

How Do Weather Patterns and Exchange Rate Fluctuations in West African Countries Influence Cocoa Futures Prices?

Sameeck Bhatia, Arnav Dey, Sofia Duzha, Tanya Pandey

2025-04-04

1 Introduction

Cocoa is a key commodity in the global market; it plays a crucial role in chocolate production, which is one of the largest commodity markets in the world. It is derived from the seeds of the cacao tree (*Theobroma cacao*), which thrives in tropical climates with high humidity and consistent rainfall. Its price is shaped by various factors like weather, geopolitics, exchange rate, and supply chain disruptions. Understanding and forecasting cocoa and its price movements is of great interest to producers, traders, and policymakers. Modelling commodity prices is essential in various industries like agriculture, finance, and trade. However, modelling and forecasting commodity prices can be challenging; multiple factors influence commodity prices and often exhibit characteristics such as seasonality, volatility, and external shocks.

The significance of cocoa price forecasting extends beyond financial speculation; it directly impacts economic planning in cocoa-exporting nations and procurement strategies for global food and beverage companies. 70% of the world's cocoa originates from countries such as Ghana, Ivory Coast, Nigeria, and Cameroon, making the West African region critical to global cocoa. This project focuses on forecasting cocoa futures prices through time series analysis, exploring the question: **How do weather patterns and exchange rate fluctuations in West African countries, mainly Ghana and Cameroon, influence global cocoa futures prices?**

This research question is explored using two forecasting models. First, a baseline ARIMA(9,1,1) model captures the autocorrelation and moving average structure in historical cocoa price data. Second, a more sophisticated ARDL model is constructed to assess the short-run and long-run relationships between cocoa prices and macro-climatic variables such as exchange rates and average temperatures. The ARDL approach is particularly appropriate given the presence of cointegration among the variables, allowing for a meaningful interpretation of economic relationships over time.

Time series forecasting involves analyzing historical data points collected over time to predict future values. It is particularly suited to financial and economic datasets, which often exhibit trends, seasonality, and autocorrelation. In this project, two key time series models are employed. The **ARIMA (Autoregressive Integrated Moving Average)** model is a classical statistical approach that captures patterns in a single time series by accounting for its own past values and past forecast errors. It is widely used for its simplicity and effectiveness in short-term forecasting (Shumway & Stoffer, 2017). The **ARDL (Autoregressive Distributed Lag)** combines short-run dynamics with long-run equilibrium relationships among multiple time series. This model is particularly powerful when variables are integrated of different orders (I(0) or I(1)) and potentially cointegrated, making it suitable for macroeconomic modeling involving both climatic and financial predictors (Natsopoulos et al., 2001).

Several methodological and practical challenges were encountered during the study. Cocoa futures data exhibited non-stationarity, needing differencing and unit root testing. Inconsistencies and missing observations in daily data also posed a concern, which was addressed by aggregating the series to monthly averages. Fur-

thermore, aligning exchange rates and climate data from multiple countries required extensive preprocessing. These challenges highlight the inherent complexity in modeling real-world commodity prices, which often violate the assumption of being independently and identically distributed (i.i.d.). In time series contexts, the i.i.d. assumption is rarely satisfied because observations tend to exhibit serial correlation, seasonality, and evolving variance over time. Such dependencies must be explicitly modeled, as ignoring them can lead to biased estimates and unreliable forecasts. Furthermore, the presence of structural breaks and external shocks introduces additional layers of difficulty in capturing stable relationships across time (Shumway & Stoffer, 2017).

This paper delves into the usage of the baseline ARIMA and advanced ARDL model for the forecasting of the global cocoa futures pricing in Cameroon and Ghana, mainly examining two factors: exchange rate fluctuations and weather patterns.

2 Literature Review

2.1 ARIMA Literature

Sukiyono et al. (2018) conducted a study on international and Indonesian cocoa prices using data from 2008 to 2016. The authors applied three forecasting models: Autoregressive Integrated Moving Average (ARIMA), exponential smoothing, and decomposition. The study found that ARIMA outperformed the other models based on three evaluation metrics: Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Squared Deviation (MSD). While the dataset is relatively old, it remains stable and provides valuable insights. Additionally, the study acknowledged Ghana's significant role in the global cocoa market.

Similarly, Assis, Amran, and Remali (2010) analyzed cocoa price forecasting using data from 1992 to 2006 on Bagan Datoh cocoa, a local Malaysian variety. The authors examined four models: exponential smoothing, ARIMA, Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and a combination of ARIMA and GARCH. Their analysis incorporated multiple goodness-of-fit criteria, including Akaike Information Criterion (AIC), as well as evaluation metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and MAPE. The study concluded that the GARCH model provided the most accurate forecasts for Malaysian cocoa prices.

2.2 ARDL Literature

The central modeling hypothesis of this paper is that cocoa prices can be explained by a set of regressors, specifically the exchange rates and average temperatures of four countries. Additionally, the model assumes that lagged values of these predictor variables also influence cocoa prices. It is likely that buy and sell decisions in the cocoa market are based on historical data, where speculators use past exchange rates and temperature trends to predict future price movements. This justifies the use of the Autoregressive Distributed Lag (ARDL) model, which effectively incorporates dynamic lag structures.

Commodity price forecasting has traditionally sought to examine the long-term relationships between the target variable and other explanatory factors through cointegration analysis. Understanding these long-term dynamics and their influence on short-term fluctuations has been instrumental in improving the prediction and interpretation of commodity price movements. In their study *Forecasting Commodity Prices: Futures Versus Judgment*, Bowman and Husain employ an error correction model (ECM) to model long-term relationships and forecast commodity prices. Their findings indicate that ECM-based forecasts outperform other methods for eight out of fifteen commodities. This result motivates the present study's exploration of an ECM framework for modeling cocoa prices. (Husain, & Bowman, 2004)

Building on insights from both ARDL and ECM, this paper adopts an ARDL-ECM approach. According to Hassler and Wolters in *Autoregressive Distributed Lag models and Cointegration*, the ARDL model is particularly well-suited for cointegration analysis involving multiple variables. The ARDL-ECM framework is advantageous as it captures both short-run dynamics and long-term relationships, provided that all

time series variables are either integrated of order zero, $I(0)$, or integrated of order one, $I(1)$, in terms of stationarity.(Hassler & Wolters, n.d.)

In the context of cocoa price forecasting, Coulibaly et al. in *Climate Change Effects on Cocoa Export: Case Study of Côte d’Ivoire* employ an ARDL-ECM model to assess the impact of climate variables on cocoa prices. Their findings indicate that temperature indirectly affects cocoa prices through yield loss, as unfavorable temperatures reduce cocoa production, leading to higher prices. Additionally, while exchange rates also exhibit an indirect influence on cocoa prices, this effect is not always statistically significant. Nevertheless, given its potential relevance, exchange rate dynamics will be incorporated into the present analysis.(K. Coulibaly et al., n.d.)

3 Data

This study utilizes three main datasets to model and forecast cocoa futures prices. The primary dataset is the Daily Prices ICCO.csv file, which contains daily closing prices for cocoa futures contracts traded on major commodity exchanges. Spanning from March 10, 1994, to February 27, 2025, this dataset was obtained from the International Cocoa Organization (ICCO) and serves as the target variable in all forecasting models (International Cocoa Organization, 2025).

Ghana, being the second-largest cocoa producer globally, plays a critical role in shaping global cocoa supply and pricing dynamics. The Ghanaian dataset used in this study comprises two main components: exchange rate data and climate data. The exchange rate series, included in the file ghana_final.csv, records daily Ghanaian cedi to U.S. dollar conversion rates from July 13, 2007, to November 28, 2024. This data was sourced from Google Finance and merged with corresponding climate data obtained from the National Centers for Environmental Information (NCEI). Although the original dataset also included daily average temperature values from the Ghana data.csv file, these were ultimately excluded from the final model due to low correlation with cocoa futures prices during exploratory analysis. The exchange rate was retained as a key predictor due to its potential impact on export revenues and domestic producer incentives in Ghana’s cocoa sector.

Cameroon, though a smaller producer compared to Ghana, contributes significantly to West African cocoa exports and serves as a relevant comparison in analyzing regional price dynamics. The Cameroonian dataset, saved as cameroon_final.csv, contains daily exchange rate data for the Central African CFA franc to U.S. dollars and daily average temperature readings from July 13, 2007, to November 28, 2024. The exchange rate data was obtained from Google Finance, while the weather data was retrieved using the Open-Meteo historical weather API. Both variables were retained in the final ARDL model as they exhibited relatively strong associations with the cocoa price series. Aggregating this information into monthly averages allowed for better alignment with the cocoa futures dataset and helped reduce volatility caused by short-term fluctuations or missing observations in the raw data.

After preprocessing and aggregation (further discussed in Methods), the datasets were combined into a monthly time series to facilitate model fitting and reduce noise due to missing values and reporting inconsistencies in daily data. Summary statistics and correlation analysis guided the selection of relevant predictors, with only Ghana and Cameroon variables retained in the final ARDL model due to their stronger association with cocoa price movements.

Note: Initial data for Ivory Coast (Google Finance, 2025d)(Open-Meteo, 2025) and Nigeria (Google Finance, 2025c)(Open-Meteo, 2025), specifically the exchange rate fluctuations and weather were also used, but upon assessing the correlational matrix, Ivory Coast and Nigeria’s average temperature and exchange rate data were highly correlated with each other, making them redundant in the model. If added, it would lead to highly inflated and unstable coefficients in the ARDL model, which risks overfitting and is thus excluded from the final forecasting model (further discussed in Forecasting and Results).

Table 1: Summary Statistics of Key Variables

Variable	Min	Q1	Median	Mean	Q3	Max	NA.
ICCO Price	774.1000	1559.0000	2202.3000	2338.1000	2735.4000	11984.7000	3324
Ghana Ex Rate	0.0600	0.1700	0.2500	0.3667	0.5700	1.0600	1325
Ghana Avg Temp	71.0000	79.0000	82.0000	81.5700	84.0000	90.0000	353
Cameroon Ex Rate	0.0015	0.0017	0.0017	0.0018	0.0019	0.0022	2140
Cameroon Avg Temp	68.7000	73.8000	75.8000	76.5700	78.9000	87.4000	0
Ivory Coast Ex Rate	0.0015	0.0017	0.0017	0.0018	0.0020	0.0022	2142
Ivory Coast Avg Temp	70.6000	76.6000	77.9000	78.1500	79.7000	87.5000	0
Nigeria Ex Rate	0.0006	0.0024	0.0028	0.0038	0.0061	0.0068	2150
Nigeria Avg Temp	70.4000	75.2000	77.3000	77.2900	79.2000	85.8000	0

The summary statistics in Table 1 provide an overview of global cocoa prices (ICCO Price) and exchange rates and average temperatures in Ghana, Cameroon, Ivory Coast, and Nigeria. The ICCO price has a wide range, from a minimum of 774.1 to a maximum of 11,984.7, with a median of 2,202.3 and a mean of 2,338.1, suggesting potential skewness due to extreme values. Exchange rates vary significantly across countries, with Ghana's having the highest median (0.25) and Nigeria's the lowest (0.0028), indicating differences in currency valuation. The average temperatures for all four countries are relatively stable, with median values between 75.8°F and 82°F, suggesting warm climates with slight variations. However, missing values (NA's) are present, especially in exchange rates, which could impact analyses involving these variables.

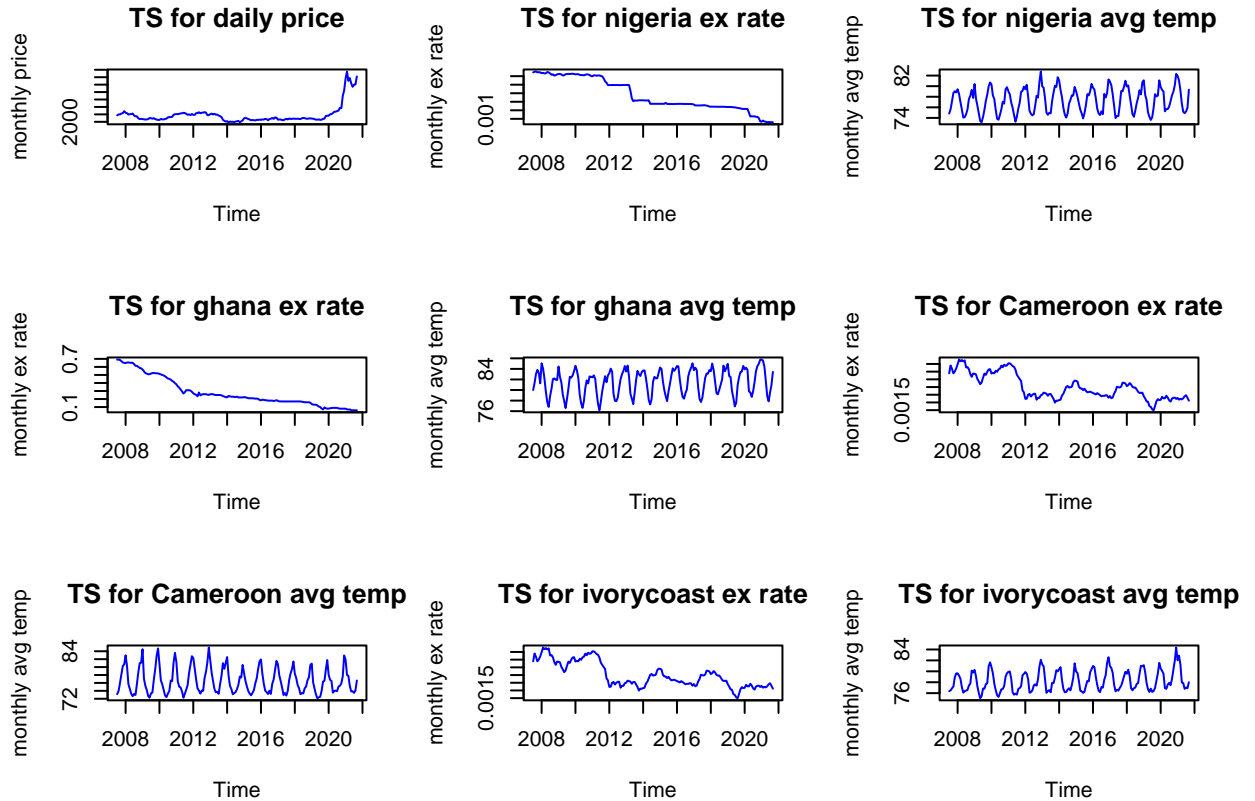


Figure 1: Time Series Plots for Price, Temperature, and Currency Datasets

4 Methodology

4.1 Data Preprocessing

The initial stage of data preprocessing involved compiling and aligning datasets by country, ensuring temporal consistency across all series. Given the extended time span of the data, ranging from approximately 1990 to 2025, many variables contained missing or irregularly spaced observations. An early attempt to impute missing values using overall averages proved ineffective due to the high variability and long historical range of the data, which rendered global averages unrepresentative. Usually, the global exchanges are closed on weekends and public holidays as they are non-trading days; there were a lot of missing values (also seen in the summary statistics in the Data section). As an alternative, a more robust strategy was adopted: computing monthly averages based only on the available daily values for each month. For instance, if at least five daily observations were available within a given month, their mean was taken as the representative monthly value, and missing values were simply excluded rather than imputed. This method significantly reduced noise and preserved the underlying data structure, allowing for clearer trends to emerge in the resulting time series. The monthly aggregation not only smoothed out short-term volatility but also facilitated alignment across variables and improved model performance in subsequent analysis.

4.2 ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model was implemented as a baseline approach for forecasting cocoa futures prices. The analysis began by importing the ICCO daily cocoa price data, which was then aggregated into monthly averages to address missing data and stabilize short-term fluctuations. The resulting series was converted into a univariate time series object. The stationarity of the price series was evaluated using the Augmented Dickey-Fuller (ADF) test, which indicated non-stationarity at level but stationarity after first differencing, justifying the use of an integrated model with $d=1$. Autocorrelation and partial autocorrelation plots of the differenced series informed the choice of lags for the autoregressive (AR) and moving average (MA) components. A model with order (9,1,1) was selected based on its performance and fit, capturing the long-memory structure and noise patterns observed in the data. The fitted ARIMA(9,1,1) model was then summarized to assess parameter significance and residual diagnostics.

4.3 SARIMA Model and Spectral Analysis

To explore the possibility of seasonal patterns in cocoa prices, a seasonal decomposition was initially conducted using exponential smoothing state space modeling (ETS). The seasonal characteristics were further investigated using spectral density estimation, which identified dominant cyclical components in the frequency domain. Although this analysis revealed some periodicity, the visual inspection of the decomposed ETS model and the results of spectral analysis did not indicate strong or consistent seasonality that warranted the addition of seasonal parameters in the ARIMA framework. As a result, a full Seasonal ARIMA (SARIMA) model was not implemented in the final forecasting pipeline. However, this exploratory step provided valuable insight into the frequency structure of cocoa prices and supported the decision to proceed with a non-seasonal ARIMA model.

4.4 ARDL Model

The Autoregressive Distributed Lag - Error Correction Model (ARDL-ECM) was employed as a more explanatory framework to analyze the impact of macroeconomic and climate variables on cocoa futures prices. The model incorporated monthly averaged data from Ghana and Cameroon, including exchange rates and average temperatures, as well as a time trend to capture structural climate effects. Prior to estimation, each series was tested for stationarity using the ADF test and found to be either $I(0)$ or $I(1)$, satisfying the conditions for ARDL modeling. Cointegration between cocoa prices and predictor variables was verified using the Engle-Granger two-step method, confirming the existence of a long-run equilibrium relationship. Optimal lag orders were selected using the `auto_ardl()` function, and several model variations were estimated on both full and truncated datasets to account for potential structural breaks. Diagnostic tests were conducted to

evaluate normality, autocorrelation, and homoskedasticity of residuals. An ARDL-ECM structure was then specified to isolate short-run dynamics and long-run effects, offering interpretable insights into how exchange rate fluctuations and climate variations influence cocoa prices over time. Ultimately, the ARDL-ECM model and ARDL model were equivalent hence, ARDL was the final version used for forecasting. For ARDL forecasting, ARIMA models for each predictor variable were constructed and used to generate forecasts from December 2024 to July 2025 (training split). These forecasted values were then fed into the ARDL model, along with the lagged values of the predictors from July 2024 to November 2024 (testing split), to predict monthly prices from December 2024 to July 2025 (forecasting split). Initially, data for Ivory Coast (Google Finance, 2025d)(Open-Meteo, 2025) and Nigeria (Google Finance, 2025c)(Open-Meteo, 2025), specifically the exchange rate fluctuations and weather were also used, but upon assessing the correlational matrix, Ivory Coast and Nigeria's average temperature and exchange rate data were highly correlated with each other, making them redundant in the model. If added, it would lead to highly inflated and unstable coefficients in the ARDL model, which risks overfitting and is thus excluded from the final forecasting model (further discussed in Forecasting and Results).

5 Forecasting and Results

When the data was cleaned, it aligned from 2007 to 2024. As a result, the model was trained on data from July 2007 to July 2024, tested on data from August 2024 to November 2024, and used to generate forecasts from December 2024 to July 2025. Since data is available for December 2024 to February 2025, cross-validation can be performed during that period to assess the model's performance, given that the price of cocoa is already known up to the present date.

5.1 ARIMA Model

An ARIMA(p, d, q) model follows the formula:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Y_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \epsilon_t$$

Where:

- Y_t is the time series value at time t
- B is the backshift operator (shifts values back in time)
- d is the number of times the series is differenced to make it stationary
- ϕ terms represent the autoregressive (AR) part
- θ terms represent the moving average (MA) part
- ϵ_t is random error (white noise)

For an ARIMA(1,1,1) model, the formula simplifies to:

$$(1 - \phi_1 B)(1 - B)Y_t = (1 + \theta_1 B)\epsilon_t$$

And the forecasting equation is:

$$Y_t = Y_{t-1} + \phi_1(Y_{t-1} - Y_{t-2}) + \theta_1 \epsilon_{t-1} + \epsilon_t$$

Where:

- Y_t is the predicted value at time t
- Y_{t-1} is the previous value
- ϕ_1 is the autoregressive (AR) coefficient
- $(Y_{t-1} - Y_{t-2})$ represents the differencing step to remove trends
- θ_1 is the moving average (MA) coefficient
- ϵ_t and ϵ_{t-1} are random error terms

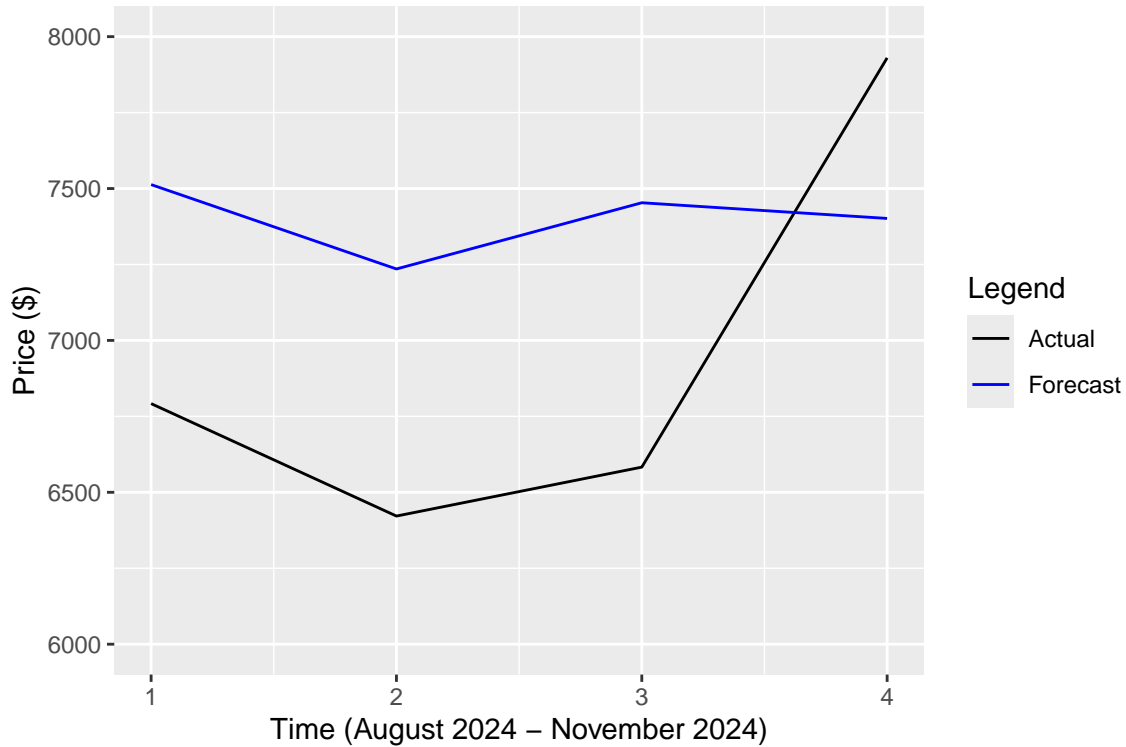


Figure 2: ARIMA Forecast on Testing Data from August 2024 to November 2024

Figure 2 illustrates the actual and forecasted cocoa prices from August to November 2024, generated using an ARIMA model. The forecast (blue line) demonstrates a relatively stable trend within a narrow price range, while the actual prices (black line) exhibit more significant fluctuations, including a sharp decline in September and a substantial surge in November. Notably, the ARIMA model fails to capture this late upward trend, leading to a considerable divergence between the forecast and the actual observed prices. This suggests a limitation in the model's ability to predict sudden shifts or external factors influencing the cocoa market during this period, warranting a potential re-evaluation of the model's parameters or consideration of alternative forecasting methodologies for improved accuracy.

Figure 3 illustrates a significant failure of the ARIMA model to accurately forecast cocoa prices for November 2024, as the actual price exhibits a dramatic upward surge immediately preceding and presumably continuing into the forecast period, while the model predicts a relatively flat trajectory at a much lower price point. This stark divergence highlights the ARIMA model's inability to capture the sudden and substantial shift in market dynamics, suggesting a potential structural change or the influence of unmodeled external factors driving the recent price explosion. Consequently, the forecast lacks practical value and underscores the need for a more responsive forecasting approach capable of adapting to rapidly evolving market conditions.

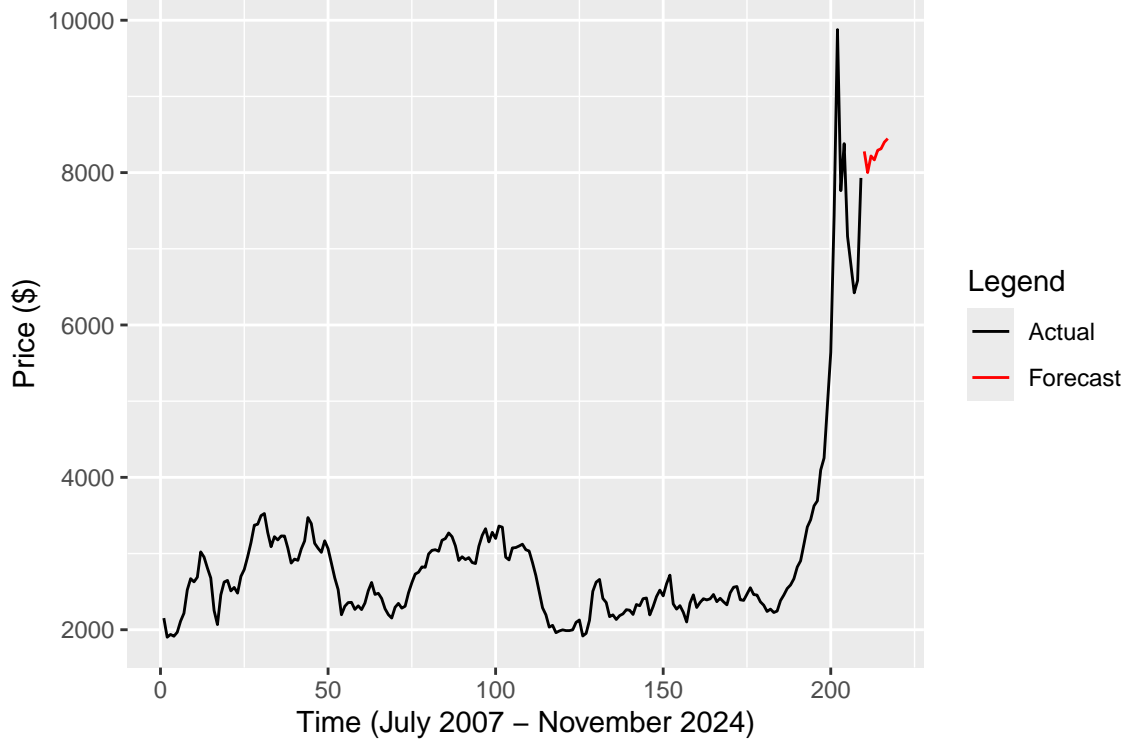


Figure 3: ARIMA Forecast for Cocoa Price from December 2024 to July 2025

5.2 ARDL Model

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_5 Y_{t-5}$$

$$+ \gamma_0 X_{1,t} + \dots + \gamma_5 X_{1,t-5} + \delta_0 X_{2,t} + \dots + \delta_5 X_{2,t-5}$$

$$+ \theta_0 X_{3,t} + \dots + \theta_5 X_{3,t-5} + \lambda_0 X_{4,t} + \dots + \lambda_5 X_{4,t-5} + \tau_0 T_t + \dots + \tau_5 T_{t-5} + \epsilon_t$$

- Y_t = Daily price
- X_1 = Cameroon exchange rate
- X_2 = Cameroon average temperature
- X_3 = Ghana exchange rate
- X_4 = Ghana average temperature
- T_t = Trend variable
- ϵ_t = Error term

$$Y_t = 1.486 \times 10^3 + 1.199 Y_{t-1} + 1.249 Y_{t-2} + 8.837 \times 10^{-2} Y_{t-3}$$

$$+ 1.150 Y_{t-4} + 7.431 \times 10^1 Y_{t-5} + (3.543 \times 10^5) X_{1,t} + (2.050 \times 10^5) X_{1,t-1} - (7.765 \times 10^3) X_{1,t-2}$$

$$- (6.272 \times 10^5) X_{1,t-3} - (1.482 \times 10^5) X_{1,t-4} - (1.339 \times 10^5) X_{1,t-5} - (2.995 \times 10^1) X_{2,t} + 1.569 X_{2,t-1} - 1.579 X_{2,t-2}$$

$$\begin{aligned}
& -6.612X_{2,t-3} + 5.887X_{2,t-4} - 3.244X_{2,t-5} - (2.766 \times 10^2)X_{3,t} + 1.871X_{3,t-1} - 1.862X_{3,t-2} \\
& + 1.693X_{3,t-3} - 1.964X_{3,t-4} - 1.290X_{3,t-5} - (9.584 \times 10^2)X_{4,t} + 8.297X_{4,t-1} + 2.917X_{4,t-2} \\
& + 7.165X_{4,t-3} - 3.507X_{4,t-4} + 1.150T_t + \epsilon_t
\end{aligned}$$

Table 2: ARIMA Models for Cocoa Price Predictors

Predictor	Model
Cameroon Average Temperature	ARIMA(0,0,3)
Cameroon Exchange Rate	ARIMA(0,1,1)
Ghana Average Temperature	ARIMA(0,0,3)
Ghana Exchange Rate	ARIMA(0,1,3)

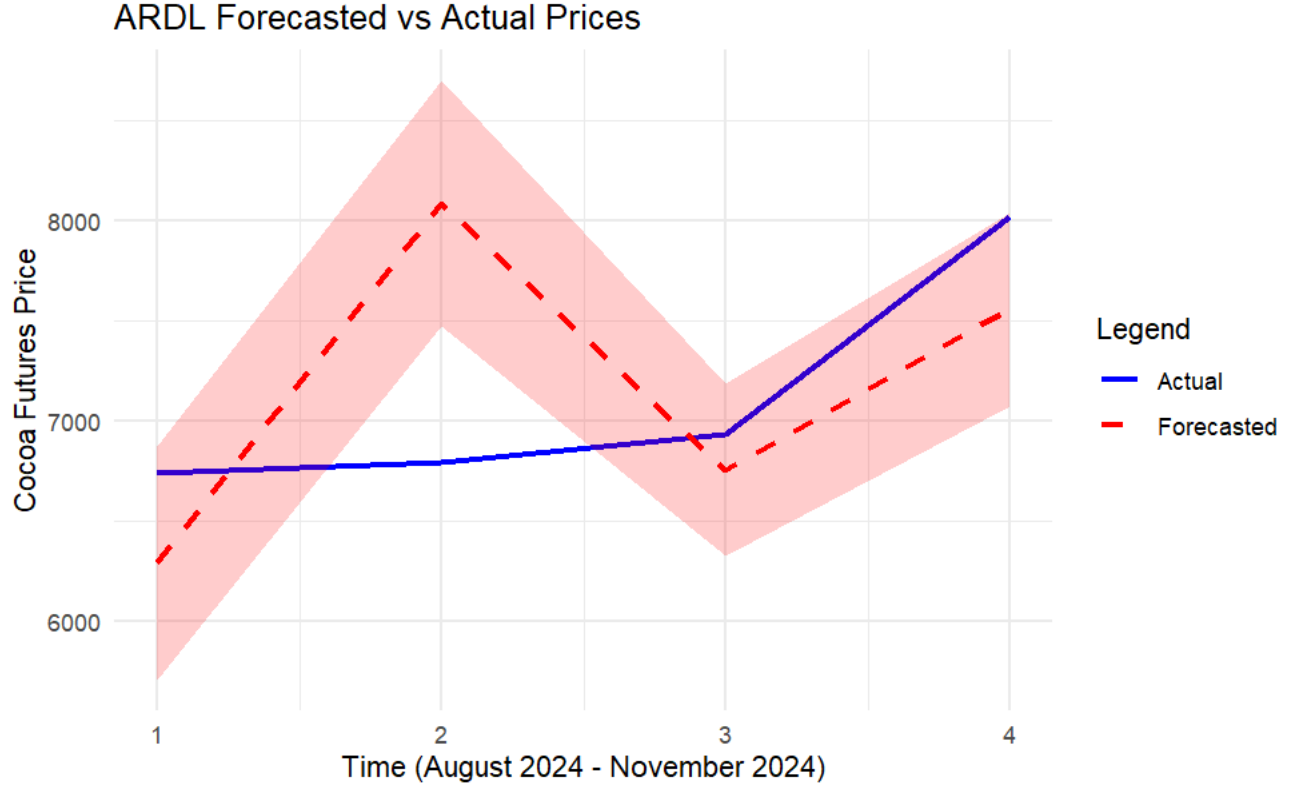


Figure 4: ARDL Forecast on Testing Data from August 2024 to November 2024

In the forecast plot comparing predicted and actual prices during the test period, the model initially overestimates prices in August and September 2024. Then, there is a phase of underestimation in October and November. The initial overestimation is likely due to the model being trained on the 2023–2024 spike, which is not representative of typical market behavior.

The ARDL forecast in Figure 5 exhibits a rapid decline followed by a rebound in prices. This pattern likely reflects the influence of the immediate pre-November 2024 data, which displays a similar trend.

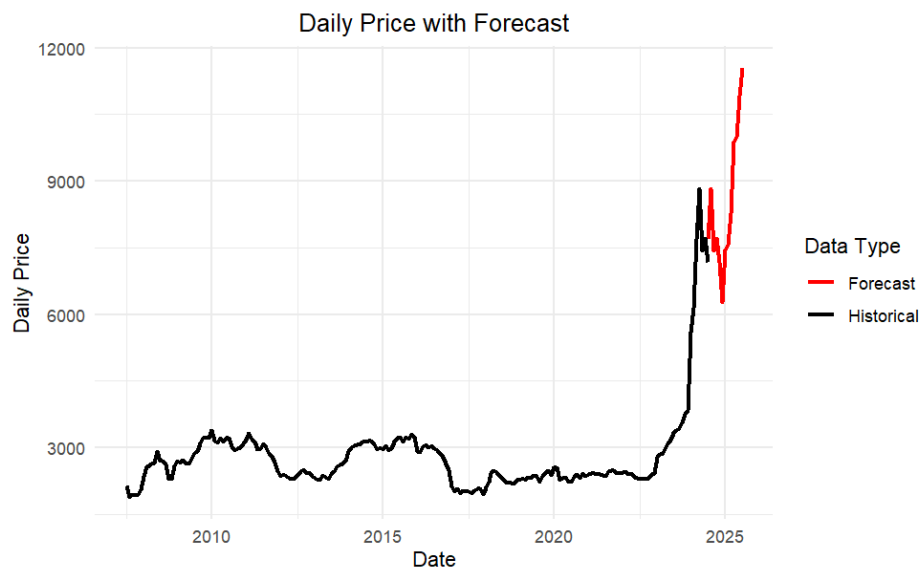


Figure 5: ARDL Forecast on Price Data from December 2024 to July 2025

Table 3: Forecast Comparison Table

Month	Actual.Price	ARIMA.Results	ARDL.Results
December 2024	9426.31	8278.464	6262.458
January 2025	10842.69	8000.739	7462.340
February 2025	9642.11	8218.715	7551.912

Table 4: RMSE Comparison Between Models

Model	RMSE.on.Training.Data	RMSE.on.Testing.Data
ARIMA	2959.442	744.716
ARDL	188.430	731.709

Table 5: Evaluation Metrics and Tests for ARIMA and ARDL

Test_Metric	ARIMA_Model	ARDL_Model
Ljung-Box	p = 0.811	p = 0.999
Shapiro-Wilk	p < 2.2e-16	p = 1.893e-15
GARCH	p < 2.2e-16	
Breusch-Pagan		p = 3.062e-05
AIC	2918.660	2727.068
Testing RMSE	744.72	731.707

6 Discussion

The results table presents three residual diagnostic tests. According to the Ljung-Box test, the null hypothesis that residuals are uncorrelated cannot be rejected at any conventional significance level, with a p-value of 0.99. This indicates that the residuals exhibit no significant autocorrelation. In contrast, the Breusch-Pagan test rejects the null hypothesis of homoscedasticity, suggesting the presence of heteroscedasticity in the residuals. Similarly, the Shapiro-Wilk test rejects the null hypothesis of normality, indicating that the residuals deviate from a normal distribution. These results are likely driven by a large spike in the data during 2023–2024, which introduces non-normality and variance instability.

Despite these issues, the lack of autocorrelation in the residuals is the most critical result. It indicates that the model adequately captures the dependence structure of the time series, which is the primary objective in time series modeling. The estimated model includes three significant lags of daily cocoa prices, along with one significant lag each for Cameroon’s average temperature, Ghana’s average temperature, and Ghana’s exchange rate. Although many lag terms are individually insignificant, this does not necessarily undermine the model’s validity.

The trend term is statistically insignificant, likely due to the extreme variation introduced by a sharp price spike in the 2023–2024 period. This outlier event appears to have influenced both the trend estimation and model performance. Additionally, because the ARDL model relies heavily on lagged values, it may struggle to capture sharp price movements driven by unexpected market shocks. This limitation leads to under or overestimation during periods of heightened volatility. The confidence intervals also fail to encompass these sharp fluctuations, reinforcing the interpretation that the ARDL model is less responsive to sudden, short-term shocks. By the time the model adjusts to the impact of such shocks, actual prices may have already diverged onto a new trend path. The evaluation of model performance based on RMSE indicates potential overfitting. While the training RMSE is relatively low at 188, the testing RMSE rises sharply to 731, suggesting that the ARDL model fits the training data well but may generalize poorly to unseen data. This discrepancy is likely due to the model interpreting the sharp price disruption in cocoa prices during 2023–2024 as a deterministic trend rather than an outlier event.

The ARDL-ECM used in this analysis is an unrestricted error correction model that performs similarly to the standard ARDL model in terms of AIC. While it offers no gain in predictive accuracy, it provides insight into long-run equilibrium relationships. The error correction term has a coefficient of +0.07, suggesting divergence rather than mean reversion—prices tend to correct upward over time instead of stabilizing. The short-run coefficient on lagged price (Y_{t-1}) is 0.013, indicating a weak but positive short-term effect. Both positive coefficients are likely influenced by the sharp price spike in 2023–2024, which the model may misinterpret as a long-run equilibrium. This distorts the ECM’s understanding of long-term dynamics. From a practical standpoint, this result could be interpreted as suggesting that cocoa prices are persistently driven upward by strong demand, with supply failing to adjust sufficiently to restore equilibrium. However, such a conclusion is inconsistent with real-world supply-demand dynamics, where supply typically responds to price signals over time. The model’s behavior therefore highlights the sensitivity of long-run interpretation to outlier events and structural breaks in the data.

A key limitation of the ARDL model is its sensitivity to lag selection. The `auto_ardl` function, which selects lags based on AIC/BIC, may overfit by including too many lags—many of which are statistically insignificant. This overparameterization can lead to multicollinearity, distorting coefficient estimates and reducing interpretability. It may also explain the omission of early lags of the trend variable. Moreover, ARDL assumes structural stability and does not handle structural breaks well. As seen with the cocoa price disruption in 2023–2024, such events can degrade forecast accuracy and mislead the model’s long-run interpretation.

Forecasting cocoa prices is particularly difficult due to the sharp spike observed between 2023 and 2025 and the strong climate sensitivity of cocoa production. Ghana, a major global producer, is especially affected by weather variability linked to La Niña, part of the El Niño-Southern Oscillation (ENSO). The irregularity of these events adds significant uncertainty to production forecasts and their effect on futures markets. The price surge has also been driven by a global cocoa shortage caused by disease, climate change, and extreme weather. Additionally, chronic underinvestment in cocoa farms has contributed to a sluggish supply response

to rising demand (“Impact of rising cocoa prices | J.P. Morgan research,” 2024).

The global cocoa shortage, which has contributed to the price surge, can also be attributed to a combination of environmental factors and structural issues within the cocoa industry. Diseases like black pod and climate change-related issues, such as rising temperatures and unpredictable rainfall, have negatively impacted crop yields. Furthermore, chronic underinvestment in the cocoa sector, particularly in farm maintenance, infrastructure, and technology, has led to a slow response in increasing production to meet the rising global demand for cocoa. As a result, supply has not been able to keep pace with the growing demand, further exacerbating price increases. According to J.P. Morgan research (“Impact of rising cocoa prices,” 2024), these issues underscore the vulnerability of the cocoa market to both environmental and economic pressures, making it even more difficult to accurately forecast future price trends. The uncertainty in production and the volatility in the market due to these compounded factors make reliable cocoa price predictions a challenging task for economists and market analysts.

Between the ARIMA and ARDL models, the ARDL model demonstrates superior performance in terms of both AIC and test RMSE. This is likely due to the additional information it captures from the inclusion of data on Cameroon and Ghana. However, an interesting observation emerges in the forecast of monthly prices from December 2024 to February 2025: the ARIMA model generates forecasts that are closer to the actual values than those produced by the ARDL model. There are two possible explanations for this result. First, it is possible that the ARDL model performs better over a longer forecast horizon, such as 10 to 15 months, while a three-month comparison window may be too short to yield conclusive evidence about relative model performance. Second, and more plausibly, the ARDL model’s inferior short-term performance may be attributed to uncertainty propagation. Specifically, because the ARDL model relies on forecasts of several predictor variables, each generated by separate ARIMA models, it accumulates additional layers of uncertainty. These include not only model uncertainty inherent to the ARDL itself but also the compounded forecast errors from the individual predictor models. In contrast, the ARIMA model forecasts the target variable directly and thus only contends with its own model uncertainty. Given these findings, if the primary objective is to maximize predictive accuracy across the entire forecast period, the ARDL model is the more suitable choice. However, for short-term forecasting over the next few months, the ARIMA model appears to offer greater stability. Moreover, if the goal is to balance predictive performance with model simplicity (i.e., minimizing the number of predictors), the ARIMA model may again be the preferred option.

7 Conclusion

This paper investigates how weather patterns and exchange rate fluctuations in West Africa (Ghana and Cameroon) impact global cocoa futures prices. By applying two forecasting models, ARIMA and ARDL, the study aimed to determine whether incorporating weather and exchange rate variables from Ghana and Cameroon could enhance predictive accuracy for forecasting cocoa futures pricing, specifically exploring the question: **How do weather patterns and exchange rate fluctuations in West African countries, mainly Ghana and Cameroon, influence global cocoa futures prices?** Findings show that while both models captured price movements reasonably well, the ARDL model demonstrated a lower AIC and RMSE. Overall, it suggests superior predictive performance and incorporates external datasets and more variables like weather and exchange rate.

Despite this apparent advantage, the ARIMA model outperforms the ARDL model in short-term forecasts for December 2024 to February 2025, largely because the ARDL forecasts accumulate additional uncertainty from the separate ARIMA models used to predict its explanatory variables. Thus, for longer forecast horizons where capturing broader trends in exchange rates and temperature is crucial, the ARDL model is preferable. Over shorter windows, especially those fraught with sudden shocks or limited lead times, a univariate ARIMA may yield more stable short-term predictions.

These findings are highly relevant to stakeholders throughout the cocoa supply chain. Manufacturers and traders can leverage the ARDL forecasts to manage price risk and make more informed decisions regarding procurement timing, hedging strategies, and futures contracting. Likewise, governments in cocoa-producing countries can use these insights for more effective policy interventions, addressing issues such as climate

mitigation, farm investment, and exchange rate stability to secure both local producer livelihoods and global supply. For consumers, a better understanding of price fluctuations can offer transparency about the factors driving potential cocoa-product price increases.

However, the study’s narrow geographic focus on West Africa, driven by the fact that approximately 70% of global cocoa originates from this region, introduces its own set of limitations. The inherent correlations among West African exchange rates, climates, and structural market factors can blur causal interpretations and potentially inflate standard errors in regression models, challenging the assumption of independent and identically distributed (i.i.d.) data. Future research could expand the sample to other major cocoa-producing regions, such as Latin America and Southeast Asia, to capture a broader set of market conditions and reduce multicollinearity. Incorporating additional climate variables (e.g., rainfall, humidity, and soil conditions) and macroeconomic indicators (e.g., global demand metrics, interest rates in consuming countries) could also improve the model’s explanatory power and help account for structural breaks or sudden shocks.

Cocoa futures contracts remain a critical tool for hedging and price discovery in this volatile market, underscoring the importance of robust forecasting models. This study, while illustrating some of the complexities of commodity price modeling, shows that ARDL approaches provide nuanced insights and a modest edge in predictive capacity, especially when lagged financial and weather-related variables are included. In a world of increasingly unpredictable climatic events and global economic shifts, continued refinement of such models can aid both cocoa-producing nations and the wider supply chain in navigating risk, stabilizing livelihoods, and ensuring a more resilient market for cocoa consumers and companies alike.

8 Appendix

All accompanying code and graphs can be found at:

<https://github.com/tanya-pand3y/FuturesPricingForecasting>

9 Bibliography

- Assis, K., Amran, A., & Remali, Y. (2010). Forecasting cocoa bean prices using univariate time series models. *Researchers World*, 1(1), 71.
- Google Finance. (2025a). CFA franc BEAC to US dollar (XAF/USD) exchange rate. Retrieved March 13, 2025, from <https://www.google.com/finance/quote/XAF-USD>
- Google Finance. (2025b). Ghanaian cedi to US dollar (GHS/USD) exchange rate. Retrieved March 13, 2025, from <https://www.google.com/finance/quote/GHS-USD>
- Google Finance. (2025c). Nigerian naira to United States dollar (NGN/USD) exchange rate. Retrieved March 13, 2025, from <https://www.google.com/finance/quote/NGN-USD>
- Google Finance. (2025d). West African CFA franc to United States dollar (XOF/USD) exchange rate. Retrieved March 13, 2025, from <https://www.google.com/finance/quote/XOF-USD>
- Hassler, U., & Wolters, J. (n.d.). Autoregressive distributed lag models and Cointegration. *Modern Econometric Analysis*, 57-72. https://doi.org/10.1007/3-540-32693-6_5
- Husain, A. M., & Bowman, C. (2004). Forecasting commodity prices: Futures versus judgment. IMF Working Papers, 04(41), 1. <https://doi.org/10.5089/9781451846133.001>/ International Cocoa Organization. (2025). Statistics. <https://www.icco.org/statistics>
- J.P. Morgan. (2024, April 3). The impact of rising cocoa prices | J.P. Morgan research. <https://www.jpmorgan.com/insights/global-research/commodities/cocoa-prices>
- K. Coulibaly, Mekongcho Terence, Cae Erbao, & Zhang Ya Bin. (n.d.). Climate Change Effects on Cocoa Export: Case study of Cote d'Ivoire. Allied Social Science Association (ASSA)/ American Economic Association (AEA).
- National Centers for Environmental Information. (2025). Home. NOAA. <https://www.ncei.noaa.gov>
- Natsiopoulou, K., & Tzeremes, N. G. (2022). ARDL bounds test for cointegration: Replicating the Pesaran et al. (2001) results for the UK earnings equation using R. *Journal of Applied Econometrics*, 37(5), 1079–1090. <https://doi.org/10.1002/jae.2919>
- Open-Meteo. (2025). Historical weather API documentation. <https://open-meteo.com/en/docs/historical-weather-api>
- Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: With R examples (4th ed.). Springer.
- Sukiyono, K., Nabiu, M., Sumantri, B., Novanda, R. R., Arianti, N. N., Yuliarso, M. Z., ... & Mustamam, H. (2018, November). Selecting an accurate cacao price forecasting model. In *Journal of Physics: Conference Series* (Vol. 1114, No. 1, p. 012116). IOP Publishing.