditions (Sherstinsky 2020; Zhao et al. 2017). An LSTM cell includes memory units that retain information across sequences, regulated by three gates: the forget gate (f_t), input gate (i_t), and output gate (o_t). These gates operate as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, y_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, y_t] + b_i)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, y_t] + b_o)$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, y_t] + b_c)$$

$$h_t = o_t * \tanh(c_t)$$

where the hidden state h_t and cell state c_t manage information retention and update across time steps. This architecture enables the LSTM to effectively capture dependencies within sequences, which is crucial for modeling the temporal behavior of carbon prices.

To configure the LSTM model, we perform hyperparameter tuning using `RandomizedSearchCV`, optimizing batch size, epochs, optimizer, and learning rate. This search yield an optimal multi-layer LSTM configuration with dropout layers for regularization and either the `Adam` or `RMSprop` optimizer, contingent on the chosen learning rate.

For model training, we preprocess the carbon price data by normalizing features and target values, then structure the dataset with a look-back period of 1 time steps to capture relevant historical patterns. Using backpropagation through time, we train the LSTM model to update its weights based on the sequential structure of the data. The training process employ early stopping, with `mean_absolute_error` as the loss function to optimize prediction accuracy.

4. Descriptive Statistics

In the analysis of carbon pricing trends within the European Union's Emissions Trading System (EU ETS), this study draws upon a meticulously compiled dataset encompassing the period from January 1, 2021, to January 31, 2024. During this time frame, a total of 763 data points were collected, providing a granular view of carbon price movements. The carbon price data, sourced from Investing.com, is critical for understanding the economic implications of regulatory shifts and market dynamics under the EU's climate policy framework.

This timeframe is deliberately aligned with the initiation of Phase 4 of the European Union's Emissions Trading System (EU ETS), marking a significant transition

from the system's previous phases. The focus on this specific period is due to the substantial regulatory shifts introduced in Phase 4, which differ markedly from the earlier phases in terms of emissions reduction targets and allowance allocation mechanisms.

Notably, in Phase 3 of the EU ETS (2013-2020), the Union-wide cap for stationary installations was subject to an annual decrease by a linear reduction factor of 1.74%, with the 2013 cap determined based on the average total quantity of allowances issued annually during 2008-2012. Conversely, Phase 4 (2021-2030) introduces a more aggressive approach to emissions reduction, with the cap on emissions decreasing annually at an increased linear reduction factor of 2.2%. This heightened ambition reflects the EU's commitment to achieving more substantial emissions reductions and necessitates an in-depth analysis of carbon pricing trends under the new regulatory framework. The study aims to isolate the impacts of these changes on carbon pricing, underscoring the importance of understanding the dynamics introduced by Phase 4's enhanced regulatory measures.

The descriptive statistics for the dataset, which includes carbon pricing trends within the European Union's Emissions Trading System (EU ETS) from January 1, 2021, to January 31, 2024, are presented in Table 1. These statistics provide insights into the behavior of the carbon market during Phase 4 of the EU ETS. The carbon price exhibits a mean of 72.98 and a standard deviation of 16.45, reflecting market volatility over the 763 observations in the dataset.

Table 1: Descriptive Statistics of the Dataset

Variable	Mean	Standard Deviation	Maximum	Minimum
Carbon	72.98	16.45	98.01	31.96
Coal	176.47	93.46	439.00	63.75
Crude Oil	83.82	14.26	127.98	51.09
Natural Gas	72.45	54.61	339.20	15.53
Fit_for_55	0.83	0.38	1	0
Law	0.22	0.41	1	0
Compromise	0.25	0.43	1	0
EPU	276.88	69.80	418.62	141.39
DAX	14980.43	1082.54	16972.34	11975.55
SPGTCLNT	1699.21	295.68	2720.79	1088.34
FTMIB	25692.05	2414.81	30744.24	20352.98
CAC40	6722.63	526.14	7677.47	5399.21
AEX	732.69	44.97	827.57	630.58
STOXX	448.01	22.92	494.35	382.89
Temperature	12.94	6.64	31.1	-3.2
Precipitation	1.53	4.05	36.03	0.0
Solar Energy	6.75	5.88	27.1	0.1
Wind Speed	11.39	3.82	30.7	3.8

The time series plot of the carbon price, depicted in Figure 1, emphasizes the fluctuations over the observed period. Furthermore, the study will explore the patterns in related energy commodities, policy indicators, European stock indices, and weather variables, all of which could influence emission levels and subsequently, carbon market dynamics.

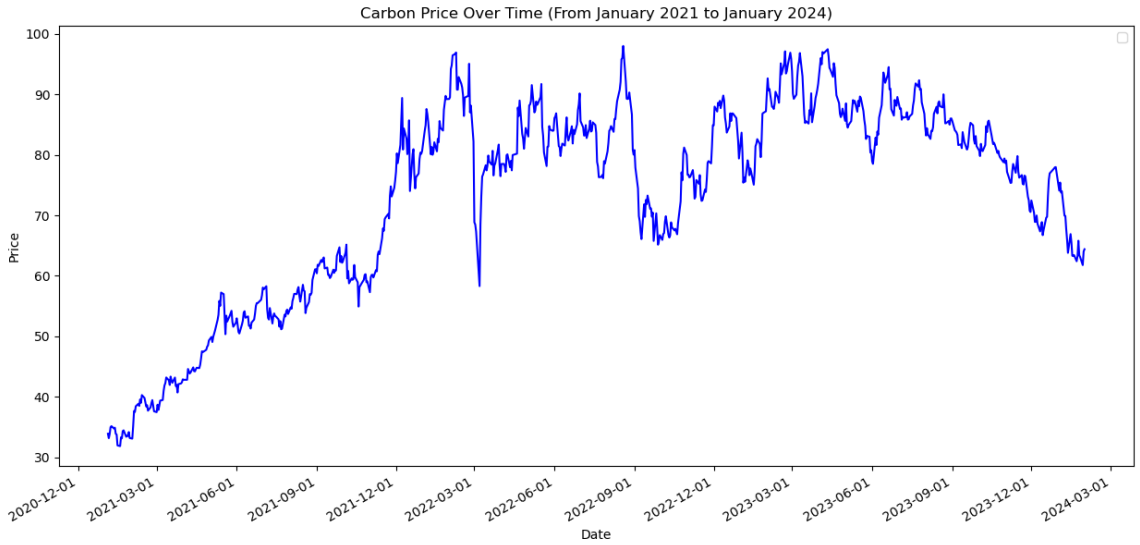

Figure 1: Time Series Plot of Carbon Price from January 2021 to January 2024.

5. Performance and Results

5.1. Evaluation Criteria

To evaluate the accuracy and reliability of our predictive models in forecasting carbon prices, we utilize several statistical metrics. The primary measure is Root Mean Squared Error (RMSE), which captures the magnitude of prediction errors:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

where y_t represents the observed value, $\hat{y}_t$ the forecasted value, and n is the number of observations. RMSE is complemented by Mean Absolute Percentage Error (MAPE), offering insights into the relative error size:

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|$$

To capture additional dimensions of forecast accuracy, we also include Mean Absolute Error (MAE) and Mean Squared Error (MSE):

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$$

The MAE provides the average absolute error, while the MSE further emphasizes larger errors due to its quadratic form. We also assess model fit with the Coefficient of Determination (R^2), which quantifies the proportion of variance explained by the model:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

where $\bar{y}$ is the mean of actual values y_t. An R^2 value closer to 1 indicates a better fit.

5.2. Model Performances

The table below (Table 2) summarizes the optimal parameters for each model, detailing the chosen settings for ARIMA, Random Forest, and LSTM based on our hyperparameter tuning and cross-validation processes.

Table 2: Optimal Model Parameters for Carbon Price Forecasting Models

Model	**Parameter**	**Optimal Value**
ARIMA	Autoregressive Order	1
	Differencing Order	1
	Moving Average Order	1
Random Forest (Without Differencing)	Number of Trees	150
	Maximum Depth	10
	Minimum Samples Split	5
	Minimum Samples Leaf	2
	Maximum Features	`sqrt`
Random Forest (With Differencing)	Number of Trees	200
	Maximum Depth	10
	Minimum Samples Split	5
	Minimum Samples Leaf	4
	Maximum Features	`sqrt`
LSTM	Number of Layers	2 (50 units each)
	Batch Size	1
	Epochs	50
	Optimizer	`rmsprop`

The ARIMA model, which incorporates exogenous variables such as PCA-transformed stock indices and energy prices, demonstrated strong in-sample and out-of-sample predictive capabilities. The in-sample RMSE, MAPE, and MAE values were low, with a high R^2 of 0.986, while out-of-sample evaluations further indicated the models robustness, achieving an RMSE of 1.46 and an R^2 of 0.967.

The Random Forest (RF) model was evaluated in two configurations: one trained with the original undifferenced data and another with a differenced series to address non-stationarity in the carbon price data. The model trained on the original, undifferenced data showed strong in-sample performance, achieving an R^2 of 0.991, an RMSE of 1.69, and a MAPE of 1.58%. This high in-sample R^2 suggests that the model was able to capture the majority of the variance within the training set. However, the out-of-sample results indicated a drop in performance, with an RMSE of 3.33 and an R^2 of 0.826, pointing to overfitting in the model when applied to unseen data. The elevated out-of-sample error metrics suggest that the undifferenced RF model struggled to generalize well, likely due to the presence of trends or other non-stationary elements in the carbon price series.

In contrast, the differenced Random Forest model demonstrated better generalization across both in-sample and out-of-sample evaluations. In-sample performance remained robust, with an R^2 of 0.984, an RMSE of 2.22, and a MAPE of 2.17%. While these metrics indicate slightly lower performance than the undifferenced model in the training set, the differenced model performed significantly better out-of-sample, with an RMSE of 1.54 and an R^2 of 0.963. The lower out-of-sample errors and higher R^2 for the differenced model suggest that differencing effectively stabilized the model, enhancing its ability to make accurate predictions on new data and reducing overfitting.

The LSTM model, configured with a look-back period of 1 time step and optimized through hyperparameter tuning, displayed competitive results for both in-sample and out-of-sample forecasting. In-sample performance showed a good fit, with an R^2 of 0.950, an RMSE of 3.92, and a MAPE of 4.04%. Out-of-sample results were also strong, with an RMSE of 1.88, MAPE of 1.78%, and an R^2 of 0.944, demonstrating that the LSTM model was able to capture temporal dependencies effectively in the carbon price series.

The performance of the ARIMA model, Random Forest models, and LSTM model for forecasting carbon prices is summarized in Table 3.

Table 3: Performance Summary of ARIMA, Random Forest, and LSTM Models

Model	Metric	In-Sample	Out-of-Sample
ARIMA(1,1,1)	RMSE	2.07	1.46
	MAPE	2.08%	1.53%
	MAE	1.50	1.19
	MSE	4.28	2.13
	R^2	0.986	0.967
Random Forest (No Differencing)	RMSE	1.69	3.33
	MAPE	1.58%	3.60%
	MAE	1.17	2.72
	MSE	2.85	11.11
	R^2	0.991	0.826
Random Forest (With Differencing)	RMSE	2.22	1.54
	MAPE	2.17%	1.55%
	MAE	1.56	1.21
	MSE	4.93	2.37
	R^2	0.984	0.963
LSTM	RMSE	3.92	1.88
	MAPE	4.04%	1.78%
	MAE	2.93	1.40
	MSE	15.33	3.53
	R^2	0.950	0.944

The figures comparing actual and predicted carbon prices for in-sample and out-of-sample periods show that all three models performed well. The ARIMA model closely matched actual carbon prices with only minor deviations in out-of-sample forecasts, which were within acceptable margins. The differenced Random Forest model showed consistent alignment with actual prices across both in-sample and out-of-sample periods, effectively capturing non-stationary characteristics in the time series. The LSTM model also achieved a high level of accuracy, closely aligning with actual values in the test datasets.

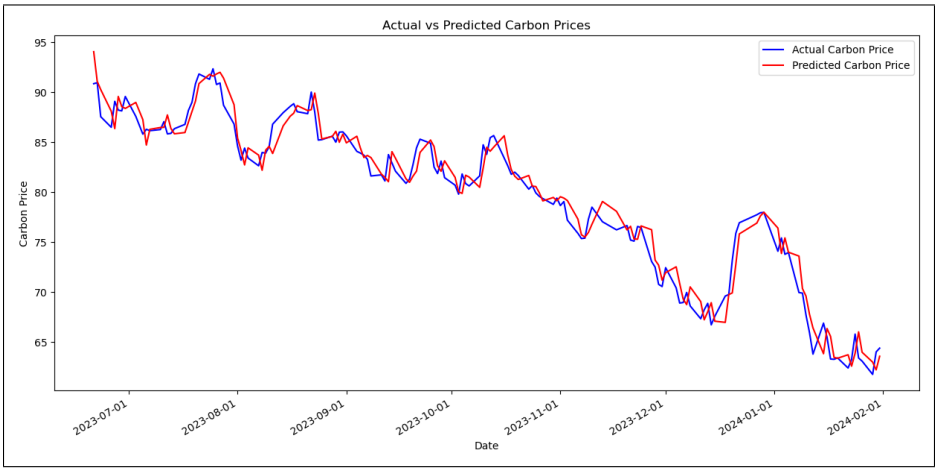

Figure 2: Out-of-sample Actual vs Predicted Carbon Prices using ARIMA(1,1,1).

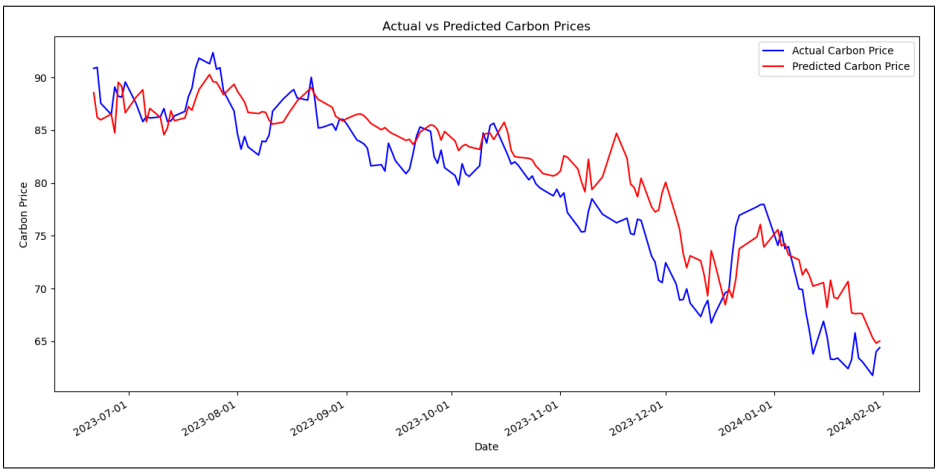

Figure 3: Out-of-sample Actual vs Predicted Carbon Prices using Random Forest (Without Differencing).

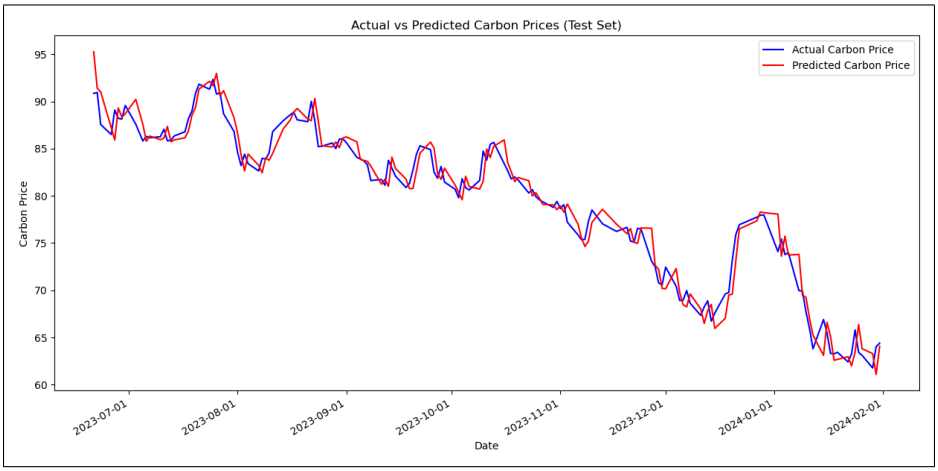

Figure 4: Out-of-sample Actual vs Predicted Carbon Prices using Random Forest (With Differencing).

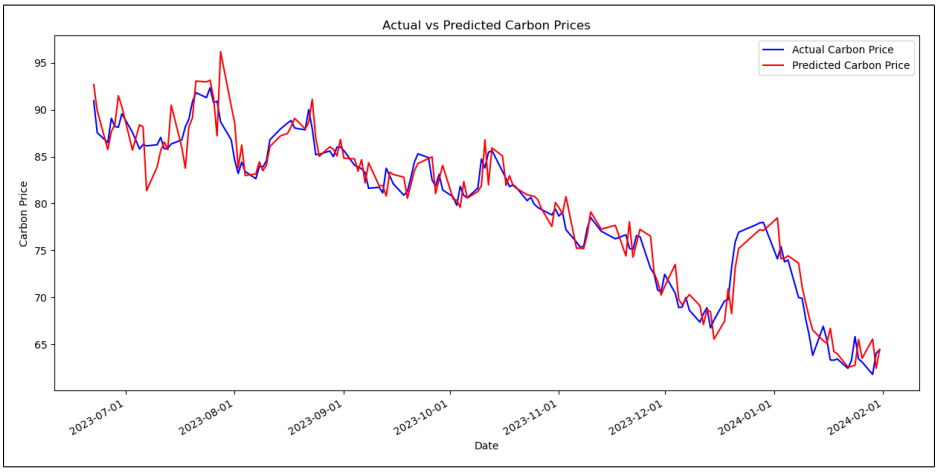

Figure 5: Out-of-sample Actual vs Predicted Carbon Prices using Long Short Term Memory.

5.3. Discussion

In this study, the ARIMA model exhibited the best overall performance in forecasting carbon prices, closely followed by the differenced Random Forest (RF) model. And the LSTM model is generally effective in capturing sequential dependencies.

The data used in this study, spanning from 2021 to 2024, presents specific characteristics that influence model performance. Unlike prior studies that incorporate data from 2005 to 2021encompassing phases 1 through 3 of the carbon marketour

dataset includes only phase 4. This distinction is significant because earlier phases were marked by substantial policy shifts, volatility, and structural changes, making the data from those periods more variable and unpredictable. Phase 4, however, has seen increased regulatory stability, which likely reduces volatility and simplifies the underlying price patterns. This reduced variability may favor simpler models, such as ARIMA, that are well-suited to linear and relatively stable time series data.

Additionally, the datas non-stationary nature presented challenges (Sun 2018). Non-stationarity, often manifested as trends or seasonality, can disrupt models that assume data stability over time. Differencing techniques can partially mitigate non-stationarity by stabilizing the mean, which explains why the differenced RF model achieved better out-of-sample performance than the undifferenced version.

The ARIMA models effectiveness is consistent with findings from previous studies that emphasize its utility for short term forecasting on linear datasets (Kontopoulou 2023). ARIMAs simplicity allows it to capture predictable patterns without overfitting. This models success in capturing carbon price trends is indicative of the relatively smooth and stable structure within the limited dataset.

The RF model, particularly the undifferenced version, performed well in-sample but struggled with generalization. While RFs ensemble structure can capture complex relationships, non-stationarity can introduce instability, leading to overfitting (Grushka 2017). Differencing helped stabilize the RF model, improving its out-of-sample accuracy by removing some of the trend components and allowing the model to focus on the underlying pattern rather than random fluctuations.

LSTM, a deep learning model specifically designed to handle sequential dependencies, did not outperform ARIMA or the differenced RF model, which aligns with studies suggesting that machine learning methods may not always surpass traditional time-series approaches for carbon price forecasting (Fan 2017; Yamak 2019). Hua (2020) found ARIMAs simpler structure often yielded comparable or better short-term predictions than LSTM, particularly in volatile series like Bitcoin. Similarly, Latif et al. (2023) observed that ARIMA could outperform LSTM for short-term forecasts due to its emphasis on recent historical data. Several factors may explain this outcome. First, the LSTMs performance may be limited by the short time horizon (2021-2024) and lack of high-frequency data. LSTMs generally perform best with abundant data to learn long-term dependencies; thus, the limited dataset size may have restricted its ability to capture nuanced patterns effectively.

Another contributing factor is the selection of features. Although LSTM models are powerful, their performance can be improved by incorporating more granular features or alternative data sources, such as high-frequency trading data or additional macroeconomic indicators (Yamak 2019). However, the studys limited set

of features, which focused primarily on macroeconomic and energy factors, might have constrained the model's capacity to leverage its inherent strengths in sequential learning.

6. Conclusion

Summarize the findings, their implications for

References

[1] Adebiyi A Ariyo, Adewumi O Adewumi, and Charles K Ayo. Stock price prediction using the arima model. In *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*, pages 106–112. IEEE, 2014.

[2] Javier Contreras, Rosario Espinola, Francisco J Nogales, and Antonio J Conejo. Arima models to predict next-day electricity prices. *IEEE transactions on power systems*, 18(3):1014–1020, 2003.

[3] John Hua Fan and Neda Todorova. Dynamics of chinas carbon prices in the pilot trading phase. *Applied Energy*, 208:1452–1467, 2017.

[4] Yael Grushka-Cockayne, Victor Richmond R Jose, and Kenneth C Lichtendahl Jr. Ensembles of overfit and overconfident forecasts. *Management Science*, 63(4):1110–1130, 2017.

[5] Yiqing Hua. Bitcoin price prediction using arima and lstm. In *E3S web of conferences*, volume 218, page 01050. EDP Sciences, 2020.

[6] Yumeng Huang, Xingyu Dai, Qunwei Wang, and Dequn Zhou. A hybrid model for carbon price forecasting using garch and long short-term memory network. *Applied Energy*, 285:116485, 2021.

[7] Luckyson Khaidem, Snehanshu Saha, and Sudeepa Roy Dey. Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*, 2016.

[8] Vaia I Kontopoulou, Athanasios D Panagopoulos, Ioannis Kakkos, and George K Matsopoulos. A review of arima vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, 15(8):255, 2023.

[9] Ali Lahouar and J Ben Hadj Slama. Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management*, 103: 1040–1051, 2015.

[10] Navmeen Latif, Joseph Durai Selvam, Manohar Kapse, Vinod Sharma, and Vaishali Mahajan. Comparative performance of lstm and arima for the short-term prediction of bitcoin prices. *Australasian Accounting, Business and Finance Journal*, 17(1):256–276, 2023.

[11] Jules Sadefo Kamdem and James B Njong. Time-frequency analysis and machine learning models for carbon market forecasting. *Annals of Operations Research*, pages 1–20, 2023.

[12] Alex Sherstinsky. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404: 132306, 2020.

[13] Wei Sun and Chongchong Zhang. Analysis and forecasting of the carbon price using multiresolution singular value decomposition and extreme learning machine optimized by adaptive whale optimization algorithm. *Applied energy*, 231: 1354–1371, 2018.

[14] Jujie Wang, Xin Sun, Qian Cheng, and Quan Cui. An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *Science of the Total Environment*, 762:143099, 2021.

[15] Yingying Xu, Yifan Dai, Lingling Guo, and Jingjing Chen. Leveraging machine learning to forecast carbon returns: Factors from energy markets. *Applied Energy*, 357:122515, 2024.

[16] Peter T Yamak, Li Yujian, and Pius K Gadosey. A comparison between arima, lstm, and gru for time series forecasting. In *Proceedings of the 2019 2nd international conference on algorithms, computing and artificial intelligence*, pages 49–55, 2019.

[17] Ping Yang, Yelin Wang, Shunyu Zhao, Zhi Chen, and Youjie Li. A carbon price hybrid forecasting model based on data multi-scale decomposition and machine learning. *Environmental Science and Pollution Research*, 30(2):3252–3269, 2023.

[18] Xi Zhang, Kailing Yang, Qin Lu, Jingyu Wu, Liang Yu, and Yu Lin. Predicting carbon futures prices based on a new hybrid machine learning: Comparative study of carbon prices in different periods. *Journal of Environmental Management*, 346:118962, 2023.

[19] Zheng Zhao, Weihai Chen, Xingming Wu, Peter CY Chen, and Jingmeng Liu. Lstm network: a deep learning approach for short-term traffic forecast. *IET intelligent transport systems*, 11(2):68–75, 2017.

[20] Bangzhu Zhu, Julien Chevallier, Bangzhu Zhu, and Julien Chevallier. Carbon price forecasting with a hybrid arima and least squares support vector machines methodology. *Pricing and forecasting carbon markets: Models and empirical analyses*, pages 87–107, 2017.