Forecasting Carbon Prices: A Comparative Study of ARIMA, Random Forest, and LSTM Models*

Erica Zhao[†‡]

December, 2024

Abstract

Accurate carbon price forecasting within the European Union's Emissions Trading System (EU ETS) is essential for guiding policy and investment decisions in the transition to a low-carbon economy. This study evaluates and compares the performance of ARIMA, Random Forest (RF), and Long Short-Term Memory (LSTM) models in predicting carbon prices during Phase 4 (2021–2024) of the EU ETS. Trained with a comprehensive dataset that incorporates macroeconomic indicators, energy prices, weather conditions, and stock indices, the models are assessed for their ability to capture the dynamics of carbon pricing. ARIMA demonstrates strong performance in modeling linear trends and stable time series, while RF effectively handles non-linear relationships and addresses non-stationarity with differencing techniques. LSTM captures sequential dependencies in the data, leveraging its temporal modeling capacity. The results highlight that all three models achieve robust forecasting accuracy, demonstrating their suitability for the relatively stable regulatory conditions of Phase 4. This study provides practical insights into model selection for carbon price forecasting, offering guidance for stakeholders navigating the complexities of carbon markets and climate policy.

*I am grateful to my honors thesis supervisor, Professor Haiqing Xu, and my second reader, Professor Xinyue Bei, for their invaluable guidance and feedback throughout this research.

[†]Dr. Haiqing Xu. h.xu@austin.utexas.edu

[‡]Dr. Xinyue Bei. xinyue.bei@austin.utexas.edu

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. The EU ETS, a cornerstone of the EU's climate policy, operates as a cap-and-trade system that regulates carbon emissions through market-based mechanisms. Forecasting carbon prices in this context is challenging due to the interplay of diverse factors, including regulatory changes, macroeconomic indicators, energy prices, and weather conditions, which contribute to market dynamics and volatility.

Traditional time series models, such as ARIMA, have been widely applied in carbon price forecasting. These models are valued for their interpretability and ability to capture linear trends and autocorrelations in stable market conditions (Zhu et al. 2017). However, carbon markets are increasingly influenced by nonlinearity and complex interactions, particularly in later phases of the EU ETS. This raises questions about whether traditional models are sufficient for capturing the intricacies of carbon price movements (Huang et al. 2021).

In response to these challenges, machine learning (ML) models, such as Random Forest (RF) and Long Short-Term Memory (LSTM) networks, have been proposed as alternatives due to their ability to manage nonlinear dependencies and sequential patterns in data. RF is particularly noted for its ensemble structure, which can capture complex relationships among predictors, while LSTM leverages its recurrent architecture to model temporal dependencies (Xu et al. 2024; Zhang et al. 2023). Despite these advancements, the relative effectiveness of traditional and ML-based models remains an open question, particularly under the evolving market conditions of the EU ETS.

This study conducts a comparative analysis of ARIMA, RF, and LSTM models to forecast carbon prices within the EU ETS during Phase 4 (2021 to 2024). By applying consistent predictor variables, including economic indicators, energy prices, weather conditions, and regulatory factors, across all models, I ensure a fair evaluation of their predictive capabilities. ARIMA is included for its simplicity and suitability for stationary data, RF for its robustness in modeling nonlinear relationships, and LSTM for its ability to capture sequential dependencies. To assess model performance, I employ time-series cross-validation and rolling forecasting techniques, preserving the temporal structure of the data.

The contributions of this study are twofold. First, it provides an empirical com-

parison of ARIMA, RF, and LSTM models under the same market conditions, offering insights into their relative strengths and limitations in forecasting carbon prices. Second, it highlights the influence of market characteristics, such as regulatory stability and reduced volatility in Phase 4, on model performance. Unlike prior studies, which often emphasize the superiority of one approach, this study demonstrates that all three models perform well when appropriately calibrated to the dataset. These findings provide practical guidance for policymakers and investors in selecting suitable forecasting methods for 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 model's 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 dependencies—a 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 ARIMA's 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 evaluates the performance of three models: ARIMA, Random Forest, and LSTM, in forecasting carbon prices. Similar exogenous variables have been utilized to train the three models: energy prices (Carbon, Coal, Crude Oil, and Natural Gas), macroeconomic indicators (Fit_for_55, Law, Compromise, and the Economic Policy Uncertainty (EPU) index), stock indices (DAX, SPGTCLNT, FTMIB, CAC40, AEX, and STOXX), and environmental factors (Temperature, Precipitation, Solar Energy, and Wind Speed). The models are trained and tested using time-series data split into 80% training and 20% testing sets.

To evaluate the accuracy and reliability of the predictive models in forecasting carbon prices, the evaluation criteria include Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Determination (R^2):

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

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

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

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

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

where y_t represents the observed values, $\hat{y}_t$ the predicted values, $\bar{y}$ the mean of observed values, and n the number of observations.

3.1 ARIMA Model

The ARIMA model is frequently applied in carbon and energy price forecasting due to its capability to capture various types of patterns in time series data. Studies demonstrate ARIMA's effectiveness in emissions and financial markets, particularly in relatively stable conditions (Contreras 2003; Ariyo 2014). This study uses 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 exogenous variables, respectively. X_{t-k} denotes the exogenous variables, and ε_t represents the error term.

I conduct 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 choose parameters p, d, and q. I then estimate the optimal parameters, including ϕ_i, θ_j, and β_k, using maximum likelihood estimation (MLE). This process 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 the variance of the error term.

To address the issue of multicollinearity, I apply Principal Component Analysis (PCA) to two feature groups: stock indices (DAX, STOXX, FTMIB, CAC40, AEX, SPGTCLNT) and energy prices (Crude Oil, Coal, and Natural Gas). This dimensionality reduction condenses these variables into two principal components, $PCA_{\text{Stock\_Index}}$ and $PCA_{\text{Energy\_Prices}}$, which I include as exogenous variables in the ARIMA model.

3.2 Random Forest

Unlike linear models, Random Forest does not assume a linear relationship between input features and target variables, enabling it to capture complex, non-linear relationships in time series forecasting. Studies, including Khaidem et al. (2016) and Lahouar and Slama (2015), demonstrate RF's 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 allows individual instances to repeat or be excluded in subsets.

The RF algorithm constructs a sequence of decision trees $(T_1, T_2, \ldots, T_N)$ during training. Each tree T_i is trained on a bootstrap sample D_i, drawn with replacement from the full training dataset D. At each node, a random subset of m features is selected for splitting. For a new input x, each tree T_i produces a prediction y_i, and the RF's final prediction $\hat{y}(x)$ is calculated as the average of these predictions:

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

Within each tree, prediction error is minimized using the Mean Squared Error (MSE) objective:

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

where n_t is the number of instances in node t, and k_t is the average value of instances

within the node.

This study employs two variations of the Random Forest model: one using the original undifferenced carbon price data and another using a differenced series to address non-stationarity. Differencing is performed as:

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

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. Differencing removes trends and other non-stationary components, while exogenous variables remain undifferenced.

To optimize the Random Forest model, I tune hyperparameters including the number of trees, maximum tree depth, minimum samples required to split a node, minimum samples required at a leaf node, and the number of features considered at each split. Hyperparameter tuning is performed via grid search with Mean Squared Error (MSE) as the objective.

In addition to the standard RF model, I explore a modified version where the final prediction is the median of individual tree predictions, rather than the mean. This adjustment helps improve robustness against outliers.

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, including 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, I perform hyperparameter tuning using grid search, optimizing batch size, epochs, optimizer, and learning rate. This search yield an optimal multi-layer LSTM configuration with dropout layers for regularization and the `RMSprop` optimizer, contingent on the chosen learning rate.

For model training, I 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, I train the LSTM model to update its weights based on the sequential structure of the data.

4 Descriptive Statistics

This study analyzes carbon pricing trends in the European Union's Emissions Trading System (EU ETS) using a dataset spanning January 1, 2021, to January 31, 2024. The dataset comprises 763 observations of carbon prices sourced from Investing.com. This timeframe aligns with the start of Phase 4 of the EU ETS, a period marked by significant regulatory changes.

Phase 4 (2021 to 2030) of the EU ETS introduces stricter emissions reduction measures compared to Phase 3 (2013 to 2020). While Phase 3 applied a linear reduction factor of 1.74% to the annual cap on emissions, Phase 4 increases this factor to 2.2%, reflecting the EU's heightened climate ambitions. These adjustments necessitate a focused analysis of their impact on carbon price trends and market behavior.

Descriptive statistics for the dataset are presented in Table 1. The carbon price exhibits a mean of €72.98 and a standard deviation of €16.45.

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

Figure 1 illustrates the carbon price trends over the study period, capturing fluctuations influenced by regulatory adjustments and market responses. The analysis also incorporates related variables, including energy commodities, policy indicators, European stock indices, and weather conditions, to explore their potential effects on emissions and carbon market dynamics.

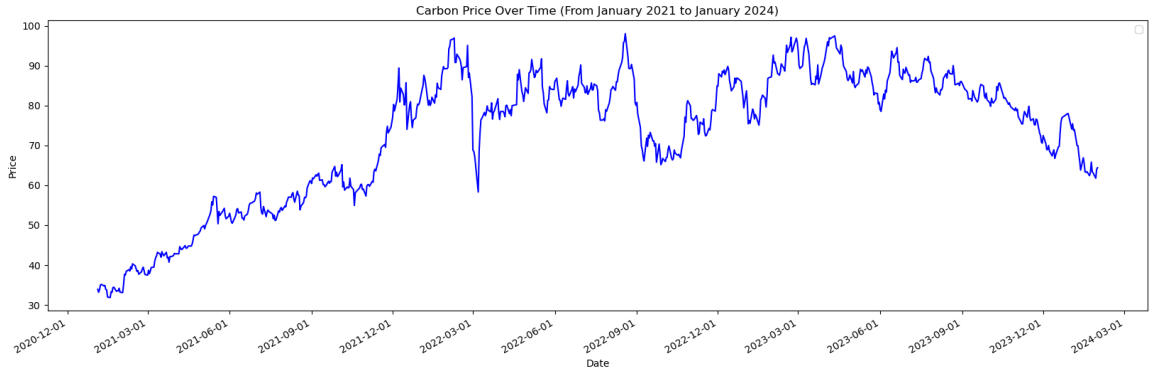

Figure 1. Time Series Plot of Carbon Price from January 2021 to January 2024.

5 Performance and Results

5.1 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 my 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
RF (Without Differencing)	Number of Trees	150
	Maximum Depth	10
	Minimum Samples Split	5
	Minimum Samples Leaf	2
	Maximum Features	`sqrt`
RF (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, incorporating exogenous variables such as PCA-transformed stock indices and energy prices, demonstrates strong predictive capabilities both in-sample and out-of-sample. In-sample evaluation shows low RMSE, MAPE, and MAE values, with a high R^2 of 0.986. Out-of-sample performance further highlights the model's robustness, achieving an RMSE of 1.46 and an R^2 of 0.967.

The Random Forest (RF) model is evaluated in two configurations: one using the original undifferenced data and another employing a differenced series to address non-stationarity in carbon price data. The undifferenced model exhibits strong in-sample performance, achieving an R^2 of 0.991, an RMSE of 1.69, and a MAPE of 1.58%. These metrics indicate that the model captures most of the variance within the training set. However, out-of-sample performance shows a decline, with an RMSE of 3.33 and an R^2 of 0.826, suggesting overfitting and difficulty in generalizing to unseen

data. The elevated out-of-sample errors likely result from trends or non-stationary elements in the carbon price series.

In contrast, the differenced Random Forest model achieves better generalization across in-sample and out-of-sample evaluations. In-sample results remain robust, with an R^2 of 0.984, an RMSE of 2.22, and a MAPE of 2.17%. Although slightly lower than the undifferenced model in the training set, the differenced model performs significantly better out-of-sample, achieving an RMSE of 1.54 and an R^2 of 0.963. The improved out-of-sample performance highlights the effectiveness of differencing in stabilizing the model, enhancing its predictive accuracy, and reducing overfitting.

The LSTM model delivers competitive results in both in-sample and out-of-sample forecasting. In-sample evaluation shows a strong fit, with an R^2 of 0.950, an RMSE of 3.92, and a MAPE of 4.04%. Out-of-sample results further validate the model's capability to capture temporal dependencies, achieving an RMSE of 1.88, a MAPE of 1.78%, and an R^2 of 0.944.

Table 3 summarizes the performance of the ARIMA model, Random Forest models, and LSTM model in forecasting carbon prices.

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
RF (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
RF (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.

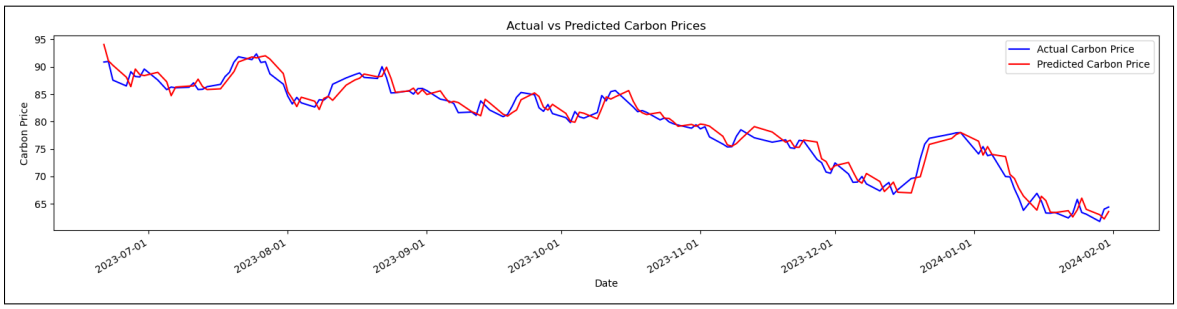

Figure 2. Out-of-sample Actual vs Predicted Carbon Prices using ARIMA(1,1,1).

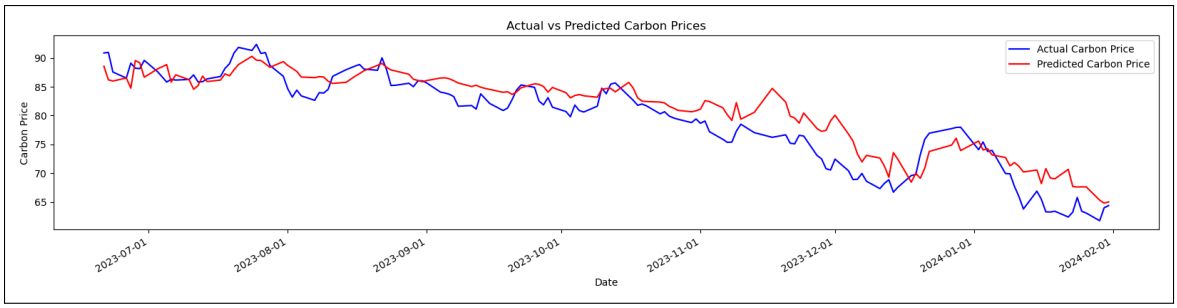

Figure 3. Out-of-sample Actual vs Predicted Carbon Prices using Random Forest.

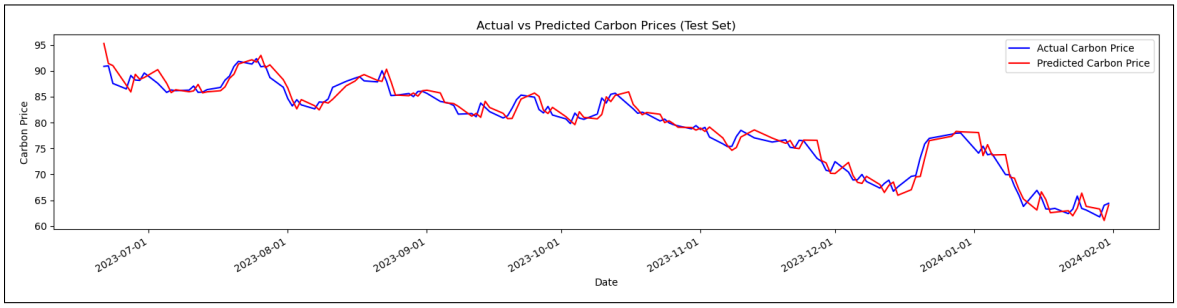

Figure 4. Out-of-sample Actual vs Predicted Carbon Prices using Random Forest (Differencing).

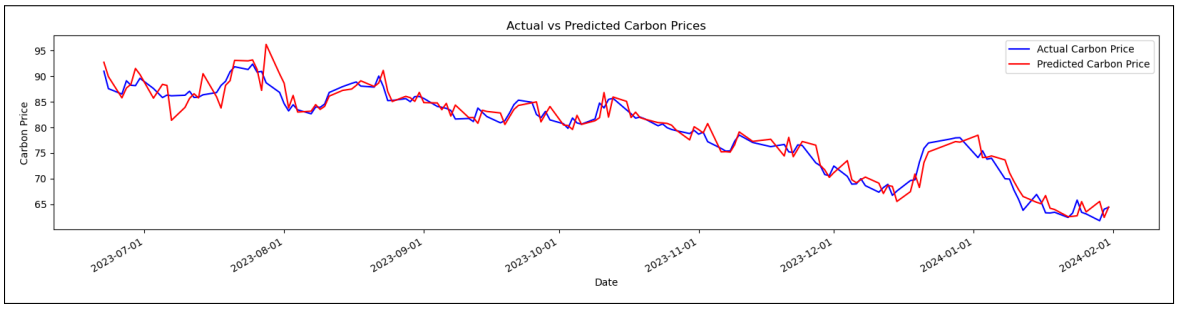

Figure 5. Out-of-sample Actual vs Predicted Carbon Prices using Long Short Term Memory.

5.2 Discussion

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 2021, encompassing Phases 1 through 3 of the carbon market, my 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 models like ARIMA, which are well-suited to linear and relatively stable time series data.

The ARIMA model's performance aligns with findings from previous studies, which highlight its utility for short-term forecasting on linear datasets (Kontopoulou 2023). Its simplicity allows it to capture predictable patterns without overfitting. In this study, the ARIMA model successfully captures carbon price trends, benefiting from the relatively smooth and stable structure of the Phase 4 dataset.

The RF model demonstrates its capability to handle complex relationships, benefiting from its ensemble structure. But the data's 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 likely contributes to the effective performance of the differenced Random Forest (RF) model. This stabilization enables the RF model to focus on underlying patterns while avoiding overfitting caused by trends or random fluctuations.

The LSTM model leverages its design for capturing sequential dependencies and performs well on the dataset. However, its potential may be constrained by the

dataset's relatively short time span and limited frequency. While LSTMs are powerful tools for learning temporal relationships, their performance can improve with larger datasets or more granular features, as noted in previous studies (Yamak 2019; Hua 2020). Despite these limitations, the LSTM model effectively captures the temporal dynamics present in the carbon price series.

Overall, the dataset characteristics, such as regulatory stability in Phase 4 and the presence of non-stationarity, shape the strengths and challenges for each model. But all three models: ARIMA, Random Forest (RF), and LSTM exhibit strong predictive performance, reflecting their ability to adapt to the characteristics of the dataset.

6 Conclusion

This study investigates the application of ARIMA, Random Forest (RF), and LSTM models for forecasting carbon prices within the European Union's Emissions Trading System (EU ETS) during Phase 4 (2021 to 2024). Leveraging a dataset that includes macroeconomic indicators, energy prices, weather conditions, and stock indices, the analysis demonstrates that all three models: ARIMA, RF, and LSTM exhibit strong predictive performance under the EU's current regulatory framework. Each model provides unique strengths that align with specific characteristics of the dataset, reflecting the importance of model selection based on data properties.

The ARIMA model effectively captures linear and stable price patterns, benefiting from the relatively smooth and regulated nature of Phase 4 carbon markets. Its simplicity and interpretability make it well-suited for datasets with reduced volatility. The RF model, especially with differencing, proves robust in managing non-linear relationships and handling non-stationary time series components. Meanwhile, the LSTM model showcases its strength in capturing sequential dependencies, although its performance could benefit from larger datasets or higher-frequency data to better exploit its potential.

Despite the robust results, this study has limitations that present opportunities for future research. The analysis focuses on one-step-ahead forecasts (y_{t+1}), which may inherently favor simpler models like ARIMA, given the rich information set available. Extending the prediction horizon to multi-step forecasts, such as y_{t+7} (weekly) or y_{t+30} (monthly), could reveal more significant performance distinctions between statistical and machine learning models. Longer-term predictions often amplify the complexity of the underlying relationships, providing a clearer test of each model's ability to

generalize and adapt.

Additionally, the dataset is limited to Phase 4 of the EU ETS, which benefits from enhanced regulatory stability. While this ensures consistency in the analysis, it may not capture the volatility and structural changes observed in earlier phases. Incorporating data from these periods or exploring high-frequency trading data could offer deeper insights into the adaptability of each model. Future research could also consider hybrid models that combine statistical and machine learning techniques to balance interpretability with flexibility, further enhancing forecasting accuracy.

By highlighting the relative strengths of ARIMA, RF, and LSTM models within a stable regulatory context, this study provides valuable insights into the trade-offs involved in carbon price forecasting. These findings underscore the need for continued refinement of predictive methods to support policy and investment decisions in the evolving landscape of carbon markets.

References

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.

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.

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

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.

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

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.

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.

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.

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

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

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.

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.

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.

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.

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.

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.

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.