

# Cocoa Price Forecasting in Volatile Markets\*

## Why XGBoost Outperforms Classical Time Series Models

Jessica Lu, Solai Ramusarayanan, Cristina Su Lam

April 4, 2025

### Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Literature Review</b>	<b>2</b>
<b>3</b>	<b>Data</b>	<b>4</b>
3.1	Overview . . . . .	4
3.2	Exploratory Data Analysis . . . . .	5
<b>4</b>	<b>Methodology</b>	<b>10</b>
4.1	Model Training and Validation . . . . .	10
4.2	Model Diagnostics . . . . .	11
<b>5</b>	<b>Forecasting and Results</b>	<b>11</b>
5.1	Forecast Performance and Model Comparison . . . . .	11
5.2	Residual Diagnostics . . . . .	13
<b>6</b>	<b>Discussion and Conclusion</b>	<b>15</b>
6.1	Implications . . . . .	15
6.2	Limitations and Areas for Improvement . . . . .	16
6.3	Conclusion . . . . .	16
	<b>Appendix</b>	<b>17</b>
6.0.1	Correlation Matrix of Cocoa Predictor Variables . . . . .	17
6.0.2	Residual Diagnostics for Classical Time Series Models . . . . .	18
	<b>References</b>	<b>20</b>

---

\*Code and data are available at: [https://github.com/cristinaasu/cocoa\\_prices](https://github.com/cristinaasu/cocoa_prices)

# 1 Introduction

Commodity prices play a crucial role in the global economy, influencing industries ranging from agriculture to finance. Cocoa, in particular, is essential to chocolate production and other food-related sectors. However its price is notoriously volatile, driven by weather variability, supply chain disruptions, geopolitical tensions, and shifts in global demand. Accurately predicting these fluctuations is vital for stakeholders such as farmers, traders, policymakers, and businesses that rely on stable pricing for procurement, budgeting, and strategic planning.

The motivation for this analysis lies in the need to evaluate which forecasting approaches are best suited for commodities influenced by both climate and currency shifts. This study responds to that challenge by comparing models that integrate exogenous variables, such as rainfall and exchange rates, to better anticipate rapid price movements and provide more reliable insights for decision-making.

In light of this, the objective is to assess the effectiveness of different forecasting strategies under conditions of high volatility and structural complexity. To do this, we combine historical cocoa price data from the International Cocoa Organization (2025) with weather records from Ghana—the world’s largest cocoa producer (National Centers for Environmental Information 2025) and exchange rate data between the US dollar and the Ghanaian cedi (Investing.Com Canada 2025).

We compare four models: three statistical approaches (ETS, ARIMAX, SARIMAX) and one machine learning technique (XGBoost). These were selected to capture both linear and non-linear patterns in the data. Feature engineering techniques, including log transformation, differencing, and lag creation, were applied to enhance model performance. Preliminary results suggest that XGBoost outperforms the other models, achieving the lowest RMSE (278.41) and capturing the 2023–2024 price surge more effectively. This reinforces the value of incorporating nonlinear dynamics and external drivers when forecasting complex commodity markets.

The remainder of the paper is structured as follows: Section 2 reviews relevant literature on commodity price forecasting. Section 3 describes the data sources and preprocessing steps. Section 4 outlines the modeling approaches. Section 5 presents model performance and findings. Section 6 concludes with key insights and directions for future research.

## 2 Literature Review

Forecasting approaches for agricultural commodities have expanded in recent years to better accommodate the irregular patterns and sudden disruptions characteristic of global markets. Researchers have applied a range of time series techniques, from traditional statistical models to more flexible machine learning methods, to capture different aspects of price behavior.

This section outlines the importance of incorporating external variables such as weather conditions and currency exchange rates, and introduces several relevant models for cocoa price forecasting.

In Ghana, climate variability, particularly shifts in rainfall and temperature, has been shown to significantly affect yields. While projections vary by region, global studies suggest that cocoa-growing areas in West Africa may experience yield declines of up to 40% by the end of the century due to climate change. Excessive rainfall also increases the risk of fungal diseases, lowering crop quality (Bomdzele and Molua 2023). On the economic side, fluctuations in the USD/GHS exchange rate impact the profitability of cocoa exports; depreciation of the cedi raises local cocoa prices and affects producer margins (Zimwara 2024). Including these variables in forecasting models enables more realistic simulations of future price behavior and enhances their ability to capture real-world market dynamics.

To address these challenges, we evaluate the following models. First, the Exponential Smoothing State Space (ETS) model is a widely used statistical method that captures level, trend, and seasonal patterns. Due to its computational efficiency and simplicity, it is especially suited for relatively stable time series. However, it lacks the ability to include external variables, which limits its effectiveness in settings influenced by weather and economic shifts. In this study, ETS serves primarily as a baseline model to assess the value of incorporating exogenous regressors.

ARIMA is a classical approach that captures univariate time series patterns through autoregressive terms, differencing, and moving averages. Dooley and Lenihan (2005) demonstrated its effectiveness in modeling historical price trends for metals. Although not used independently here, it provides the structural basis for two extended forms applied in our analysis.

The first of these, ARIMAX, incorporates exogenous variables to account for influences beyond the internal structure of the series. This makes it especially relevant for price forecasting when environmental or macroeconomic drivers are involved. In this project, ARIMAX is used to examine how cocoa prices respond to rainfall and exchange rate fluctuations, without accounting for seasonality.

SARIMAX extends this framework further by modeling seasonality alongside external drivers. This added complexity makes it well-suited to crops like cocoa that follow distinct agricultural cycles. Alharbi and Csala (2022) found the model effective in forecasting electricity sector performance; by incorporating both economic and seasonal indicators, it achieved higher accuracy than simpler ARIMA-based techniques. By accounting for recurring patterns and structural influences, SARIMAX is expected to provide more reliable forecasts during cyclical phases of cocoa production.

XGBoost is a machine learning algorithm based on gradient-boosted decision trees. It is designed to handle high-dimensional and nonlinear datasets, making it a powerful alternative to classical models. Trinh et al. (2021) demonstrated XGBoost’s effectiveness in financial time series forecasting, particularly when combined with deep learning models. Similarly, Lago, Marcjasz, and Weron (2021) showed that it consistently outperformed classical time

series approaches in the context of electricity markets. Its ability to model interactions and sudden shifts in data makes it especially useful when markets are driven by complex, nonlinear forces.

What distinguishes our analysis is its side-by-side comparison of four modeling strategies applied to a single dataset enriched with weather and exchange rate variables. While previous studies often focus on one or two approaches, this broader evaluation reveals how each model performs under different levels of volatility and external disruption. Consistent with findings from Lago, Marcjasz, and Weron (2021), our results reinforce the strength of XGBoost in capturing complex, nonlinear market dynamics. In contrast to earlier work that prioritizes accuracy in one modeling framework, our comparative design provides insights into how different tools behave across real-world shocks and cycles. By combining exogenous predictors with a range of modeling techniques, this study extends prior research and underscores the value of machine learning in forecasting commodity price behavior.

## 3 Data

### 3.1 Overview

This analysis uses three main datasets: daily cocoa futures prices, weather data from Ghana, and USD to Ghanaian Cedi (GHS) exchange rates.

Data preparation and visualization were conducted in R (R Core Team 2022) using the tidyverse (Wickham et al. 2019) and lubridate (Grolemund and Wickham 2011) packages. Exploratory analysis included correlation plots generated with corrplot (Wei and Simko 2021) and spectral analysis performed using astsa (Stoffer 2025).

The cocoa price data from International Cocoa Organization (2025) spans daily prices from 1994 to 2025 and reports prices in US dollars per tonne. Before analysis, formatting symbols such as commas were removed, and values were converted to numeric. The date column was parsed into proper date format and arranged in chronological order. To address non-stationarity and heteroscedasticity, we applied a logarithmic transformation to stabilize variance—a common preprocessing step for time series models.

Weather data from National Centers for Environmental Information (2025) includes daily observations from 1990 to 2024, with measurements for rainfall (PRCP), average temperature (TAVG), maximum temperature (TMAX), and minimum temperature (TMIN) collected across multiple weather stations in Ghana. These were aggregated by date using daily means across all stations. The date column was standardized to match the cocoa dataset.

Exchange rate data were obtained from Investing.Com Canada (2025), providing daily USD to GHS rates from 1994 to 2024. Only relevant columns—Date and Price—were kept, and the price column was renamed to ExchangeRate to match the variable naming conventions of the other datasets.

The datasets were then merged by date to create a unified time series. Rows with missing values in any of the key columns (price, weather, or exchange rate) were removed to ensure model compatibility—this constituted our handling of missing data. The result was a clean dataset with 2,911 daily observations, containing complete information for all three sources. This cleaned dataset was used to train the ETS, ARIMAX, and SARIMAX models.

To support feature engineering for the XGBoost model, a second version of the dataset was created. This version included lagged features representing cocoa prices from 1 to 7 days earlier, allowing the model to learn temporal patterns from past values. All numeric values were also rounded to three decimal places to ensure consistency. After accounting for lagged data loss, the final version used for XGBoost included 2,904 observations.

These preparation steps—cleaning, organizing, and creating new features—lay the groundwork for the modeling described in Section 4.

Samples of both datasets are shown in Table 1 and Table 2. For brevity, only lags 1 to 2 are shown from the lagged dataset.

Table 1: Sample of Cocoa Dataset

Date	Price	PRCP	TAVG	TMAX	TMIN	Exchange Rate	Log Price
1994-10-12	1411.61	0.71	75.67	81.00	69.00	0.1005	7.2525
1994-10-14	1415.50	0.30	80.40	86.50	71.67	0.1005	7.2552
1994-10-27	1497.14	0.27	79.40	85.40	71.50	0.1020	7.3113
1994-11-01	1464.60	0.33	80.00	88.25	70.40	0.1020	7.2893
1994-11-07	1426.50	0.00	78.00	96.50	71.00	0.0972	7.2630

Table 2: Sample of Cocoa Dataset with Lagged Features

Date	Price	PRCP	TAVG	TMAX	TMIN	Exchange Rate	Log Price	Lag 1	Lag 2
1994-11-18	1455.18	0.00	81.29	91.00	73.14	0.102	7.283	7.288	7.267
1995-03-09	1507.64	0.96	81.75	90.62	72.00	0.105	7.318	7.283	7.288
1995-03-13	1500.62	0.00	80.50	92.00	76.00	0.107	7.314	7.318	7.283
1995-03-15	1488.67	0.52	79.29	89.75	72.80	0.107	7.306	7.314	7.318
1995-03-16	1466.61	0.01	85.00	93.29	73.29	0.107	7.291	7.306	7.314

### 3.2 Exploratory Data Analysis

To understand the underlying structure of the cocoa price series and its associated predictors, we conducted a set of visual and statistical analyses.

In Figure 1, we begin with the time series plot of cocoa prices from 1994 to 2025, which shows long periods of relative stability followed by an unprecedented surge after 2023. This sharp spike underscores the need for models that can accommodate structural breaks, volatility, and external shocks. According to Table 3, cocoa prices ranged from approximately \$778 to \$10,690 per tonne, with a mean of \$2,590.

Table 3: Summary Statistics

Statistic	Log Price	Price	PRCP	TAVG	TMAX	TMIN	ExchangeRate
Min.	6.66	778.43	0.00	73.60	76.50	61.00	0.10
1st Qu.	7.43	1689.41	0.00	78.56	85.50	72.57	0.89
Median	7.75	2330.95	0.08	80.50	88.67	73.67	3.45
Mean	7.77	2589.71	0.25	80.62	88.44	73.86	4.19
3rd Qu.	7.98	2931.30	0.30	82.60	91.25	75.00	5.74
Max.	9.28	10690.70	10.28	88.00	101.00	82.00	16.30



Figure 1: Cocoa Price Trend Over Time

Cocoa prices are highly skewed, with extreme values pulling the mean well above the median. Hence, a logarithmic transformation was applied to address this. As shown in Figure 2, the resulting distribution is more symmetric and better suited for modeling.

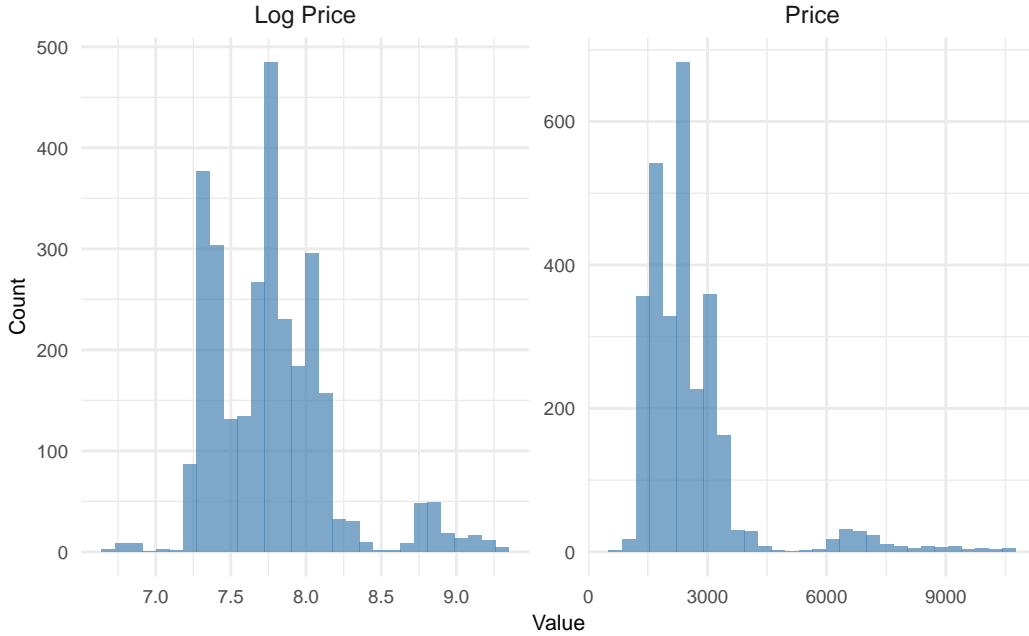


Figure 2: Histograms of Cocoa Prices

To further explore the characteristics of the predictors, Figure 3 presents their distributions. PRCP is heavily skewed toward zero, reflecting the prevalence of dry days. In contrast, the temperature variables are more normally distributed. The exchange rate distribution displays multiple peaks, pointing to periods of currency instability. Based on Table 3, the exchange rate fluctuated dramatically—from as low as 0.10 to over 16.3—capturing Ghana’s significant macroeconomic shifts during the study period.

See in Section 6.0.1, the figure illustrates the linear relationships among variables. Temperature metrics, particularly average and maximum, are closely linked, with a coefficient of 0.89. This makes sense since hotter days typically push up both values. The exchange rate shows a notable association with cocoa prices at 0.73, highlighting its potential as a useful predictor. In contrast, precipitation has a much weaker connection at just 0.05. This suggests its effects may be indirect or delayed, and likely better captured by non-linear or time-lagged models.

To break down structural components in the price series, we applied STL decomposition (Seasonal and Trend decomposition using Loess). This method separates the series into three parts: long-term trend, recurring seasonal patterns, and short-term residuals. Figure 4 presents that the seasonal component displays a clear and consistent cyclical pattern, while the trend remains relatively flat before accelerating upward. The residuals also show periods of heightened volatility toward the end of the series. These features reinforce the use of models like SARI-MAX, which explicitly incorporate both seasonality and external predictors to capture such patterns.

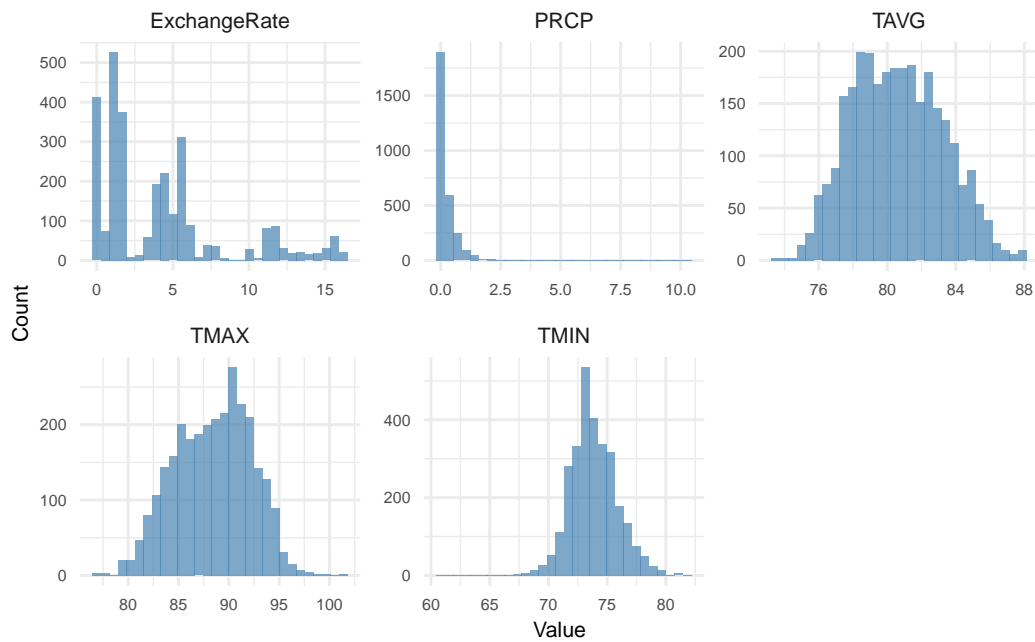


Figure 3: Histograms of Cocoa Predictor Variables

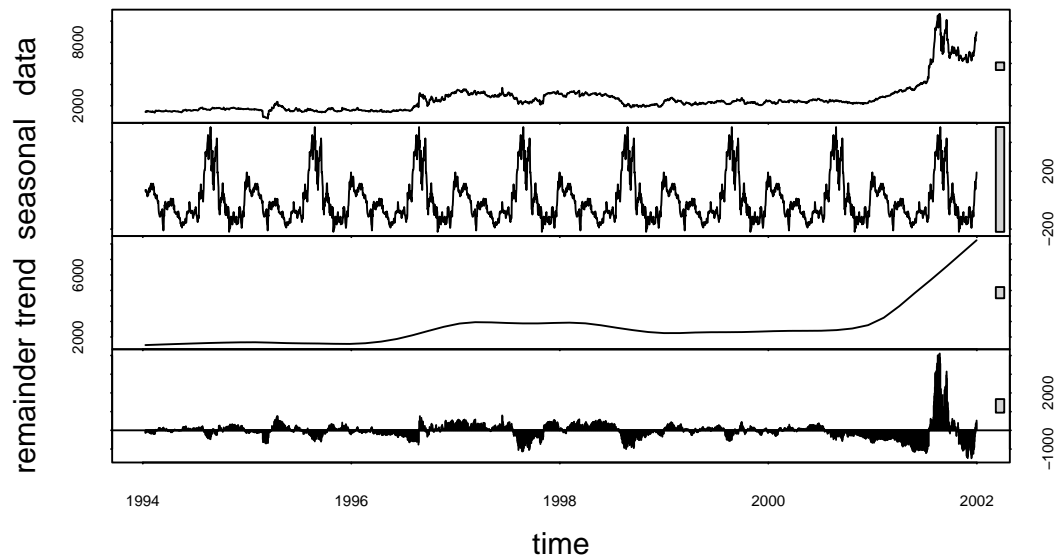


Figure 4: Seasonal Decomposition of Cocoa Price



We then examined temporal dependence in cocoa prices using an autocorrelation function (ACF) plot. In Figure 5, autocorrelation remains high even beyond 30 lags, with a slow and steady decay. This indicates strong persistence, where past prices continue to influence future values over time. As noted in the final part of Section 3.1, we addressed this by creating lagged versions of the log-transformed prices, allowing the XGBoost model to better capture recent momentum.

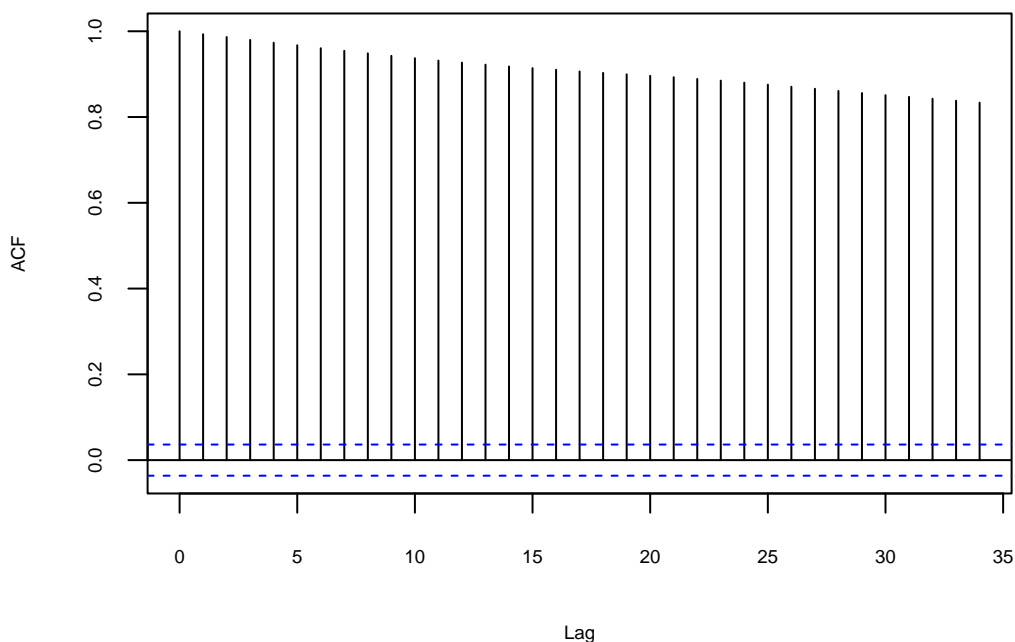


Figure 5: Autocorrelation of Cocoa Prices

To explore the frequency structure of the series, we used both a periodogram and a smoothed periodogram. Figure 6 reveal strong signals at low frequencies, suggesting that price movements tend to evolve gradually rather than through short, abrupt swings. This supports the use of models like ETS and SARIMAX, which are well suited for capturing long-term trends and recurring seasonal patterns

Together, these exploratory steps informed our model selection and structure by confirming the importance of long-term trends, recurring seasonal behavior, autocorrelation, and external drivers—all of which are essential for building robust and interpretable forecasting models.

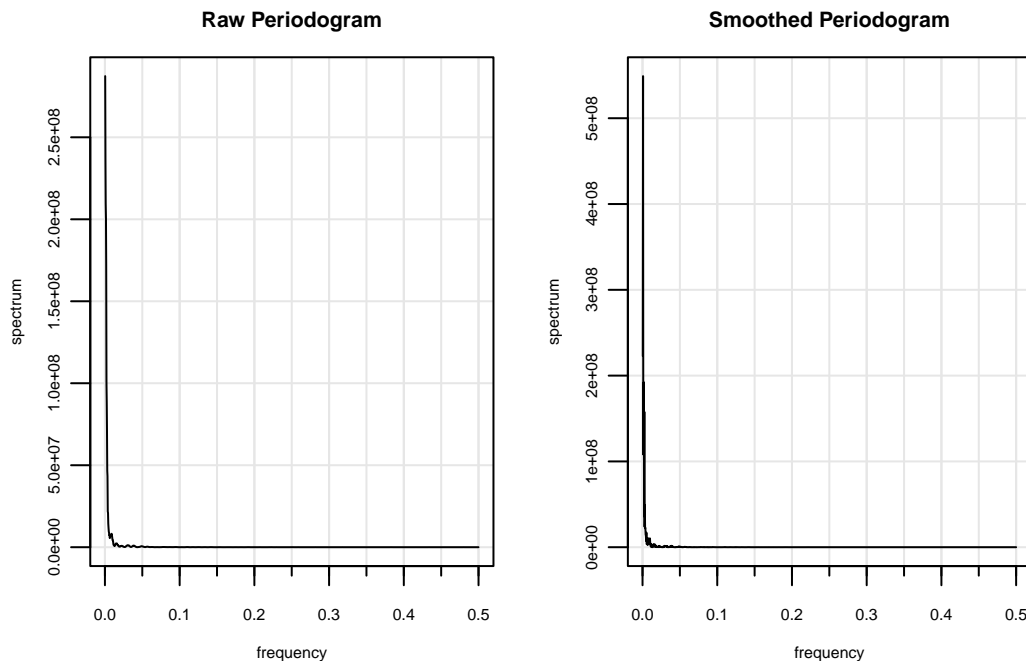


Figure 6: Both have a similar pattern

## 4 Methodology

As outlined in Section 1, we forecast daily cocoa prices using a comparative modeling approach that includes three classical time series models—ETS, ARIMAX, and SARIMAX—alongside the machine learning algorithm XGBoost. Model selection was guided by theoretical considerations and prior research, as discussed in Section 2. This section details our model setup and evaluation procedures.

We implemented the models using the following R packages: `forecast` (Hyndman et al. 2024) for classical time series methods, and `xgboost` (Chen et al. 2025) for machine learning.

### 4.1 Model Training and Validation

Model training was performed separately for each approach using a time-based 80/20 train-test split to preserve temporal ordering. The final 20% of the data was held out to evaluate out-of-sample forecasting performance.

ETS was trained on raw (non-log-transformed) cocoa prices and served as a benchmark model. It captured level, trend, and seasonality, without incorporating external predictors.

ARIMAX included five exogenous variables: precipitation (PRCP), average (TAVG), maximum (TMAX), and minimum (TMIN) temperatures, along with the USD/GHS exchange rate. These regressors were input as matrices during both training and forecasting. SARIMAX extended this setup by incorporating seasonal components, supported by patterns observed in the STL decomposition and periodogram.

For these classical models, the response variable was kept on the original price scale to maintain interpretability. The `auto.arima()` function automatically applied necessary transformations (e.g., differencing or drift) and selected the optimal model structure based on AIC/BIC.

XGBoost was trained on log-transformed cocoa prices to stabilize variance, reduce the impact of extreme values, and focus the model on relative changes. It used lagged price features (lags 1 through 7), weather variables, and the exchange rate as predictors. Training followed an expanding walk-forward validation approach: at each time step, the model was retrained using all data up to that point and used to predict the next value—mimicking real-time forecasting. The resulting predictions were exponentiated to return to the original price scale.

Performance for ETS, ARIMAX, and SARIMAX was assessed using the forecast package’s built-in metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). For XGBoost, RMSE was computed on exponentiated predictions to ensure comparability with the classical models.

## 4.2 Model Diagnostics

To evaluate statistical reliability and check for model violations, we conducted residual diagnostics for each approach. Residuals were plotted over time to inspect for visible patterns or systematic biases. Autocorrelation was assessed using ACF plots and formally tested with Ljung–Box tests to evaluate independence. To assess normality, a key assumption in many statistical forecasting models, histograms and Q–Q plots were used.

These diagnostic checks helped identify potential weaknesses in model fit and supported more reliable interpretation of results. By combining predictive performance with diagnostic soundness, this framework enabled a more robust and well-rounded comparison across modeling approaches.

# 5 Forecasting and Results

## 5.1 Forecast Performance and Model Comparison

While all models aimed to track overall price trends, their ability to handle sharp movements varied. As seen in Figure 7, the classical models failed to keep up with the 2024 surge, producing forecasts that smoothed over major shifts. ARIMAX incorporated external variables, and

SARIMAX added seasonality on top—but their results were identical. Both returned the exact same test metrics, offering only a marginal improvement over ETS, as shown in Table 4.

XGBoost, by contrast, closely tracked both the rapid acceleration and the subsequent volatility in prices. Its ability to integrate lagged features and nonlinear relationships allowed it to react more dynamically to recent changes. This adaptability translated into a dramatic performance gain: XGBoost reduced RMSE by nearly 90% compared to the classical models, signaling a substantial improvement in predictive accuracy.

Table 4: Forecast Accuracy on Test Set

Model	RMSE	MAE	MAPE
ETS	2795.25	1768.68	48.56
ARIMAX	2639.67	1661.65	27.73
SARIMAX	2639.67	1661.65	27.73
XGBoost	285.92	NA	NA

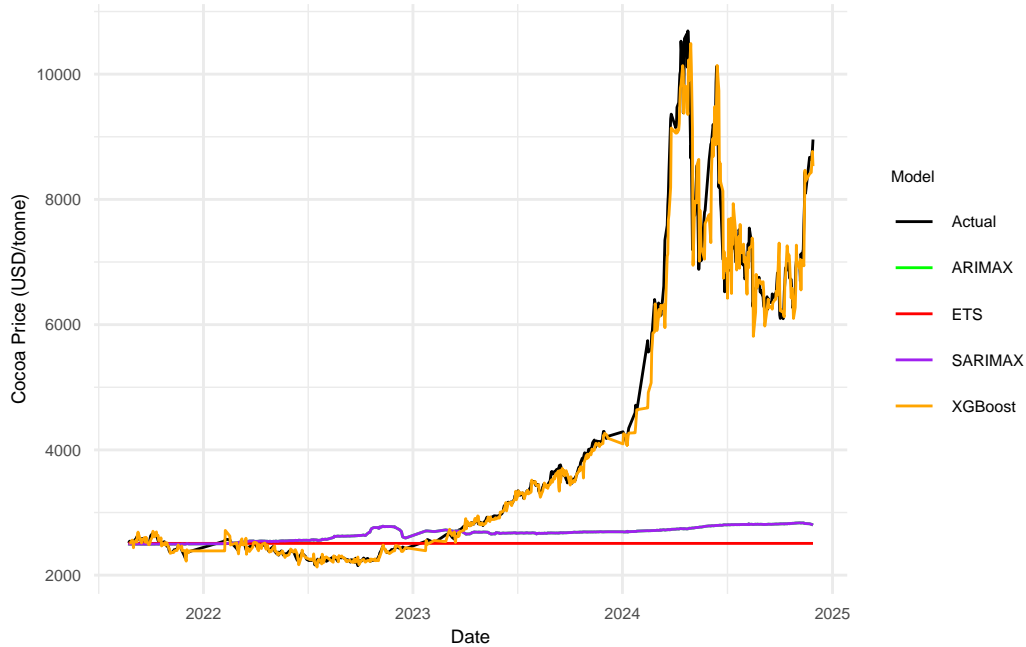


Figure 7: Forecast vs Actual Cocoa Prices

## 5.2 Residual Diagnostics

Given XGBoost’s superior accuracy, we focus diagnostics on this model. Visualizations and metrics for the remaining models are provided in the Appendix.

As represented in Figure 8, residuals stayed close to zero during stable periods but widened significantly following the 2024 price spike—reflecting greater volatility in both prices and prediction errors. In Figure 9, mild autocorrelation is observed at short lags (1–2), but it quickly decays, indicating that most short-term dependencies were successfully captured.

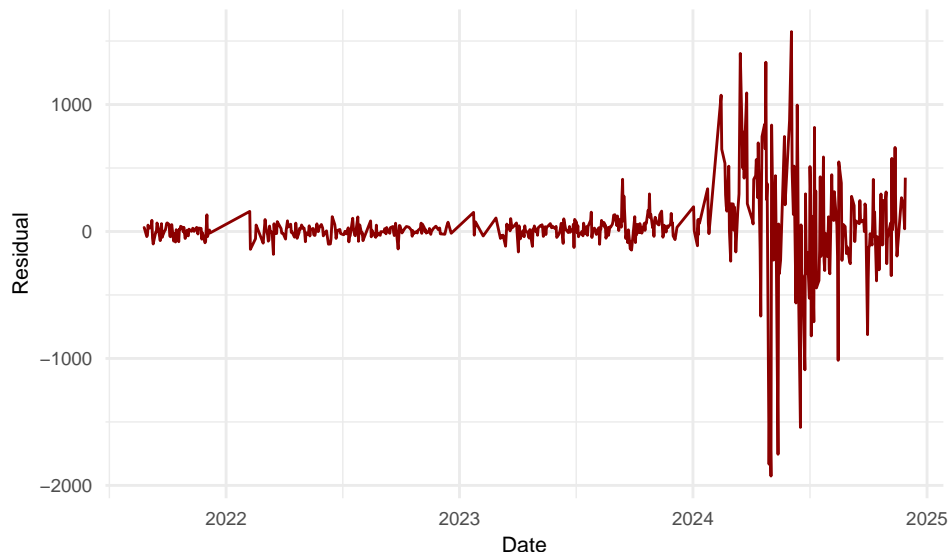


Figure 8: XGBoost Residuals Over Time

However, the Ljung–Box test revealed statistically significant autocorrelation ( $p < 2.2\text{e-}16$ ), suggesting some structure remains unexplained.

Model	Statistic	p.value
ETS	6633.900	$< 2.2\text{e-}16$
ARIMAX	6491.000	$< 2.2\text{e-}16$
SARIMAX	6491.000	$< 2.2\text{e-}16$
XGBoost	96.668	$2.44\text{e-}15$

Normality diagnostics in Figure 10 show a residual distribution that peaks sharply near zero but has heavy tails. While the Q–Q plot aligns with the normal line at the center, deviations in the tails confirm occasional large forecasting errors.

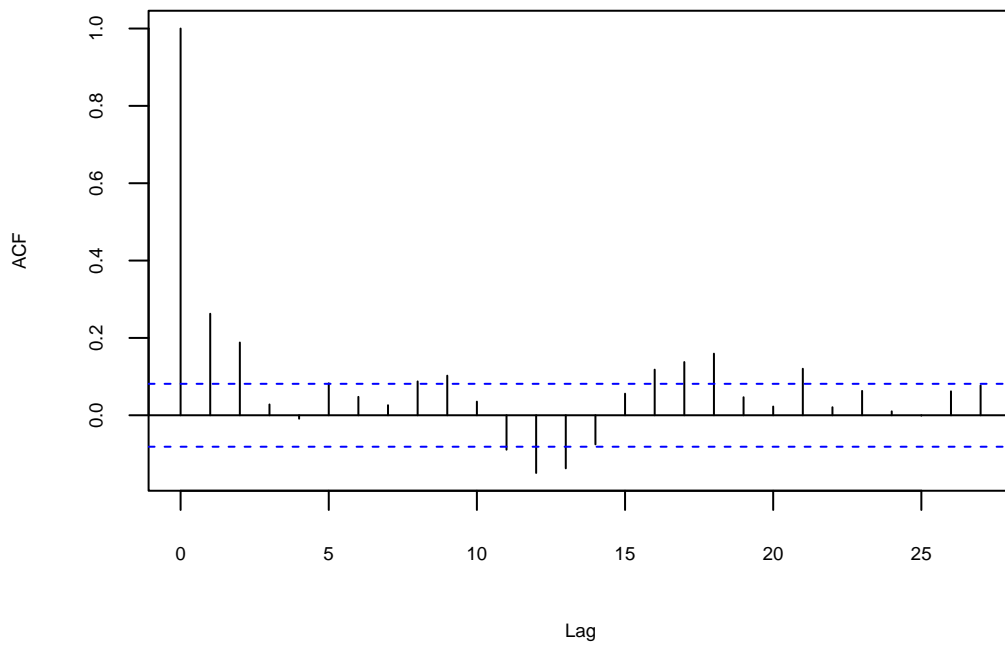


Figure 9: Autocorrelation of XGBoost Residuals

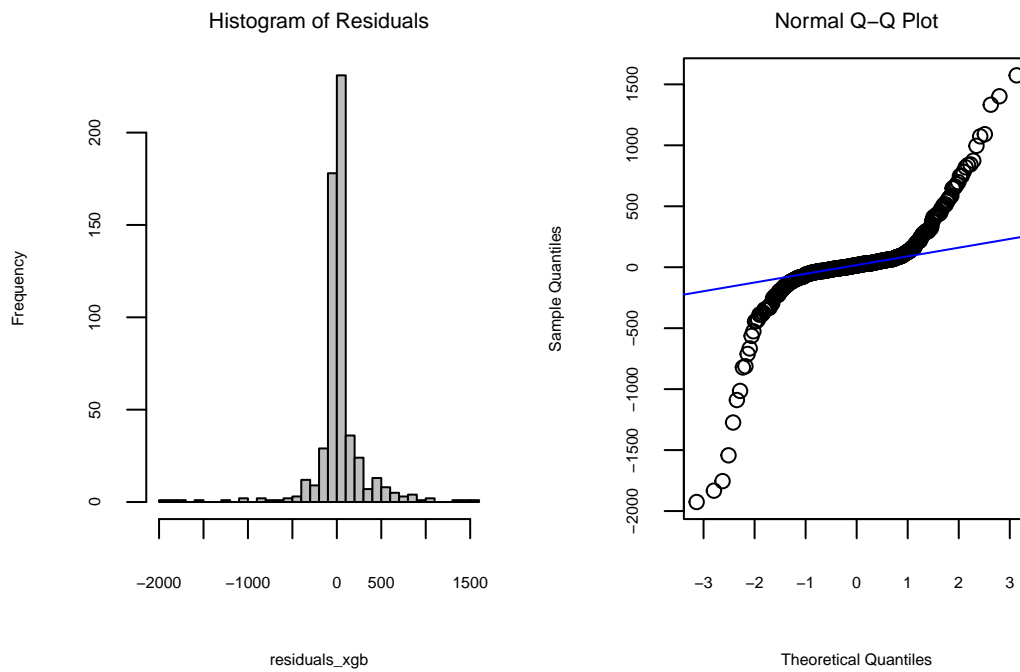


Figure 10: Evaluating Normality of XGBoost Residuals

Altogether, these results indicate that XGBoost delivered highly accurate forecasts and captured key nonlinearities, but further refinement may be needed to fully model extreme market movements.

## 6 Discussion and Conclusion

As is evident from Section 5, there is a clear divide between classical time series models and machine learning approaches in handling commodity price forecasting under volatile conditions. While ETS, ARIMAX, and SARIMAX served as reasonable benchmarks, XGBoost clearly outperformed them—demonstrating a far greater ability to respond to real-time shifts, especially during the steep 2024 surge.

### 6.1 Implications

- Forecast Accuracy

The significant gap in RMSE between XGBoost and the classical models emphasizes how important model flexibility is in volatile settings. Unlike traditional methods that tended to smooth over sharp changes, XGBoost was able to adjust forecasts as new data came in—something crucial when markets are unstable.

- Integration of External Factors

XGBoost's edge also came from how well it handled multiple external inputs. Daily weather conditions and currency fluctuations were incorporated directly into the model, helping it adjust dynamically in ways ETS and even SARIMAX could not. While ARIMAX and SARIMAX tried to address this through exogenous regressors, the results were nearly identical—suggesting that their rigid structure limited how much external information could actually help.

- Residual Structure and Predictive Limits

Despite its strong performance, diagnostic tests showed that XGBoost's residuals were not perfect. There was some autocorrelation, and the error distribution had heavy tails—signs that the model missed some extreme fluctuations. This suggests future improvements could come from adding new features, refining lag structure, or even using ensemble methods to better handle tail risks.

- Real-World Relevance

The approach used here reflects what real-world forecasting often looks like: walk-forward validation, daily retraining, and reliance on publicly available signals. That makes these findings not only academically interesting, but also actionable for use in commodity risk management, trading, or agricultural planning.

## 6.2 Limitations and Areas for Improvement

- Model Interpretability vs. Performance

While XGBoost delivered strong predictive performance, interpretability remains a key trade-off. In contrast to ETS or ARIMAX—where model structure and component behavior are more transparent—tree-based ensembles function more like black boxes. This lack of interpretability may be a limitation when forecasts need to be communicated to stakeholders or policymakers. Future research could explore hybrid methods that strike a better balance between explainability and predictive power.

- Parameter Tuning and Diagnostic Refinement:

Although hyperparameters were selected using walk-forward validation, the residual diagnostics suggest further tuning is warranted. The presence of autocorrelation and heavy tails indicates that certain temporal or structural patterns were not fully captured. Additional tuning, paired with more targeted out-of-sample testing, may help reduce residual noise and improve generalizability.

- Expanding Feature Space

While weather and exchange rates enhanced model performance, other external factors—such as macroeconomic indicators, commodity sentiment, or supply chain disruptions—may help capture extreme market shifts. Incorporating such features could further improve the model’s responsiveness to shocks and improve accuracy during periods of instability.

## 6.3 Conclusion

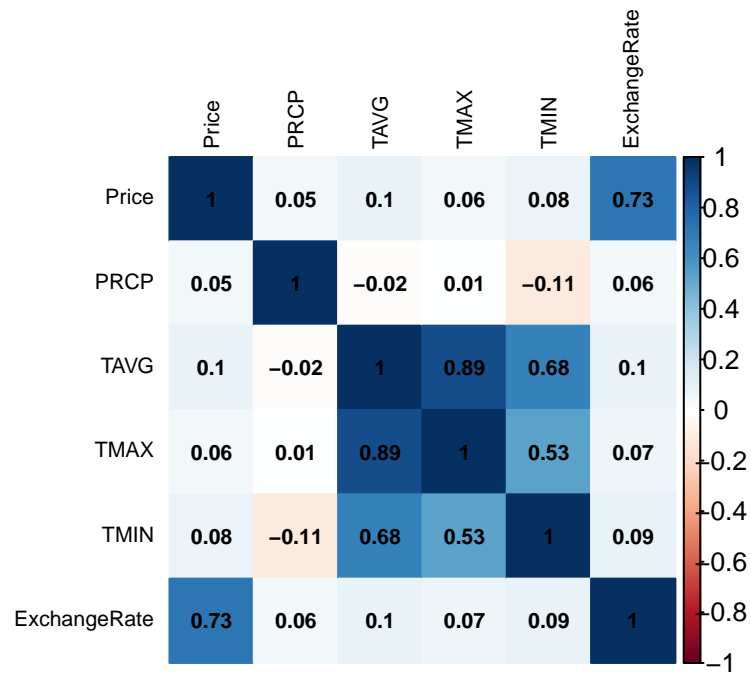
This study shows that cocoa price forecasting benefits significantly from flexible modeling strategies that integrate both recent historical behavior and external drivers. Among the four models evaluated, XGBoost proved most effective, substantially reducing prediction error and better capturing sharp market movements.

While classical time series models remain valuable for their simplicity and interpretability, machine learning approaches like XGBoost offer clear advantages in volatile environments. These findings hold practical relevance for traders, analysts, and decision-makers who rely on timely and accurate forecasts to navigate uncertainty in global commodity markets.



## Appendix

### 6.0.1 Correlation Matrix of Cocoa Predictor Variables



## 6.0.2 Residual Diagnostics for Classical Time Series Models

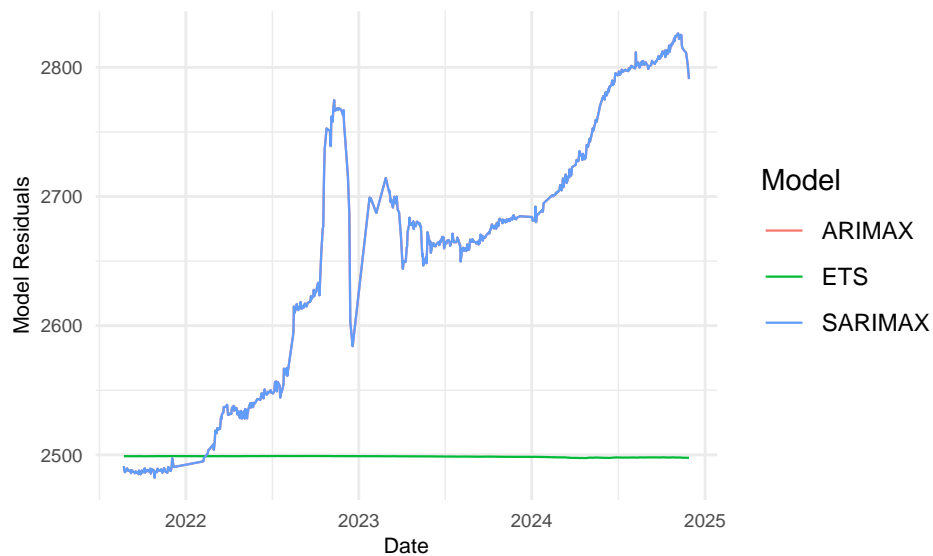
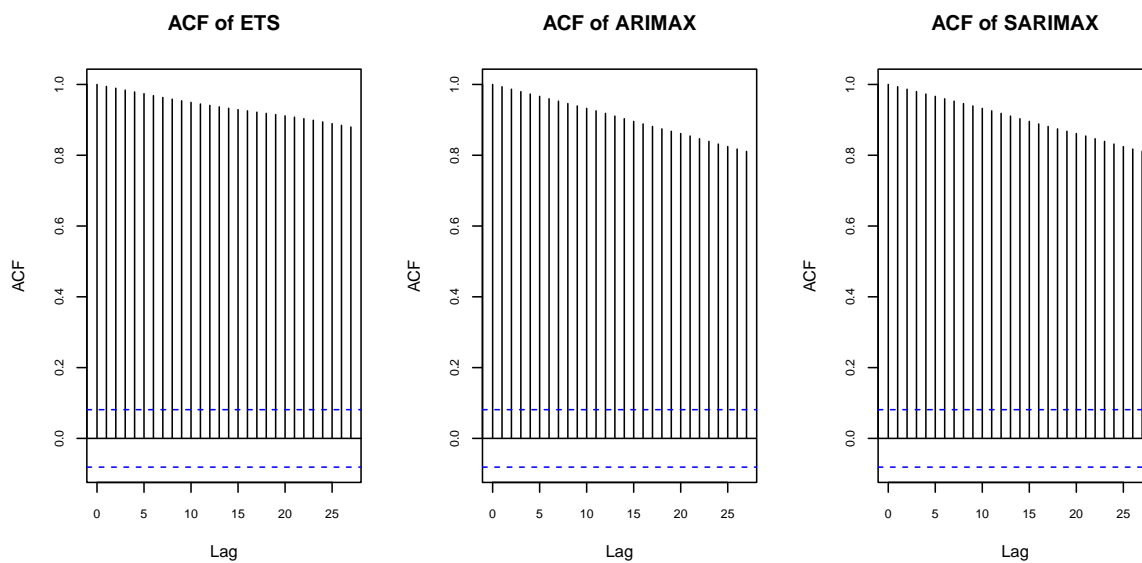


Figure 11: Residuals Over Time



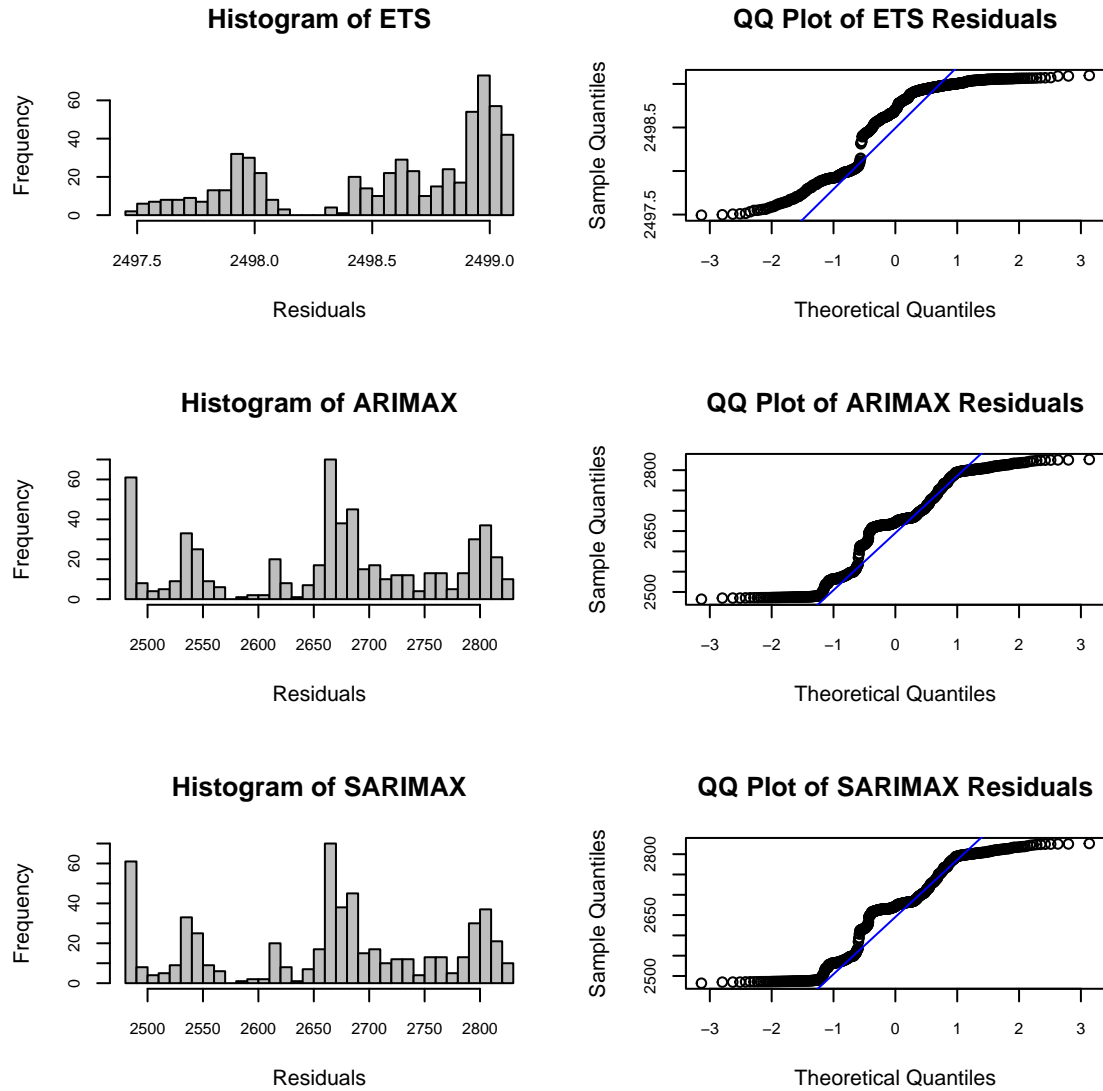


Figure 12: Residual Diagnostics for ETS, ARIMAX, and SARIMAX

## References

- Alharbi, F. R., and D. Csala. 2022. “A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach.” *Inventions* 7 (4): 94. <https://doi.org/10.3390/inventions7040094>.
- Bomdzele, E., and E. L. Molua. 2023. “Assessment of the Impact of Climate and Non-Climatic Parameters on Cocoa Production: A Contextual Analysis for Cameroon.” *Frontiers in Climate* 5. <https://doi.org/10.3389/fclim.2023.1069514>.
- Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, et al. 2025. *Xgboost: Extreme Gradient Boosting*. <https://CRAN.R-project.org/package=xgboost>.
- Dooley, G., and H. Lenihan. 2005. “An Assessment of Time Series Methods in Metal Price Forecasting.” *Resources Policy* 30 (3): 208–17.
- Grolemund, G., and H. Wickham. 2011. “Dates and Times Made Easy with Lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Hyndman, R. J., G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O’Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmien. 2024. *Forecast: Forecasting Functions for Time Series and Linear Models*. <https://pkg.robjhyndman.com/forecast/>.
- International Cocoa Organization. 2025. “Home.” <https://www.icco.org/>.
- Investing.Com Canada. 2025. “USD GHS Historical Data.” <https://ca.investing.com/currencies/usd-ghs-historical-data>.
- Lago, J., G. Marcjasz, and R. Weron. 2021. “Forecasting Day-Ahead Electricity Prices: A Review of State-of-the-Art Algorithms, Best Practices and an Open-Access Benchmark.” *Applied Energy* 293: 116983. <https://doi.org/10.1016/j.apenergy.2021.116983>.
- National Centers for Environmental Information. 2025. “National Centers for Environmental Information (NCEI).” <https://www.ncei.noaa.gov/>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Stoffer, D. 2025. *Astsa: Applied Statistical Time Series Analysis*. <https://CRAN.R-project.org/package=astsa>.
- Trinh, T. D., P. H. Vuong, T. K. Mai, P. H. Uyen, and P. T. Bao. 2021. “Stock-Price Forecasting Based on XGBoost and LSTM.” *Journal of Information Science and Engineering* 37 (5): 1149–65. [https://www.researchgate.net/publication/357486637\\_Stock-price\\_forecasting\\_based\\_on\\_XGBoost\\_and\\_LSTM](https://www.researchgate.net/publication/357486637_Stock-price_forecasting_based_on_XGBoost_and_LSTM).
- Wei, T., and V. Simko. 2021. *R Package ‘Corrplot’: Visualization of a Correlation Matrix*. <https://github.com/taiyun/corrplot>.
- Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Zimwara, T. 2024. “Ghana’s Cedi Hits Record Low Against US Dollar Amid Cocoa Output Decline – Economics Bitcoin News.” <https://news.bitcoin.com/ghanas-cedi-hits-record-low-against-us-dollar-amid-cocoa-output-decline/>.