Introduction

Cocoa is the primary material for producing chocolate. The market is so large that many countries' economies depend highly on cocoa production. Therefore, it is especially helpful for companies and countries that aim to maximize profit based on forecasted cocoa prices. Among these countries, Ghana, Brazil, and Indonesia are three representers of corresponding regions: West Africa, Latin America, and Southeast Asia. West Africa accounts for 70% of global cocoa production, so Ghana can well represent the global cocoa production. (Terminal8 GmbH, B. (n.d.)) Exchange rate between the Ghanaian Cedi and the US dollar also helps build the model as the cocoa market uses the US dollar. (Adegunsoye, E. A., Tijani, A. A., & Kolapo, A. (2024)). In addition, these three countries have similar climate and geographic characteristics including precipitation and temperature. The average temperature is between 20 and 30 degrees, and there is enough precipitation for cocoa trees' growth.

Notably, it is challenging to precisely and effectively predict the price because of the frequent fluctuations brought by cocoa price's sensitivity to the environment and economy. This includes climate change, currency exchange rate, supply chain, and global demand. (Beg, M. S., Ahmad, S., Jan, K., & Bashir, K. (2017).).

In following paragraphs, we will analyze a couple of papers that used similar statistical techniques to forecast cocoa prices. We will discuss the similarity and difference between our model and existing models. We will use multiple time series models and select the best one based on theoretical concerns. We then use multilinear regression model and machine learning algorithm to make the prediction. Before presenting our methodology, we will provide descriptions with visualizations that can directly reflect statistical characteristics of the data. We will also fit each model from the real cocoa price dataset to find the best fitted model and present our results. Finally, draw conclusions with potential limitations and improvements of our analysis for future research.

Literature Review

Since the market of cocoa is worth to study due to its economic influence, there are existing papers that have predicted cocoa prices with time series models. The first paper (Sukiyono et al., 2018) concludes that ARIMA is the best model by using statistical indicators including MAD, MAPE and MSD. Based on the monthly cocoa price for both the world and Indonesian from Jan. 2008 to Dec. 2016, ARIMA has the lowest value of all indicators for world cocoa price and domestic cocoa price, so it is the most appropriate model for forecasting the world. Besides ARIMA, we can also use GARCH to make prediction, as the second paper (Kamu et al., 2010) states. Similarly, the author uses RMSE, MAE, MAPE, and U-statistics as indicators based on the average monthly data from January. 1992 to December.2006. Result shows the value of these indicators is the lowest for GARCH. Predicted cocoa price are increasing linearly in 2007. The third article (Sari, Duran, Kutlu, & Arslan, 2024) shows that machine learning methods are another way of forecasting. It uses daily price data of eleven distinct agricultural commodities goods including cocoa from start date to Mar. 11, 2022. LSTM is used to predict, which shows that predicted cocoa price is stable in Mar. 2022 to Jun. 2022. Hence, all papers emphasize time series model's characteristics which are the capability to take historical data into account to make accurate prediction.

Additionally, the first difference between our research and the previous papers is the dataset we used. First two papers used monthly price data. Instead, we use daily dataset from 1994 to 2025 for making the prediction more accurate by using more observations. The second difference is that we use ARIMAX, the model takes climate change and the exchange rate into the model. This increases accuracy of model. Similar to the previous paper, we use RMSE, AIC, MAE, and MAPE to help select our model. In addition, cocoa price is relatively stable for all previous dataset, while recently there was sudden jump. It makes our prediction different from the second and third articles' result, and also makes prediction challenging. Eventually, we also include machine learning algorithm and multilinear regression model in comparison with in-class traditional models, which assist us to find the limitations and potential improvements of our model.

Data

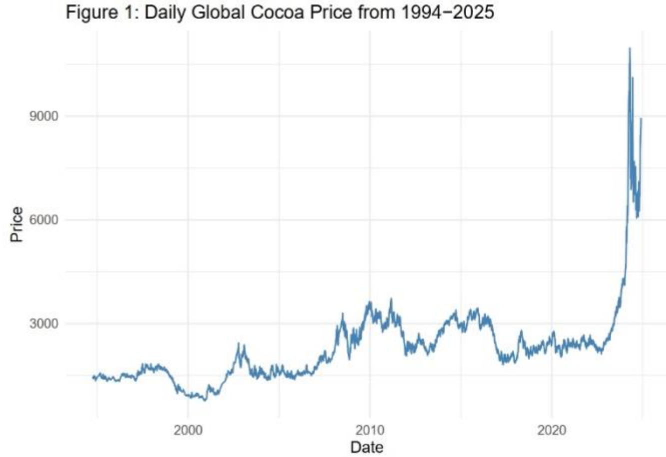
a) Data Description

As previously discussed, the global cocoa price is influenced by other factors, including climate conditions and exchange rates. For this project, forecasting global cocoa prices involved three primary datasets: daily cocoa price data from the International Cocoa Organization (ICCO) (International Cocoa Organization, 2025), Ghana climate data from the National Centers for Environmental Information (NCEI.Monitoring.Info@noaa.gov, n.d.), and historical exchange rates between the US dollar (USD) and the Ghanaian Cedi (GHS) sourced from Investing.com (USD GHS Historical Data - Investing.com Canada, n.d.).

Cocoa Price Data:

This dataset includes daily closing cocoa prices (Price), standardized to the “yyyy-mm-dd” format. The data was cleaned by removing duplicate date observations.

Figure 1: Daily Global Cocoa Price from 1994–2025



As illustrated in Figure 1, global cocoa prices generally exhibited an upward trend with moderate fluctuations from 1994 to 2023, followed by a notable spike between 2024 and 2025. The price series also displays day-to-day noise, represented by random fluctuations without clear or predictable patterns.

Ghana Climate Data:

This dataset comprises two main climate variables: daily precipitation (PRCP) and daily average temperature (TAVG). These variables are critical indicators of cocoa growing conditions in Ghana.

USD/GHS Exchange rate Data:

Historical USD/GHS exchange rate data was compiled from two datasets, covering periods before and after October 2013, and then merged into a single coherent dataset.

Integrated Dataset (cocoa_data):

Three datasets are merged by "Date" using left join. The resulting dataset contains Date, PRCP, TAVG, Rate. In the following content, the cocoa_data dataset was used for cocoa price forecasting via Time Series Model, Multi-Linear Regression Model and Machine Learning method. Based on our initial assessment of the data, we decided to first apply a log transformation followed by first differencing to address the issue of heteroscedasticity. Both transformed versions of the data will be retained separately. This approach allows flexibility in meeting the needs of different models—for example, traditional in-class models like ARIMA can automatically determine the appropriate differencing order, whereas models like MLR and machine learning algorithms may prefer to work directly with the log-transformed prices.

Figure 2: Log-transformed Daily Global Cocoa Price from 1994–2025

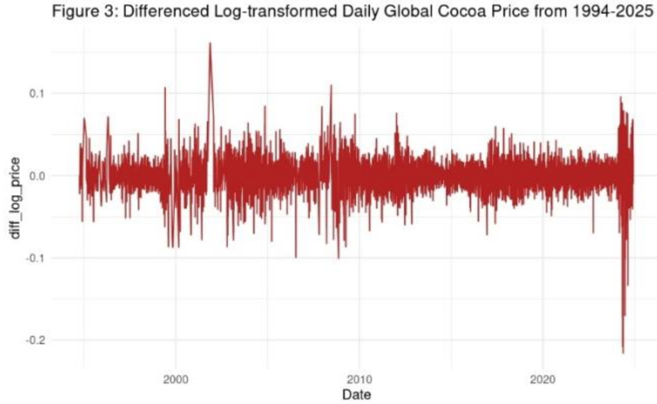


Figure 3: Differenced Log-transformed Daily Global Cocoa Price from 1994-2025

b) Data Splitting

Based on Figure 1, a significant spike occurred starting in 2023. Therefore, the training dataset (6434 observations) was selected from October 3, 1994, to January 1, 2024, encompassing historical cocoa prices and the initial part of the spike. The testing dataset (230 observations) spans from January 2, 2024, to November 28, 2024. Subsequent methodologies will investigate the most effective models for predicting the observed spike and fluctuations.

Methodologies：

First, the traditional in-class models including ETS, ARIMAX and SARIMAX will be applied to predict cocoa beans price after 2024. Particularly, ‘X’ in the models represents other factors that may influence the cocoa bean price, such as precipitation and average daily temperature in the Ghanaian region, and the exchange rate of the U.S. dollar to the Ghanaian Cedi. The differenced log-transformed price in the train data is then introduced into the model so that R can automatically fit the status of ETS model with respect to Error, Trend and Seasonality. Also, the train data is brought into the model allowed R to automatically select the ARIMAX (p, d, q) model with respect to the AR order p, difference order d and MA order q; and the SARIMAX (p, d, q) (P, D, Q, s) where (p, d, q) are non-seasonal AR, d and MA orders, and (P, D, Q, s) are the seasonal AR, d, MA orders and seasonal period s. Then, the fitted model is used to predict the price of cocoa beans after 2024, and the test data are combined to generate RMSE, MAE, MAPE, and AIC to further select the model with the best prediction ability. Meanwhile, by checking whether the plots of standardized residuals fluctuate around zero, the ACF images of residuals all cut off after the lag = 0, the Q-Q plot of residuals follows a straight line without large deviations, and the P-Values of residuals for Ljung-Box are greater than 0.05 to fail to reject H0: No autocorrelation up to lag = h to confirm that the best model is statistically significant. Finally, reconstruct the forecasted price (measured in USD/tonne) from diff(log(price)) and evaluate the plot of the predicted price against the actual price.

Second, Generalized Autoregressive Conditional Heteroscedasticity model is further introduced to predict the future price of cocoa beans by combining the classic ARIMAX model for analyzing the past price movements and volatility forms. Then, the statistics of the model are summarized to verify whether the predicted value of the coefficient $\alpha 1$ is statistically significant. Similarly, the residuals assumption is checked to confirm that the model is also statistically significant. Finally, if the model passes all tests, the predicted prices from reconstruction (measured in USD/tonne), and RMSE, MAE and MAPE are used to compare with the best model selected above.

Third, an out-of-class statistical model — Multilinear Regression (MLR) model with Recursive Rolling Forecast — is introduced. This is not merely an alternative model that incorporates multiple potential factors influencing cocoa price changes, but also a tool through which the statistical significance of each factor can be examined — specifically, by evaluating whether the p-values of each factor are below 0.05. This allows for investigation into whether key variables such as precipitation, average daily temperature and exchange rate are statistically significant factors of cocoa price fluctuations. To mitigate the high-frequency noise and fluctuations present in the diff(log(price)), we use the log(price) directly as the target variable. This offers a more interpretable and economically meaningful representation of price levels. The model predicts one-time step ahead by linearly combining lagged log(price) and external covariates. Specifically, lag_1 represents the most recent known or predicted log(price) at time t, while lag_2 corresponds to the log(price) at time t-1. More importantly, this recursive walk-forward strategy prevents data leakage because at time t+1 the model will only be trained on information available up to time t, and when the data at time t+1 is successfully predicted, its prediction value is saved in the historical data and updated to be the new value at time t for assisting to forecast the value at time t+2. Finally, test the residuals assumptions of the MLR model, and reconstruct the results to produce RMSE, MAE and MAPE, and the plot of forecasted prices vs. the actual prices to further demonstrate the differences between the MLR model and the best classical model got above.

Fourth, the last extracurricular model, a special RNN algorithm in machine learning: the Long Short-Term Memory Neural Network (LSTM), will be introduced in order to predict financial data such as cocoa bean prices with sequential temporal dynamics and external factors that can affect it. Firstly, to avoid data leakage, the previously cleaned data is divided into two files for model training and testing respectively, which further ensures that the model cannot read the test data before measuring the prediction results. Then, a historical feature is produced manually for each time step to simplify training by combining classical autoregressive modeling with deep learning, which reduces the burden on R required the importation of Python environments. Also, enable LSTM to automatically encode static lag features (lag1, lag2, etc.) into state representations and then memorize the patterns through hidden layers, and build structures that allow the model to learn across sample structures. All these steps allow the LSTM model to handle the long-term dependency problem better than traditional models (ARIMA, MLR), benefiting from its special cell state and gate mechanisms. Finally, let the model automatically train and predict, and measure the prediction results with test data. It is worth noting that in order to find the relatively optimal results, the model is set to randomly initialize the train data rather than a strict starting point and find the relatively optimal results after several training sessions with different weights.

Results and Discussions

a) Classic Time Series Models

The following tables summarize the results for *ETS1, ETS2, ARIMA*, and *SARIMA* based on the implemented methods using R Studio.

Model	RMSE (Test Set)	MAE (Test Set)	MAPE (Test Set)	AIC
ETS Model 1	0.04250931	0.0291992	Inf	3613.470
ETS Model 2	0.04250931	0.0291992	Inf	3613.470
ARIMAX	0.04246658	0.0290799	Inf	-34561.51
SARIMAX	0.04246659	0.0290799	Inf	-34561.88

Table 1: Forecast Accuracy Metrics Across Different Models for differenced log-price (Based on Test Set Performance Metrics)

Table 1 compares the testing dataset square root of mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) for each model for the differenced log-price variable based on the training dataset. Over all models, ARIMAX shows the least RMSE 0.04246658, slightly smaller than the SARIMA model, and same MAE for both models. RMSE and MAE for ETS1 and ETS2 are higher than ARIMA and SARIMA. It's trivial to compare MAPE for these four models since they're all infinite. Based on the comparison of RMSE and further validation of AIC value, ARIMAX (0, 0, 1) has the overall lowest RMSE and AIC value, indicating it is the best model within the four classic models.

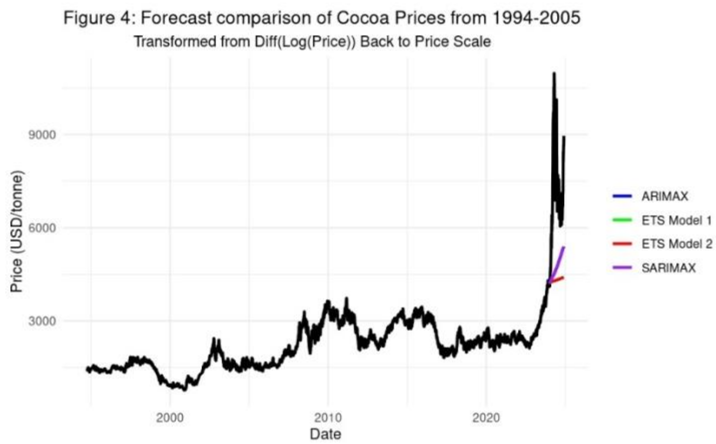
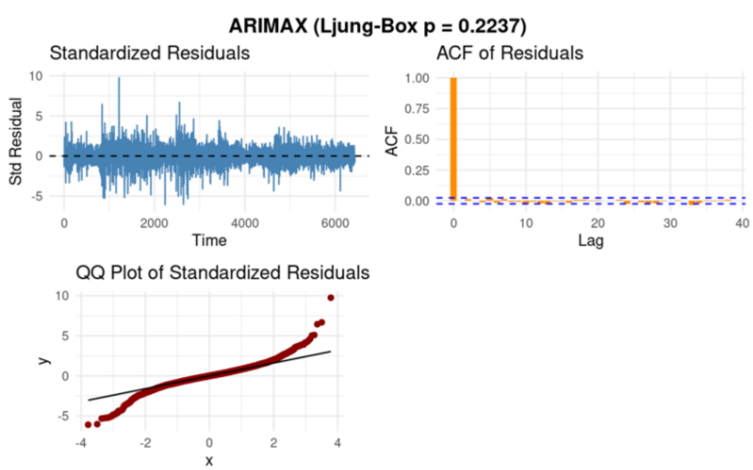


Figure 4 shows the forecasted testing data plot for all four models, further justifying that ARIMAX is the best model since it presents the closest increasing trend to actual data. ARIMAX and SARIMAX present a more robust increase than ETS1 and ETS2, sticking more tightly with actual spikes. It is notable that ARIMAX and SARIMAX present very similar forecasting data except for slightly different RMSE, MAE and AIC. This may be due to the global daily cocoa price dataset not exhibiting obvious seasonality trends, as shown in the complete dataset. On the other hand, ETS1 and ETS2 models completely overlap, both presenting worse forecasts than ARIMAX and SARIMAX.

Figure 5: Diagnostic Plots for ARIMAX Model



Since ARIMAX is found to be the best model, it is necessary to check the following four assumptions: The standardized residuals randomly scatter around zero with no visible pattern. ACF plot quickly decays after lag = 0, indicating no correlation between residuals. QQ plot exhibits a roughly straight line with slightly deviations because of some extreme outliers. Ljung-

Box p-value is 0.2237, greater than 0.05, indicating that there is no evidence against null hypothesis, therefore it is statistically significant that residuals are uncorrelated white noise. ARIMAX model satisfies all residual checks. ARIMAX was found to be in the following formula:

$$\phi(B)Y_t = \mu + \theta(B)\varepsilon_t + \sum_{i=1}^{k} \gamma_i X_{i,t}$$

$$\hat{Y}_t = \mu(0) + 0.0454 \cdot \varepsilon_{t-1} - 0.0003 \cdot PRCP_t + 0 \cdot TAVG_t + 0.0001 \cdot Rate_t$$

~~Miu 是 0 因为没有 intercept 显示在 arimax$coef 里面~~

GARCH + ARIMAX Model

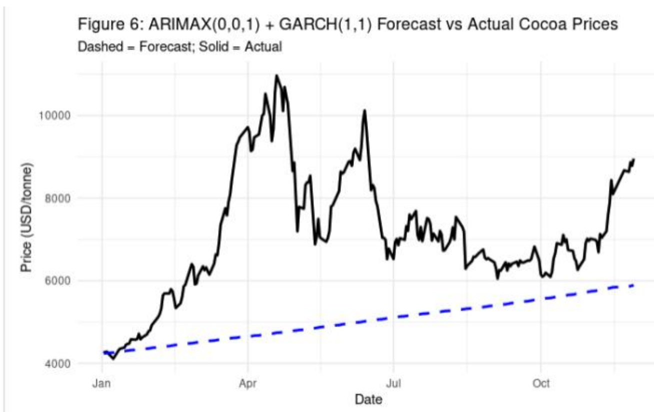


Figure 6: ARIMAX(0,0,1) + GARCH(1,1) Forecast vs Actual Cocoa Prices
Dashed = Forecast; Solid = Actual

In the forecasting graph, the GARCH + ARIMAX model exhibits an increasing trend which corresponds to the overall trend, but not precise fluctuation. Compared to all classical time series models, the forecasted differenced log prices from ETS, SARIMAX, and ARIMAX have been transformed back to the original price scale. The table below presents the residual error metrics for ETS, ARIMA, SARIMA, and GARCH+ARIMAX models.

Model	**RMSE (Test Set)**	**MAE (Test Set)**	**MAPE (Test Set)**
ARIMAX	2767.417	2327.700	30.03839 %
ETS Model 1	3137.737	2753.987	36.04764 %
ETS Model 2	3137.737	2753.987	36.04764 %
SARIMAX	2767.628	2327.944	30.04189 %
GARCH + ARIMAX	2535.402	2053.318	26.19934 %

Table 2: Forecast Accuracy Metrics Across Different Models

Based on the testing dataset error table, GARCH + ARIMAX achieves the lowest RMSE, MAE, MAPE values, indicating the best performance among these models. This suggests that incorporating GARCH efficiently enhances ARIMAX through capturing volatility feature of price data. While GARCH adds complexity, it improves overall predictive accuracy in this case. The following residuals assumptions for GARCH are checked.

Figure 7: ACF plot of standardized residuals for GARCH model

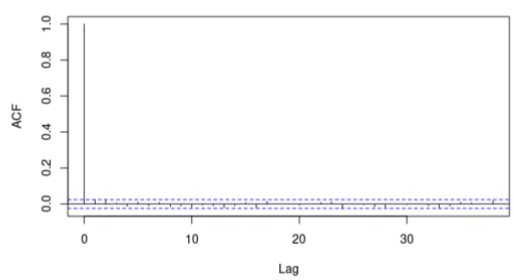
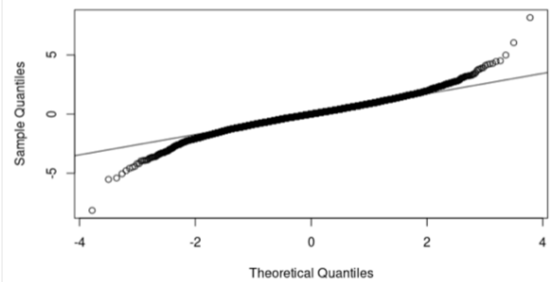


Figure 8: Q-Q plot of standardized residuals from GARCH model



ACF graph exhibits quick decay after lag = 0, indicating the residuals are uncorrelated. QQ plot exhibits a roughly straight with some heavy-tailed trait since the last two years' significant spikes were outliers far away from past data. The p value for Box-Ljung test is 0.3764, greater than significance level 0.05, indicating that residuals are uncorrelated statistically significantly. Overall GARCH satisfies the residual assumption test. Variance for GARCH (1,1) follows the formula:

$$\sigma_t^2 = 0.000001 + 0.022935 \cdot \varepsilon_{t-1}^2 + 0.972457 \cdot \sigma_{t-1}^2$$

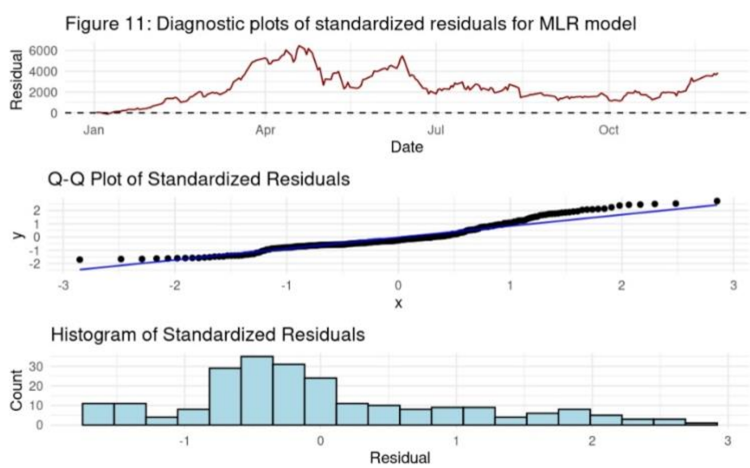
b) Out-Course Methods

Recursive MLR model:

Figure 10: Summary Statistic of MLR Model

Predictor	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0080	0.0076	1.0503	0.2936
lag_1	1.0457	0.0122	85.3925	0.0000
lag_2	-0.0476	0.0122	-3.8836	0.0001
PRCP	-0.0002	0.0004	-0.5220	0.6017
TAVG	0.0001	0.0001	1.0091	0.3130
Rate	0.0002	0.0001	2.5278	0.0115

In the forecasting graph, the MLR model presents a similar pattern as the GARCH model, but with a ~~stronger~~ weaker increasing trend, ~~however~~ moreover, it still doesn't capture more specific increasing trend that corresponds to the actual datasets.



From Figure 11, the following residual assumptions are checked: The residuals plot exhibits are mostly positive but are also not around zero with some large positive deviations particularly between March and July. This is because the initial part of the spike in 2024 April is too significant. The QQ plot shows a roughly straight line that indicates normal distribution for residuals. But the histogram shows a slightly left-skewed normal distribution, probably due to the early-stage outlier. Overall, the recursive MLR model satisfies residual assumptions. By examining the performance metrics for the recursive MLR model, we found that it has an RMSE of 2834.67, an MAE of

2417.75, and a MAPE of 31.33. Compared to the ARIMAX model, the RMSE of the MLR model is slightly higher.

The MLR formula is given below:

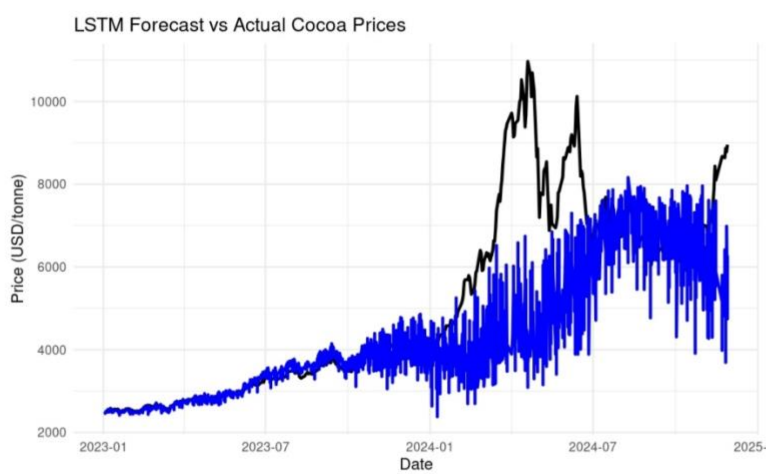
$$Y_t = \beta_0 + \beta_1 \cdot Y_{t-1} + \beta_2 \cdot Y_{t-2} + \beta_3 \cdot PRCP + \beta_4 \cdot TAVG + \beta_4 \cdot Rate + \varepsilon$$

$$\widehat{Y_t} = 0.00802 + 1.04573 \cdot Y_{t-1} - 0.04755 \cdot Y_{t-2} - 0.00023 \cdot PRCP + 0.000069 \cdot TAVG + 0.00017 \cdot Rate$$

Thus, MLR provides more intuitive and interpretable result than classic time series models. However, based on the Figure 10, it is notable that the coefficients for all external predictors are substantially small, and p-value for daily precipitation and daily average temperature are larger than 0.05, indicating that there is no statistically significant relationship between daily precipitation, temperature average and log(price). The Ghana climate dataset doesn’t have a significant impact on this MLR model. This further exposes the limitation of the available dataset used for forecasting, which will be discussed in the limitation part. The MLR model captures almost only rolling past two days data, which leads the forecast of testing dataset not accurately since there is no previous surge.

LSTM Algorithm (Machine Learning Method):

Figure 12: Comparison of LSTM Forecast and Actual Cocoa Prices Over Time



As shown in the forecasting graph, the LSTM method predicts fluctuation before the initial stage of spiking extremely accurately. However, after around 2023-09, LSTM failed to predict the sudden increase and instead presented a slight increase following the decreasing trend. This may be due to some limitations regarding the LSTM method when it deals with datasets and leads to further explanation in the limitation part. The forecasting continues oscillating with wide variance, partially due to LSTM is very sensitive to high frequency, day-to-day noise without strong seasonality and smooth structure. The testing dataset error is presented in table below:

Model	LSTM
RMSE	1729.54
MAE	952.62
MAPE	13.43%

Table 3: Forecasting Accuracy Metrics for the LSTM Model

Compared to other models, LSTM provides an overall more accurate forecast, with at least RMSE 1729.54 due to its ability to process complex lag relationships, its excellent forgetting gate mechanism and ability to handle long sequential data. LSTM model produces finer forecasts each time since it continually learns and adapts through optimization process. However, the LSTM algorithm still provides a relatively unstable prediction like any other model, and it is hard to capture those sharp price changes well. This is mainly due to the huge day-to-day noise of the data and the absence of datasets that can represent such drastic rises. This would be classified in the limitation part.

Conclusions and Limitations

Overall, this report predicts changes in cocoa bean prices by building different models and combining different influencing factors. It is effective to demonstrate the prediction ability and estimable results of the different models in the face of these types of data. Based on the presentation and discussion in the result section, it is realized that ARIMAX (0, 0, 1) is the best forecasting model among all traditional models in the class, which at least captures the increase in the data, despite being affected by the volatility and noise of the high-frequency daily trading market of cocoa beans. So, in combination with the results of the literature review on the prediction of cocoa bean prices before huge price changes, such models may be more appropriate to predict the data with stable economic cycles such as a country's GNP changes and futures with stable price changes.

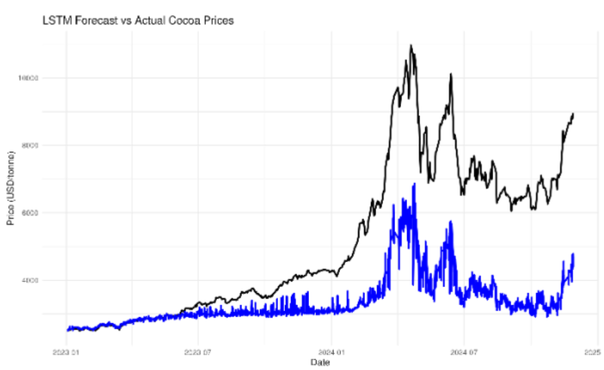
Based on different data, a GARCH model can be combined to further capture volatility, when necessary, but be aware of the model complexity and risk of overfitting as shown in this report. People can realize through this report that such models have the limitation of not being able to predict well on non-linear high-frequency trading markets with severe noise and volatility such as stocks or futures. Meanwhile, it is necessary to find more factors with their datasets that may cause huge data (price) variations and train the models further to address the imprecision.

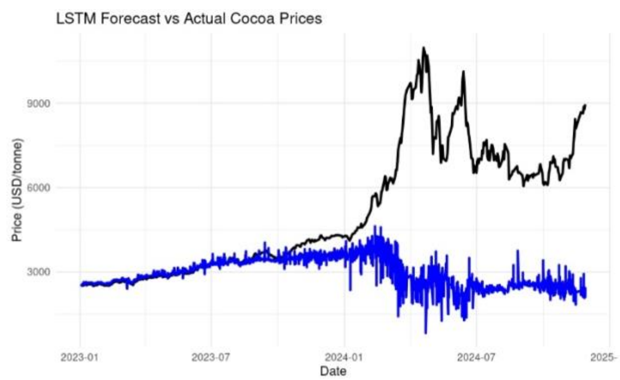
Additionally, the results of the MLR model indicate that linear models are also generally not quite suitable for forecasting a non-linear and strongly volatile market like cocoa beans, although it demonstrates a slight upward trend. For example, the translation of the statistically significant variable USD/GHS exchange rate suggests that log (cocoa-beans price) grows by an average of \$0.00017 for every additional dollar of strengthening in the US dollar when all other variables are held constant. But in fact, the basic economic theorem explains that for cocoa beans that are traded in US dollars, the strengthening of the US dollar causes local Ghanaian farmers to earn more Ghanaian dollars for the same one US dollar, which stimulates them to produce, and further leads to more production and a decrease in the price of cocoa beans. Thus, through this report it can be noted that the traditional linear regression model is more suitable for predicting data which have a strong linear relationship between the dependent and independent variables such as the price of housing in an area related to variables such as its location and the size of the house.

Finally, the Long-Short Term Memory algorithm in machine learning was introduced to utilize its cellular structure of forgetting, input and updating mechanisms for better capturing the nonlinear patterns of cocoa bean price changes and the problem of long-term data dependence. Consistent with Prof. Sari's conclusions in the literature, the LSTM algorithm exhibits constantly fluctuating upward predictions, especially around July 2024, and demonstrates the strength of machine learning models in learning and predicting similar financial data. However, the absence of a dataset that can more perfectly quantify sudden changes in the price of cocoa beans and the training of more similar financial data also becomes a limitation to the capability of the LSTM algorithm to predict price changes as dramatic as those in early 2024. With this report, while noting the effectiveness of the LSTM algorithm, one should not overlook the fact that this algorithm gives a relatively good prediction attributed to not setting a strict learning starting point in the train data (Appendix 2). The changes in the high-frequency prediction results (accompanied by the best model parameters) are still affected by the learning starting point in train data and by the daily noise. Therefore, it is necessary to extend the data type and combine different machine learning algorithms, such as GA-LSTM, to train the model in the formal Python environment even further.

Appendix

1. The following GitHub Link includes all R codes used for this project.
2. The following two plots indicate the different forecasting results with the same code. The major difference is due to different learning starting points of train data, so it's better to train the model again and again for the potential best model shown in the result part.





Reference

1. International Cocoa Organization. (2025, March 17). *Statistics - International Cocoa Organization*. https://www.icco.org/statistics/
2. NCEI.Monitoring.Info@noaa.gov. (n.d.). *Past weather | National Centers for Environmental Information (NCEI)*. https://www.ncei.noaa.gov/access/past-weather/ghana%20climate
3. *USD GHS historical data - Investing.com Canada*. (n.d.). Investing.com Canada. https://ca.investing.com/currencies/usd-ghs-historical-data
4. Kamu, A., Yusoff, R., & Ahmed, A. (2010). (pdf) forecasting cocoa bean prices using univariate time series models. https://www.researchgate.net/publication/285068778_Forecasting_Cocoa_Bean_Prices_Using_Univariate_Time_Series_Models
5. Sukiyono, K., Nabiu, M., Sumantri, B., Novanda, R. R., Arianti, N. N., Sriyoto, Zulkarnain Yuliarso, M., Badrudin, R., Mustopa Romdhon, M., & Mustamam, H. (2018). Selecting an accurate cacao price forecasting model. *Journal of Physics: Conference Series*, *1114*, 012116. https://doi.org/10.1088/1742-6596/1114/1/012116
6. Beg, M. S., Ahmad, S., Jan, K., & Bashir, K. (2017). Status, supply chain and processing of Cocoa - A Review. *Trends in Food Science & Technology*, *66*, 108–116. https://doi.org/10.1016/j.tifs.2017.06.007
7. 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). https://doi.org/10.1016/j.heliyon.2024.e3274
8. Terminal8 GmbH, B. (n.d.). *Cocoa facts and figures*. Zur Startseite. https://www.kakaoplattform.ch/about-cocoa/cocoa-facts-and-figures#:~:text=Most%20cocoa%20is%20produced%20in%20West%20Africa&text=C%C3%B4te%20d%27Ivoire%20and%20Ghana%20are%20by%20far%20the%20two,is%20the%20largest%20producer%20country.

9. Sari, M., Duran, S., Kutlu, H., Guloglu, B., & Atik, Z. (2024). Various optimized machine learning techniques to predict agricultural commodity prices. *Neural Computing and Applications*, *36*(19), 11439–11459. https://doi.org/10.1007/s00521-024-09679-x