**Forecasting International Cocoa Prices Using ETS, ARIMAX and SARIMAX, GARCH, and Multiple Linear Regression Models**

## 1. Introduction

Cocoa is one of the most economically significant agricultural commodities, especially for major producers in West Africa. Its international market is marked by sharp price fluctuations, driven by supply shocks, geopolitical instability, climate variability, and financial speculation. In recent years, price volatility has increased, underscoring the need for robust forecasting tools that can inform producers, policymakers, and investors.

Given the importance of climate to cocoa yield, particularly in countries like Ghana, there is growing interest in understanding how environmental factors influence short-term price dynamics. However, while climate indicators are theoretically important for agricultural commodities, their empirical forecasting value remains debated.

This study aims to forecast monthly cocoa futures prices using a range of time series models, including ARIMA, SARIMA, ETS, GARCH, and climate-augmented regressions. We evaluate the models based on out-of-sample forecast accuracy and examine the explanatory power of Ghanaian climate variables. Our analysis spans over 30 years of monthly price and weather data and focuses on the 2023–2024 period, when cocoa prices experienced historic spikes.

By comparing baseline time series models with those incorporating external climate inputs, we aim to assess whether environmental variation contributes meaningfully to short-run price prediction. Our findings contribute to ongoing research on volatility modeling in commodity markets and offer practical insights for decision-makers navigating uncertain supply conditions.

## 2. Literature Review

Recent studies on commodity price forecasting have primarily relied on ARIMA and GARCH-family models to capture trend and volatility structures. Assis and Caldeira (2021) showed that GARCH-type models outperform standard ARIMA models when volatility is time-varying, especially in commodity markets. Similarly, Kumar and Pandey (2013) applied ARIMA-GARCH models to agricultural prices and found improved performance during periods of high uncertainty.

Blanco et al. (2022) focused on income volatility in Argentina using relative income measures and rolling windows, which informs our decision to examine structural shifts and instability over time. Although their study was not forecasting-focused, it influenced our approach to exploring volatility patterns prior to modelling.

In terms of exogenous variables, Mensah and Alagidede (2017) explored how macroeconomic indicators influence cocoa prices in Ghana. Their findings support the inclusion of external factors, such as climate, although they also caution that global price movements often overshadow local effects. Similarly, Opoku et al. (2022) integrated weather and production data in machine learning models but found that climate variables alone had limited forecasting power.

Building on these insights, our study combines ARIMA, SARIMA, and GARCH models with a comparison to linear and regression-augmented alternatives. We contribute by evaluating the real-world performance of climate-augmented time series models on recent cocoa price shocks, using a robust multi-model framework.

### 3. Methodology

To capture various facets of cocoa price fluctuations, this study employed several forecasting models. These models, which were chosen to represent particular patterns or features seen in the data, include the ETS, ARIMAX and SARIMAX, ARIMA and GARCH, and regression models. We seek to determine the most effective method for predicting cocoa prices under various structural assumptions by contrasting models from several statistical families.

### ETS Model

In order to accommodate for additive errors without trend or seasonality, we employed an ETS(A,N,N) model. This model represents time series that oscillate around a fluctuating level without significant directional movement. Hyndman and Athanasopoulos (2018) claim that ETS models can be constructed as innovative state space models, in which the extent of recent prediction errors and the preceding level are used to update each forecast. Because it offered the greatest match among the potential exponential smoothing models, the ETS(A,N,N) model was chosen using AICc.

### ARIMAX and SARIMAX

We first looked at the data's structure in order to create ARIMA-type models. A steady increasing trend can be seen in the cocoa price time series plot, and strong seasonal patterns were found via STL decomposition(Figure 3.), suggesting consistent, recurrent swings throughout time. These findings support the idea that the series is nonstationary and affected by both seasonality and trend.

For nonstationary time series that can be differenced into stationary series, the ARIMAX(p,d,q) model works especially well. To eliminate trends and stabilize the variance, first differencing is often enough to make series modelling easier (Montgomery et al., 2024, Chapter 5.6). We expanded the model to a seasonal version since the data showed obvious seasonality. Additionally, we used the ARIMAX and SARIMAX models since we had access to external variables like temperature and precipitation. These enable us to simulate both internal structure and external influences on cocoa prices by including exogenous regressors into the forecasting process.

### ARIMA + GARCH

Although ARIMA and exponential smoothing techniques well capture seasonality and trends, they make the assumption that data variance doesn't change over time. Our initial research, however, revealed indications of volatility clustering, which denotes times when variance is abnormally high or low. Montgomery et al. (2024) assert that typical ARIMA techniques are ineffective when variance varies over time, underscoring the necessity of models that explicitly address volatility, such as the GARCH family.

We included a GARCH(1,1) component in our ARIMA model to address volatility clustering. The GARCH model permits volatility to fluctuate over time based on historical errors and volatility patterns. Because of this, it can be used to model cocoa prices, which frequently see volatile fluctuations. Our selection of the ARIMA + GARCH strategy is further supported by the fact that extensions of the GARCH model can also handle more complex behaviors, such as asymmetric responses to market shocks (Box, Jenkins, Reinsel, & Ljung, 2015).

**Multiple Linear Regression Model**

In addition, we employed a multiple linear regression model to investigate the relationship between cocoa prices and a range of explanatory variables, including time and weather-related factors such as precipitation, average temperature, maximum temperature, and minimum temperature. This approach enabled us to quantify the linear influence of each variable on cocoa price movements and provided a baseline for assessing the extent to which external environmental factors alone could account for observed price variations.

## 4. Data

This study integrates two primary datasets to support forecasting models for international cocoa prices. The first consists of daily cocoa futures prices from the International Cocoa Organization (ICCO), covering March 1994 to February 2025. The second is daily climate data from Ghana, obtained from the National Centers for Environmental Information (NCEI), including precipitation and minimum, average, and maximum temperature. Since Ghana is one of the world's top cocoa producers, weather conditions may influence supply and, indirectly, prices.

### 4.1 Data Sources and Cleaning

The raw datasets required several preprocessing steps. For the climate data, multiple entries existed for the same calendar day. After inspection, we retained only the last record per day, as it consistently reported the highest precipitation and widest temperature range—suggesting a complete daily summary.

Daily cocoa price changes were negligible, and log-transformed values clustered tightly, offering little signal. We therefore aggregated both datasets to monthly frequency by computing monthly means. This clarified seasonal patterns and aligned with agricultural production cycles. We then parsed numeric formats, merged the datasets by date, dropped missing values, and applied first-order differencing to the cocoa price series to address non-stationarity.

### 4.2 Monthly Aggregation and Exploratory Analysis

Figure 1 shows three broad regimes: stability through 2020, a gradual rise to 2022, and a sharp price surge in 2023–2024. These movements highlight volatility and possible structural breaks common in commodity markets.



Figure 1: Monthly Average Cocoa Futures Prices (1994–2025)
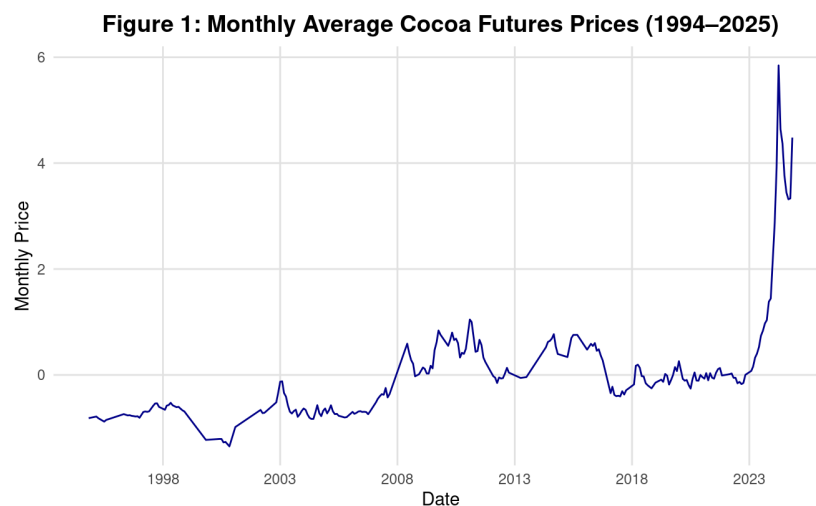
**Figure 1:** Monthly Average Cocoa Futures Prices (1994–2025)

*Monthly cocoa prices show long-term stability followed by a sharp surge in 2023–2024, highlighting increased market volatility.*

**Table 1:** Summary Statistics for Cocoa Futures Prices
*Monthly cocoa prices (1994–2025) show high variability, with a wide range and large standard deviation indicating strong market volatility.*

Table 1: Summary Statistics for Cocoa Prices (USD)

| Mean | SD | Min | Max |
|---|---|---|---|
| 2407.191 | 1354.329 | 780.51 | 10639.55 |

Table 1 reflects high variability, with a wide range and large standard deviation relative to the mean. This confirms the volatility observed visually and supports the use of models like GARCH.

### 4.3 Transformation and Seasonality

ACF and PACF plots of the raw price series (Figure 2) show strong persistence and a clear spike at lag 1, indicating non-stationarity and the need for differencing.
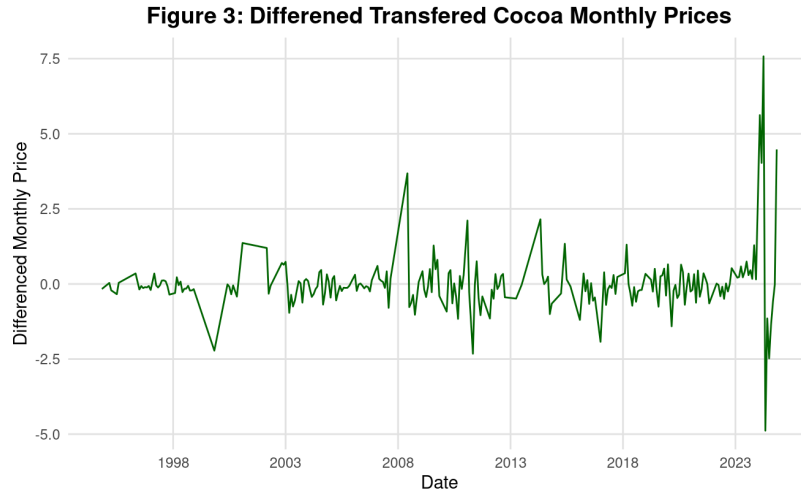
**Figure 2 : ACF and PACF of Monthly Cocoa Prices**
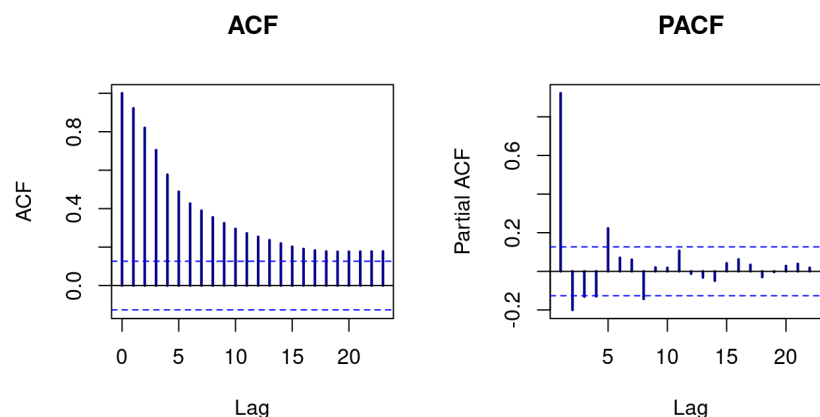


**Figure 2:** ACF and PACF of Monthly Cocoa Prices
*The ACF shows slow decay and the PACF spikes at lag 1, indicating non-stationarity and the need for differencing.*

Differencing the series stabilized the mean but revealed volatility bursts—especially post-2022—consistent with conditional heteroskedasticity (Figure 3).

**Figure 3: Differenced Transfered Cocoa Monthly Prices**



**Figure 3:** First-Differenced Monthly Cocoa Futures Prices
*The differenced series highlights local volatility and improved stationarity, suitable for ARIMA modeling.*

The ACF and PACF of the differenced series (Figure 4) show fast decay and weaker partial autocorrelations, supporting its stationarity and suitability for ARIMA-type modeling.

**Figure 4: ACF and PACF of Differenced Cocoa Prices**
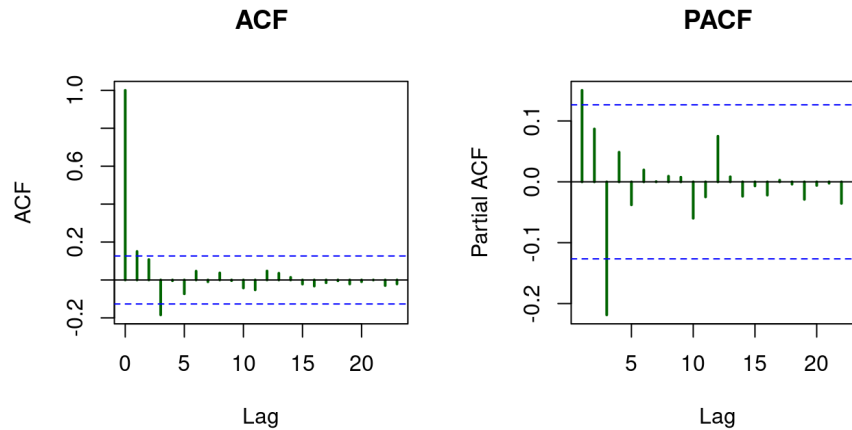


*Rapid ACF decay and weak PACF structure confirm stationarity, supporting ARIMA-family modeling.*

STL decomposition of the differenced series (Figure 5) reveals a strong annual cycle matching cocoa harvest patterns in West Africa, with a relatively stable residual. This justifies the use of SARIMA models with seasonal terms.
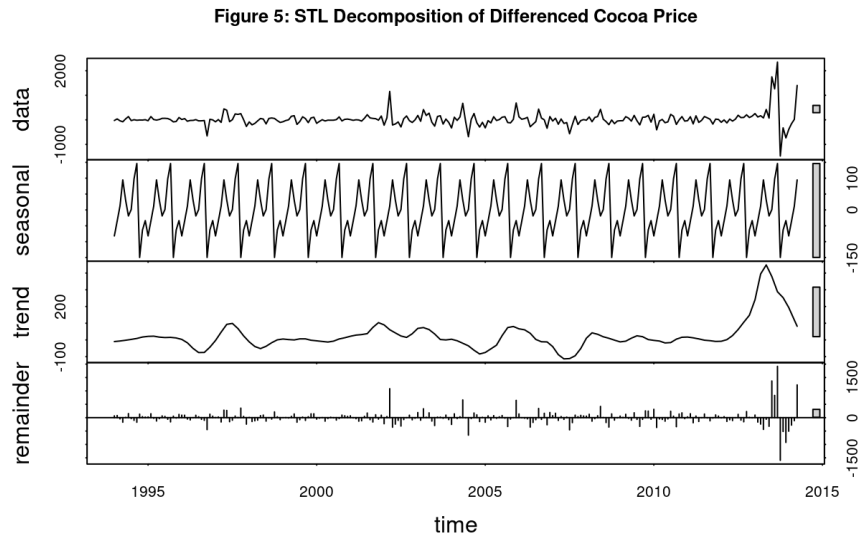
**Figure 5:** STL Decomposition of Differenced Cocoa Price
*STL separates the differenced series into seasonal, trend, and irregular components for clearer pattern analysis.*

## 4.4 Climate Variables as External Regressors

To align with the price series, we aggregated the Ghana climate data to monthly frequency. Table 2 summarizes precipitation and temperature values. While temperature shows regular seasonal patterns and low variance, precipitation is highly skewed, with many zero values and occasional spikes.

**Table 2:** Summary Statistics for Ghanaian Climate Variables
*Monthly averages of precipitation and temperature (min, avg, max) for Ghana's climate data.*

Table 2: Summary Statistics for Ghana Climate Variables

|  | Mean | SD | Min | Max |
|---|---|---|---|---|
| PRCP | 0.2784684 | 0.5949306 | 0 | 7.48 |
| TAVG | 79.4496883 | 2.7093137 | 72 | 89.00 |
| TMAX | 88.0249332 | 4.2047954 | 76 | 103.00 |
| TMIN | 71.8370436 | 2.7052433 | 60 | 82.00 |

These variables were included in ARIMAX and SARIMAX models to test their predictive value. While theoretically relevant, especially for local agricultural supply, their impact on short-term price variation was limited. One limitation is the narrow geographic scope—cocoa prices are shaped by global production, whereas our climate data only reflects conditions in Ghana.

## 5. Forecasting and Results

Building on the cleaned and aggregated monthly dataset described in the previous section, we now evaluate several time series models for forecasting cocoa futures prices. The

goal is to compare models with varying assumptions about trend, seasonality, volatility, and exogenous influences. Specifically, we implement and assess five models: Exponential Smoothing (ETS), ARIMAX and SARIMAX, ARIMA-GARCH, and a multiple linear regression (MLR) model with climate variables. Each model is fit to the training data and evaluated on its out-of-sample forecast performance using metrics such as RMSE, MAE, and MAPE. In the sections that follow, we present the structure, forecast accuracy, and diagnostic results for each model, and conclude with a comparison to determine which model best captures the observed price dynamics.

## 5.1 Model Training and Validation Process

The four primary forecasting models that we employed in our study were trained and validated in this part. We explain the procedure used to arrive at the final fitted model, including parameter selection and data modifications. We also describe how we divided the data into training and testing sets and the measures we employed to assess the accuracy of our forecasts. We outline the steps we used to build each model and apply it to our data on cocoa prices.

To evaluate the forecasting models, we split the time series data into a training set and a test set in chronological order. In addition to four climate-related variables—total precipitation (PRCP), average temperature (TAVG), maximum temperature (TMAX) and minimum temperature (TMIN). The dataset included 244 monthly observations of cocoa prices from November 1994 to November 2024 (after missing values were eliminated). The training set consisted of the first 240 months (November 1994–July 2024), while the test set consisted of the final four months (August–November 2024).

This arrangement represents a realistic forecasting scenario in which only historical data is used to estimate future values. Four months were set aside for evaluation, while the remaining 240 months were used for training. Similarly, external regressors were divided into training and prediction values, with values up to July 2024 being used for training. This historical data was used to train each model, and forecasts for August through November of 2024 were used to test them. In our analysis, the following forecasting models were used and trained:

## ETS Model

In order to address the nonstationary trend found in the original data, we trained the ETS model using the differenced cocoa price series (diff_price). The best-fitting exponential smoothing model was automatically chosen using the corrected Akaike Information Criterion (AICc) when we used R's ets() function.

ETS(A,N,N) model's structure works well for time series that shift around a fluctuating level without obvious seasonal or trend components. Maximum likelihood was used to estimate the parameters, and the fitted model was then used to predict the variations in cocoa prices throughout the test period (August–November 2024).

**Decomposition by ETS(A,N,N) method**
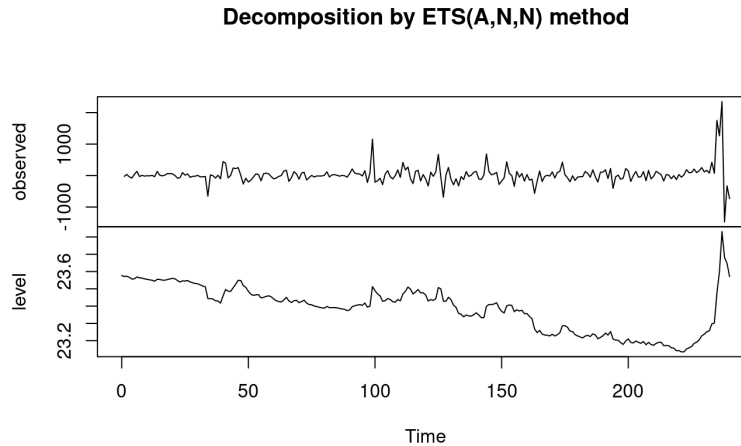


**Figure 6:** ETS(A,N,N) Decomposition of Monthly Cocoa Prices

*The decomposition separates observed prices into level and residual components using the ETS(A,N,N) model.*

The projected differences were converted back into actual price forecasts by adding them cumulatively to the last observed price from the training set in order to provide predictions on the original price scale.

**ARIMA-family models**

Here, we trained ARIMAX and SARIMAX, two models based on ARIMA. Both models made use of the original (non-differenced) cocoa price series, and R's auto.arima() function handled differencing automatically when needed. During the fitting process, this function minimized the corrected Akaike Information Criterion (AICc) and chose the best ARIMA orders (p, d, and q).

Maximum temperature (TMAX), average temperature (TAVG), minimum temperature (TMIN), and total precipitation (PRCP) were all included as external regressors. These factors were included to enhance forecast accuracy by accounting for the possible influence of weather on cocoa prices.

Seasonal terms were not used in the training of the ARIMAX model (seasonal = FALSE). It concentrated on the external factors and time-based autocorrelation. To represent yearly patterns such as harvest seasons, the SARIMAX model used a 12-month cycle with seasonal components (seasonal = TRUE).

The maximum likelihood method was used to fit both models. We created forecasts for August through November 2024 using the external factors from the previous four months after training.

**ARIMA + GARCH Model**

To model both price levels and volatility in cocoa prices, we combined ARIMA and GARCH methods. We started by using the differenced price data to create an ARIMA(0,1,1) model. Because of its simplicity and capacity to represent fundamental temporal patterns, this structure was selected. On the other hand, residual diagnostics showed evidence of volatility clustering. The residual time series displayed notable shifts during specific times, and the

Ljung-Box test yielded a p-value of 0.01497, indicating autocorrelation persisted. These results showed that volatility could not be fully explained by ARIMA(0,1,1) alone (Figure 7).
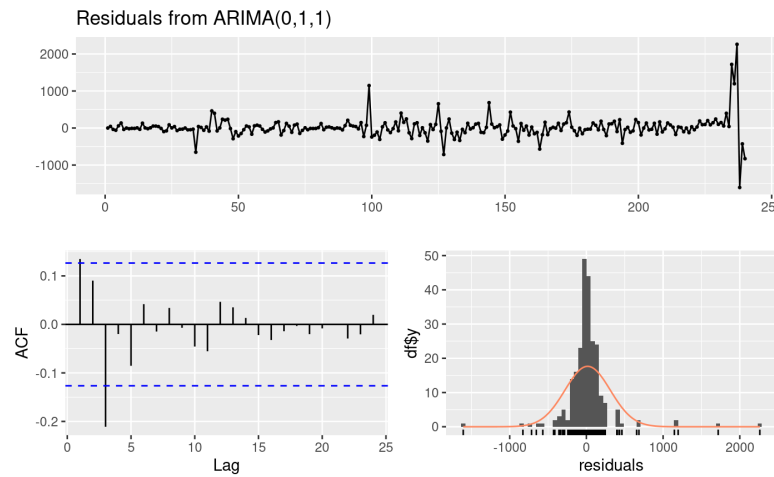


**Figure 7:** Residual Diagnostics for ARIMA(0,1,1) Model
*Residuals show volatility clustering and non-normality, supporting the need for a GARCH extension.*

In order to solve this, we expanded the model to include a GARCH(1,1) component, which represents volatility clustering and permits the variance to fluctuate over time. The combined ARIMA(0,1,1) + GARCH(1,1) model was then fitted to the data that had changed. We experimented with ARIMA(0,1,1), ARIMA(1,1,1), and ARIMA(1,1,2) in conjunction with GARCH(1,1) to investigate different configurations. With the lowest AIC (13.608), the ARIMA(0,1,1) + GARCH(1,1) model provided the best balance between simplicity and fit. Table 3 provides a summary of the AIC values for each model.

**Table 3:** AIC Comparison of ARIMA + GARCH(1,1) Models
*The ARIMA(0,1,1) + GARCH(1,1) model has the lowest AIC, indicating the best training fit.*

Table 3: Forecast Accuracy Comparison Across Models

| ARIMA_Model | GARCH_Model | AIC |
|---|---|---|
| ARIMA(0,1,1) | GARCH(1,1) | 13.60842 |
| ARIMA(1,1,1) | GARCH(1,1) | 13.61140 |
| ARIMA(1,1,2) | GARCH(1,1) | 13.62346 |

The final ARIMA(0,1,1) + GARCH(1,1) model was estimated using the garchFit() function from the Garch package in R. The model was able to capture important price dynamics and time-varying volatility, according to residual diagnostics. When compared to using ARIMA alone, this combined model offered a more thorough and accurate method of modeling monthly cocoa prices.

**Multiple Linear Regression Model**
Lastly, we constructed a regression model to examine how external variables might influence cocoa prices. The model was trained using R's built-in lm() function with the following formula:

$$Monthly\ Price = \beta_0 + \beta_1 t + \beta_2 PRCP_t + \beta_3 TAVG_t + \beta_4 TMAX_t + \beta_5 TMIN_t + \epsilon_t$$

The model incorporates four climate-related regressors to account for potential weather-related effects on cocoa yields and prices: total precipitation (PRCP), average temperature (TAVG), maximum temperature (TMAX), and minimum temperature (TMIN). The linear time variable (time(Monthly_price)) captures the overall long-term trend. Ordinary Least Squares (OLS) was used to estimate the regression parameters using only the training data. This method, which complemented the other time series techniques used in our analysis, enabled us to clearly establish correlations between cocoa pricing and external factors.

## 5.2 Performance Evaluation of the Forecasting Models

In order to further evaluate the forecasting accuracy of our models, we used several standard forecasting evaluation metrics. Before interpreting the results, we will briefly explain what each metric represents.

Firstly, the "Mean Error (ME)" gives us a sense of the model's overall bias, which measures the average difference between predicted and actual values. And a ME value close to zero means the model is generally balanced and not consistently over-predicting or under-predicting. Secondly, "Root Mean Squared Error (RMSE)" is used for evaluating forecast accuracy. Since we are calculating the square root of the average squared differences between predicted and actual values, the larger mistakes of the model appear to be penalized more heavily. On the other hand, "Mean Absolute Error (MAE)" is more straightforward, as it averages the absolute size of the forecast errors with treating all mistakes equally and regardless of direction. Moreover, the "Mean Percentage Error (MPE)" expresses the average forecast error as a percentage of the actual values. And it's helpful when we want to understand how far off predictions are in relative terms. Fifthly, "Mean Absolute Percentage Error (MAPE)" expresses forecast accuracy as a percentage, which makes it easier to interpret how far off the predictions were relative to the actual values. Generally, a MAPE under 10% is generally considered excellent, 10–20% is good, 20–50% is reasonable and above 50% indicates poor forecasting accuracy. Finally, the "Mean Absolute Scaled Error (MASE)" measures the forecast accuracy by comparing the model's performance to its baseline(typically one that assumes the next value will be the same as the previous one), and a MASE value below 1 indicates that the model performs better than this basic approach, while a value above 1 suggests it performs worse.

To evaluate how well each model could predict unseen data, we trained them on historical monthly data from November 1994 to July 2024. Then, we used each model to forecast cocoa prices for the final four months—August to November 2024. By comparing these predictions with the actual prices using standard accuracy metrics like RMSE, MAE, and MAPE, we were able to assess how each model performed in a real-world forecasting scenario. The results are summarized in Table 4.

**Table 4:** Forecast Accuracy Comparison Across Models
*GARCH outperforms ETS, ARIMAX, SARIMAX, and regression across RMSE, MAE, and MAPE, capturing both trend and volatility most effectively.*

Table 4: Forecast Accuracy Comparison Across Models

|             | model      | RMSE      | MAE       | MAPE        |
|-------------|------------|-----------|-----------|-------------|
| Test set...1 | ETS        | 724.7028  | 494.1472  | 84.34063    |
| Test set...2 | ARIMAX     | 7149.7841 | 7095.2161 | 11190.51346 |
| Test set...3 | SARIMAX    | 7149.7841 | 7095.2161 | 11190.51346 |
| Test set...4 | GARCH      | 703.4451  | 431.3802  | 70.96833    |
| Test set...5 | Regression | 5571.8638 | 5520.4952 | 80.12304    |

**ETS Model**

According to the output of training set performance (see Table 5 in Appendix) of the ETS(A, N, N) model, which is fitted on the first-differenced monthly cocoa price series, it suggests that the model captures a simple level component without trend or seasonality. An alpha value extremely close to zero ($\alpha$ = 1e-04) indicates that the model barely adjusts its level in response to new observations. Moreover, the model's residual standard deviation ($\sigma$) was 295.37, which suggests a relatively widespread in the fitted values and reflects the high volatility of the cocoa price data even after differencing. Furthermore, the information criteria values of AIC = 4049.69 and AICc = 4049.79 can also indicate low risk of overfitting. The RMSE of 294.13 and MAE of 150.80 tell us that, on average, the model's predictions were off by quite a bit in absolute terms. The MAPE, which came in at a high 162.51%, shows the model's percentage errors were large, especially when the actual prices were lower. Finally, the MASE value of 0.75 suggests that the model did outperform a baseline forecast. Overall, while the ETS model captures some structure in the data, its performance is limited in the face of strong volatility, which is common in commodity markets like cocoa.

When we turn to the ETS model's testing set performance (see Table 4), these limitations become even more apparent. The ETS model produced a RMSE of 724.70, which means that on average, the model's predictions deviated from the actual cocoa prices by over 724 USD per tonne. This is more than double its training set sample RMSE, which shows that the model struggled significantly to generalize outside the training data. Moreover, the MAE of 494.15 further confirms this issue; on average, the forecast was off by nearly 500 USD per tonne cocoa. Meanwhile, the MAPE of 84.34% indicates that the model's forecast errors were, on average, 84% of the actual values, which is still relatively high and indicates large proportional misestimates, particularly during volatile months.

Therefore, together, these results tell us that the ETS(A, N, N) model lacked the complexity to adapt to dramatic shifts or spikes in the test period (such as those seen in late 2024).

**ARIMAX & SARIMAX Model**

According to the output of training set performance (see Table 6 in the Appendix), both ARIMAX and SARIMAX models were fitted on the monthly cocoa price data using the same ARIMA(1,1,2) specification. The autoregressive coefficient (ar1 = -0.9458) and two moving average terms (ma1 = 1.1753, ma2 = 0.3682) all have relatively small standard errors, which suggests that the model successfully captured autocorrelation and recent error shocks in the training set. However, the added weather variables (precipitation and temperature) appear statistically insignificant. For example, PRCP = 11.75 (s.e. = 38.60), TAVG = -16.12 (s.e. = 10.35). These high standard errors suggest that the climate variables didn't meaningfully improve predictive performance. The log-likelihood of -1684.81 and AIC = 3385.61 are not particularly low. Therefore, it can only indicate a moderate fit.

When we shift focus to test set performance (see Table 4), the weaknesses of the ARIMAX and SARIMAX models become much more apparent. Both models produced an RMSE of 7149.78 and MAE of 7095.22, which means that the average forecast error was over 7000 USD per tonne, which is an extremely poor result, especially when compared to the ETS model's RMSE of 724.70. The MAPE value of 11,190.51% is alarmingly high, and this suggests that the forecasts were orders of magnitude off from the actual values during the volatile test

period. This enormous error can be attributed to the models' failure to anticipate the sharp spike in cocoa prices from August to November 2024.

Therefore, both ARIMAX and SARIMAX models performed extremely poorly on the test set, and the climate variables did not meaningfully improve forecasting power. Both models failed to adjust to structural breaks and extreme price shifts.

**GARCH(1, 1) Model**

According to the training output of the GARCH(1,1) model fitted on the first-differenced monthly cocoa price series (see Table 7 in the appendix), the model effectively captures the time-varying volatility present in the data. As the two key volatility parameters $\alpha_1 = 0.6272$ (ARCH term) and $\beta_1 = 0.5297$ (GARCH term) are both statistically significant (p-values < 0.001). These indicate that recent and past shocks play a major role in shaping future variability. In other words, the model recognizes that large price movements tend to cluster, which is a known feature in commodity markets like cocoa. Additionally, the significant moving average term ($ma_1 = 0.2941$, p = 0.0034) also shows that recent forecast errors help adjust the model's short-term predictions, which improving model's adaptability.

When evaluating the GARCH model on the test set (see Table 4), its performance stands out across all three accuracy metrics. It achieved the lowest RMSE of 703.45, which indicates that, on average, the predictions deviated by around 703 USD/tonne, which was better than all other models tested. The MAPE of 70.97%, though still high due to the extreme volatility of cocoa prices, was the lowest among all five models, suggesting that the GARCH(1,1) model handled percentage-based forecast deviations most effectively.

Altogether, the GARCH(1,1) model demonstrated a strong ability to capture both the shifts and clustering of price fluctuations, making it the most reliable forecasting model.

**Multiple Linear Regression Model**

Based on the regression model fitted to the monthly cocoa price data, the model seemed straightforward and incorporated logical predictors at first glance. However, the training set performance revealed some key limitations. As seen in the model's training set output (see Table 7 in appendix), the residuals ranged widely, from -1181.1 to 5907.5 USD per tonne, which means the model missed the actual price by thousands of dollars, especially for a commodity as sensitive as cocoa. Even though we included climate variables that should theoretically impact cocoa production, none were statistically significant based on their standard errors and t-values (also in Table 7). This suggests that in this model, the weather data didn't really help explain price movements during the training period.

When we tested the model on new data from August to November 2024 (Table 4), its performance dropped off even more sharply. The RMSE was over 5500 USD, and the MAE was nearly the same, which means the model was consistently off by thousands of dollars. The MAPE of 80.12% shows that, on average, the forecasts were about 80% away from the true values, which is a massive miss in proportional terms.

Therefore, the regression model struggled to adapt to the sharp volatility in late 2024.

After comparing all models, the GARCH(1,1) model stood out. While ETS, ARIMAX SARIMAX, and regression models each captured parts of the price patterns, they struggled during the sharp volatility of late 2024. GARCH(1,1), on the other hand, could adapt to the ups and downs in volatility, giving it a clear edge. It delivered the most accurate forecasts overall, making it the best fit for this unpredictable and real-world cocoa market.

### 5.3 Forecasted Values and Observed Patterns

After assessing each model's forecasting accuracy with standard error metrics, we further validated the models by graphically comparing their predicted values against the actual cocoa prices. These visualizations help us understand how well each model captures key patterns(such as upward trends, volatility spikes and overall price direction)  in the data, particularly during the test period from August to November 2024. By analyzing the alignment between forecasted and observed values, we can then more intuitively assess each model's real-world effectiveness.

### ETS Model

According to Figure 8, while actual prices fluctuate sharply, especially toward the end, the ETS model's forecasts remain flat and fail to respond to these changes. The graph reinforces the error metrics by illustrating that the ETS model performs adequately under stable conditions but fails to capture real-time data fluctuations. This limitation makes it less suitable for forecasting in volatile markets like cocoa, where abrupt changes are frequent.
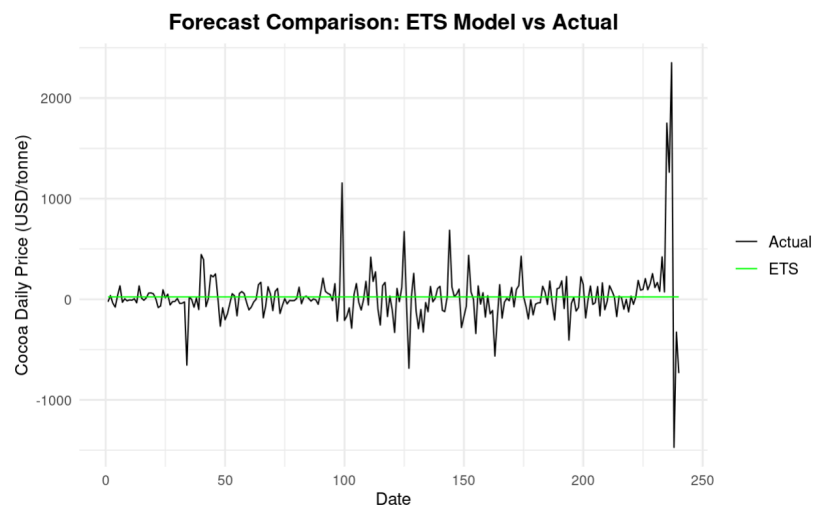


**Figure 8:** Forecast vs. Actual Cocoa Prices: ETS Model
*ETS forecasts remain flat, failing to capture the sharp price spikes in late 2024.*

### ARIMAX & SARIMAX Model

According to Figure 9, while actual cocoa prices rise sharply toward the end of the series, both ARIMAX and SARIMAX models closely track historical trends but fail to anticipate the extreme price spike. Despite aligning well with past movements, their forecasts lag dramatically during volatile periods. This suggests that although these models can capture long-term patterns, they struggle with sudden structural breaks that limit their effectiveness in forecasting under rapidly changing market conditions.
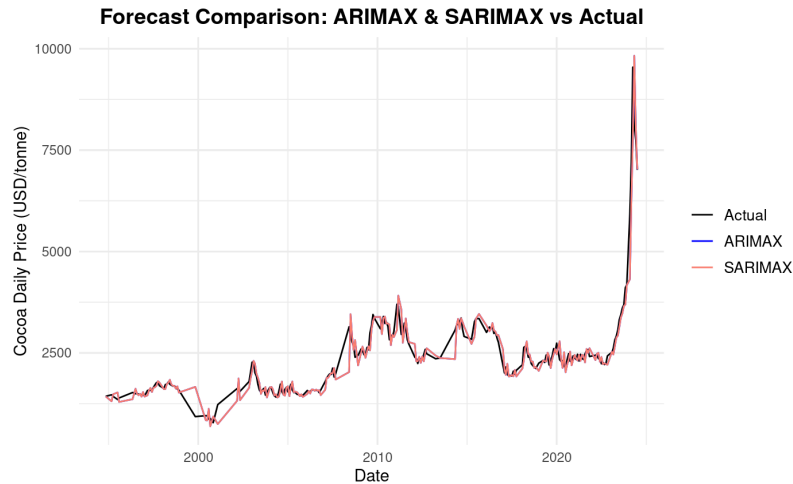
**Figure 9:** Forecast vs. Actual Cocoa Prices: ARIMAX & SARIMAX Model

*Both models track long-term trends but fail to capture the sharp 2024 price spike, highlighting limitations in adapting to sudden shifts.*

**GARCH Model**

According to Figure 10, the GARCH model tracks the direction and magnitude of volatility more closely than other models, especially during periods of heightened price fluctuation. While not perfectly aligned, the GARCH forecasts (blue line) show responsiveness to extreme changes in the actual differenced cocoa prices (black line), particularly toward the right end of the plot. This visual alignment supports the model's superior performance in volatile settings.
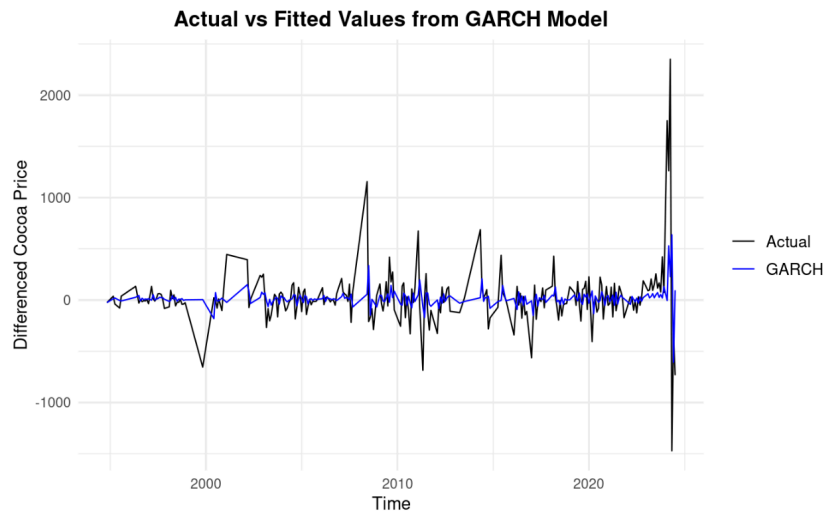


**Figure 10:** Forecast vs. Actual Cocoa Prices: GARCH Model

*The GARCH model closely tracks price fluctuations, capturing volatility patterns more effectively—especially in late 2024.*

**Multiple Linear Regression Model**
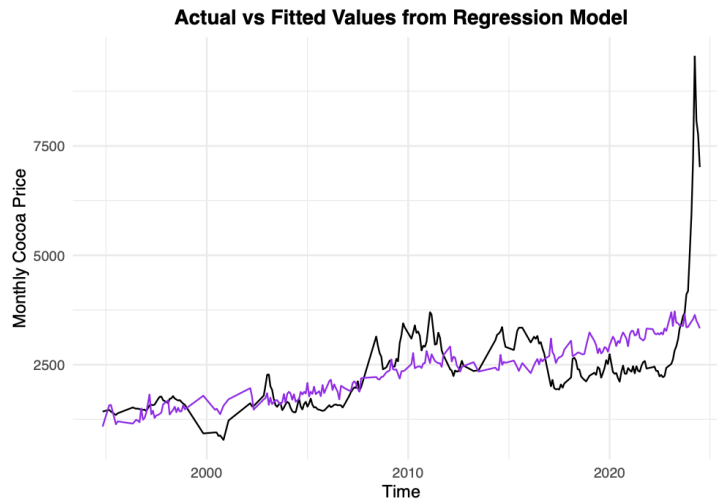
**Figure 11:** Forecast vs. Actual Cocoa Prices: Regression Model

*The regression model captures broad trends but fails to reflect sharp fluctuations, including the late 2024 surge.*

According to Figure 11, the regression model captures the general upward trend in cocoa prices but fails to respond to sharp short-term fluctuations, especially the extreme spike near the end of the series. The predicted values remain relatively smooth compared to the actual prices, which indicates the model's limited ability to adapt to sudden market changes. This visual pattern reinforces the regression model's weaker performance in volatile conditions.

**5.4 Forecast Comparison of All Models**

Figure 12 reveals apparent performance gaps among the forecasting models. The GARCH model is the most adaptive, closely tracking the spikes and shifts in actual cocoa prices during the test period. In contrast, the ETS model remains flat and unresponsive, indicating poor adaptability to volatility. ARIMAX and SARIMAX show some trend-following behaviour but significantly overshoot the actual values. The regression model underpredicts the surge entirely. GARCH demonstrates the best alignment with actual market behaviour, making it the most reliable model in this volatile setting.
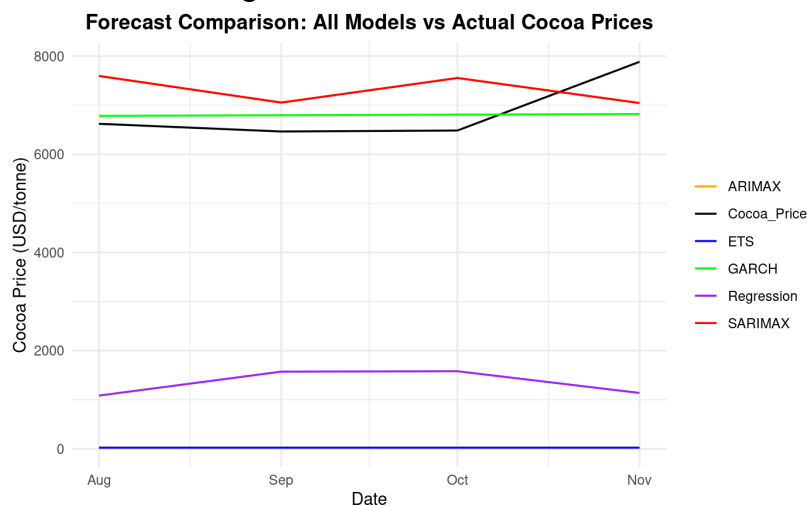


**Figure 12:** All Models Forecast vs Actual Cocoa Prices (Aug–Nov 2024)

*GARCH closely followed observed values, while ARIMAX and SARIMAX overestimated November prices. ETS and Regression underperformed, missing the upward spike.*

## 6. Discussion and Conclusion
### 6.1 Key Findings
Our results show that the GARCH model outperforms all other specifications in forecasting cocoa prices across all evaluation metrics. As shown in Table 4, GARCH achieves the lowest RMSE, MAE, and MAPE, confirming its effectiveness in capturing the volatility and sharp price shifts that occurred in late 2023 and early 2024. This model's ability to account for conditional heteroskedasticity makes it especially well-suited for commodity markets prone to abrupt shocks.

The baseline ARIMA(0,1,1) and seasonal SARIMA models performed moderately well but failed to capture volatility clustering. Although SARIMA incorporated seasonal components identified in the STL decomposition, its forecast accuracy lagged behind GARCH's. The ETS(A, N, N) model struggled to adapt to recent structural breaks and performed worst across all metrics.

Models incorporating exogenous climate variables—ARIMAX, SARIMAX, and multiple linear regression—did not improve forecasting performance. While these variables have theoretical relevance due to Ghana's role in global cocoa supply, their short-term impact on price dynamics appears limited. This result is consistent with prior studies that found macro and financial shocks often dominate agricultural commodity pricing in the short run.

### 6.2 Limitations
While the models presented perform well in capturing key features of the cocoa price series, several limitations must be acknowledged.

First, our use of monthly data—while adequate for smoothing noise and revealing seasonal patterns—may obscure short-term dynamics, such as sudden shocks or intra-month volatility, that daily data could capture.

Second, although Ghana is a major cocoa producer, our climate dataset reflects only local weather conditions. Global cocoa prices are influenced by production across multiple countries, so our climate variables may only partially represent environmental supply shocks affecting international markets.

Third, while ARIMA-family models and the GARCH model are potent tools for linear and volatility modelling, they may fall short in capturing non-linear or structural break patterns, particularly those observed during the 2023–2024 price surge. More flexible models, such as regime-switching or machine-learning methods, may improve accuracy.

Finally, the out-of-sample evaluation window is relatively short, which may limit the generalizability of our forecast performance results. Expanding the test period or conducting rolling forecasts could offer a more robust assessment in future work.

### 6.3 Future Work
Future research could build on this study by expanding the scope of climate data beyond Ghana. Since cocoa prices are a reflection of worldwide output, adding meteorological data from other major producers, such as Indonesia and Côte d'Ivoire, could increase the exogenous regressors' explanatory power.

Furthermore, the current price spike in 2023–2024 raises the possibility of structural disruptions. There may be greater flexibility in capturing non-linearities and abrupt transitions with techniques like regime-switching models or machine learning techniques (e.g., random forests, gradient boosting). Additionally, these models could be able to identify intricate relationships between market and climate variables.

Finally, our evaluation used a single four-month forecast window. Future studies could use rolling forecasts or longer out-of-sample periods to assess model stability and robustness over time, especially under different market conditions.

**6.4 Real-World Implications**

Our research has a number of applications for those involved in the cocoa industry.

First, it is important to take time-varying volatility into consideration, as demonstrated by the GARCH model's good performance. Models that adapt dynamically to market uncertainty are useful for exporters, policymakers, and commodities dealers, particularly during times of geopolitical or climate-related shocks.

Secondly, although Ghanaian climate variables had little ability to predict the short term, their presence is still important for tracking agricultural risk. Policymakers may be better equipped to foresee supply disruptions and stabilize domestic markets with more geographically broad climatic data.

Last but not least, precise short-term projections can help with better decision-making regarding hedging tactics, export scheduling, and inventory management. Strong forecasting methods are becoming more and more important for reducing financial risk in the cocoa supply chain as price volatility keeps getting worse.

# References

Assis, K., Amran, A., & Remali, Y. (2010). Forecasting cocoa bean prices using univariate time series models. *Researchers World, 1*(1), 71.

Asante, P. A., Rahn, E., Anten, N. P., Zuidema, P. A., Morales, A., & Rozendaal, D. M. (2025). Climate change impacts on cocoa production in the major producing countries of West and Central Africa by mid-century. *Agricultural and Forest Meteorology, 362*, 110393.

Benneh, E. Y., & Anaman, K. A. (2022). Economic shocks and the growth of the Ghanaian cocoa industry from 1975 to 2019. *Ghana Journal of Development Studies, 19*(2). https://doi.org/10.4314/gjds.v19i2.3

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. https://otexts.com/fpp2/ets.html

Kumar, K. A., Spulbar, C., Pinto, P., Hawaldar, I. T., Birau, R., & Joisa, J. (2022). Using econometric models to manage the price risk of cocoa beans: A case from India. *Risks, 10*(6), 115.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2024). *Introduction to time series analysis and forecasting* (2nd ed.). John Wiley & Sons.

Oduro-Gyimah, F. K. (2012, December). Forecasting cocoa production and the effects on the economy of lower middle income country: A case study of Ghana. In *1st Applied Research Conference in Africa* (Vol. 1, pp. 1–11).