A Comparative Approach to Forecasting Cocoa Price using GAM and SARIMA Models

STA457 Final Project

Group Members

Peize Zhang
Jinting Liu
Haoyuan Chen
Binhe Jia
University of Toronto

2025 Winter STA457: Time Series Analysis

1 Introduction

Cocoa, Ghana's core agricultural industry and a pillar of its economy, considerably influences employment and export. As an internationally traded commodity priced in US dollars, changes in cocoa price directly affect Ghana's economic stability, foreign exchange reserves, and national unemployment rate. Given the cocoa's importance, accurate price forecasting is crucial for supporting Ghana's social welfare and informing policy decisions. However, cocoa prices are inherently volatile due to factors such as seasonality, product demand, and climate variability, necessitating the use of advanced statistical tools for reliable predictions.

This analysis utilizes a comparative approach to forecast future cocoa prices and improve the prediction accuracy. One research demonstrates that changes in precipitation due to seasonal variation significantly impact cocoa-growing areas and yields (Asante et al., 2025). Thus, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model would be suitable for capturing the underlying seasonal patterns in cocoa prices. To ensure precision, multiple SARIMA models are developed and compared, with the optimal model selected for final forecasting. Additionally, previous research highlights the statistical significance of exchange rates in affecting cocoa supply and thus pricing in Ghana (Adegunsoye et al., 2024). To account for potentially impactful factors, specifically, exchange rate, temperature and precipitation, a Generalized Additive Model (GAM) is applied, allowing for the inclusion of nonlinear relationships between potential influential factors and cocoa prices.

The significance of this research lies in addressing challenges associated with potential seasonality, and external variables that impact cocoa prices. By developing various forecasting models, we can enhance the prediction accuracy. The results would be beneficial for the Ghana government, cocoa suppliers, and laborers to make decisions regarding economic policy, production amount, and employment, ensuring stability in the cocoa sector.

2 Literature Review

Effective modeling techniques are required for forecasting, especially for the price of commodity goods such as cocoa. A previous study applied the SARIMA method to forecast weekly price indices of red lentils for 2019 (Divisekara et al., 2021). Researchers identified the optimal SARIMA model by analyzing ACF/PACF plots of differenced data and comparing models using AIC, BIC, and error metrics. This study highlighted SARIMA's effectiveness in forecasting commodity price with cyclical fluctuations in the short term. Another study combined SARIMA and computational intelligence approaches, developing sophisticated models to forecast the prices of major food crops (Shao et al., 2018). The experimental results emphasize the limitations of SARIMA in handling nonlinear predictors, reinforcing the advantage of advanced models that can capture nonlinear relationships within the data. Kharin et al. (2023) applied GAM to analyze nonlinear price transmission between feed maize, eggs and chicken. Results show that by accommodating smooth and flexible functional forms, GAM provides a more precise representation of intricate price interactions, emphasizing its effectiveness in modeling nonlinear dependencies.

Prior research has highlighted the power of SARIMA and GAM models under different contexts in time series forecasting. Following the ideas established from prior studies, we develop multiple SARIMA models and select the optimal one based on statistical criteria to forecast cocoa prices. Similarly, to capture nonlinear relationship between price and external variables, multiple GAM models are built, and the best-performing one is selected. Then a direct comparison is made between SARIMA and GAM in forecasting. This comparative analysis has contributed to the growing literature on time series modeling by providing empirical evidence of the applicability of linear and nonlinear methods to real-world commodity price forecasting.

3 Methodology

3.1 Model Architecture

Our group considers Seasonal Autoregressive Integrated Moving Average (SARIMA) model and Generalized Additive Model (GAM) as potential models. SARIMA is popular due to its model structure in capturing trend and seasonality of a time series explicitly (Kumar Dubey et. al., 2021). GAM is powerful in modeling non-linear relationships using smoothing functions, with no mandatory stationary assumption required (Hastie & Tibshirani, 1990). Each model brings unique strengths, making them both suitable candidates for our forecast purpose.

3.2 SARIMA

SARIMA is a generalization of the Autoregressive Moving Average (ARMA) model to trend-stationary series with periodic variation (Hastie & Tibshirani, 1990). The model can be decomposed as:

$$\Phi_P(B^s)\phi_p(B)\nabla_s^D\nabla^d x_t = \delta + \Theta_Q(B^s)\theta_q(B)w_t$$

Here $\phi_p(B)$ and $\theta_q(B)$ are ordinary AR(p) and MA(q) components respectively, while $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are seasonal AR(P) and MA(Q) components with seasonality s. ∇^d and ∇_s^D are ordinary and seasonal differencing operators with differencing orders d and D respectively. δ is the constant drift term and w_t represents the leftover Gaussian white noise. In SARIMA model, (p, d, q) are standard ARIMA parameters capturing non-seasonal patterns, while (P, D, Q) are additional seasonal components. The differencing operators enable SARIMA to handle time series data with inconsistent mean. However, SARIMA may generate biased estimates and unreliable forecasts if data is heteroscedastic.

3.3 GAM

GAM is a generalized linear model (GLM) in which the response variable depends linearly on some smoothing functions of predictor variables (Box, 2015). In time series forecasting, the response Y_t is specified with an exponential family distribution, along with a link function $h(\cdot)$. Y_i is associated with linear predictor Z_i and nonlinear predictors $\{X_{1,t}, X_{2,t}, \dots, X_{m,t}\}$. We are interested in building smoothing functions $f_1(\cdot), f_2(\cdot), \dots, f_m(\cdot)$ with respect to nonlinear predictors $\{X_{1,t}, X_{2,t}, \dots, X_{m,t}\}$ and combine them with linear predictor Z_i, to describe the behavior of Y_t:

$$h(\mathbb{E}[Y_t]) = \hat{\beta}_0 + \hat{\beta}_1 Z_t + \hat{f}_1(X_{1,t}) + \hat{f}_2(X_{2,t}) + \cdots + \hat{f}_m(X_{m,t})$$

As for forecasting purpose, we will apply inverse-h transform to get the prediction function:

$$\hat{Y}_t = h^{-1}\left[\hat{\beta}_0 + \hat{\beta}_1 Z_t + \hat{f}_1(X_{1,t}) + \hat{f}_2(X_{2,t}) + \cdots + \hat{f}_m(X_{m,t})\right]$$

3.4 Evaluation Metric

Various evaluation metrics are used in our report. But in order to compare the forecast performance of models, our group uses Root Mean Square Error (RMSE) as the primary evaluation metric. If a vector of n predictions $\hat{Y}$ is generated from n testing samples, and Y is a vector of observed values of the variable being predicted, RMSE of the predictor is computed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

where Y_i represents observed value and $\hat{Y}_i$ represents predicted value.

Since large deviations in price forecasts can be costly, it is important that our model avoids major deviations in price predictions. RMSE squares the errors, which penalizes large deviations more than smaller ones. This aligns with our criteria for evaluating the quality of an ideal model. Therefore in part 4 of the report, we will compare RMSE of various models and the one with the least RMSE is considered optimal under our evaluation metric.

4 Experimental Setup

4.1 Dataset

Cocoa price data and weather variation data are provided by datasets "Daily Price_ICCO" and "Ghana_data", respectively. "Daily Price_ICCO" dataset contains daily cocoa price in USD per tonne from 1994-10-03 to 2025-02-27. "Ghana_data" dataset contains daily precipitation, daily average temperature, maximum daily temperature and minimum daily temperature in Ghana from 1990-01-01 to 2024-11-28. Considering the significant short-term fluctuations in daily financial market data, we convert the daily data into monthly averages while accounting for the varying number of days in each month, to enhance the robustness and interpretability of models. To ensure time consistency across datasets, we aligned the timeline of the "Ghana_data" to February 2025 by filling in missing values using data from the same time in the previous year.

Exchange Rate Dataset is obtained from World Bank Group (World Bank, n.d.) which contains data from 1994 to 2024 and Google Finance (Google, n.d.) which contains data after 2024. They contain the official annual exchange rate of Ghanaian Cedi (GHS) against USD. For simplicity, we assume that the exchange rate remains constant throughout each year. To maintain consistency with the timeline of the "Daily Price_ICCO" dataset, we only used exchange rate data from October 1994 to February 2025. To facilitate model training, the last three months of "Daily Price_ICCO" dataset is our testing set, with all prior data serving as the training set.

4.2 Implementation Framework

SARIMA model is implemented to explore the relationship between future cocoa price and past cocoa price only. It assumes that the underlying time series is trend-stationary. Therefore, we follow a data transformation pipeline before implementing SARIMA model:

1. Inspect the cocoa price time series data to check if there's trend or increasing variance, apply log-transformation to stabilize the variance and differencing to remove trend.
2. If there is still persistence in the seasons, apply twelfth-order differencing to capture seasonal effect.
3. If the transformed data shows stationarity, plot ACF and PACF plots to identify sequencing orders.
4. Implement various SARIMA models and make comparisons between them using metrics.
5. Perform model diagnostics checks, and select optimal SARIMA model.

While GAM allows for non-stationary data, we first use GAM to evaluate the significance of monthly average precipitation, monthly average temperature and monthly average exchange rate on cocoa price. Then, we fit GAM on those significant predictors and make GAM model comparisons. A final model comparison is made between the best SARIMA and the best GAM by comparing their RMSE on testing set. Furthermore, we employ both models to generate a 6-month ahead forecast utilizing the complete set of all data available. The 6-month horizon is chosen as it reflects a practical time period often used in commodity markets, helping people make informed decisions.

5 Result

5.1 Time Series Approach (ARIMA/SARIMA Model)

First, we convert our data into monthly average to maintain the data consistency and divide the data into two parts: the first 362 data points are used for fitting the model, referred to as training data, and the last three data points are used for forecast validation, RMSE computation, referred to as test data. To fit a SARIMA model for the training dataset, we set the seasonal period to 12 since we assume the price follows a periodic pattern which repeats yearly. There are 12 months in a year and month is the most representative unit for observing seasonality. Seasonal conditions such as temperature and demand cycles would affect the price. We can now handle data in the form of monthly prices, ensuring consistency with our data processing.

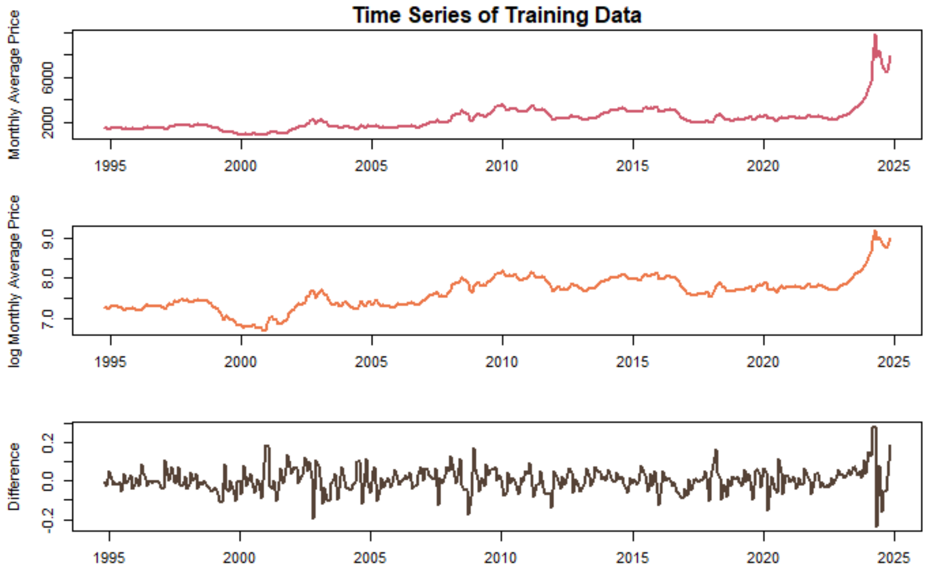


Figure 1: Time series plots in order from top to bottom: the original price, price with log-transformation, and price with both log-transformation and differencing.

Then we need to decide whether the time series is stationary. To do so, we run the augmented Dickey-Füller Test (ADF) which outputs a p-value of 0.7005. The large p-value provides strong evidence that the original data is not stationary and thus requires further transformation and differencing. According to Figure 1, original time series shows higher variance around year 2010, 2015, and 2024. To stabilize the variance, we apply a logarithm transformation to monthly average price. An increasing trend is then observed in the log plot, and thus we apply differencing to the data, obtaining the plot of first differencing.

The new time series representing the first differencing shows a relatively stationary shape: an approximately constant mean around 0 and constant variance is observed, and no trend or seasonality is observed. The augmented Dickey-Füller Test now shows a p-value less than 0.01, which is an indication that we successfully reached stationarity. Therefore, the differencing order d is set to 1. Since no clear seasonality is observed, seasonal differencing is not applied. We then examine the ACF and PACF plots of the differenced series:

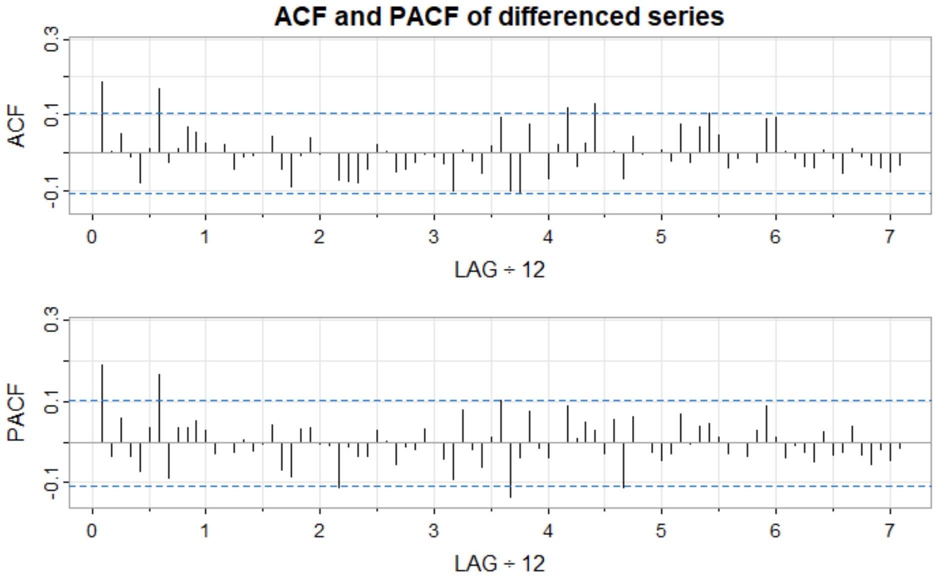


Figure 2: Plots of Sample Autocorrelation Function and Partial Autocorrelation Function of Differenced time series.

Figure 2 shows the ACF and PACF of the differenced time series, excluding values when the lag is 0. The x-axis, labeled LAG÷12, represents time in years, where each spike corresponds to a one-month lag. Each integer on the x-axis represents a yearly lag, equivalent to 12 monthly lags, corresponding to the 12 months in a year. For non-seasonal component, ACF approximately cuts off after 1 monthly lag, indicating an MA(1) process. PACF also approximately cuts off after 1 monthly lag, indicating an AR(1) process. No clear tail-off pattern is observed. Thus, we select $p = 1, d = 1$ and $q = 1$ as the non-seasonal sequencing order. For seasonal component, no obvious spike are present at yearly lags in either ACF or PACF plot. Since there is no clear cut-off or tail-off pattern, we select P=0, D=0, and Q=0 for seasonal sequencing order, indicating that no significant seasonality is observed in the series.

(a) ARIMA(1,1,1)

Coefficient	Estimate	Std Error	t-statistics	p-value
AR(1)	-0.1251	0.2954	-0.4236	0.6721
MA(1)	0.3267	0.2826	1.1560	0.2485
Constant	0.0048	0.0038	1.2731	0.2038

(b) ARIMA(0,1,1)

Coefficient	Estimate	Std Error	t-statistics	p-value
MA(1)	0.2056	0.0534	3.8508	0.0001
Constant	0.0048	0.0039	1.2462	0.2135

(c) ARIMA(1,1,0)

Coefficient	Estimate	Std Error	t-statistics	p-value
AR(1)	0.1928	0.0522	3.6930	0.0003
Constant	0.0048	0.0040	1.2159	0.2248

(d) ARIMA(0,1,1) $\times$ (0,0,1)$_{12}$

Coefficient	Estimate	Std Error	t-statistics	p-value
MA(1)	0.2053	0.0535	3.8366	0.0001
SMA(1)	0.0260	0.0605	0.4302	0.6673
Constant	0.0048	0.0040	1.2259	0.2210

(e) ARIMA(0,1,1) $\times$ (1,0,0)$_{12}$

Coefficient	Estimate	Std Error	t-statistics	p-value
MA(1)	0.2052	0.0535	3.8362	0.0001
SAR(1)	0.0250	0.0593	0.4205	0.6744
Constant	0.0048	0.0040	1.2259	0.2210

Table 1: Coefficient estimates and t-test results of SARIMA Models.

Model	RMSE	AIC
ARIMA$(0, 1, 1)$	2096.408	-2.74592
ARIMA$(1, 1, 0)$	2059.942	-2.74378
ARIMA$(0, 1, 1) \times (0, 0, 1)_{12}$	2061.678	-2.74089
ARIMA$(0, 1, 1) \times (1, 0, 0)_{12}$	2062.720	-2.74087
ARIMA$(1, 1, 1)$	2127.545	-2.74083

Table 2: RMSE and AIC comparison between models, sorted by AIC in ascending order from top to bottom.

The SARIMA function in R helps us estimate the coefficients using maximum likelihood estimation and perform t-test for the estimates. According to Table 1a, the p-value for AR coefficient estimate is greater than 0.1 in our first model, indicating that it is not statistically significant. This implies that adding AR(1) to the model might not be useful. Nonetheless, we can't directly remove it since we have to consider the interaction effect between AR and MA components. The combined effect of AR and MA components might not be clear. This problem often occurs due to the interdependence between AR and MA terms. Thus, we build 2 separate models, ARIMA$(0, 1, 1)$ and ARIMA$(1, 1, 0)$, to help show the unique contribution of each parameter. Table 1b and Table 1c both show statistical significant results with p-values of MA(1) and AR(1) less than 0.01. This indicates that 2 new models are suitable for the model selection.

In addition, we are interested in finding seasonal component's effect. Using R method AUTO.ARIMA, which automatically selects SARIMA models with the lowest AIC, we find the top 5 models with the lowest AIC (Table 2). Hence, ARIMA$(0, 1, 1) \times (0, 0, 1)_{12}$ and ARIMA$(0, 1, 1) \times (1, 0, 0)_{12}$ are also included in our model selection process since the AIC indicates a good fit, even though the t-test indicates that the seasonal components SAR(1) and SMA(1) show no significance (Table 1d,Table 1e). Five models in Table 2 are compared against each other using different statistical criteria. Among all models, ARIMA$(1, 1, 1)$ has the largest RMSE and AIC, indicating the weakest prediction accuracy, and it is the first to be excluded. The next models considered for elimination are ARIMA$(0, 1, 1) \times (0, 0, 1)_{12}$ and ARIMA$(0, 1, 1) \times (1, 0, 0)_{12}$. Though they have low RMSE, the t-tests indicate that the seasonal components are not significant. Keeping insignificant coefficients could lead to overfitting problem and reduced model interpretability. Moreover, while they have lower RMSE compared to ARIMA$(0, 1, 1)$, ARIMA$(1, 1, 0)$ demonstrates even slightly lower AIC and RMSE values. Therefore, these models with seasonal components are also excluded due to model model complexity and risk of overfitting. Then we would only consider ARIMA$(0, 1, 1)$ and ARIMA$(1, 1, 0)$.

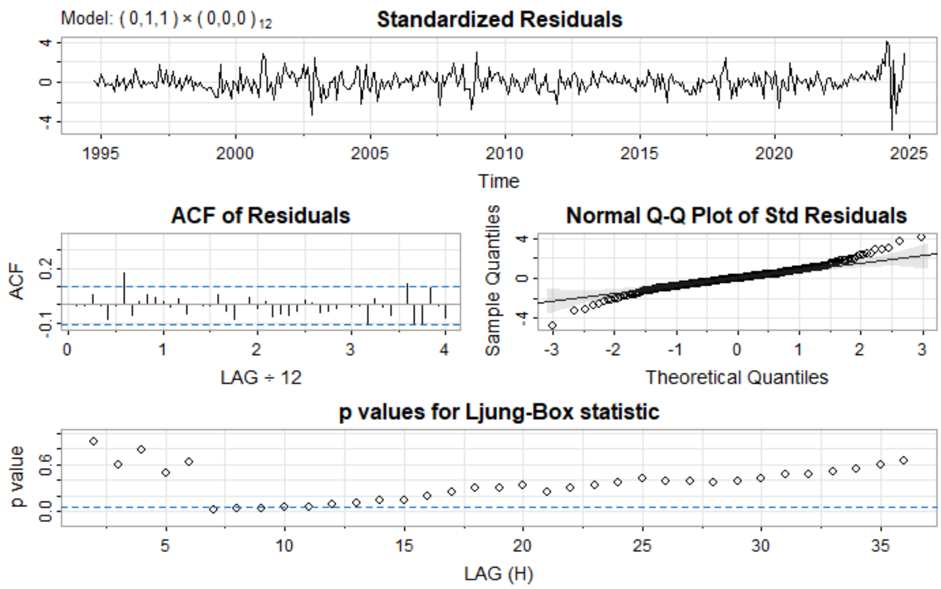


Figure 3: Model Diagnostics of ARIMA$(0, 1, 1)$.

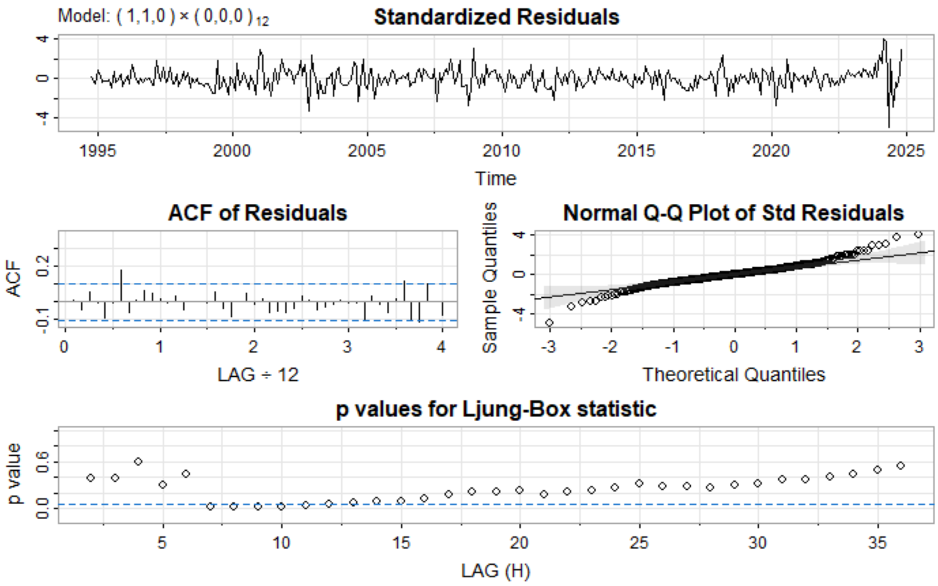
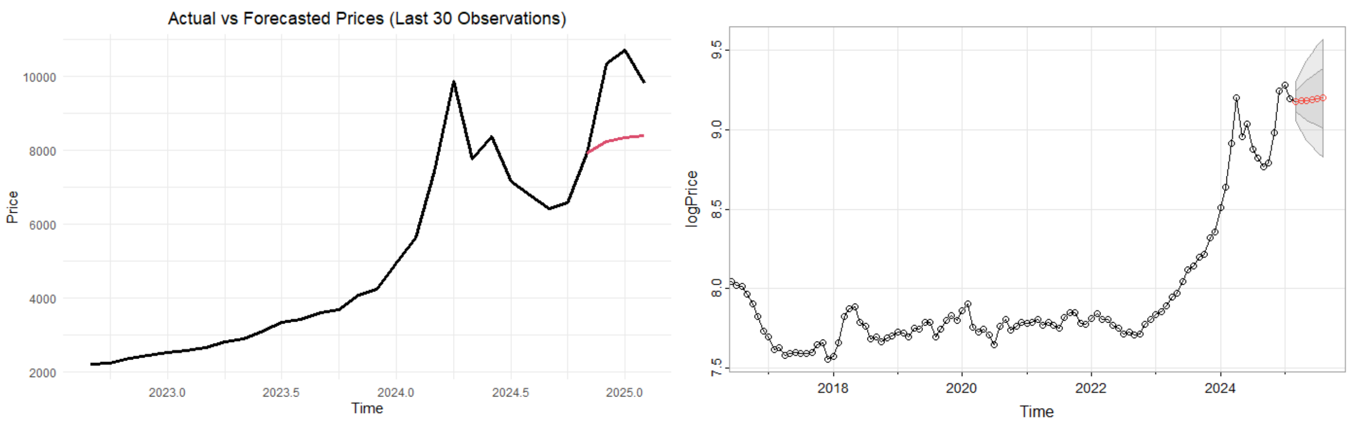


Figure 4: Model Diagnostics of ARIMA$(1, 1, 0)$.

ARIMA$(0, 1, 1)$ has a lower AIC while ARIMA$(1, 1, 0)$ has a lower RMSE. We first perform model diagostic checks for both models, containing residual plots, ACF plots, normal Q-Q plots, and Ljung-Box tests to see whether the residual assumptions are met. Figure 3 presents the diagnostics of ARIMA$(0, 1, 1)$. The residuals are fluctuating around 0, except for a time period around year 2024, suggesting that mean and variance are approximately constant. The ACF plot shows that most ACF fall within the error bounds of $\pm\frac{2}{\sqrt{n}}$, implying little significant autocorrelation. The normal Q-Q plot shows that most of the residuals are lying on the straight line, indicating approximately normally distributed residuals. The Ljung-Box test shows that most p-values are higher than 0.05, with 3 to 4 exceptions, so the residuals are mostly uncorrelated. Overall, ARIMA$(0, 1, 1)$ satisfies the assumptions well.

Figure 4, which displays the diagnostics for ARIMA$(1, 1, 0)$, provides similar results. The residuals are centered around zero and exhibit approximately constant variance. The ACF plot shows most ACF fall within the threshold, and the normal Q-Q plot indicates approximate normality. The Ljung-Box test results show that most p-values are above 0.05. However, compared to in Figure 3, Ljung-Box test in Figure 4 shows slightly more p-values below 0.05, around 1 to 2 additional points, suggesting a slightly weaker fit in terms of ACF compared to the other model.

In general, both models satisfy the residual assumptions decently. Therefore, we won't select the model directly based on the diagnostic tests.



(a) 3 months Validation vs. Actual values using test data.

(b) 6 months forecasting ahead using all data.

Figure 5: Forecasting Result

Since the differences in AIC, RMSE, and model diagnostics across the two models are relatively minor, we prioritize RMSE as our primary selection standard. The objective of this study is to generate accurate short-term forecasts, specifically focusing on six-month-ahead predictions. Given that our aim is not to interpret or explain the underlying model structure in depth, but rather to optimize predictive performance, RMSE, which directly represents the measure ability of forecast, is the most suitable and proper choice for model selection. Since ARIMA$(1, 1, 0)$ has the lowest RMSE, we believe it is the best fit to our data.

Figure 5 (a) shows the comparison of forecast and actual values from our last 3 observations. The red line represents the predicted cocoa price (in natural scale) using ARIMA$(1, 1, 0)$, and it is much lower than the actual values. We believe this is due to the high price volatility after year 2024. Therefore huge errors occur. Finally, we apply the selected model, ARIMA$(1, 1, 0)$, to forecast the natural logarithm of cocoa's monthly average prices for the next 6 months using all data, from March 2025 to August 2025. Figure 5 (b) presents these forecasts as red dots (in log scale). A slightly increasing trend is observed. The darker grey area represents the 80% prediction interval, and the lighter grey area represents the 95% prediction interval, meaning that we are 95% confident that an actual future observation will fall within this interval. The prediction intervals tend to become larger as we predict further into the future. This happens because uncertainty accumulates over time; the further ahead the forecast, the more change can influence the potential outcomes. Therefore, longer-term forecasts are less precise.

5.2 Generalized Additive Model

5.2.1 Model training and validation process

This section presents the results of the forecasting analysis using four Generalized Additive Models (GAM 1 to 4), each is constructed with various combinations of covariates. The models are trained on monthly cocoa price data ranging from October 1994 to November 2024, implying that the first 362 months are used as the training set. The last three months (December 2024, January and February 2025) are reserved for out-of-sample testing, and further extrapolated forecasts beyond February are generated for visual inspection of predictive trends.

Each GAM is fitted using the REML (Restricted Maximum Likelihood) approach, with smoothed components been modeled via penalized splines. While all models share a common structure, they vary in the inclusion of covariates such as monthly precipitation, average temperature, and exchange rates. Their forecasting performance are evaluated by comparing the predicted values for the last three months with the actual observed values.

Initially, we fit GAM 1 that includes month index (as indexed integer from 1 to 362 to represent months as a timeline), monthly total precipitation, average temperature, and the exchange rate. We treat all of these variables as smoothing terms. This is because GAMs are useful in situations where the relationship between predictors and the response is unknown or potentially nonlinear. If the effective degrees of freedom (EDF) for a smoothing term is close to 1, it usually means the relationship is approximately linear. But if the EDF is much larger than 1, it suggests a more complex, nonlinear pattern. Here we are using GAM 1 to evaluate potential nonlinearities between each covariate and monthly average cocoa price.

Based on Table 3, we observe that the effective degrees of freedom (EDF) for monthly average temperature and total monthly precipitation are close to one, and their p-values are relatively large. The large p-values suggest that the smoothing terms for these two variables are not statistically significant in the model. Therefore in GAM 2, we treat monthly average temperature and total monthly precipitation as fixed (linear) effects rather than smoothing terms.

According to the second part of Table 3, we observe that when monthly average temperature and total monthly precipitation are included as linear terms in the model, their p-values still remain above 0.05. This suggests that these two predictors are not statistically significant either as a linear or nonlinear predictor. Therefore, we choose to remove them from the model in the next specification.

To further refine the model, we construct GAM 3, which excludes monthly average temperature and total monthly precipitation as predictors. Therefore, GAM 3 includes only two predictors: month index and the exchange rate, both work as smoothing terms. This choice allows us to capture the potentially nonlinear temporal trend and exchange rate dynamics without the noise introduced by irrelevant covariates. As shown in Table 3, both smoothing terms in GAM 3 have relatively high effective degrees of freedom and extremely low p-values, suggesting that they are statistically significant and exhibit non-linear relationships with the log-transformed monthly cocoa price.

Finally, we construct GAM 4, which includes month index as the only predictor. The purpose of constructing GAM 4 is to compare its prediction results with those of GAM 3, in order to examine whether incorporating exchange rate would improve model's forecast performance.

Generalized Additive Model 1

$\ln(\mathbb{E}[\text{Average Price}]) = \beta_0 + f_1(\text{Month}) + f_2(\text{Precipitation}) + f_3(\text{Temperature}) + f_4(\text{Exchange Rate})$

Coefficient	**Estimate**	**Standard Error**	**t-statistic**	**p-value**
(Intercept)	7.6565	0.2564	-0.4256	0.6721
Smoothing Term	**Effective df**	**Reference df**	**F-statistic**	**p-value**
s(Month)	8.380	8.810	27.330	$< 2 \times 10^{-16}$
s(Precipitation)	1.002	1.003	0.326	0.569
s(Temperature)	2.935	3.714	1.239	0.370
s(Exchange Rate)	7.654	8.472	12.394	$< 2 \times 10^{-16}$

Generalized Additive Model 2

$\ln(\mathbb{E}[\text{Average Price}]) = \beta_0 + \beta_1(\text{Precipitation}) + \beta_2(\text{Temperature}) + f_1(\text{Month}) + f_2(\text{Exchange Rate})$

Coefficient	**Estimate**	**Standard Error**	**t-statistic**	**p-value**
(Intercept)	7.489	0.311	24.112	$< 2 \times 10^{-16}$
Precipitation	0.001	0.001	0.840	0.401
Temperature	0.002	0.004	0.521	0.603
Smoothing Term	**Effective df**	**Reference df**	**F-statistic**	**p-value**
s(Month)	8.367	8.805	27.290	$< 2 \times 10^{-16}$
s(Exchange Rate)	7.656	8.474	12.510	$< 2 \times 10^{-16}$

Generalized Additive Model 3

$\ln(\mathbb{E}[\text{Average Price}]) = \beta_0 + f_1(\text{Month}) + f_2(\text{Exchange Rate})$

Coefficient	**Estimate**	**Standard Error**	**t-statistic**	**p-value**
(Intercept)	7.657	0.008	940.300	$< 2 \times 10^{-16}$
Smoothing Term	**Effective df**	**Reference df**	**F-statistic**	**p-value**
s(Month)	8.402	8.820	29.530	$< 2 \times 10^{-16}$
s(Exchange Rate)	7.685	8.494	13.220	$< 2 \times 10^{-16}$

Generalized Additive Model 4

$\ln(\mathbb{E}[\text{Average Price}]) = \beta_0 + f_1(\text{Month})$

Coefficient	**Estimate**	**Standard Error**	**t-statistic**	**p-value**
(Intercept)	7.661	0.009	820.200	$< 2 \times 10^{-16}$
Smoothing Term	**Effective df**	**Reference df**	**F-statistic**	**p-value**
s(Month)	8.778	8.986	173.600	$< 2 \times 10^{-16}$

Table 3: Coefficient estimates and statistical significance tests for four Generalized Additive Models (GAM 1–4). Each model includes parametric and/or smoothing terms for selected predictors. Coefficients are reported with standard errors, test statistics (T or F), and corresponding p-values. Smoothing terms are evaluated using approximate F-tests.

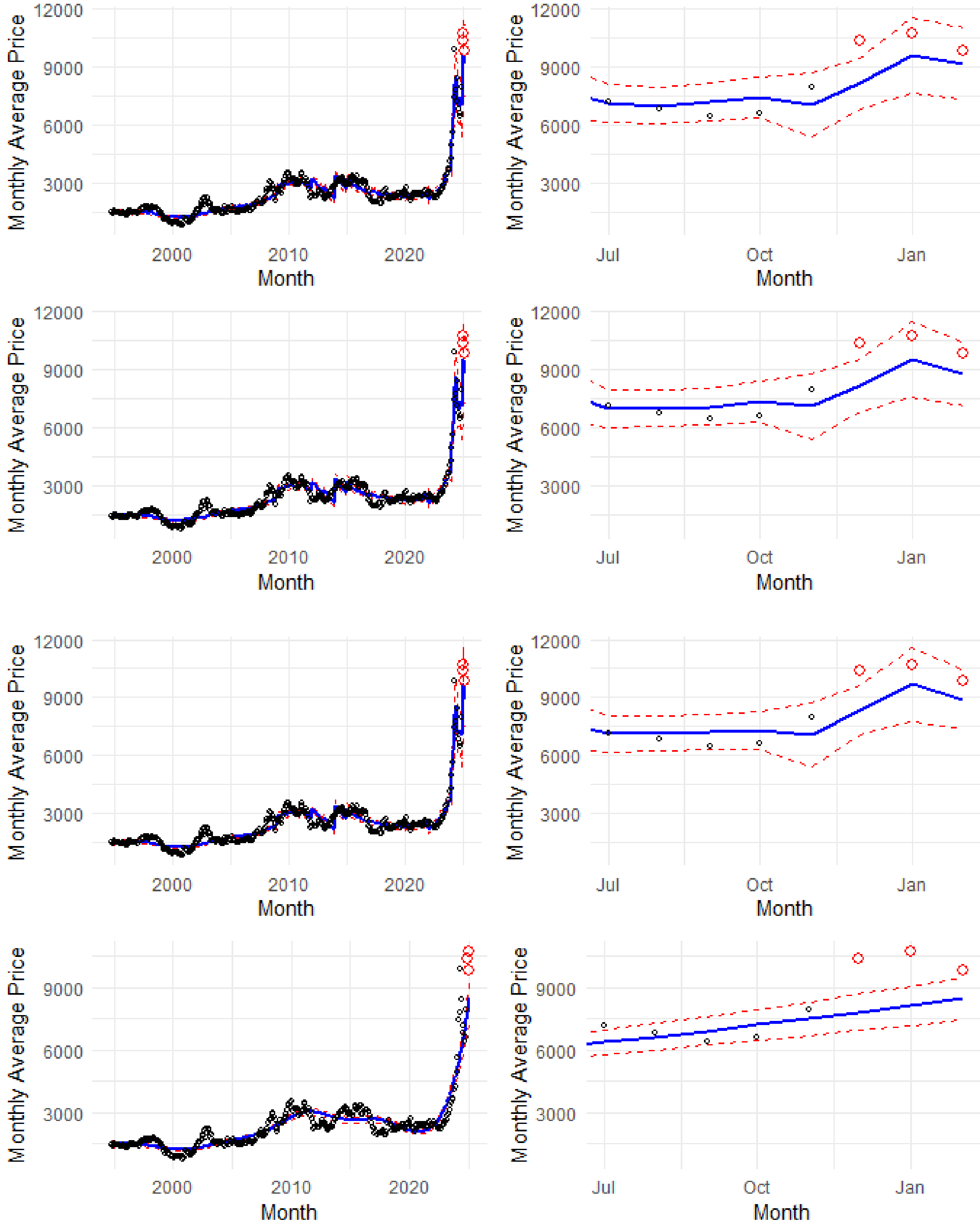


Figure 6: Forecast comparison plots of Generalized Additive Models (GAM 1–4). Each row corresponds to one model, where the left panel displays predicted prices (blue solid line) versus actual prices from the training set (black dots), and the right panel zooms into the forecast window with actual test data (red circles), predicted values (blue line), and 95% confidence intervals (red dashed lines). The time range spans from October 1994 to February 2025.

5.2.2 Performance Evaluations

We now evaluate the forecast comparison plots of all the four GAMs. In addition, we assess their forecasting performance using key metrics, including adjusted R^2 , deviance explained, AIC, MSE, and RMSE.

Based on the left panels of Figure 6, all four GAMs successfully capture the overall trend in monthly average cocoa future prices. However, the forecasted curves in the last row (GAM 4) appear smoother compared to the first three models. This increased smoothness is likely due to the exclusion of the exchange rate variable. When fewer covariates are included in the model, the fitted trend often appears smoother because the model has fewer patterns or fluctuations to account for.

Turning to the right panels of Figure 6, we observe that the confidence intervals of the first three models (GAM 1–3) all contain the actual prices for January and February 2025. However, none of these models accurately capture the spike in December 2024. This discrepancy may be due to unexpected fluctuations in cocoa prices during that month, which the models are not able to anticipate, even though they are designed to be flexible. It's possible that the smoothing method makes the model less responsive to sudden and sharp price changes.

For GAM 4, the forecast interval doesn't include any of the actual prices from January and February 2025. This raises concerns about its forecasting capability. One potential explanation is that GAM 4 is over-smooth and excludes exchange rate information, which may be an important driver of cocoa price dynamics during this period.

Model	$\boldsymbol{R^2_{\text{adj}}}$	**Deviance Explained (%)**	**AIC**	**MSE**	**RMSE**
GAM 1	0.9146	87.80	5228.039	2223574	1491.165
GAM 2	0.9143	87.58	5228.232	2403256	1550.244
GAM 3	0.9144	87.57	5224.480	2047209	1430.807
GAM 4	0.8235	83.23	5317.067	4991039	2234.063

Table 4: Comparison of four Generalized Additive Models (GAM 1–4) based on adjusted R^2, deviance explained, and prediction accuracy metrics (AIC, MSE, RMSE).

While GAM 1 and GAM 2 show slightly higher adjusted R^2 values and deviance explained than GAM 3, the differences are relatively small. Similarly, although GAM 4 has the lowest AIC among the four models, the improvement compared to GAM 3 is minimal. In practice, what we care about most is the RMSE, since it directly reflects the average prediction error between the model's estimates and the actual prices. Among all models, GAM 3 clearly achieves the lowest RMSE, suggesting that it performs best in terms of forecasting accuracy. This result supports the use of GAM 3 for predicting future monthly cocoa prices. It also highlights that including exchange rate as a smoothing term helps improve model performance, as shown by the lower RMSE when compared to GAM 4, which does not include it.

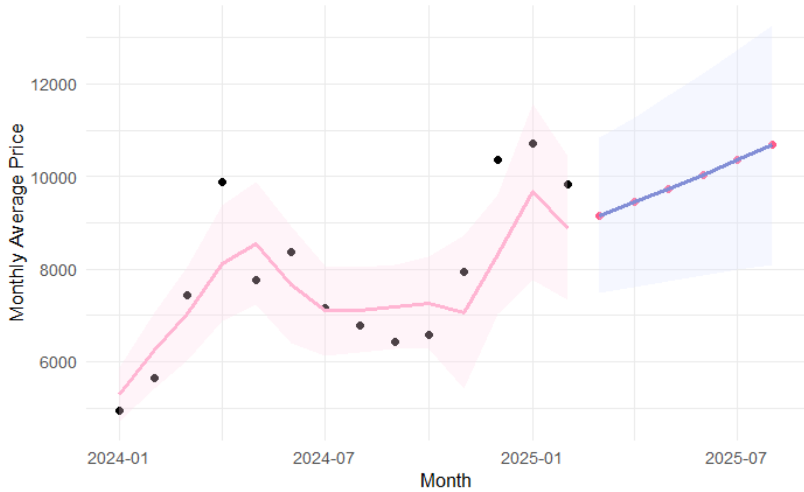


Figure 7: Forecast visualization using GAM 3. The black dots show observed cocoa prices (Jan 2024–Feb 2025), with the pink line and band indicating fitted values and 95% confidence intervals. The purple line and red dots represent 6-month forecasts, with the lavender band showing the 95% forecast interval.

Date	Exchange Rate	Month Index	Forecasted Price
2025-03-01	15.27	366	9159.386
2025-04-01	15.27	367	9444.210
2025-05-01	15.27	368	9737.891
2025-06-01	15.27	369	10040.704
2025-07-01	15.27	370	10352.994
2025-08-01	15.27	371	10674.873

Table 5: Forecasted monthly average cocoa prices from March to August 2025 using GAM 3.

Using GAM 3, we generate a 6-month forecast of monthly cocoa prices from March to August 2025. While in reality, we will not have the access to future exchange rate. Therefore in our forecasting, we assume the exchange rate will remain stable at its February 2025 level, without unexpected short-term shocks.

The forecast reveals a steady upward trend in cocoa prices over the next half-year. This rising pattern may reflect continued pressure on the market, possibly due to limited supply or increasing demand. Interestingly, this trend is also consistent with the fluctuations seen during the most recent observed months, suggesting that the model effectively captures both seasonal effects and the impact of exchange rate changes. Overall, the results highlight the importance of including meaningful predictors, such as the exchange rate, when modeling monthly cocoa future prices.

5.3 Final Model Selection

	ARIMA(1,1,0)	GAM 3
RMSE	2059.942	1430.807

Table 6: RMSE comparison of ARIMA and GAM models

To select the best model for forecasting, we evaluated their accuracy and interpretability of the prediction, focusing on short-term prediction. As shown in Table 6, GAM indicates better accuracy as it has a substantially lower RMSE. This means GAM's predictions are closer to the actual values. Although AIC values are also available for both models, they were not chosen for direct comparison. Different packages compute AIC differently for different models and different packages have different formulas for AIC computation.

Hence, comparing GAM with ARIMA using AIC is not suitable, as they belong to different model types.

GAM provides deeper insight into the potential variables affecting price. GAM 3 includes exchange rate and time index as non-linear predictors, identifying significant non-linear relationships. This enhances the explanatory power of the model. It can adjust the prediction according to recent exchange rate, indicating a more practical advantage. In addition, unlike ARIMA, which relies on differencing, GAM estimates a smooth trend directly from the data, and can capture fluctuation that may not follow a fixed pattern, modeling trend.

Thus, GAM 3 is selected as the best model for forecasting cocoa prices. However, it's important to notice that we use the true exchange rate when doing prediction and evaluation on testing dataset. However, when we are doing forecast in reality, future exchange rates are unknown. So we choose to assume the future exchange rate stays constant. This will in some ways affect forecast accuracy.

6 Discussion

Our study provides important insights into the effectiveness of SARIMA and GAM in forecasting commodity prices. Experimental results reveal the nonlinear effect of exchange rate in cocoa price, highlighting the strength of GAM as a strong forecasting tool capable of capturing nonlinear patterns. However, several limitations remain and warrant further discussion. These limitations also suggest potential directions for future improvements.

6.1 Limitation

6.1.1 Dataset and Data Processing

The original dataset consists of daily observations, but we transformed it into monthly averages to reduce complexity. This transformation reduces the number of observations dramatically, limiting the model's ability to capture complex patterns. Thus it might be difficult to find seasonal pattern. After 2024, cocoa prices suddenly increased and became highly volatile. However, since we cannot obtain future data for this period, it may be difficult for the model to accurately predict future prices given this abnormal behavior.

6.1.2 Limitations of SARIMA

SARIMA is a classic univariate model, as it only requires the historical data of one single variable, and it assumes linearity and stationarity. Therefore, it can't directly incorporate the effect of external variables. This property becomes problematic when the cocoa price is heavily influenced by other variables. These non-random effects are smoothed out by the SARIMA model. As a result, when dramatic changes occur, such as the cocoa price surge around 2024, SARIMA fails to respond to these unstable variations. Since most of our data belongs to 1994-2022 during which the price only increased gradually without shocks, our model might overfit to the data during this period and therefore cause low prediction accuracy. An article suggests that ARIMA models are best suited to low-noise, linear systems (Kontopoulou et al., 2023). Their accuracy decreases when dealing with volatile and structurally complex environments.

6.1.3 Exchange Rate Assumption and Forecast Limitations of GAM

In our GAM, we cannot make future predictions without assigning values to key predictors like the exchange rate. Since we don't know what the exchange rate will be in the future, we choose to hold it constant at its February 2025 level. While this is a practical approach, it does have some limitations. Unlike weather variables such as temperature or precipitation, which usually follow seasonal patterns, exchange rates are influenced by many unpredictable factors like political events or economic policy. Hence, assuming that the exchange rate won't change could lead to huge forecast errors, especially if there are sudden market shifts.

6.2 Possible Model Improvement

6.2.1 Future Values of Predictors in GAM

As discussed in 6.1.3, the key problem of using GAM is the need for future values of predictor variables. To address this, one solution is to collect forecasted exchange rate data online through trustful agencies. Or we can review other articles regarding exchange rate forecast, using their models or building our own model to forecast the exchange rate itself. The forecasted rates can then be input into GAM. This method introduces uncertainty but allows the model to operate in reality. Alternatively, we can assume a range of reasonable future exchange rates. For instance, for simplicity, our analysis assumes constant rate when forecasting. We could also assume a constantly increasing or decreasing exchange rate. By forecasting or simulating future covariates, we can extend the applicability of GAM to real-world applications.

6.2.2 Data Size and Machine Learning Approach

One of the most straightforward way to improve the model performance is to incorporate more data and conduct multivariate time series analysis rather than univariate time series. Another way to improve the model performance is to apply machine learning approaches. While it poses danger of potentially overfit the data, it is possible to avoid it by fine-tuning the hyperparameters, assigning regularization term and data drop-out. However, this requires huge time commitment as well as compute power which is unfortunately not permitted by the available resources and time. An example of such applications can be found in an article (Shao et. al., 2018), in which the authors were able to achieve much better performance using simple integration of ARIMA and Artificial Neural Network (Rice Price Forecast RMSE: 0.12037) compared to ARIMA alone (Rice Price Forecast RMSE: 0.53077).

7 Conclusion

In this analysis, we explored two different statistics models, SARIMA and Generalized Additive Models, to forecast monthly average cocoa prices. Cocoa plays an important role in Ghana's economy, and cocoa price is affected by different variables such as weather and exchange rate. This creates challenges that requires both accuracy and flexibility. We compared SARIMA, a classic univariate time series model, with GAM, a more flexible approach which can detect the non-linear relationship between external variables and cocoa price.

Through a comprehensive SARIMA selection process, we ultimately identified ARIMA$(1, 1, 0)$ as the best-performing SARIMA model, noticing that no seasonality was found. While this model effectively captures trends, its inability to incorporate external factors, such as exchange rate, limits its prediction accuracy, especially during highly volatile periods (around 2024). In contrast, GAM is able to capture the relationships between cocoa price and external variables using smoothing functions. We built four GAMs and evaluated the significance of precipitation, temperature, exchange rate and month index. According to the comparison among the GAMs, GAM 3 indicates the best overall performance across the metrics, especially RMSE. It includes month index and exchange rate as smoothing predictors, with both showing statistically significant results. This confirms GAM's effectiveness in capturing non-linear relationship.

We select GAM 3 as the best model for forecasting cocoa prices over the next six months. GAM incorporates the effects of predictors, captures nonlinear trends, and provides relatively more accurate forecasts. While SARIMA remains a useful tool for time series forecasting, its limitations become apparent in complex situations, such as our cocoa price case, and it is therefore not chosen for final forecasting. Nevertheless, obtaining future values for predictors, such as the exchange rate in our example, remains a vital challenge when applying GAM in reality.

8 Reference

Adegunsoye, E. A., Tijani, A. A., & Kolapo, A. (2024). Liberalization vis-à-vis non-liberalization trade policy: Exploring the impact of price volatility on producer share price and cocoa supply response in Nigeria and Ghana. *Heliyon*, *10*(12), e32741-. https://doi.org/10.1016/j.heliyon.2024.e32741

Arlt, J., & Trcka, P. (2021). Automatic SARIMA modeling and forecast accuracy. *Communications in Statistics. Simulation and Computation*, *50*(10), 2949–2970. https://doi.org/10.1080/03610918.2019.1618471

Asante, P. A., Rahn, E., Anten, N. P. R., Zuidema, P. A., Morales, A., & Rozendaal, D. M. A. (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-. https://doi.org/10.1016/j.agrformet.2025.110393

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

Divisekara, R. W., Jayasinghe, G. J. M. S. R., & Kumari, K. W. S. N. (2021). Forecasting the red lentils commodity market price using SARIMA models. *SN Business & Economics*, *1*(1). https://doi.org/10.1007/s43546-020-00020-x

Google. (n.d.). *USD/GHS currency exchange rate & news*. Google Finance. https://www.google.com/finance/quote/USD-GHS?sa=X&ved=2ahU KEwjDyLiNoauMAxUov4kEHZtlJ0AQmY0JegQIARAs

Hastie, T. J., & Tibshirani, R. J. (1990b). *Generalized additive models*. Chapman & Hall/CRC.

Kharin, S., Kapustova, Z., & Lichner, I. (2023). Price transmission between maize and poultry product markets in the Visegrád Group countries: What is more nonlinear, egg or chicken? *Agricultural Economics (Praha)*, *69*(12), 510–522. https://doi.org/10.17221/320/2023-AGRICECON

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet*, *15*(8), 255-. https://doi.org/10.3390/fi15080255

Kumar Dubey, A., Kumar, A., García-Díaz, V., Kumar Sharma, A., & Kanhaiya, K. (2021). Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies and Assessments*, *47*, 101474-. https://doi.org/10.1016/j.seta.2021.101474

Official Exchange Rate (LCU per US$, period average) - ghana. World Bank Open Data. (n.d.). https://data.worldbank.org/indicator/PA.NUS.FCRF?end=2023&locations=GH&start=1960&view=chart

Serinaldi, F. (2011). Distributional modeling and short-term forecasting of electricity prices by Generalized Additive Models for Location, Scale and Shape. *Energy Economics*, *33*(6), 1216–1226. https://doi.org/10.1016/j.eneco.2011.05.001

Shao, Y. E., Dai, J.-T., & Koczy, L. T. (2018). Integrated Feature Selection of ARIMA with Computational Intelligence Approaches for Food Crop Price Prediction. *Complexity (New York, N.Y.)*, *2018*(2018), 1–17. https://doi.org/10.1155/2018/1910520

9 Appendix

Our R code and processed data can be found in the Github repository (https://github.com/Binhe-Jia/A-Comparative-approach-to-Forecasting-Cocoa-Price-Using-SARIMA-and-GAM-Models)