Forecasting Cocoa Prices in Ghana
A Comparative Study of Time Series and Machine Learning Models

Zheyu Lu, Junsong Wang, Yuhe Wang, Yinuo Yang

1. Introduction

Cocoa is a globally traded agricultural commodity vital to the production of chocolate and other food products. Its market price has far-reaching implications—not only for multinational corporations involved in food manufacturing but also for the millions of smallholder farmers who depend on cocoa cultivation for their livelihoods. Therefore, understanding and forecasting cocoa price dynamics is of critical importance for both economic planning and social welfare.

Cocoa prices exhibit both recurrent patterns and unexpected shocks. Internally, seasonal harvest cycles and periodic spikes in global consumption, such as those surrounding Easter, often lead to short-term price increases due to heightened demand for chocolate products (Guardian, 2025). Externally, macroeconomic and environmental factors significantly affect supply and pricing. Among these, weather is a crucial determinant of cocoa production, as cocoa trees rely heavily on specific rainfall and temperature conditions for healthy growth (Yoroba et al., 2019). Ghana's climate data is selected for analysis due to its relevance and accessibility, as the country is the world's second-largest cocoa producer, shares a border with Côte d'Ivoire—the largest producer—and serves as a representative case given the availability and quality of its meteorological data.

In addition to climate factors, macroeconomic influences—such as exchange rate fluctuations—can also impact cocoa pricing. For example, changes in the exchange rate between the Ghanaian cedi and the US dollar may affect export values and producer income (JP Morgan, 2024; Amoah & Vowles, 2010). Recent research further supports the inclusion of such variables in predictive models, suggesting that rainfall, currency changes, and broader demand factors all hold forecasting value (Kamu et al., 2010; Xia, 2023; Huang et al., 2023).

In this study, traditional models were initially applied for multi-step forecasting, but they failed to adequately capture seasonal and structural changes present in the data. This observation led to the consideration of seasonal-trend decomposition using Loess (STL) as a means to extract and model these components more effectively. The resulting STL-ARIMA hybrid model demonstrated a clear improvement in forecasting accuracy and successfully captured key variations in the price series, including the abrupt surge in 2024. When compared with both conventional statistical approaches and machine learning models such as Long Short-Term Memory (LSTM), the STL-ARIMA model consistently delivered superior results. To ensure that model performance could be evaluated under both typical and abnormal conditions, the dataset was extended to include the recent 2024 price surge, providing a meaningful test case for assessing how well different models respond to abrupt market shifts and complex seasonal structures.

2. Literature Review

Time series forecasting, especially for agricultural commodities, has made extensive use of conventional statistical models including ARIMA, GARCH, and ETS. Assis, Ahmed, and Yusoff (2010) forecasted monthly cocoa prices in Malaysia using Holt-Winters Addtive Model, GARCH, and ARIMA. GARCH was acknowledged for its capacity to model price volatility,

even though ARIMA had the lowest prediction error overall. All of the models ignored outside factors and only used historical data. Furthermore, Huang et al. (2023) pointed out that although ARIMA-GARCH models are good at capturing structured patterns like trend and volatility, they are not able to capture multivariate or nonlinear interactions. These methods provide a standard for evaluating the benefits of combining policy-based and climate predictions.

To address seasonality and structural variation in time series data, decomposition methods like STL (Seasonal and Trend decomposition using LOESS) are commonly used. Introduced by Cleveland et al. (1990), STL separates a series into trend, seasonal, and residual components, which allows for targeted modeling of recurring patterns—especially in agricultural markets with regular production and consumption cycles. Cocoa prices, for example, often spike around holidays such as Easter and Valentine's Day due to increased demand (Topping, 2025). STL is also robust to outliers and adaptable to shifting seasonal behavior, making it suitable for forecasting settings influenced by both climatic and holiday-related factors.

Recent advances in deep learning have introduced models capable of capturing nonlinear and multi-dimensional dynamics in time series forecasting, with Long Short-Term Memory (LSTM) networks being one of the most prominent examples. Huang et al. (2023) demonstrated that LSTM models are highly effective in agricultural price forecasting tasks due to their ability to learn complex temporal patterns from large datasets. Their study highlighted LSTM's flexibility in incorporating external factors such as weather and market demand. Xia (2023) similarly applied LSTM to forecast cocoa prices and found that the model provided superior forecasting accuracy over ARIMA, particularly for long-term predictions. These findings support the use of deep learning methods in environments where market behavior is shaped by interacting external drivers, and where nonlinearity plays a substantial role in price formation. LSTM thus serves as a modern benchmark against which the performance of hybrid models—particularly those combining STL and ARIMA—can be evaluated when integrating climatic and socioeconomic variables.

Among the external factors affecting cocoa prices, climate variability—particularly rainfall—plays a crucial role in shaping supply dynamics. Yoroba et al. (2019) examined the effects of rainfall and temperature on cocoa yields in Côte d'Ivoire and found that rainfall was a more significant predictor of yield variation than temperature. This reinforces the importance of using rainfall data as a seasonal indicator in decomposition methods like STL, which are sensitive to periodic agricultural and climatic cycles. In parallel, global market events such as El Niño have triggered extreme weather and contributed to recent cocoa price spikes, especially in major producing countries like Ghana (JP Morgan, 2024). These environmental drivers justify the inclusion of climate-related variables in forecasting models, particularly those designed to accommodate seasonal disruptions. Given these findings, Ghanaian temperature and precipitation data are incorporated to improve model responsiveness to climate-induced production shocks.

Beyond environmental considerations, institutional factors such as export restrictions and farmgate pricing schemes can limit the transmission of global price gains to local producers. According to the International Food Policy Research Institute (IFPRI, 2024), farmers in Ghana and Côte d'Ivoire often fail to fully benefit from international cocoa price increases due to domestic regulatory controls. This highlights the need for forecasting models to account not only for market and climate forces but also for policy-induced distortions that influence how prices behave at the local level. By incorporating these socioeconomic influences—particularly currency reforms such as Ghana's 2007 redenomination (Dzokoto, Young, & Mensah,

2010)—this study aims to evaluate whether exchange rates and policy variables can meaningfully improve predictive performance.

Together, these studies suggest that effective cocoa price forecasting requires a combination of traditional statistical models, decomposition techniques, and machine learning tools. More importantly, they underscore the value of integrating both socioeconomic factors—such as currency dynamics and policy regimes—and climatic variables like rainfall and temperature. This integrated perspective aligns closely with the present study's objective: to evaluate whether incorporating such external influences can improve the prediction of cocoa prices, particularly during periods of structural change and environmental stress.

3. Methodology

3.1 Data wrangling and preparation

This project explores various time series forecasting methods to model and predict cocoa futures prices, beginning with traditional approaches and gradually incorporating more advanced techniques. The primary objective is to develop models that can effectively capture both short-term fluctuations and long-term trends in Ghana cocoa prices, while also enabling a comparison of different modeling strategies in terms of forecasting performance.

The cocoa price data from Ghana was first cleaned and merged, and then the distribution of the price series was analyzed to understand its overall pattern. Given the pronounced volatility in recent years—particularly the sharp increase beginning in 2024—the log return rate was used as the primary preprocessing method to stabilize the variance and normalize the distribution. The transformation is equivalent to the first difference of the natural log of series:

$$r_t = \log(P_t) - \log(P_{t-1}) = \log\left(\frac{P_t}{P_{t-1}}\right)$$

where rt is the log return at time t, Pt is the price at time t, and Pt-1 is the price at time t - 1.

After preprocessing, the data was arranged in chronological order and split into training and testing sets for forecasting. With training set covers the first 95% of the data so that the model can learn from the full historical trend including the 2024 spike, and the last 5% is reserved for evaluation. A Q-Q plot of the log return series was also examined to assess the normality of the distribution. Points closely following the reference line indicate a normal distribution, while large deviations suggest the presence of outliers or heavy-tailed behavior.

3.2 Holt-Winters Additive Method

An ETS (Error, Trend, Seasonality) model was first fitted, as it can automatically account for trend and seasonal components without requiring the data to be stationary. After the model was fitted, the Ljung-Box test was applied to check for any remaining autocorrelation in the residuals. A significance level of p < 0.05 was used, and the p-values were plotted across different lags. If the majority of the p-values stayed above the threshold line, it indicated that the residuals resembled white noise. This suggests that the model effectively captured the temporal structure of the data, with no significant autocorrelated patterns remaining in the errors.

3.3 Arima Model

An Autoregressive Integrated Moving Average (ARIMA) model was then fitted. Prior to model fitting, the Augmented Dickey-Fuller (ADF) test was applied with a significance level of $p < 0.05$ to confirm the stationarity of the series. The best-fitting ARIMA model was selected based on the Akaike Information Criterion (AIC). After fitting, the residuals were evaluated using the Ljung-Box test to ensure that the autocorrelation structure in the series had been adequately captured.

3.4 GARCH Model

Since Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are widely used in financial time series forecasting for their ability to capture time-varying volatility, a GARCH model was also fitted.

Before model fitting, volatility clustering was assessed by visually inspecting the residuals from the ARIMA model. Based on this observation and AIC, the best-fitting GARCH was selected. The residuals were then evaluated using the Ljung-Box test to assess whether they resembled white noise.

3.5 STL-ARIMA Hybrid Model

After evaluating the performance of models based on the log return transformation, certain limitations were identified. As a result, the analysis shifted back to the original price series. To begin, the Seasonal and Trend decomposition using the LOESS (STL) method was applied to break down the cocoa price series into three components: seasonal, trend, and residual. The STL method is based on an additive model where the observed series is represented as the sum of its components:

$$y_t = T_t + S_t + R_t,$$

where yt is the observed value at time t, Tt is the trend component, St is the seasonal component, and Rt is the residual component.

STL uses LOESS (Locally Estimated Scatterplot Smoothing) to iteratively estimate the seasonal and trend components while filtering out short-term fluctuations (Cleveland et al. 1990). It is particularly useful for handling non-linear trends and flexible seasonal patterns. After decomposition, each component—trend, seasonal, and residual—was processed through differencing to improve stationarity. For each differenced series, the ADF test and the Ljung-Box test were used to assess stationarity and the presence of autocorrelation. To support model specification, ACF and PACF plots were examined for each component. Based on these diagnostics, appropriate ARIMA or Seasonal Autoregressive Integrated Moving Average (SARIMA) models were fitted to the respective components. The resulting forecasts from each component model were then combined to construct a hybrid ARIMA model.

3.6 LSTM Model

Finally, an advanced forecasting approach was also applied using a Long Short-Term Memory (LSTM) neural network to capture nonlinear relationships and long-term dependencies in Ghana cocoa price data. The model was trained on monthly cocoa price data from Ghana, along with selected explanatory variables created for machine learning. These include precipitation (PRECTOTCORR), a holiday indicator (Is_Holiday) marking Christmas, Easter,

and Halloween, the numeric month (Month_Num) to reflect seasonality, and a computed trend feature (Cocoa_Trend) represents online search trend in cocoa-related terms. The data was transformed into a supervised learning format using a sliding window method, where each input consisted of a fixed-length sequence of lagged observations used to predict the following month's cocoa price.

3.7 Model Evaluation

To compare the five forecasting approaches, model performance was evaluated on a held-out test set using both single-step rolling forecasts and multi-step forecasts. In the single-step rolling forecast, models were retrained each month using all available training data up to that point. This allows forecasts to adapt over time. For the multi-step forecast, the final model obtained at the end of the training period was used to iteratively predict prices over the test horizon without further updates. Forecast accuracy was assessed using three standard evaluation metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). In addition, 95% confidence intervals were constructed using model-specific standard error estimates to visualize forecast uncertainty. To extend the analysis, all models were also used to generate forecasts for the next 12 months beyond the test set. These future predictions help assess how well each approach responds to unseen market conditions.

4. Data

4.1 Data Collection and Preparation

The dataset analyzed in this research include three categories: Economical Data, Weather Data and Financial Data. Cocoa daily prices were acquired from the International Cocoa Organization (ICCO, 2025), while climate variables—including near-surface temperature, precipitation, and solar radiation—were sourced from NASA's (NASA, 2025) meteorological databases. Economic indicators, specifically exchange rates, were retrieved from official records published by the Bank of Ghana(Bank of Ghana, 2025). The data was arranged in chronological order to ensure consistency when merging data from different sources.

Considering significant political disruptions, including Ghana's currency reform in 2007 (Wikipedia. 2024)and associated governmental transitions, data preceding 2007 were excluded to preserve the integrity and consistency of the time series analysis. Cocoa price data used in the study were strictly derived from trading days at the New York futures market (NYBOT). Due to the absence of market activity during weekends, missing data points on weekends were imputed using the preceding Friday's closing price. For solar radiation, missing values initially coded as "-999" were transformed to NA and subsequently interpolated using a 7-day moving average. Missing precipitation data were explicitly set to zero, and any remaining missing values across variables were filled through forward imputation.

For enhanced analytical consistency and interpretability, monthly average values were adopted as the primary analysis unit. Additionally, a binary variable labeled "Is Holiday" was created, indicating periods around major holidays, specifically Christmas, Easter, and Halloween, motivated by previous empirical evidence demonstrating elevated chocolate consumption during these periods. Furthermore, the study utilized log return rates as the primary financial metric due to their stability, robustness to volatility, and widespread acceptance within financial econometrics, making them particularly effective in capturing nuanced market dynamics associated with cocoa prices.

4.2 Data Summarization and Visualization

Table 1. Descriptive Statistics Including Mean, Standard Deviation, Minimum and Maximum Value of Monthly Cocoa Price and Associated Economic and Climatic Variables (2007–2025)

Variable	Mean	Std Dev	Min	Max
Cocoa Price (USD)	2983.51	1402.74	1848.69	9806.78
Log Return Rate	0.01	0.09	-0.25	0.43
Cocoa Price Trend Index (%)	28.08	13.34	16.00	100.00
Near-Surface Air Temp (°C)	25.72	1.17	23.66	29.30
Max Air Temp (°C)	30.44	2.05	27.63	36.49
Min Air Temp (°C)	21.91	1.11	17.73	24.14
Precipitation (mm/day)	4.30	3.09	0.05	14.32
Solar Radiation (W/m²)	17.00	1.79	12.67	20.72
Exchange Rate (Mid)	4.68	3.78	0.93	15.99
Is Holiday (Binary)	0.37	0.48	0.00	1.00

Table 1 summarizes descriptive statistics of the monthly cocoa prices along with associated economic and climatic variables from 2007 to 2025. Notably, cocoa prices exhibit considerable variability, ranging significantly from USD 1848.69 to USD 9806.78 with a standard deviation of 1402.74, underscoring the volatility inherent in cocoa markets. Additionally, the exchange rate demonstrates high variability, reflecting potential macroeconomic instability influencing cocoa prices. Conversely, log return rates present much greater stability, which is further confirmed by the distribution comparisons before and after their transformation in Figure 1. Additionally, the substantial variability in Google search trends (Cocoa Price Trend Index) and exchange rates suggests that subsequent modeling processes should account explicitly for potential structural shifts or external shocks within these variables. However, certain subtleties, such as seasonal patterns and underlying trends, are difficult to discern directly from these summary statistics. Thus, graphical analyses were performed to further elucidate these patterns.
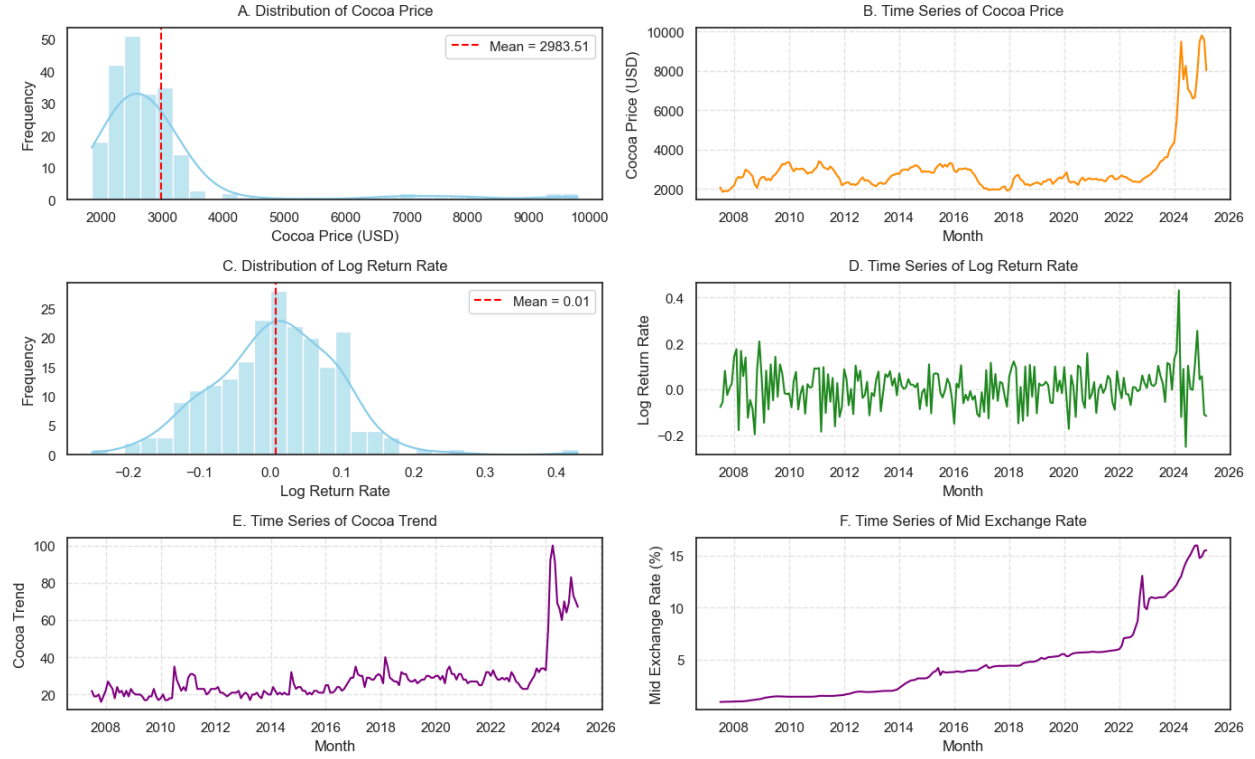
Figure 1. Distribution and Time Series Analysis of Cocoa Price and Related Metrics.
(A) Distribution of Cocoa Price (USD) with marked mean value.
(B) Time series of monthly Cocoa Prices from 2007 to 2025.
(C) Distribution of Log Return Rate with indicated mean.
(D) Monthly Log Return Rates over the analysis period.
(E) Time Series illustrating Cocoa Price Trend Index (%).
(F) Monthly Mid Exchange Rate trends from 2007 to 2025.

Data visualization was conducted to reveal the distributional characteristics and temporal dynamics of key variables. Figure 1 shows the distribution and temporal trends of cocoa prices, log return rates, cocoa price trend indices, and exchange rates. Cocoa prices demonstrate significant volatility with notable spikes post-2024, while log return rates exhibit a stable distribution centered around zero, indicative of balanced market fluctuations. Additionally, STL decomposition analysis of climate and environmental variables such as near-surface air temperature, precipitation, and solar radiation, reveals similar seasonal patterns across these climate metrics, presented in the Appendix. STL decomposition further identified minimal trend variations in temperature and solar radiation, whereas precipitation displayed a notably distinct trend, indicating meaningful climatic shifts relevant to cocoa cultivation.

5. Forecasting and Results

5.1 ETS (Holt-Winters Additive) Model Result

The Holt-Winters additive model captures the short-term dynamics of log return cocoa prices. In the full-series view (Figure 2 Panel A), the model tracks historical price trends reasonably well, with relatively narrow confidence intervals during the training period. However,

during the test period (Panel B), the rolling one-step-ahead forecast using an expanding window exhibits noticeable lag, indicating limited ability to capture abrupt upward shifts and suggesting that the model struggles to learn latent dynamics in real-time. In contrast, the multi-step forecast continues to rise steadily, suggesting that the model captured the general trend but failed to learn the seasonal pattern. The overly wide confidence intervals further highlight the uncertainty and imprecision of the prediction. Residual diagnostics support the short-term validity of the model. The Q-Q plot (Panel C) shows approximate normality, with obvious deviation in the tails due to the huge fluctuation in the dataset. The Ljung–Box test (Panel D) indicates that residuals pass the white noise test only at the first two lags, while p-values at higher lags fall below the 0.05 threshold, suggesting that residual autocorrelation remains and the model does not fully capture the temporal dependencies in the training data. Overall, the Holt-Winters additive model's additive framework fails to capture the nonlinear volatility and structural break observed in the test period, limiting its utility for long-term projections.
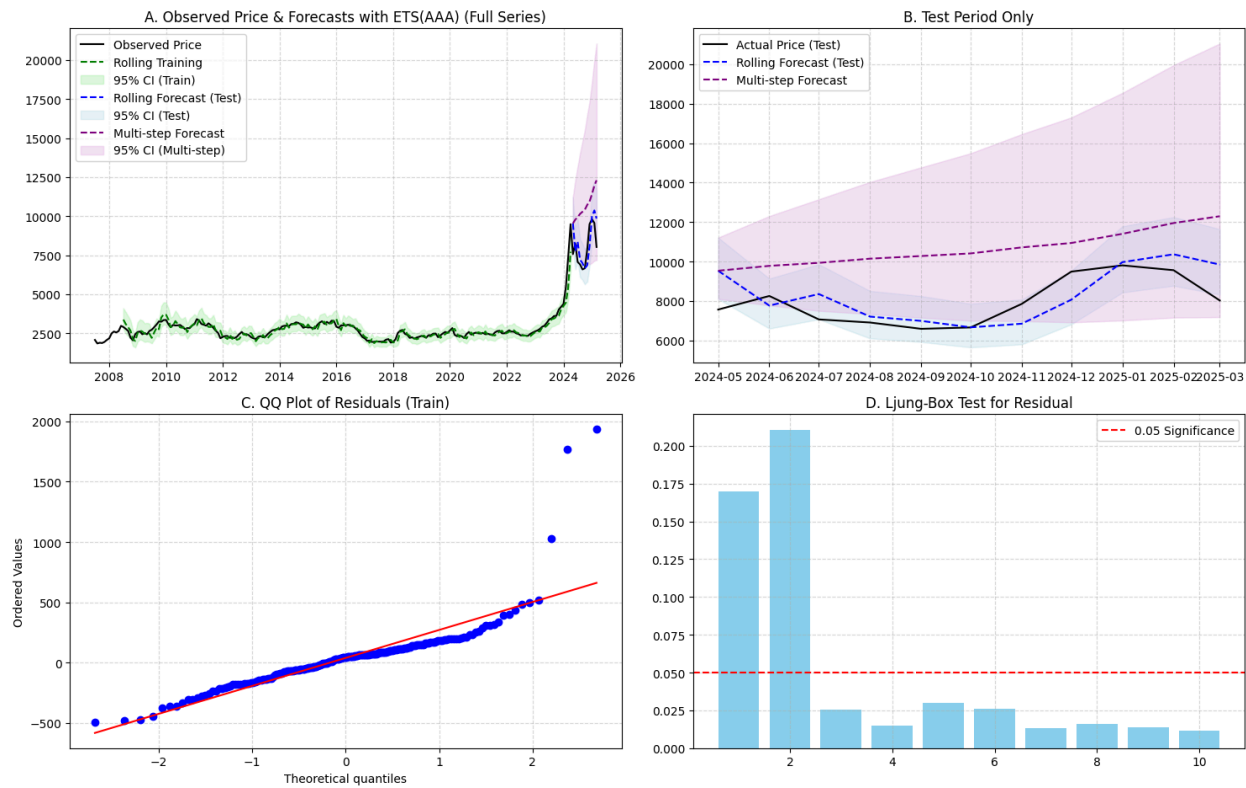

Figure 2. Forecasting Performance of ETS(AAA) Model on Cocoa Prices
(A) Observed vs. forecasted prices using ETS(AAA).
(B) Zoom-in on test forecasts; multi-step shows greater uncertainty and less accuracy.
(C) Q-Q plot shows near-normal residuals with deviation in tails.
(D) Ljung-Box test on residuals.

5.2 ARIMA Model Result

The ARIMA(1,0,1) model was fitted to the log return price series. As shown in Appendix Figure 2, the model fits the training data with relatively narrow confidence intervals (Panel A), suggesting decent in-sample stability. In the test period (Panel B), the rolling one-step forecast shows a lagging response to rapid price spikes, indicating that the model did not effectively capture the hidden nonlinear dynamics. While the forecasts generally follow the overall

direction, underestimations during peak months suggest limited adaptation to sudden price changes. However, the multi-step forecast quickly drifts upward, severely overestimating future prices with an unrealistic upward trend and increasingly wide confidence bands, indicating an uncertainty amplification over time. Residual analysis raises further concerns. The Q-Q plot (Panel C) reveals a noticeable deviation from the normal distribution in the upper tail, suggesting heavy-tailed residuals that may reflect the impact of volatility clustering or regime shifts. Meanwhile, the Ljung-Box test (Panel D) shows that most p-values exceed the 0.05, especially at higher lags, indicating that the residuals are largely uncorrelated. This suggests the ARIMA(1,0,1) model has adequately captured the temporal structure of the training data.

5.3 GARCH(1,1) Model Result

The GARCH(1,1) model demonstrates the strongest short-term performance among traditional models, but its forecasts also exhibit some lagging behavior, suggesting delayed adaptation to sudden price changes. As shown in Appendix Figure 3, the one-step rolling forecast fits closely to actual prices, with tighter confidence intervals than ARIMA, reflecting better volatility control (Panel A). In Panel B, the multi-step forecast exhibits overly smooth and fails to capture the underlying volatility, suggesting improved long-horizon stability, albeit still limited by structural shifts. Diagnostic plots confirm model refinement. Panel C shows residuals with moderate satisfaction to normality but having deviation to the default normality line with lighter tails compared to ARIMA, indicating not enough to adjust the error distribution. The Ljung-Box test (Panel D) reveals several p-values falling below the 0.05 threshold, particularly at lower lags, indicating that the model fails to fully remove the autocorrelated structure from the residuals, suggesting its conditional heteroskedasticity correction is incomplete. In conclusion, while GARCH(1,1) enhances short-term predictive accuracy by accounting for time-varying volatility, it remains inadequate in capturing structural breaks or nonlinear upward trends, which limits its long-term reliability.

5.4 STL-ARIMA Decomposition and Hybrid Model

The STL-ARIMA model demonstrated the strongest overall performance among all models tested, especially in capturing both seasonal trends and structural change. By separating the trend, seasonal, and residual components through STL decomposition and modeling each individually, the model achieved interpretable and flexible forecasts.

To determine suitable models for each component, ACF and PACF plots were generated and analyzed (Appendix Figure 4): For the seasonal component, the original series was already stationary based on the ADF test ($p < 0.05$). Although PACF exhibited a visible spike at lag 12, the autocorrelation did not persist beyond that point, and the ACF showed a small bump at lag 12. This pattern justified modeling the seasonal structure using a SARIMA(5,0,4)(1,0,1)[12] specification, where the seasonal order of 12 corresponds to the 12-month cycle in monthly data. For the residual component, after first differencing, both ACF and PACF plots showed weak but non-negligible autocorrelation, particularly with a significant PACF cutoff at lag 1. This supported the selection of an ARIMA(1,1,1) to model residual. Lastly, the trend component was differenced twice to achieve stationarity. The second-differenced series revealed a strong spike at lag 1 in the ACF and a clear PACF cutoff after lag 2, which led to fitting an ARIMA(2,2,5) model to capture the nonlinear long-term dynamics.
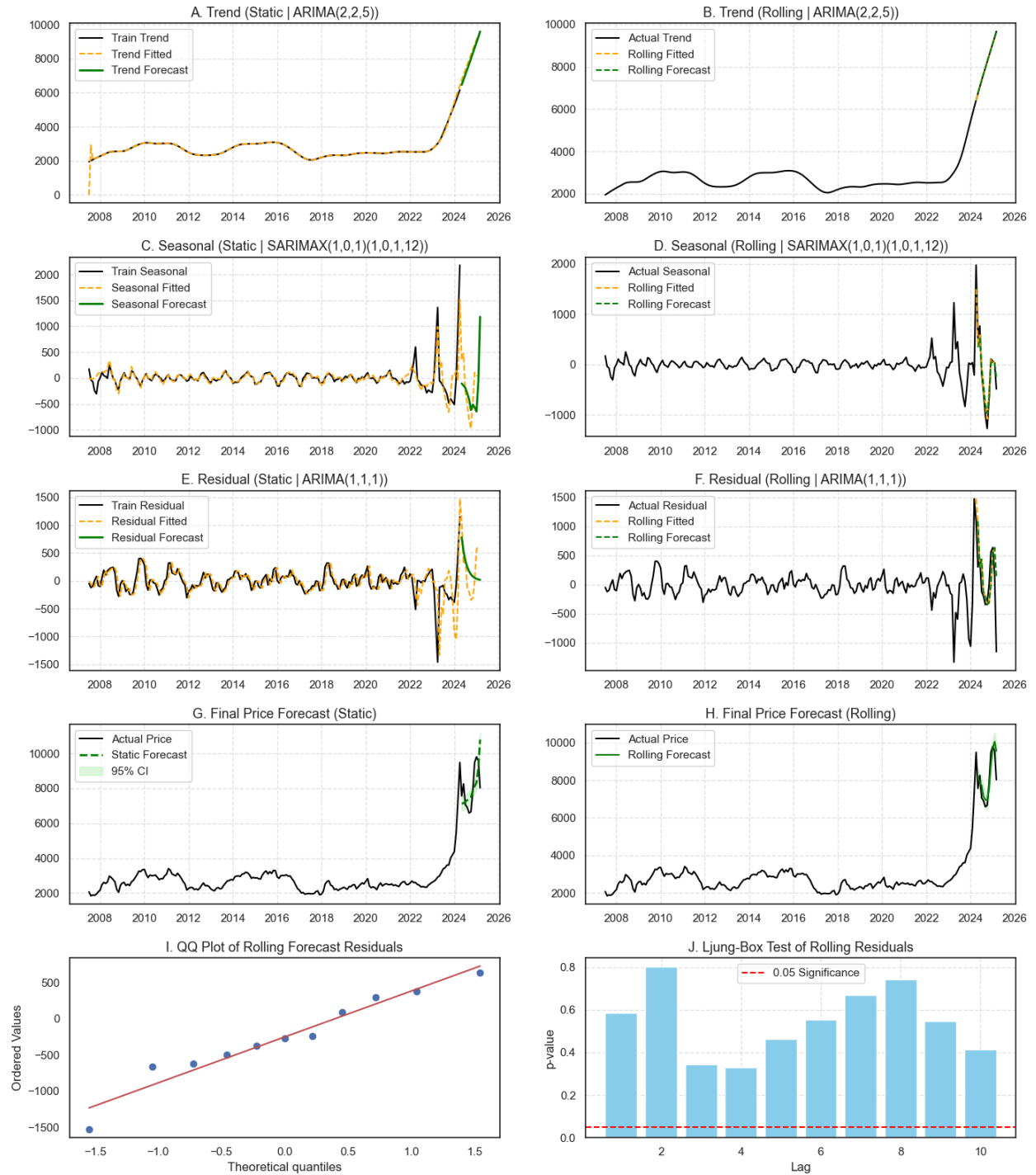
Figure 3. STL-ARIMA Decomposition and Forecasting of Cocoa Prices
(A-B) Decomposed trend (static & rolling).
(C-D) Decomposed seasonal (static & rolling).
(E-F) Decomposed residuals (static & rolling).
(G-H) Final Price Forecasting (static & rolling).
(I) QQ plot of rolling forecast residuals.
(J) Ljung-Box test of rolling residuals.

In the static forecast (Figure 3 Panels A, C, E, G), the fitted ARIMA models successfully captured each component's dynamics, especially the clear seasonality reproduced in Panel C. The final forecast (Panel G) matched the general shape of the observed prices but slightly overpredicted, particularly near the peak. The confidence intervals remained well-bounded. The rolling forecast (Panels B, D, F, H) yielded even stronger performance. The forecast followed the actual test prices closely, including both sharp spikes and seasonal dips, suggesting the component-wise rolling updates adapted well to recent information. Confidence intervals in Panel H were also narrower and more consistent than those in other models, showing stable predictive uncertainty. STL-ARIMA adapted exceptionally well to the non-linear trend shifts and complex seasonal patterns in cocoa prices. Additionally, the Ljung-Box test confirmed that the residuals resembled white noise across all lags, indicating no remaining self-correlation and suggesting that the model effectively learned the underlying pattern. This supports STL-ARIMA as the most reliable and flexible model in this analysis.

5.5 LSTM Model Result

The LSTM model demonstrates nonlinear learning capacity, capturing temporal dependencies and interactions between multiple exogenous variables. In Figure 4 Panel A, the steady decrease in training loss over epochs indicates successful model convergence without signs of overfitting. In Panel B, the model tracks recent price movements moderately well but shows slight over-smoothing, failing to fully reproduce sharp fluctuations. The multi-step future forecast (Panel C) shows a generally downward trajectory, with a smooth arc suggesting the model captured broad directional patterns but struggled with local volatility and momentum reversal — this is reflected in the relatively high MAPE of 19.44%. Panel D shows online search trends and exchange rates emerged as the most influential predictors, while rainfall, holiday indicators, and calendar month had minimal impact. This supports the hypothesis that public attention and macroeconomic signals are more predictive than periodic seasonal markers in this context.
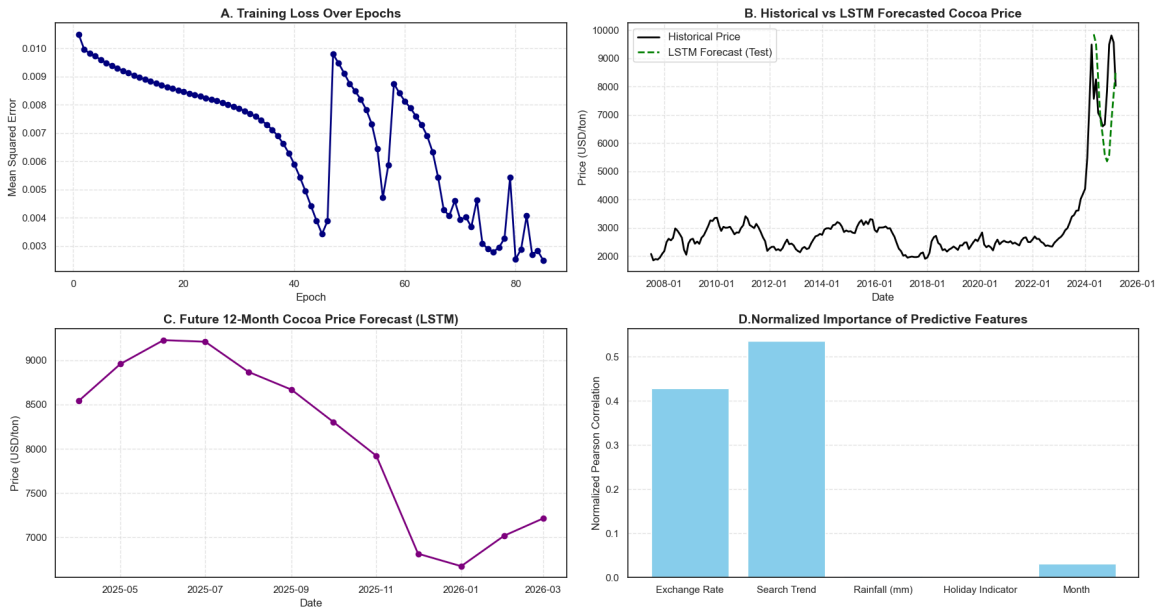

Figure 4. LSTM-Based Forecasting of Cocoa Price and Feature Analysis.
(A) Training loss curve showing model convergence across 85 epochs.
(B) Comparison of historical cocoa prices and LSTM-predicted test prices.

(C) Multi-step forecast of cocoa prices for the next 12 months.
(D) Normalized feature importance showing dominant influence of search trends and exchange rate.

5.6 Summary of Forecast Accuracy

Table 2. Forecast Accuracy Statistics of All Models Across Single-Step and Multi-Step

Model	Forecast Type	RMSE	MAE	MAPE (%)
ETS (AAA)	Single-step	1088.64	878.70	11.02
	Multi-step	2850.65	2691.88	35.50
ARIMA(1,0,1)	Single-step	1332.68	1045.96	13.37
	Multi-step	2443.91	2212.08	30.03
GARCH(1,1)	Single-step	1063.17	845.94	10.61
	Multi-step	1989.06	1682.02	23.32
STL-ARIMA Hybrid	**Single-step (rolling)**	**631.00**	**488.10**	**6.07**
	Multi-step (static)	**1052.48**	**647.82**	**7.87**
LSTM	Multi-step	2000.77	1635.79	19.44

      Table 2 summarizes forecast accuracy across different models and forecasting strategies using three standard evaluation metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). RMSE emphasizes larger errors by squaring deviations before averaging, making it sensitive to outliers. MAE offers a more interpretable measure of average error in actual units, while MAPE expresses prediction error as a percentage, enabling scale-free comparisons across models. Together, these metrics provide a comprehensive view of model performance, balancing sensitivity to error magnitude with interpretability in financial contexts.

      Overall, STL-ARIMA Hybrid (Rolling) delivers the strongest performance across all metrics, particularly achieving the lowest MAPE (6.07%), RMSE (631.00), and MAE (488.10), indicating a well-fitted model with minimal forecast error. Notably, even the multi-step (static) variant of STL-ARIMA outperforms all other multi-step forecasts, with MAPE = 7.87%, RMSE = 1052.48, and MAE = 647.82, highlighting its robustness in both short- and medium-term prediction horizons. In financial applications such as commodity price forecasting, risk assessment, or portfolio optimization, lower RMSE and MAE imply better short-term decision-making and risk control, while lower MAPE ensures that model performance remains consistently interpretable across varying price scales. The STL-ARIMA Hybrid's superior accuracy suggests its strong potential for supporting financial strategies that rely on timely and precise predictions of price dynamics.
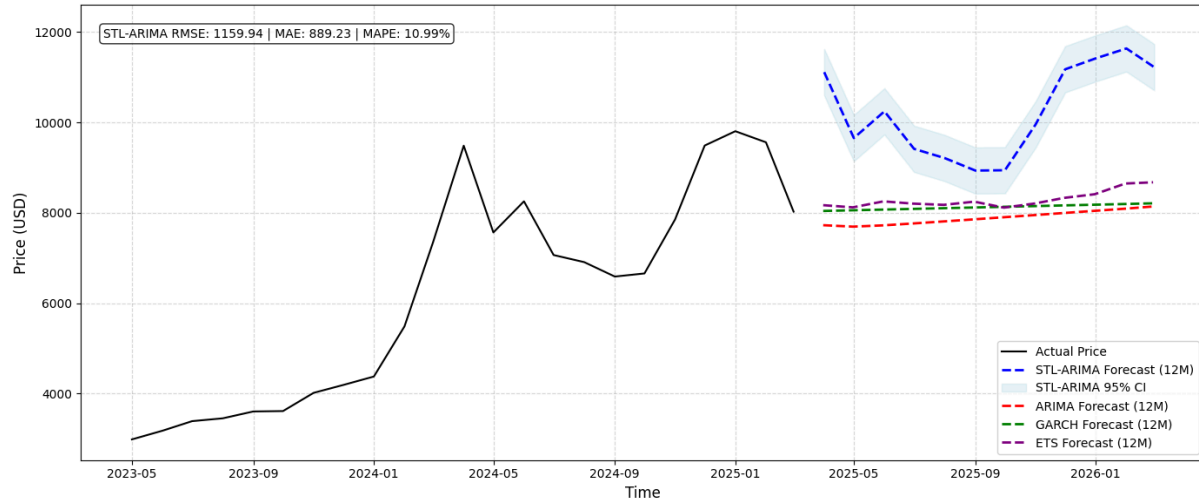
Figure 5. 12-Month Cocoa Price Forecast vs. Real Price Trend (as of March 2025)

In Figure 5, the STL-ARIMA forecast (blue dashed line) anticipates a rebound in cocoa prices with a sharp upward swing, outperforming other models in responsiveness to recent trends. Notably, actual market data has validated this forecast: Cocoa prices have risen from below 8000 USD at the end of March 2025 to 9238.6 USD at present, aligning closely with the projected surge. This not only reinforces the predictive power of the STL-ARIMA model but also underscores its practical value for short-term investment and supply chain decisions in commodity markets.

6. Discussion and Conclusion

6.1 Overview

Among the traditional models, ETS and ARIMA performed reasonably well when predicting just the following month's cocoa price, with average forecast errors around 10% compared to the actual prices. This result is consistent with Assis et al. (2010), who found that ARIMA models can be effective for short-term cocoa price forecasting using historical data. However, their accuracy dropped significantly when predicting several months ahead, with average errors rising above 30%. These larger errors suggest that while these models work for short-term forecasting, they struggle to maintain accuracy over longer time periods—especially when unexpected changes or complex trends are present. GARCH models, which are commonly used in financial forecasting to handle changing volatility over time (Huang, Wang, Zhang, 2023), showed slightly better performance in these multi-month forecasts. The GARCH model had an average error of about 23%, which is an improvement but still not ideal for capturing longer-term dynamics. One possible reason for this difficulty is the use of log return transformation, which focuses the model on short-term percentage changes. While this transformation helps stabilize variance, it also removes important long-term trends and makes it harder for the model to track larger price movements over time. This limitation is one of the main reasons the STL-ARIMA approach was adopted. By decomposing the series into interpretable components, STL helps the model better learn seasonal and trend patterns, which improves both short- and long-term forecast accuracy.

The STL-based hybrid model produced the most accurate and reliable results. In one-month-ahead forecasts, the STL model's prediction was 6.07% different from the actual

13

cocoa price on average. For multi-month forecasts, the average error only increased to 7.87%, which is still far lower than the 23%–30% error seen in the traditional models. It also performed better when looking at how far off the predictions were in actual price units—its average error stayed below 650 USD/ton, and the typical deviation from the true value was just over 1000 USD/ton. In the rolling forecast, these numbers dropped even further. The predicted price ranges were also much tighter and stayed closer to the real values, showing that the model provided not just more accurate forecasts but also more realistic uncertainty estimates. One possible reason is that it worked with the original price data and was able to focus on long-term patterns without being confused by short-term noise. It also handled seasonal changes—like weather or holiday demand spikes (Associated Press, 2024)—better than other models. As Cleveland et al. (1990) pointed out, STL is especially good for data that has both repeating patterns and big shifts over time, which makes it a strong fit for forecasting cocoa prices.

The LSTM model showed moderate performance in forecasting prices. On average, the predicted prices were about 19.44% different from the actual prices. While this is acceptable compared to traditional models, it was still less accurate than the STL-based approach. The full series and test plots show that LSTM captured general trends but tended to smooth out the sharp increase seen after 2023. This smoothing effect may be due to its tendency to minimize short-term errors and the limited amount of training data available. As noted by Xia (2022) and Huang et al. (2023), LSTM models are good at learning nonlinear patterns but may underperform when it comes to sudden structural changes in the data. Feature importance analysis also revealed that exchange rate and online cocoa search trends had the strongest correlation with price, supporting the idea that external factors play a meaningful role.

6.2 Limitations and Future Work

While the STL-based hybrid model delivered the best forecasting performance overall, it also has certain limitations. Most notably, STL treats the time series as a purely internal process—it separates the data into trend, seasonal, and residual components based only on historical price movements. As a result, it cannot directly incorporate multiple external factors such as exchange rate fluctuations, precipitation, or search trend data, which were shown to have strong correlations with cocoa prices. This limits the model's ability to react to sudden real-world shocks or complex interactions among variables. By contrast, the LSTM model incorporated multiple external predictors to enhance its forecasting ability. Despite this, it struggled with longer-term predictions, particularly during periods of structural change. Its forecasts appeared overly smooth and failed to capture sharp price spikes—likely due to the limited number of such events in the training data and the recursive nature of multi-step forecasting. The relatively small dataset may also have constrained the model's capacity to learn more complex relationships, especially given the large number of parameters in LSTM architectures. Feature importance analysis identified exchange rate and online search trends as the most influential predictors. However, the strong correlation of search interest with price may reflect public reaction to price increases rather than serving as a true leading indicator. These challenges are consistent with previous findings, which highlight that LSTM models tend to favor short-term accuracy while struggling to generalize across broader trends (Xia 2022; Huang et al. 2023).

In future work, hybrid approaches that combine STL decomposition with machine learning models like LSTM could be explored to leverage the interpretability of component-wise modeling while allowing external drivers to be included. It may also be helpful to test dynamic

decomposition methods that adjust components over time or integrate real-time external information more directly to enhance forecast responsiveness.

7. Reference List

1. Ahmed, Kamu Assis, Amran, and Remali Yusoff. (2010). Forecasting cocoa bean prices using univariate time series models. ResearchGate. Retrieved April 3, 2025, from https://www.researchgate.net/publication/285068778_Forecasting_Cocoa_Bean_Prices_Using_Univariate_Time_Series_Models
2. Bank of Ghana. (2025). Daily interbank FX rates. Retrieved April 3, 2025, from https://www.bog.gov.gh/treasury-and-the-markets/daily-interbank-fx-rates/
3. Butler, Sarah. (2025, April 1). UK food inflation increases as shoppers buy Easter eggs early.TheGuardian.RetrievedApril3,2025,fromhttps://www.theguardian.com/business/2025/apr/01/uk-food-inflation-easter-eggs-lidl-ocado-asda-kantar
4. Cleveland, Robert B., Terpenning, Irma, McRae, Jean E., & Cleveland, William S. (1990). STL: A seasonal-trend decomposition procedure based on Loess. Retrieved April 3,2025,fromhttps://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/stl-a-seasonal-trend-decomposition-procedure-based-on-loess.pdf
5. Dzokoto, Vivian Afi Abui, Young, Jessica, & Mensah, Clifford Edwin. (2010). A tale of two cedis: Making sense of a new currency in Ghana. ScienceDirect. Retrieved April 3, 2025, from https://www.sciencedirect.com/science/article/abs/pii/S0167487010000498
6. Huang, Xingdan, Chen, Dapeng, Gao, Xiaolian, & You, Panlu. (2023). Stock price prediction based on ARIMA-GARCH and LSTM. Atlantis Press. Retrieved April 3, 2025, from https://www.atlantis-press.com/article/125989735.pdf
7. International Cocoa Organization. (2025). Cocoa daily prices. Retrieved April 3, 2025, from https://www.icco.org/statistics/
8. J.P. Morgan. (2024, December 2). Rising cocoa prices: Will the chocolate crisis continue in 2025? Retrieved April 3, 2025, from https://www.jpmorgan.com/insights/global-research/commodities/cocoa-prices
9. NASA. (2025). NASA POWER: Prediction of worldwide energy resources. Retrieved April 3, 2025, from https://power.larc.nasa.gov/
10. Tabe-Ojong, Martin Paul Jr., Glauber, Joseph, & Guedegbe, Onasis Tharcisse Adetumi. (2024, May 8). Soaring cocoa prices: Diverse impacts and implications for key West Africanproducers.IFPRI.RetrievedApril3,2025,fromhttps://www.ifpri.org/blog/soaring-cocoa-prices-diverse-impacts-and-implications-key-west-african-producers
11. Xia, Chenxi. (2023). Comparative analysis of ARIMA and LSTM models for agricultural productpriceforecasting.ResearchGate.RetrievedApril3,2025,fromhttps://www.researchgate.net/publication/379531846_Comparative_Analysis_of_ARIMA_and_LSTM_Models_for_Agricultural_Product_Price_Forecasting
12. Yoroba, Fidèle, Kouassi, Benjamin K., Diawara, Adama, Assamoi, Paul, Koné, Ibrahim D., Kouadio, Yves K., Tiemoko, Dro T., Kouadio, Kouakou, & Yapo, Louis A. M. (2019, January 13). Evaluation of rainfall and temperature conditions for a perennial crop in tropical wetland: A case study of cocoa in Côte d'Ivoire. Wiley Online Library. Retrieved April 3, 2025, from https://onlinelibrary.wiley.com/doi/10.1155/2019/9405939
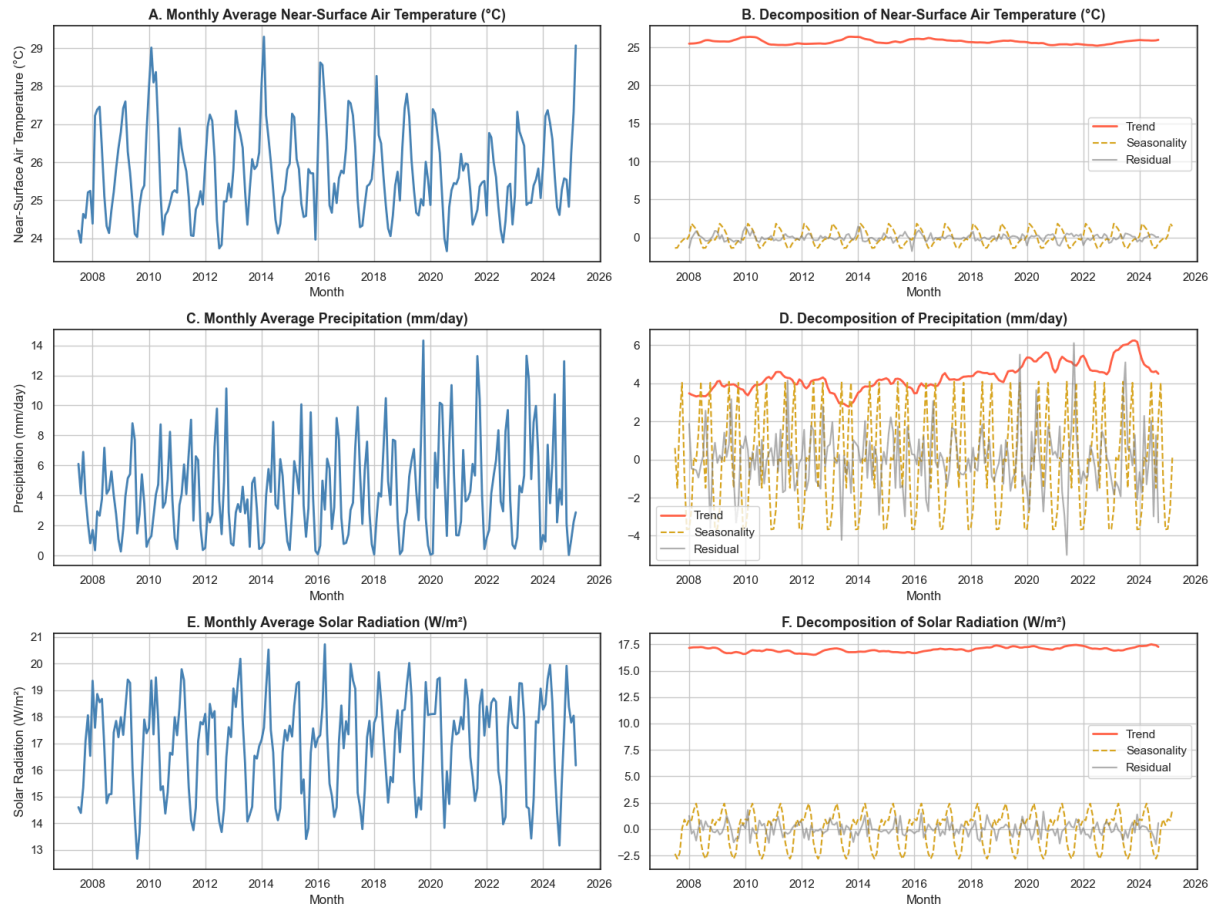
8. Appendix

8.1 Figures

Figure 1. Monthly Climate and Environmental Variables with STL Decomposition (2007–2025).
(A) Monthly average near-surface air temperature (°C).
(B) STL decomposition of near-surface air temperature into trend, seasonal, and residual components.
(C) Monthly average precipitation (mm/day).
(D) STL decomposition of precipitation illustrating clear seasonal, trend, and residual patterns.
(E) Monthly average solar radiation (W/m²).
(F) STL decomposition of solar radiation, showing stable trends and distinct seasonal cycles.
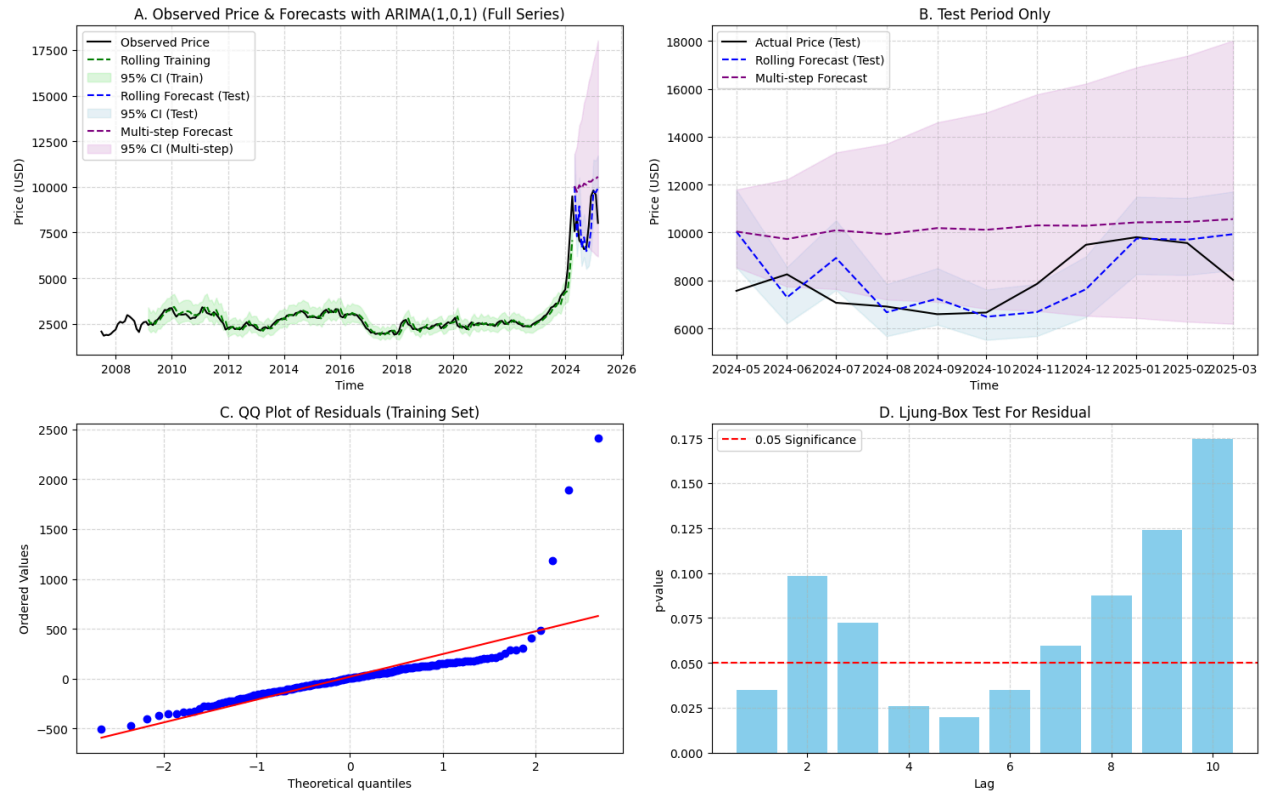
Figure 2. Forecasting Performance of ARIMA(1,0,1) Model on Cocoa Prices

(A) Observed and forecasted prices using ARIMA(1,0,1); green for training, blue/purple for test forecasts.

(B) Close-up on test period shows multi-step forecasts overestimate price with high uncertainty.

(C) Q-Q plot reveals mild heavy tails, indicating non-normal residuals.
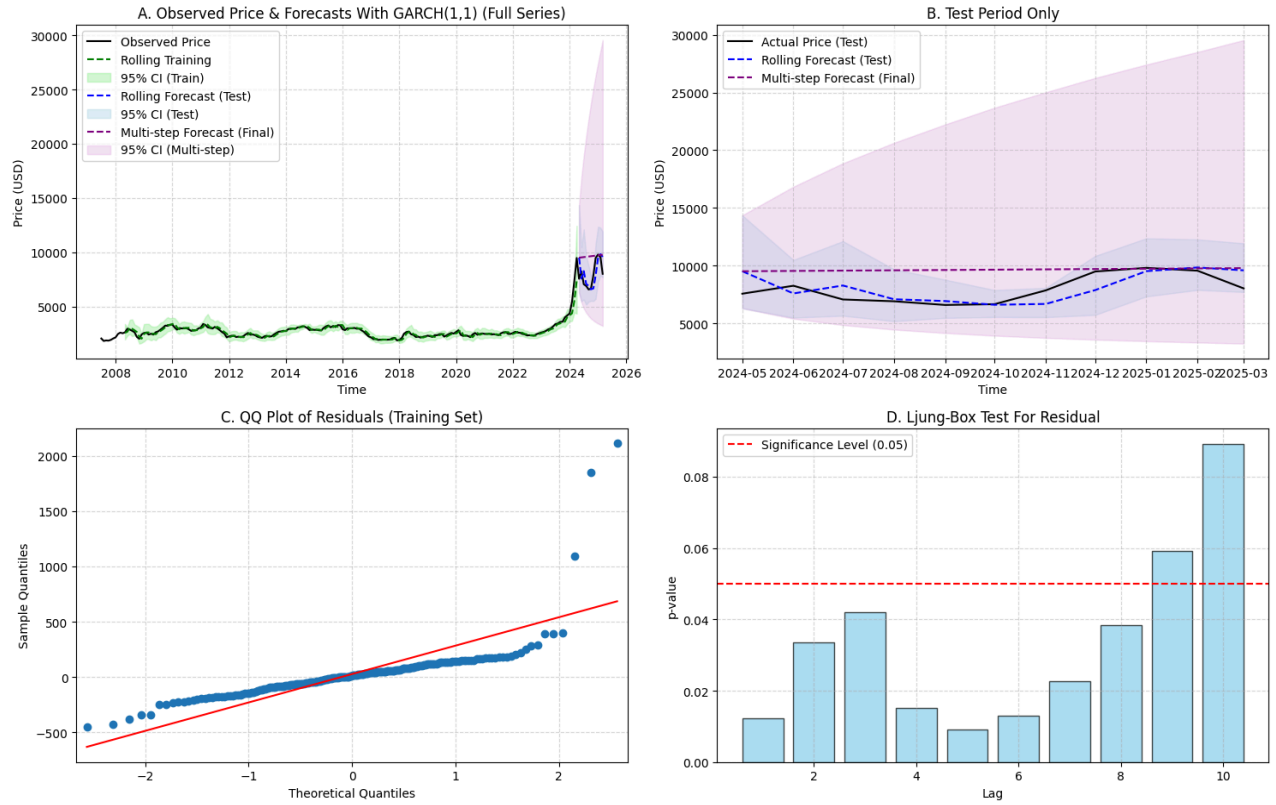
(D) Ljung-Box test on residuals.

Figure 3. Forecasting Performance of GARCH(1,1) Model on Cocoa Prices
(A) Observed monthly cocoa prices and forecasts from a GARCH(1,1) model.
(B) Zoom-in of test set, comparing rolling one-step and multi-step forecasts with actual prices
(C) Q-Q plot reveals mild heavy tails, indicating non-normal residuals.
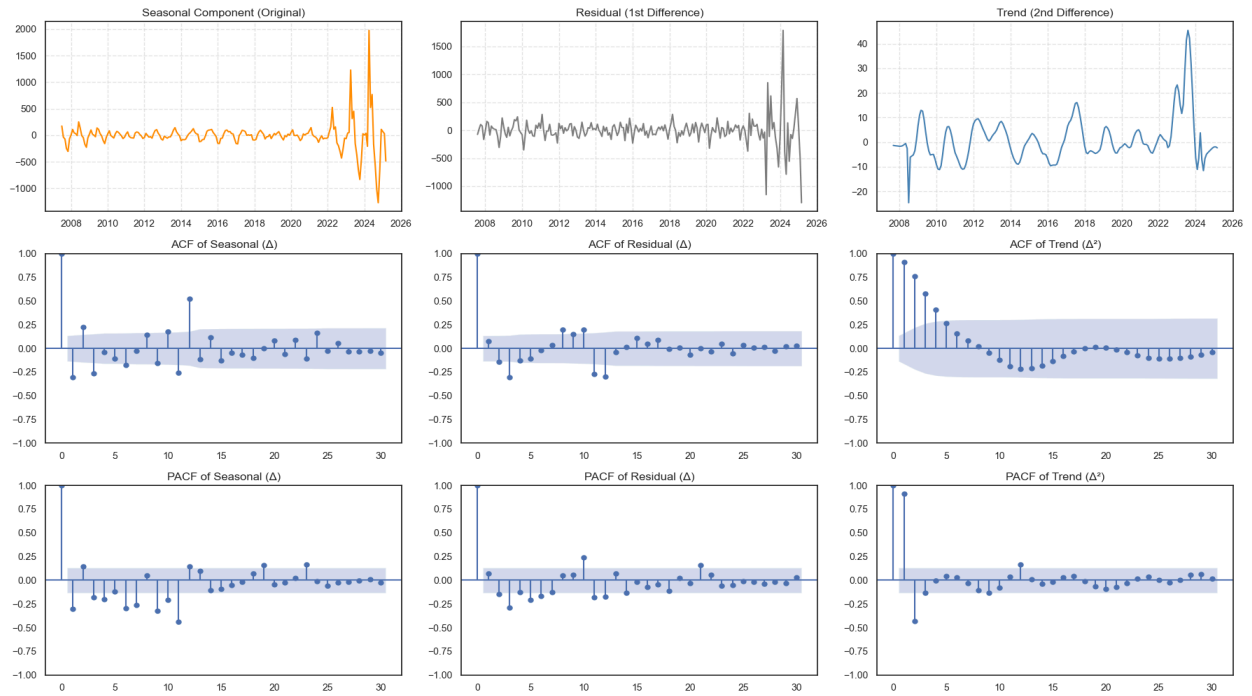**(D)** Ljung-Box test on residuals.

Figure 4. STL-ARIMA Decomposition Components ACF & PACF
(A–C) Seasonal component: original series, ACF (Δ), PACF (Δ).
(D–F) Residual component: first-differenced series, ACF (Δ), PACF (Δ).
(G–I) Trend component: second-differenced series, ACF (Δ²), PACF (Δ²).

8.3 Additional Forumla and Methodology

8.3.1 GARCH (Generalized Autoregressive Conditional Heteroskedasticity)

GARCH models are used to model time series data that exhibit time-varying volatility (heteroskedasticity). A typical GARCH(1,1) model is defined as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where:
- σ_t^2 is the conditional variance at time t,
- ϵ_{t-1} is the previous period's error term,
- $\alpha_0 > 0$ is a constant,
- $\alpha_1 \geq 0$ measures the effect of past shocks, and
- $\beta_1 \geq 0$ measures the persistence of volatility.

8.3.2 STL (Seasonal and Trend Decomposition using Loess)

STL is a robust and versatile method for decomposing a time series into three components: - **Trend:** The long-term progression of the series. - **Seasonality:** The repeating short-term cycle. - **Remainder:** The residual or irregular component.

STL's flexibility makes it suitable for series with complex seasonal patterns and varying trends.

8.3.3 ARIMA (Autoregressive Integrated Moving Average)

ARIMA models combine autoregression (AR), differencing (I for integration), and moving average (MA) to forecast time series data. An ARIMA(p, d, q) model is typically expressed as:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right) \epsilon_t$$

where:
- L is the lag operator,
- y_t is the time series,
- ϕ_i are the autoregressive coefficients,
- d is the order of differencing,
- θ_j are the moving average coefficients, and
- ϵ_t is the white noise error term.

8.3.4 LSTM (Long Short-Term Memory)

LSTM networks are a type of Recurrent Neural Network (RNN) designed to handle long-term dependencies in sequential data. Key characteristics include:

- **Gates:** LSTM cells use input, forget, and output gates to control information flow.
- **Optimization:** The network is trained using the Adam optimizer, which adjusts the learning rate adaptively.
- **Loss Function:** Mean Squared Error (MSE) is employed to measure the information loss during training.

8.3.5 Root Mean Square Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

- **Interpretation:** RMSE provides a measure of the average magnitude of the forecasting error, with larger errors being more heavily penalized.

8.3.6 Mean Absolute Percentage Error (MAPE)

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- **Interpretation:** MAPE expresses the error as a percentage, offering an intuitive measure of prediction accuracy relative to the actual values.

8.3.7 Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

- **Interpretation:** MAE calculates the average absolute difference between the actual and predicted values, providing a straightforward measure of error.