Forecasting Ghana Cocoa Prices: A Comparative Analysis of Time Series and Machine Learning Models

Author: Haochen Ding, Carrie Su, Yuxin Fan, Yixuan Gu

2025/04/04

1. Introduction

It is well known that Ghana is one of the world's largest cocoa producers, with approximately 60% of the world's cocoa grown in Ghana and Côte d'Ivoire (Nicholas Robinson, 2025). Cocoa is a crucial export commodity for Ghana, generating an estimated $5.13 billion in cocoa bean exports (OEC, 2023), and employing around 17% of the working population. It contributes an average of 3.5% to the national GDP and accounts for approximately 30% of total export earnings (Matteo Guedia, 2022). However, the cocoa sector faces growing threats. In 2024, the outbreak of the cacao swollen shoot disease devastated nearly 500,000 hectares of farmland (Olatungi Olaigbe, 2025), while climate change further reduced production levels. These disruptions led to a sharp increase in global cocoa prices, emphasizing Ghana's vulnerability to environmental and market shocks. In this context, forecasting cocoa prices is not only vital for the financial well-being of Ghanaian farmers but also crucial for maintaining a stable and efficient supply chain from production to international markets.

This study focuses on developing reliable forecasting methods for Ghanaian cocoa prices, using both historical data and key influencing factors such as global currency exchange rates and production volumes. By studying and quantifying the factors behind cocoa price fluctuations, the research aims to support effective policy formulation and improve income stability for farmers. The objective of this research is to develop a robust cocoa price forecasting methodology for Ghana. Three forecasting models were constructed which include two time series models: Autoregressive Integrated Moving Average (ARIMA) and Seasonal-Trend decomposition using Loess combined with Seasonal Autoregressive Integrated Moving Average (STL+ARIMA) in the initial phase, followed by a machine learning approach using Long Short-Term Memory (LSTM) networks to incorporate more complex and dynamic sustainability factors to test whether these factors actually influence price forecast. The models are then evaluated based on their predictive accuracy and robustness.

Despite its importance, cocoa price forecasting in Ghana presents several challenges. First, data cleaning was a major concern, particularly in handling missing values in climate data, which are critical for capturing environmental influences on cocoa production. Second, data collection was hindered by inconsistencies in variable formats, time coverage, and limited access to real-time market data for Ghana. These challenges highlight the need for not only methodological precision but also model flexibility and adaptability in the face of complex and unpredictable market behavior.

Preliminary results indicate that the LSTM model outperformed both ARIMA and STL+ ARIMA in capturing the sharp cocoa price increase in 2024, achieving higher accuracy and stronger predictive performance. This suggests that integrating advanced machine learning methods and external variables offers significant advantages in modeling volatile commodity prices, thus contributing to more informed decision-making and economic resilience in Ghana’s cocoa sector.

2. Literature Review

The relationship between cocoa production, exchange rate, and climate conditions has been demonstrated in several studies. For instance, cocoa variety, climate conditions, and soil quality all affect the quality and volume of cocoa produced (Bermudez, Voora, Larrea, & Luna, 2022). These factors contribute to yield variability, which in turn influences price trends. Oluwasegun et al. (2021) examined the impact of price volatility on cocoa supply in Ghana and Nigeria and found that in Ghana, exchange rate fluctuations and international price volatility have a long-run negative effect on producer prices.

Furthermore, Assis et al. (2010) evaluated a range of univariate and multivariate time series models, including mixed ARIMA/GARCH, GARCH, ARMA, and MARMA, and concluded that hybrid ARIMA-GARCH models offer superior performance in modeling cocoa price volatility. Building on their approach, our study incorporates ARIMA as the baseline, which models the overall trend.

The use of STL combined with ARIMA has also been explored in studies forecasting the sales volume of perishable agricultural goods. One such study applied STL decomposition to isolate seasonal patterns before modeling demand with ARIMA, resulting in improved prediction accuracy and enhanced pricing insights (Lem, K. H, 2024). Inspired by this, we apply STL+ARIMA as one of our modeling approaches to better capture seasonal fluctuations in cocoa prices, as ARIMA does not capture the seasonal effect.

Additionally, Olofintuyiet al. (2023) employed a CNN-RNN-LSTM ensemble model to predict cocoa yield in Nigeria using both climate and yield datasets. Their study showed that LSTM significantly outperformed conventional models in terms of MAE and RMSE. While their focus was on yield prediction, their findings support the use of LSTM in price forecasting. Our study extends this work by using LSTM to capture the interaction between Ghanaian cocoa prices, exchange rate movements, production, and climate conditions, offering a more flexible and responsive approach to modeling price shocks such as the 2024 surge.

Moreover, Kontopoulou et al. (2023) provide an extensive review of time series forecasting methods, comparing traditional statistical models like ARIMA with machine learning approaches. They suggest that ARIMA models are more suitable for univariate data due to their mathematical simplicity and flexibility in application, whereas machine learning methods are better equipped to capture complex and non-linear relationships in multivariate data. Based on this insight, our research applies univariate data to time series models and multivariate data to machine learning models to evaluate whether external factors significantly improve forecasting accuracy.

In general, the existing literature highlights the multidimensional nature of cocoa price dynamics and supports the inclusion of climate, exchange rate, and production variables in forecasting models. By applying both time series methods (ARIMA, STL+ARIMA) and LSTM, our study not only builds on past research but also addresses its limitations. This combined modeling strategy ensures our ability to forecast cocoa prices under uncertain market conditions in Ghana.

3. Methodology

This project applies three progressively advanced models to forecast Ghana cocoa futures prices, leveraging time series techniques, including the ARIMA and STL+ARIMA models, and machine learning approaches. The data was split chronologically, with the training set consisting of observations from January 1996 to December 2023 and the testing set comprising data from January 2024 to November 2024. A logarithmic transformation was applied to stabilize variance and address the right-skewness distribution in the raw price series.

All three models were trained on the historical data and used to generate forecasts for cocoa prices in 2024. By comparing the forecasted values with the actual observations in the test set, the predictive performance of each model was assessed, allowing for a data-driven evaluation of model suitability.

The choice of models reflects a balance between traditional time series forecasting and exploratory machine learning approaches. The univariate ARIMA and STL+ARIMA models were selected according to Kontopoulou et al. (2023), which suggests that when the relationship between additional predictors and the target variable is unclear or weak, it is often more effective to avoid including them in time series models to prevent unnecessary complexity and overfitting. In contrast, machine learning models were employed to explore the potential predictive value of external variables and uncover nonlinear patterns that traditional models may overlook. This approach allows for both rigorous time series forecasting and flexible model discovery, enabling a comprehensive evaluation of forecasting performance.

3.1 ARIMA

The ARIMA model was selected as the baseline due to its effectiveness in modeling univariate time series with clear trends. While SARIMA extends ARIMA to account for seasonality, it is not applicable since the ACF and PACF plots of the raw data did not reveal any significant seasonal patterns. The model is denoted as ARIMA(p, d, q), where p, d, and q represent the autoregressive, differencing, and moving average orders. The formula is:

$$\phi(B)(1 - B)^{d} x_t = \theta(B)\omega_t$$

Before applying the ARIMA model, stationarity was assessed. The autocorrelation function (ACF) of the log-transformed cocoa price series exhibited a slow decay (see appendix Figure 1), indicating strong autocorrelation and suggesting that the series was non-stationary. First-order differencing was applied to address this. After differencing, the ACF no longer showed a slow tail-off, supporting the conclusion that the series had become stationary. To identify the appropriate orders for the model, the ACF and partial autocorrelation function (PACF) of the differenced series were examined (see appendix Figure 1). Both ACF and PACF showed sharp cutoffs after lag 1, falling behind the 95% confidence intervals, suggesting a moving average term of order 1 and an autoregressive term of order 1, respectively. Since we are monthly data, all the seasonality we assume to be 12 observations as a period. Examining the ACF and PACF plots with the 12-period lags. No significant seasonal patterns were observed. These observations support the selection of an ARIMA (1,1,1) structure for the ARIMA model, with no seasonal component.

The ARIMA (1,1,1) model was fitted using training data from 1996 to 2023 and used to forecast cocoa prices for the year 2024. Then, compare fitted and forecasted values with actual data. We evaluate a model's fitting performances by calculating R-squared, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). A short burn-in period (first 14 months) was excluded when evaluating the model fit to avoid instability from initial predictions made with limited historical data. Models' predictive performances are evaluated by 95% confidence interval coverage, RMSE, MAE, and MAPE. All evaluations were performed on the original price scale by exponentiating the log-transformed values for interpretability.

3.2 STL + ARIMA

Seasonal patterns in the price data might still exist in a more subtle form, hard to detect by examining ACF and PACF plots. To account for this possibility and try to improve the forecasting performance, the STL + ARIMA approach was adopted. By decomposing the series into seasonal, trend, and residual components using the LOESS algorithm and fitting separate SARIMA models to each simpler series, this method allows for greater modeling flexibility and improved interpretability.

After applying STL decomposition, the seasonal component exhibited strong cyclical behavior. While we used ACF and PACF plots to manually select the ARIMA/SARIMA parameters, this approach was not feasible for this seasonal component obtained through STL decomposition. Even after applying multiple levels of differencing (up to third-order), the seasonal component remained non-stationary or highly irregular, making it difficult to interpret autocorrelation patterns reliably. As a result, we adopted a data-driven approach using Python's auto_arima function, which performs automated hyperparameter selection through stepwise search and statistical criteria such as the Akaike Information Criterion (AIC). Found that SARIMA (5, 0, 2) (2, 0, 2) [12] fit the data the most.

For the trend component, since the cyclical patterns had already been captured by the seasonal component, a non-seasonal ARIMA model was sufficient to model the remaining long-term trend. As with the earlier ARIMA modeling procedure, stationarity was assessed first. The original trend series was found to be non-stationary; after 2nd order differencing, the series appeared stationary. The ACF plot of the differenced series shows cut off at lag 5, and PACF cut off at lag 1 (see appendix Figure 3). Indicate an ARIMA(1, 2, 5) model was appropriate to fit this trend component.

For the residual component, we find it did not resemble white noise as expected. Instead of being purely random, the residuals exhibited clear autocorrelation and signs of seasonality. This might be due to the LOESS algorithm used in STL, but not the focus of this research. Due to this, we still model and add the residual component to the final result. The ACF shows stationarity, and both ACF and PACF showed a sharp drop at lag 2, indicating non-seasonal MA(2) and AR(2) (see appendix Figure 6). For lags with period 12, ACF and PACF showed a sharp drop at lag 1, indicating seasonal MA(1) and AR(1) components. As a result, a SARIMA (2, 0, 2) (1, 0, 1) [12] model was fitted for the residual component. The final forecasts were generated by summing up each component's forecast. The Model evaluation process is the same as before.

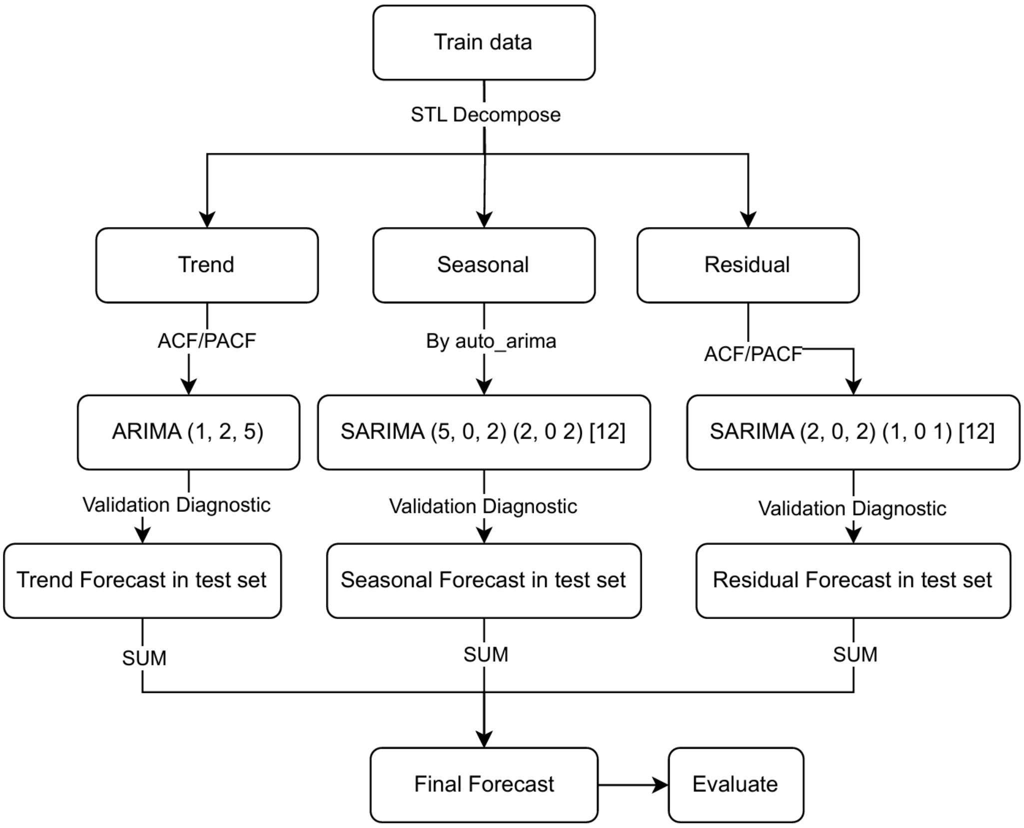

Figure 3.2.1: Flow visualization of STL+ARIMA model

3.3 LSTM

To account for the effect of exogenous factors on Ghana cocoa prices, additional variables were incorporated into the forecasting framework, including climate-related features such as average temperature and precipitation, as well as average exchange rates and average amount of production. Long Short-Term Memory (LSTM), an artificial neural network designed for long time series, was employed to integrate these variables and capture potential nonlinear relationships, thereby enhancing prediction accuracy. Unlike regular neural networks, which struggle with remembering long-term patterns, LSTM is specifically built to learn from past information over long periods by replacing the ordinary recurrent nodes with memory cells to keep important information for a long time. Thus, it is an ideal ML method for modeling the cocoa price trend because it can capture the long-term dependencies in the data.

LSTM as an ML approach does not require differencing the time series until stationary because it can ultimately learn the pattern from the original data by the structure of the neural network. In exchange for this convenience, it needs fine-tuning on the hyperparameters to achieve the best model performance. After grid-searching and evaluating the model performance, the best set of hyperparameter combinations was chosen and introduced in the table below:

Sequence length	This controls the length of history the model can see, so the model can learn long-term patterns from the data.	Choice: 30
Number of hidden units	This controls the memory capacity of the model. The more units the model has, the more past information it can remember.	Choice: 128
Number of layers	This controls the depth of the neural network. More layers allow the model to learn complex features but will take more training time.	Choice: 4
Dropout rate	This prevents the model from overfitting by randomly blocking some nodes during training, improving the generalizability of the model.	Choice: 0.015
Learning rate	This controls the speed of model learning patterns and directly affects the model's performance.	Choice: 0.001
Number of epochs	This controls the span of model learning. The model will underfit if learning too few epochs and overfit after too many.	Choice: 340

Table 3.3.1: Brief Introduction of All Hyperparameters

Before training the LSTM model on our dataset, it is essential to preprocess the data by normalizing and transforming. First of all, the dataset was clipped into training data and test data and then normalized by a scalar. Predictors were transformed into the space of standard normal distribution, and the response (i.e., cocoa price) was log-transformed and projected onto the range of [0, 1]. Normalization is important because this helps the model to handle different scales and units and prevent large fluctuations from dominating the learning updates. After that, the training and test data were wrapped into sequences where the length of sequences is a hyperparameter, and it is ready for training in the LSTM model.

After training, the predictions were transformed back to the natural scale to compare with the original data, and the model constructed a 95% confidence interval to account for variability. Finally, variable importance will be calculated for every predictor to see the relationship between predictors and the price response. The importance was measured by the increase of RMSE after randomly dropping a predictor in the model. If the RMSE increases dramatically, it implies the predictor blocked is important to the model.

4. Data

This study utilizes multiple datasets to build a forecasting model for monthly cocoa prices in Ghana. The datasets include cocoa futures prices, climate data, exchange rates, and cocoa production volumes. To ensure temporal consistency, all datasets were aggregated to monthly frequency.

4.1 Cocoa Future Price Data

Cocoa futures price data was obtained from the International Cocoa Organization (ICCO). The dataset contains daily closing prices for cocoa futures contracts traded on major global commodity exchanges, spanning the period from March 10, 1994, to February 27, 2025. This is the main variable for doing the forecast. For analysis, daily prices were aggregated to monthly averages called Monthly_Price. This transformation smooths short-term market fluctuations and aligns with the frequency of other variables in the study.

4.2 Climate Data

Daily climate data for Ghana was sourced from the National Centers for Environmental Information (NCEI). The dataset includes observations from multiple weather stations and contains variables such as:

- PRCP: Daily precipitation (mm)
- TAVG: Daily average temperature, observed at 2 meters above ground (°F)
- TMAX/TMIN: Maximum and minimum daily temperatures

To prepare this dataset for analysis, daily records were aggregated into monthly averages. Missing precipitation values (NA) were treated as no rainfall and replaced with 0. Rows with missing temperature values were removed. The temperature per day is the average of the temperature that each station observed. Both precipitation and temperature daily data are converted into monthly averages called Monthly_PRCP and Monthly_TAVG for analysis. In this study, only the monthly average temperature and precipitation were used. Since the data was aggregated from daily to monthly frequency, average temperature provides a more stable and representative measure of overall temperature trends, reducing redundancy and potential multicollinearity with max and min temperature. Climate variables are essential to this study as they directly influence cocoa yield and production quality.

4.3 Exchange Rate Data

Monthly average exchange rates between the Ghanaian cedi and the U.S. dollar were obtained from the World Bank Group. The variable Monthly_MidRate represents the monthly average mid-market exchange rate based on the new cedi. Ghana redenominated its currency in July 2007, changing from the old cedi (GHC) to the Ghanaian cedi (GHS) at a rate of 1 GHS = 10,000 GHC. To ensure consistency throughout the time series, all exchange rate values before the redenomination were divided by 10,000. This standardization allows for a continuous and comparable USD/GHS exchange rate series across the dataset. This economic indicator is included as a potential external driver of cocoa price fluctuations, particularly in relation to export market dynamics.

4.4 Cocoa Production Data

Annual cocoa production data was retrieved from Our World in Data. Since the dataset is available only at an annual level, values were converted to monthly estimates by dividing each yearly figure by 12 and named this variable Monthly_Production. Although this introduces some approximation, it enables integration with the monthly climate, price, and exchange rate data.

4.5 Data integration and Summary

All data sets were merged on a common YearMonth variable, taking the intersection of the available time ranges. The resulting dataset spans from January 1996 to November 2024, covering nearly 29 years of monthly observations. The monthly frequency was chosen to ensure the datasets' compatibility and reduce noise inherent in daily observations.

A summary of monthly cocoa prices shows a wide range, from $797.8 to $9,876.1 per metric ton (Appendix Table 1). As illustrated in Figure 4.5.1, cocoa prices remained relatively stable until around 2023, after which a sharp upward trend occurred, peaking in late 2024. The distribution of monthly prices, also shown in Figure 4.5.1, is right-skewed, suggesting that log transformation may be necessary to normalize the data during further modeling.

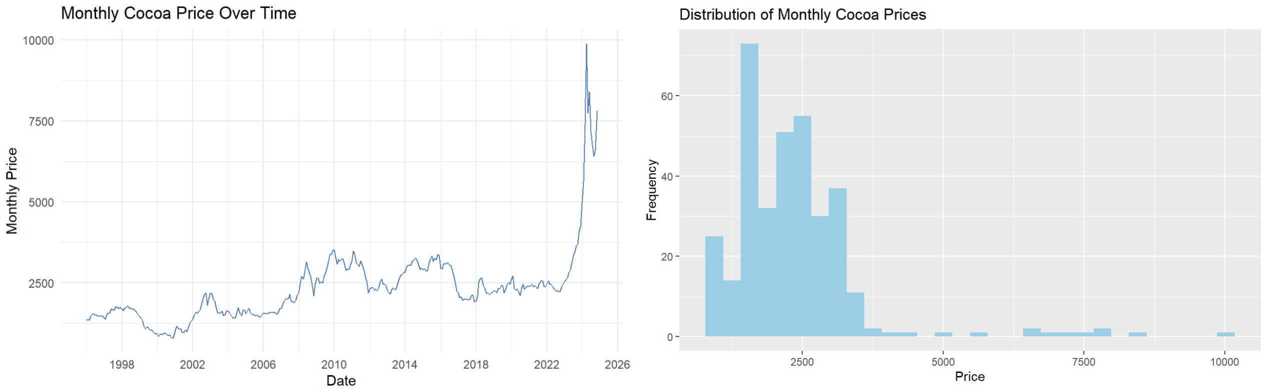

Figure 4.5.1: Trend and Distribution of Monthly Cocoa Price Over Time

The time series plots below illustrate monthly trends for key variables used in this research: exchange rate, precipitation, average temperature, and cocoa production. The exchange rate shows a long-term upward trend, particularly accelerating after 2022. similar to the monthly price. Monthly precipitation appears highly variable with occasional extreme spikes, suggesting irregular rainfall patterns. The monthly average temperature exhibits a clear seasonal cycle with consistent annual fluctuations over time. A full summary of the merged dataset is shown in Appendix Table 1, showing the variation of each predictor.

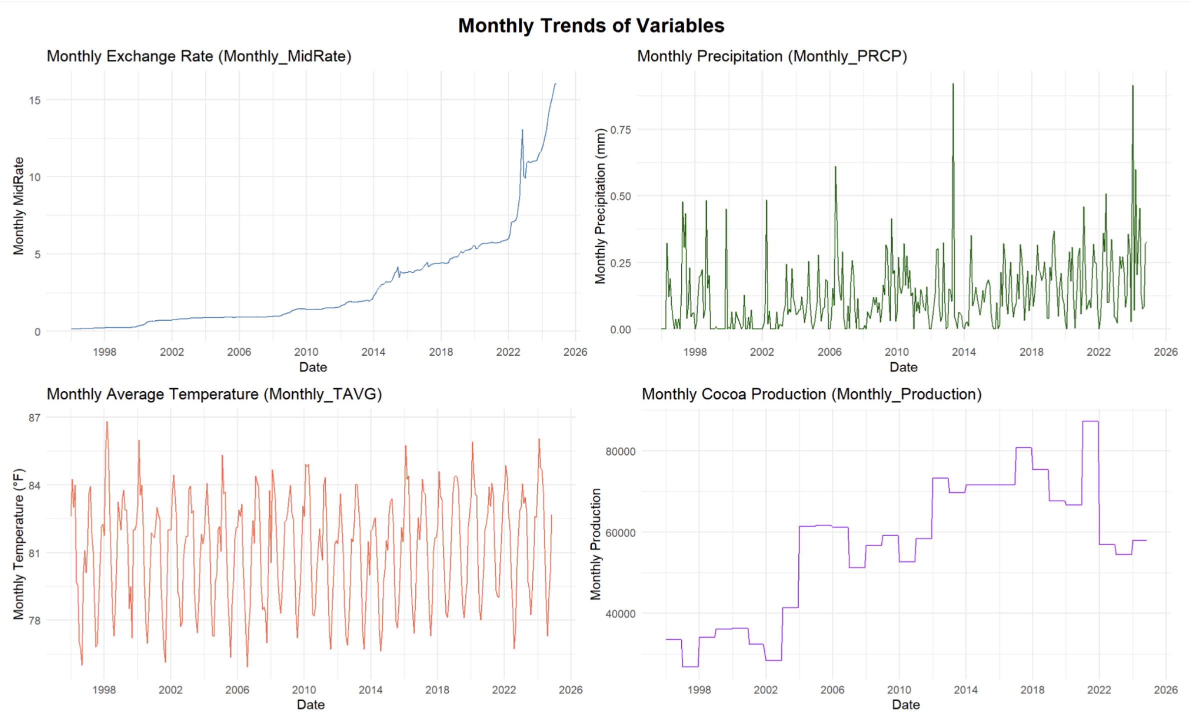

Figure 4.5.2: Monthly Trend of Variables

5. Forecast and Result

5.1 Baseline ARIMA model

From Appendix Figure 2, the diagnostic plots of the fitted ARIMA model suggest that the residuals behave approximately like white noise. The standardized residuals appear randomly scattered with no clear structure, indicating no autocorrelation left in the residuals. The histogram and Q-Q plot show that the residuals are roughly normally distributed, though with slight deviations in the tails. The ACF of residuals (correlogram) reveals no significant autocorrelation at any lag. Additionally, the Ljung-Box test p-values are mostly above the 0.05 threshold, supporting the null hypothesis that residuals are independently distributed. Together, these results indicate that the model is valid.

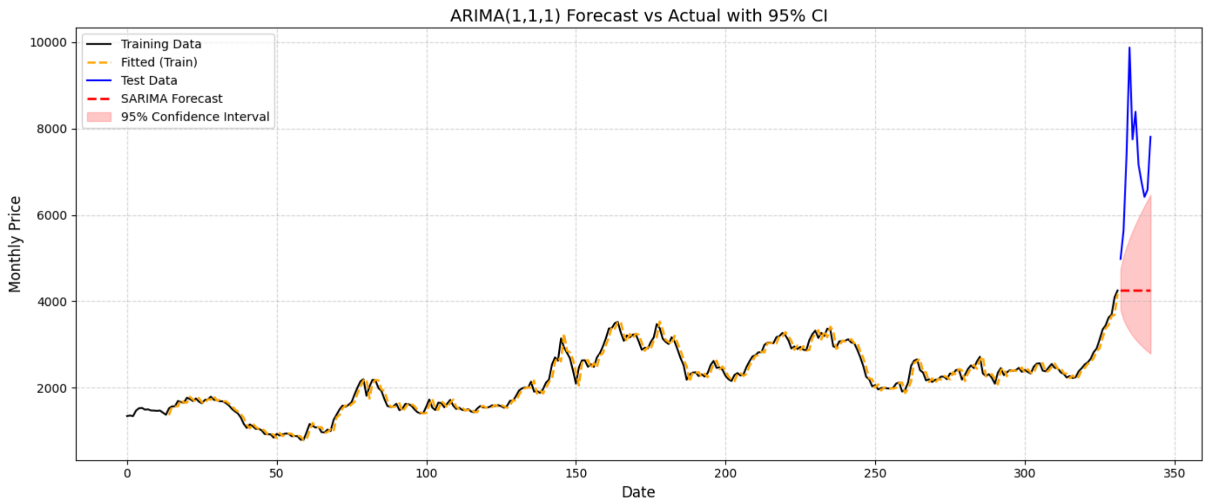

Figure 5.1.1: Forecast and Actual Trend of Price With 95% of CI form ARIMA Model

The ARIMA(1,1,1) model provides a close fit to the training data, with fitted values closely tracking the actual values throughout the training period. However, the model's performance on the test data is limited. The forecasted values remain relatively flat and fail to capture the sharp upward movements observed in the actual test period. This underestimation suggests that the base ARIMA model may not adequately reflect recent structural shifts or potential seasonality in the series. The confidence interval is wide and still didn't include the actual price in the test set. Detailed performance scores fare in Table 5.2.

5.2 STL + ARIMA

To better capture the underlying structure of the series, we applied STL decomposition followed by (S)ARIMA modeling on decomposed components. From Appendix Figure 4, 5, 7, we can see that all three models show reliable results in diagnoses, initiating their validity.

Figure 5.2.1: Actual, Fitted, and Forecast of Price with 95% CI From STL+ARIMA Model

As shown in Appendix Figure 8, the fitted values closely match the training data across all components, indicating a good in-sample fit. Figure 5.2.1 The forecasted trend captures the continued upward momentum, while the seasonal component maintains a consistent cyclical pattern. The final reconstructed forecast, combining all components, aligns closely with the actual test data and clearly outperforms the baseline ARIMA model by better capturing both the direction and magnitude of change. The model also provides a well-calibrated 95% confidence interval that contains the actual values and reflects reasonable uncertainty around future estimates.

5.3 LSTM

The LSTM model is trained on the data from 1996 to 2023 and tested the predicted cocoa price in the next 11 months of 2024. Figure 5.3.1 plots the predicted price compared with the actual price and incorporates a 95% confidence interval for the prediction. Due to the sequence feature in LSTM, the first 30th month has no prediction. The model effectively tracks the overall trend of cocoa prices from the late 1990s to 2024, as both the predicted and actual prices (dashed and solid lines, respectively) follow similar patterns. Particularly, the model captures the general fast upward movement in 2024 but fails to react to the sharp spike. The shaded region represents the 95% confidence interval of the predictions, and the interval expands significantly from 5000 to 8000, reflecting increased volatility and uncertainty in the recent price surge.

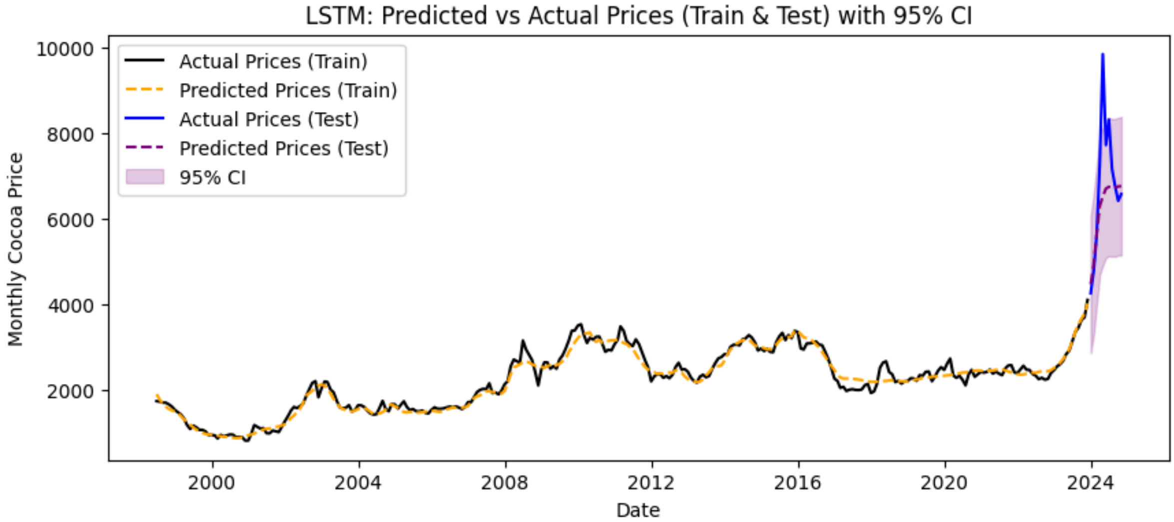

Figure 5.3.1: Actual, Fitted, and Forecast of Price with 95% CI From LSTM Model

After training the model, the variable importance for each predictor is also calculated in Figure 5.3.2. The plot displays that the exchange rate and monthly production have high importance because the RMSE rises up more than 100 when blocking these two variables. Precipitation and average temperature have less effect on the model prediction, but precipitation slightly contributes to the accurate prediction while the average temperature only has a negative effect.

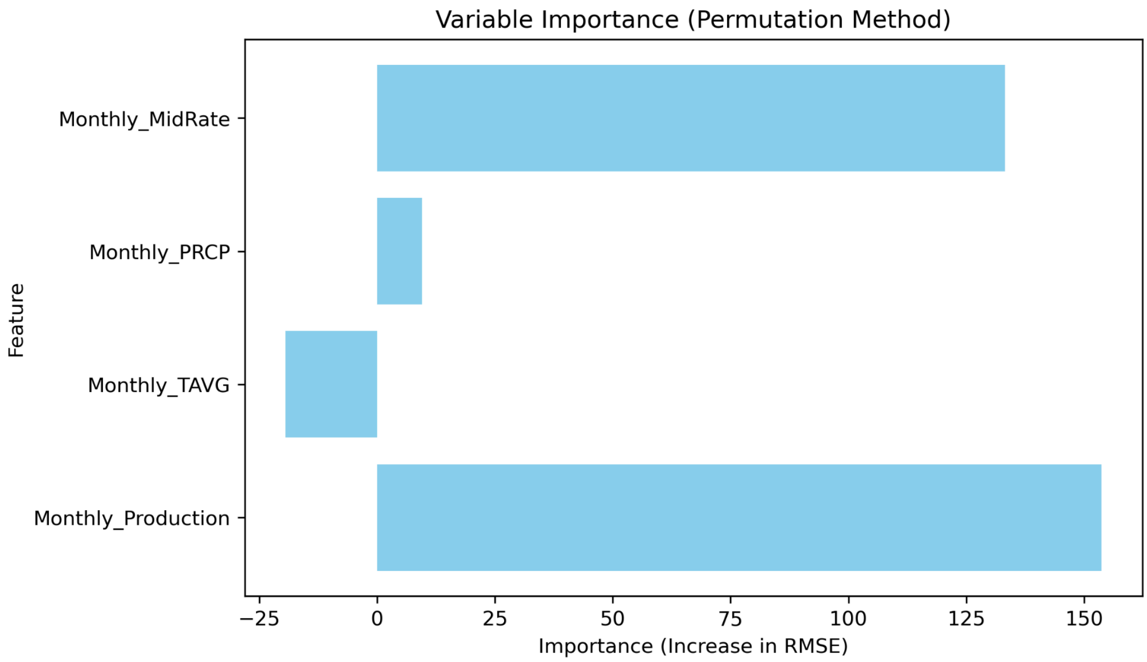

Figure 5.3.2: Variable Importance Based on LSTM Model

5.4 Models' Performance

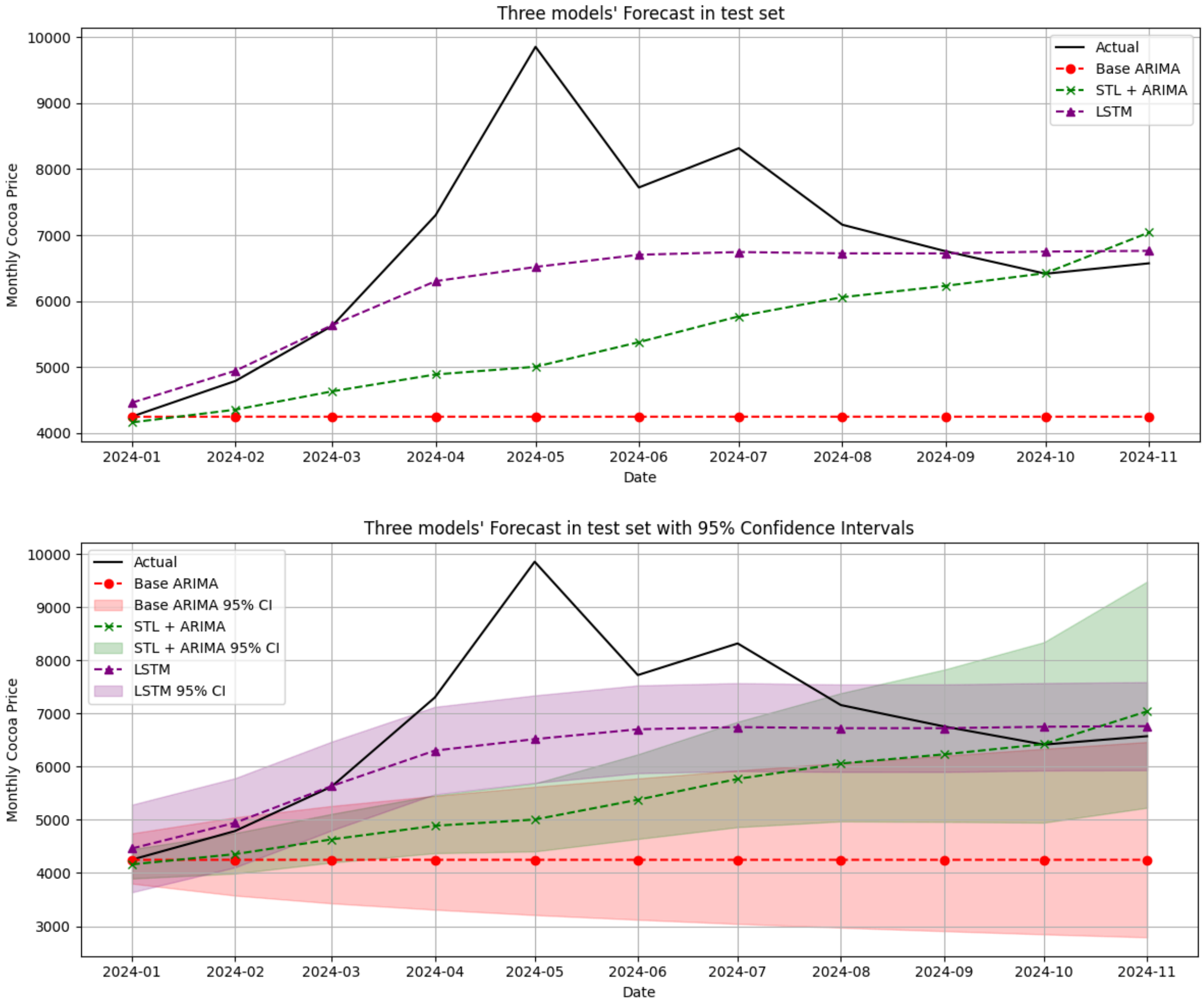

Figure 5.4.1: Forecast Performance of All Three Models on the 2024 Test Set

Train				
	R^2	RMSE	MAE	MAPE
ARIMA	0.9681	124.2860	92.3266	4.31%
STL + ARIMA	0.9942	52.7588	41.6501	1.99%
LSTM (multivariate)	0.9599	138.9132	102.8866	4.75%
Test				
	95% CI	RMSE	MAE	MAPE
ARIMA	0.00%	3177.2319	2911.5251	38.74%
STL + ARIMA	36.36%	2225.0891	1710.3778	22.12%
LSTM (multivariate)	90.91%	1207.8214	754.0692	9.33%

Table 5.4.1: Model Validations for the Three Models for Both Test and Train Sets

6. Discussion and Conclusion

The baseline ARIMA model, though statistically well-specified in its residual diagnostics, failed to anticipate the sharp price surge in 2024. Its forecast remained relatively flat, and its wide confidence interval did not contain the actual price movements in the test period. This performance suggests that ARIMA's linear structure and reliance solely on past values are insufficient for capturing sudden structural changes or volatility driven by external shocks.

The STL+ARIMA model, which decomposes the time series into trend, seasonal, and residual components, substantially improved upon the baseline. It captured both the direction and magnitude of price trends more effectively. The forecast reconstruction aligned better with the actual data, and the 95% confidence interval provided partial coverage of the observed values, reflecting a moderate improvement in uncertainty estimation.

The LSTM model significantly outperformed both traditional time series methods in the test period. It accurately responded to the sharp upward trend in 2024, and its 95% confidence interval covered the actual price in nearly all months. Notably, LSTM achieved the lowest regression metrics in the test set, confirming its robustness and adaptability under volatile conditions. These results support the hypothesis that machine learning approaches, when properly tuned and supported by relevant external data, provide greater predictive power in turbulent commodity markets.

Unlike ARIMA-based models, LSTM successfully incorporated exogenous variables—such as exchange rates, climate indicators, and production data—thus capturing the complex, nonlinear interactions driving price behavior. However, feature importance analysis suggested that the climate variables had limited predictive importance, which aligned with exploratory data analysis because their patterns did not show a strong correlation with cocoa price. The opposite effects between the climate features may suggest multicollinearity and may be improved by feature engineering in future work. Other possible approaches include additional variables like global cocoa demand, policy changes, or regional market disruptions.

Despite promising results, there are some limitations in the models. The STL+ARIMA model, while effective in trend modeling, was more heavily influenced by long-term movements and less responsive to seasonal variation, suggesting that STL decomposition may have overemphasized trend components at the cost of seasonal nuance. Although LSTM captured the sharp upward shift in cocoa prices in 2024, its performance was relatively smooth and conservative, missing extreme peaks in the price.

To improve future forecasts, possible directions include exploring transformer-based models for sequence learning, implementing feature selection techniques to refine input variables, and dynamically reweighting the contribution of trend and seasonal components in STL-based models. Incorporating domain knowledge into variable engineering could further enhance model robustness.

In conclusion, this study shows that integrating external drivers through machine learning leads to superior forecasting performance in volatile commodity markets like cocoa. While

ARIMA models serve as a reliable benchmark and STL+ARIMA adds interpretability and structural nuance, LSTM models best address the dynamic and nonlinear nature of modern price movements. For stakeholders in Ghana's cocoa sector—including policymakers, exporters, and farmers—such predictive tools can provide critical foresight, guiding strategic planning and improving resilience in the face of global shocks.

Reference

1. Robinson, N. (2025, February 13). *Cocoa prices shock 2025 financial outlook for Big Food*. FoodNavigator.com. https://www.foodnavigator.com/Article/2025/02/13/how-cocoa-prices-impact-big-food-companies-like-unilever/#:~:text=Food%20and%20drink%20majors%20have,strong%202024%20results%20this%20month.
2. *Ghana (GHA) exports, imports, and trade partners*. The Observatory of Economic Complexity. (2023). https://oec.world/en/profile/country/gha
3. Guedia, M. (2022, June 02). *"Ghana is Cocoa, Cocoa is Ghana" - OPEC fund for international development*. https://opecfund.org. https://opecfund.org/news/ghana-is-cocoa-cocoa-is-ghana
4. Olaigbe, O., & Usman, I. K. (2025, January 2). *Ghana is the second largest cocoa producer; why are its farmers still poor?*. Pulitzer Center. https://pulitzercenter.org/stories/ghana-second-largest-cocoa-producer-why-are-its-farmers-still-poor#:~:text=The%202023%2F2024%20season%20brought,50%25%20of%20harvest%20losses%20annually
5. *Cocoa bean production*. Our World in Data. (2025). https://ourworldindata.org/grapher/cocoa-bean-production?tab=chart&country=~GHA
6. Official exchange rate (LCU per US$, period average) - Ghana. World Bank Open Data. (2025). https://data.worldbank.org/indicator/PA.NUS.FCRF?view=chart&locations=GH
7. Bermudez, S., Voora, V., Larrea, C., & Luna, E. (2022, November). Global market report: Cocoa prices and Sustainability. Cocoa prices and sustainability. *https://www.iisd.org/system/files/2022-11/2022-global-market-report-cocoa.pdf*
8. Khin, A. A., Pok, W. F., Low, M. P., Rajamoorthy, Y. a/p., Tee, P. L., & Phan, Z. J. (2024, November). *Critical factors of sustainability for cocoa industry in Malaysia: Price forecasting methodology*. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-031-64140-4_6#ref-CR5
9. Assis, K., Amran, A., Remali, Y., & Affendy, H. (2010). *A comparison of univariate time series methods for forecasting cocoa bean prices*. Science Alert: Research Papers, Journals, Authors, Publishers. https://scialert.net/abstract/?doi=tae.2010.207.215
10. Lem, K. H. (2024). *ITM-Conferences*. The STL-ARIMA approach for seasonal time series forecast: a preliminary study. https://www.itm-conferences.org/articles/itmconf/pdf/2024/10/itmconf_icmsa2024_01008.pdf

11. Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023, July 30). *A review of Arima vs. Machine Learning Approaches for time series forecasting in Data Driven Networks*. MDPI. https://www.mdpi.com/1999-5903/15/8/255#:~:text=Multiple%20seasonality%20Complex%20modelling%20not,depending%20on%20the%20model
12. Olofintuyi, S. S., Olajubu, E. A., Olanike, D., c, & AbstractOne important aspect of agriculture is crop yield prediction. This aspect allows decision-makers and farmers to make adequate planning and policies. Before now. (2023, April 5). *An ensemble deep learning approach for predicting cocoa yield*. Heliyon. https://www.sciencedirect.com/science/article/pii/S2405844023024520