Modelling Cocoa Futures Prices Using ARIMA-GARCH, VARX, and ETS Methods

Unurjargal Batsuuri, Max Godfrey, Isaac Hua, Oscar (Shen) Zhao

1 Introduction

Cocoa, derived from the seeds of the cacao tree (*Theobroma cacao*), is a key agricultural commodity. Cocoa beans are the primary ingredient in chocolate production; they are also used in food and cosmetic products.

The price of cocoa is subject to fluctuations due to factors such as weather, political instability, and changes in demand. In the last two years, prices have quintupled due to reductions in cocoa yields partly caused by climate change-induced disruptions [1]. Accurately modelling cocoa's future price allows stakeholders such as producers, traders, and investors to manage and hedge against its price volatility. It also allows stakeholders to understand the market value of cocoa which is crucial for informing their decisions about production, sourcing, and investment.

This report seeks to accurately forecast monthly average cocoa futures prices up to six months in the future. Our aim is to understand how exogenous macroeconomic indicators and related commodity prices influence cocoa futures prices. Our predictors of interest are the USD/GBP and USD/GHS exchange rates, Consumer Price Indices (CPIs) from the US, EU, and Ghana, and the prices of oil, sugar, and coffee commodities. We present econometric models using Vector Autoregression with Exogenous Variables (VARX), Exponential Smoothing (ETS), and Autoregressive Integrated Moving Average (ARIMA) processes with Generalized Autoregressive Conditional Heteroskedasticity (GARCH) techniques applied to their residuals. Our analysis uses the R programming language [15].

Challenges include the complexity of external influences on cocoa prices, the need for rigorous statistical methods to account for noise and uncertainty in volatile price data, and the potential for sudden market shifts, notably the quadrupling of cocoa prices since 2023 [6].

2 Literature Review

Bilgin and Ellwanger (2017) present a dynamic factor model for commodity prices [2]. Their model imposes a block structure on factors influencing prices, notably identifying that global economic activity influences the supply and therefore the price of commodities. They also assert that every commodity has idiosyncratic factors, such as the prices of related commodities, that impact its price by affecting demand. This suggests that we can expect to see relationships between the futures price of cocoa, our macroeconomic predictors and our commodities prices of interest.

Furthermore, Shahzad et al. (2021) used the Granger Causality Test to establish the existence of time-varying causality between energy, agriculture, and precious metal commodities prices [18]. When fitting our VARX model, we follow a similar methodology, using the Granger Causality Test to determine which of our commodity price predictors significantly affect the price of cocoa futures.

Kumar et al. (2022) fitted ARIMA and VARX models to predict the response in prices of raw dry and wet cocoa beans in response to changes in cocoa futures prices. They used the Augmented Dickey-Fuller (ADF) test and correlograms to infer the stationarity of the time series which they were studying and on all of the transformations they applied to them. Box-Jenkins was used to fit their ARIMA models. They found significant ARIMA(1,1,0) and VARX(1) models for the price of dry cocoa beans, and a significant

ARIMA(1,1,2) model for the price of wet cocoa beans [13]. We build upon their methodology by using a larger set of predictors in our VARX model and by using Hyndman-Khandakar to select ARIMA models, allowing us to test a greater set of parameters. In order to achieve similar levels of robustness in model selection, we used their technique of ADF and correlograms to verify the stationarity of our transformed time series. Furthermore, we compute Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test statistics to test that our time series exhibit stationarity around deterministic trends.

The Hyndman-Khandakar algorithm [5] is an ARIMA-fitting technique which is used in commodities pricing financial and macroeconomic instruments. Scarffe (2022) uses the technique to model log-differenced Canadian export prices [17]. Furthermore, Drachal (2021) used the technique to forecast energy commodities prices, the veracity of which was measured using Mean Average Error (MAE) and Root Mean Squared Error (RMSE) [3]. We use the same metrics to determine the accuracy of our models' predictions.

As suggested by Jammeh et al. (2023), a parsimonious exponential smoothing model could be more interpretable, less prone to overfitting and more considerate of the aforementioned structural breaks [12]. Based on the seasonal nature of cocoa harvests occurring in seasons [1] and the original interpolated cocoa price plots, we expect the cocoa price to exhibit a trend and seasonality. Since cocoa price data reacts quickly to new information which can be observed from the cocoa price plot from 2023 onwards, Error, Trend, Seasonality (ETS) State-Space models are appropriate.

3 Data Description and Cleaning

3.1 ICCO Daily Cocoa Price Data

For each trading day, the ICCO Daily Cocoa Price data in United States Dollars (USD) [6] comprise observations of the mean price of cocoa futures in the next three active trading months. Prices are taken from the ICE Futures Europe (London) and ICE Futures US (New York) exchanges at the time of London close. Prices from ICE Futures Europe are converted to USD per tonne using the current six-month forward rate at the time of London close. Observations are provided from 1994-10-03 to 2025-02-27.

The data contained three weekend observations, all of which fell between June and July of 1995, a notable inconsistency with the remaining data. For the sake of consistency, the dataset was analysed from 1996-01-01 onwards. The removal of the first 4.1% of the data was justifiable, since the goal was to forecast the data's recent structure. The data also contained four duplicated dates which were removed; in the case that the duplicated dates contained different observations, the price that was most similar to neighbouring days observations was retained.

The data also missed observations on some trading days. For consistency across years, data was backfilled on all weekdays using linear interpolation, since trading days vary each year. This meant that a year of observations contained 261 points on average. The cleaned time series is shown in Figure 1 with the 119 interpolated points of 10649 observations indicated. The monthly mean of all cleaned observation months was computed. A plot of this series is shown in Figure 2.

The cleaned monthly mean time series is heteroskedastic. In order to make the time series stationary, first-order differencing was performed, yielding the time series plot, Autocorrelation Function (ACF), and Partial Autocorrelation Function (PACF) shown in Figure 3. Given a 0.05 significance level, the differenced time series has ADF test statistic -5.897 ($p < 0.01$) meaning that we reject the ADF null hypothesis of non-stationary, and KPSS test statistic 0.4987 ($p = 0.0643 > 0.05$) meaning that we fail to reject the KPSS null hypothesis of stationarity. These results, combined with the decay of the time series' ACF and PACF suggest that the first-differenced monthly price data is stationary.

Figure 1: Cleaned Daily Cocoa Prices (US$/Tonne) 1996/01/01 - 2025/02/27

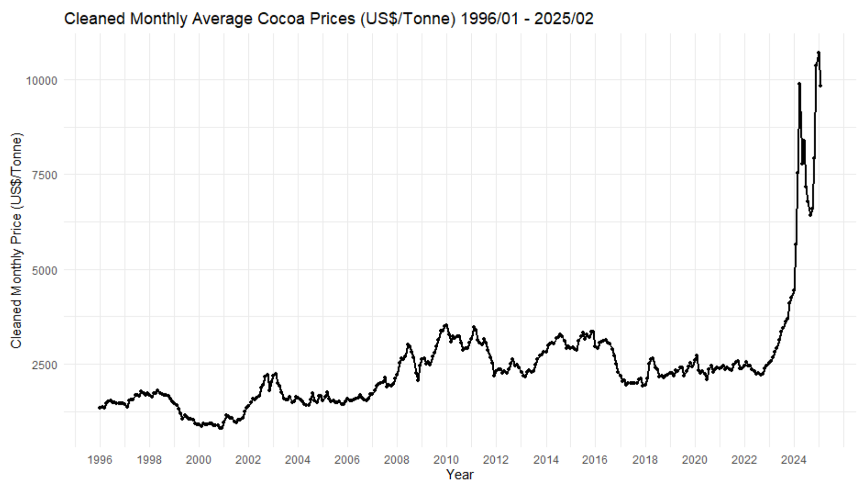

Figure 2: Monthly Interpolated Cleaned Cocoa Prices (Average US$/Tonne) 1996/01 - 2025/02

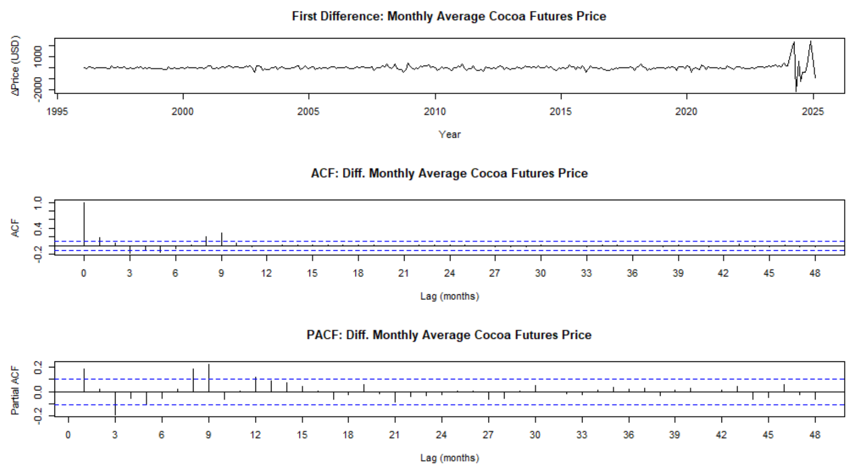

Figure 3: Plot, ACF, and PACF of first-differenced monthly average cocoa futures price data

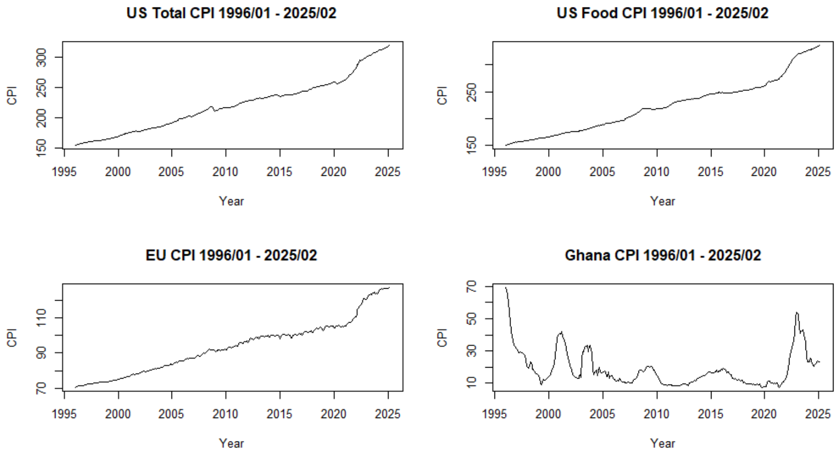

Figure 4: Plots of Monthly Average CPI data

3.2 CPI Data

CPI Data is provided on a monthly basis. Of interest were the US Urban City Average Food CPI [19], the US Urban City Average CPI [20], the Euro Area Harmonized CPI [4]. These indices are provided by the US Federal Reserve Economic Data (FRED) database. Ghanaian CPI [16] provided by the Bank of Ghana was also analysed.

Plots of the four datasets are shown in Figure 4. They are each visibly nonstationary. The Ghana CPI and EU EPI data were differenced, and a Box-Cox Transformation was applied to both the US Total CPI data ($\lambda = -0.8926298$) and US Food CPI data ($\lambda = -0.9999242$). An analysis of these transformed time series are shown in Figure 4. The ADF and KPSS test statistics for these transformed time series were all significant at the 0.05 significance level. Combined with the fact that their ACFs and PACFs tail off, this suggests that the transformed time series are stationary.

Time Series	**ADF Test Statistic (p-value)**	**KPSS Test Statistic (p-value)**
Transformed Ghana CPI	-4.950839 (< 0.01)	0.1541853 (> 0.1)
Transformed US CPI Total	-5.876088 (< 0.01)	0.2540996 (> 0.1)
Transformed US CPI Food	-4.591171 (< 0.01)	0.3439444 (> 0.1)
Transformed EU CPI	-3.606117 (0.0327)	0.2442219 (> 0.1)

Table 1: CPI Data Stationarity Test Results

3.3 Commodities Data

Commodities data is provided on a daily basis. The commodities of interest are US Coffee C Futures [9], London Sugar Futures [8], and Brent Crude Oil Futures [7]. Monthly averages of these data were computed so that they could be used in conjunction with CPI data. This is shown in Figure 6. An analysis of the first difference of these time series is shown in Figure 7; their ADF and KPSS test statistics are all significant at the 0.05 significance level. This suggested that the differenced time series were stationary. This is corroborated by each ACF and PACF cutting off after lag 2.

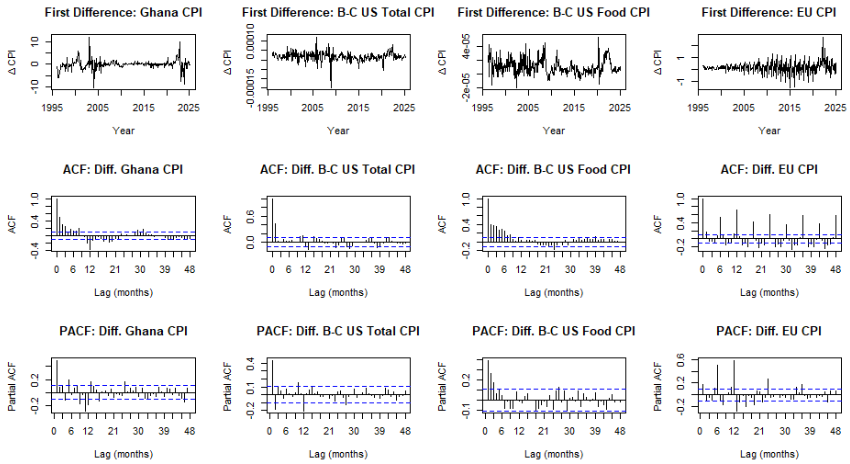

Figure 5: Plots, ACFs, and PACFs of transformed CPI data

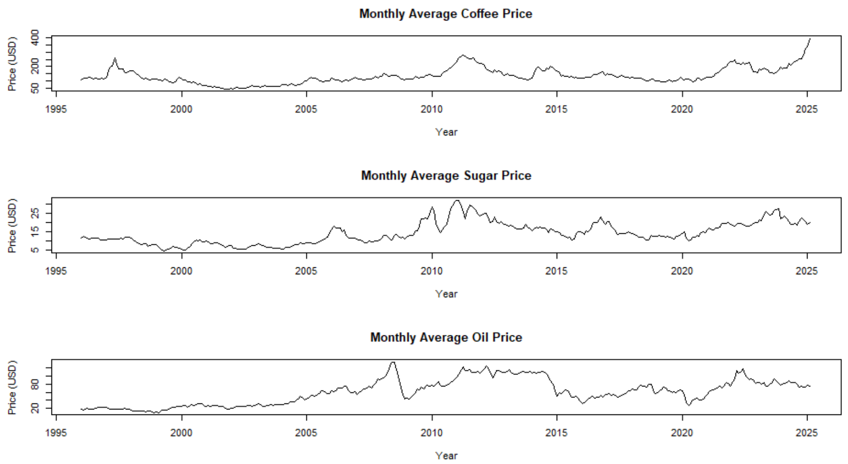

Figure 6: Plots of monthly average commodities price data (USD)

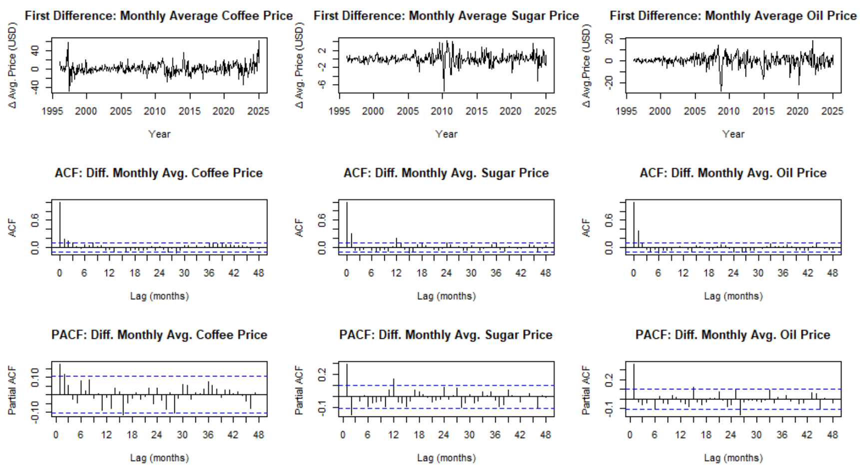

Figure 7: Plots, ACFs, and PACFs of transformed commodities data

Stationarity Test	ADF Test Statistic (p-value)	KPSS Test Statistic (p-value)
Diff. Monthly Avg. Coffee Price	-4.137491 (< 0.01)	0.255963 (> 0.1)
Diff. Monthly Avg. Sugar Price	-7.706418 (< 0.01)	0.03714485 (> 0.1)
Diff. Monthly Avg. Oil Price	-7.1064 (< 0.01)	0.05247252 (> 0.1)

Table 2: Transformed Commodity Data Stationarity Test Results

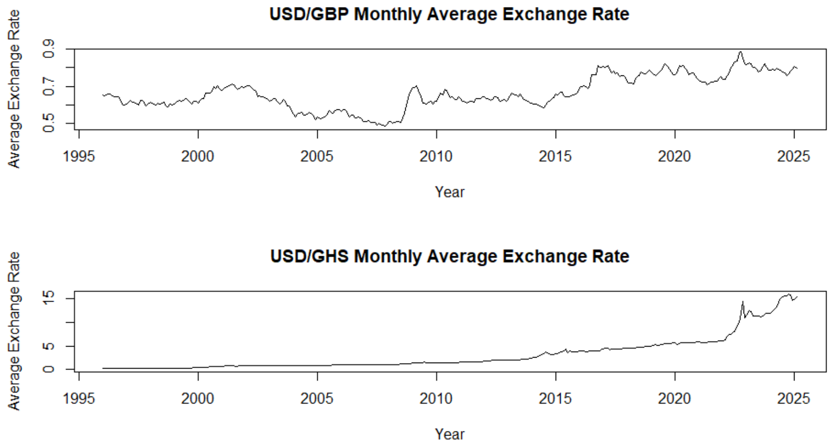

Figure 8: Plots of monthly average exchange rate data

3.4 Exchange Rate Data

The currencies of interest are the United States Dollar (USD), Pound Sterling (GBP), and Ghanian Cedi (GHS). The USD/GBP and USD/GHS exchange rate data were observed on a daily basis [10] [11]. The plots of the monthly averages of the data are shown in Figure 8. An analysis of the first difference of these time series is shown in Figure 9; The ADF and KPSS test statistics are all significant at the 0.95 significance level. This suggests that the differenced time series are both stationary, which is corroborated by both ACFs cutting off and the PACFs quickly tailing off.

Stationarity Test	ADF Test Statistic (p-value)	KPSS Test Statistic (p-value)
Transformed USD/GHS Rate	-6.4316 (< 0.01)	0.0829687 (> 0.1)
Transformed USD/GBP Rate	-7.028775 (< 0.01)	0.08990545 (> 0.1)

Table 3: Stationarity Test Results for Monthly Exchange Rates

4 Forecasting and Results

4.1 Forecasting Procedure

With the intention of forecasting accurately, each fitted model needed to be trained on some of the high-volatility data post COVID-19 pandemic. The conventional 80:20 train-to-test data ratio could not be used since the most recent fifth of the price data wholly included this high-volatility period. The testing data comprised the six transformed monthly average price observations from 2024-09 to 2025-02. Two training sets were used, the first containing all stationary monthly data from 1996-01 to 2024-08, and the second containing all stationary monthly data from 2014-09 to 2024-08. They are henceforth referred

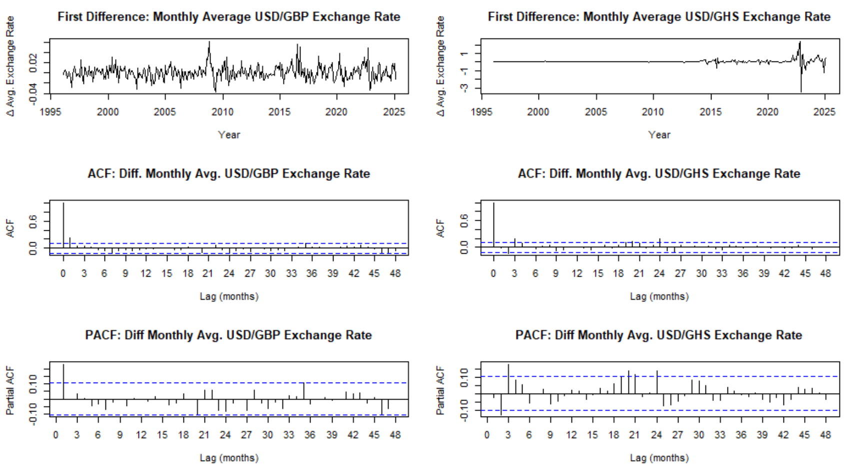

Figure 9: Plots, ACFs, and PACFs of transformed exchange rate data

to respectively as the 'full' and 'truncated' training data. This was done to investigate whether models trained on the full training set were able to generalise their predictions and forecast the more recent high-volatility portion of the price data well.

The quality of a model's forecasts were evaluated by the Mean Average Error (MAE) and Root Mean Squared Error (RMSE) of their predictions on the testing data. Focusing on how accurately each model predicted future data, ensured that the final specification not only fit the historical patterns but also generalized well in real-world forecasting scenarios.

4.2 ARIMA-GARCH Model

4.2.1 Model Equation

An ARIMA(p, d, q)-GARCH(r, s) model is specified by

$$\nabla^d CP_t = \mu + \sum_{i=1}^{p} \phi_i \nabla^d CP_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t,$$

where:

- $\nabla^d CP_t = (1 - B)^d CP_t$ denotes the d-th differenced cocoa price, with B as the backshift operator,
- μ is the intercept (mean) in the differenced space,
- ϕ_i (for $i = 1, \ldots, p$) are the AR coefficients,
- θ_j (for $j = 1, \ldots, q$) are the MA coefficients,
- ε_t is a white-noise error term with mean zero and conditional variance σ_t^2 modelled by GARCH(r, s):

$$\sigma_t^2 = \omega + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2,$$

where:

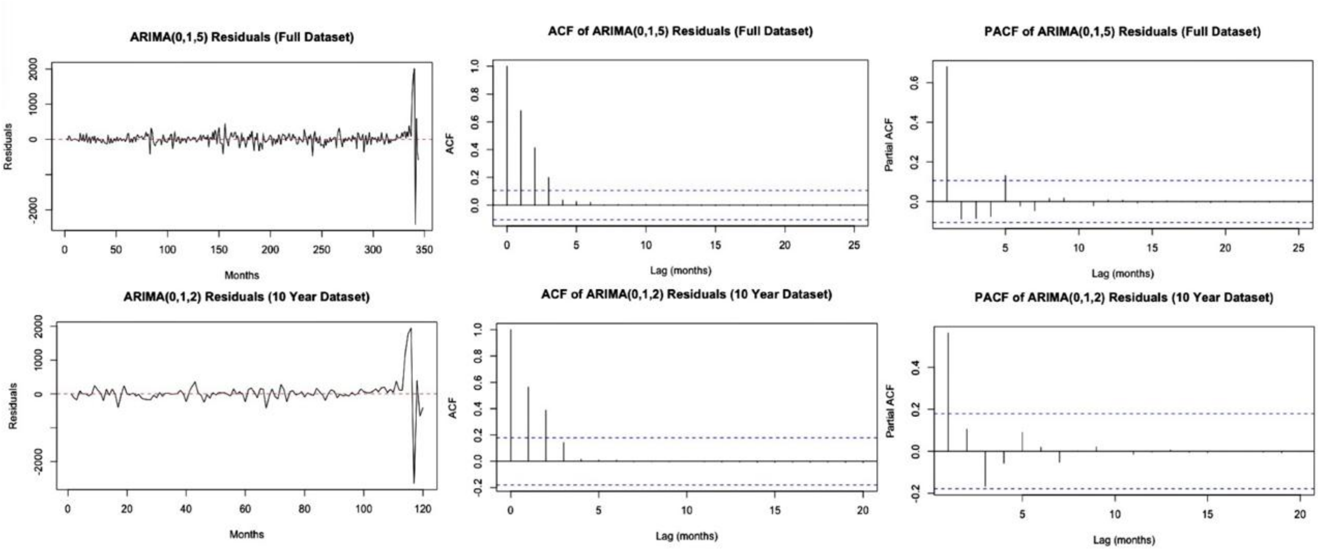

Figure 10: ARIMA-GARCH Residuals and their ACFs and PACFs

- σ_t^2 is the conditional variance of the innovation ε_t,
- ω is a constant term,
- α_i (for $i = 1, \ldots, r$) quantify the impact of past squared shocks, ε_{t-i}^2,
- β_j (for $j = 1, \ldots, s$) capture the persistence in past conditional variance, σ_{t-j}^2.

4.2.2 Model Fitting and Validation

ARIMA-GARCH models were trained solely on the stationary monthly average price data from both training sets. The full training set had an ACF which decayed after about five lags; its PACF cut off quickly. This suggested that the full training set admitted an MA(5) structure (i.e. an ARIMA(0,1,5) applied to the original time series). The truncated data set had an ACF which indicated prominent correlations up to lag two and a PACF which cut off quickly. This suggested that the truncated training set admitted an MA(2) structure (i.e. an ARIMA(0,1,2) applied to the original time series). Hyndman-Khandakar's algorithm confirmed these observations.

On the full training data, the chosen ARIMA(0,1,5) model was coupled with a GARCH(1,1) volatility process, while the truncated set gave rise to an ARIMA(0,1,2) model, again paired with GARCH(1,1). In both cases, residual checks were carried out using the adjusted Pearson Goodness-of-Fit test. High p-values in these residual diagnostics indicated that neither model exhibited significant departures from its assumed error distribution, suggesting that both were well specified and thus supporting the reliability of the ARIMA-GARCH model.

Model	Training Set	MAE	RMSE	AIC	BIC
ARIMA(0,1,5)-GARCH(1,1)	Full	2192.276	2756.551	12.48579	12.59768
ARIMA(0,1,2)-GARCH(1,1)	Truncated	2088.649	2661.740	13.06735	13.23082

Table 4: Performance metrics for ARIMA-GARCH models on both the full dataset (1996–Sep 2024) and the truncated 10-year subset (Sep 2014–Sep 2024).

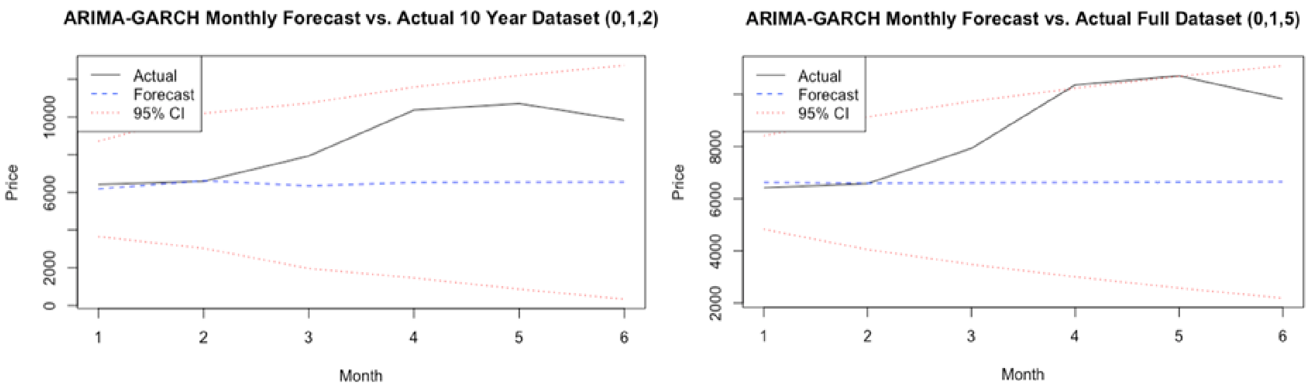

Figure 11: ARIMA-GARCH Forecasts on Testing Data

4.2.3 Results

The ARIMA(0,1,5) model trained on the full training data yielded MAE = 2192.276 and RMSE = 2756.551. Its forecasts are shown in Figure 11. The parameter estimates for the MA(5) lags were statistically significant in capturing autocorrelations in the differenced cocoa price, and the GARCH(1,1) parameters confirmed that volatility clusters persisted throughout the broader historical range. Moreover, high p-values from the Adjusted Pearson Goodness-of-Fit test of 0.8398 for the full dataset and 0.9025 for the truncated 10 year reaffirms that the residuals did not display systematic patterns left unexplained by the model.

The ARIMA(0,1,2) model trained on the truncated training data, yielded MAE = 2088.649 and RMSE = 2661.74 - marginally lower than those of the ARIMA(0,1,5) trained on the full data. The high p-values in diagnostic tests indicated a good fit, suggesting that the observed residual structure aligns with the assumptions of the ARIMA-GARCH framework.

In summary, both the ARIMA(0,1,5) and ARIMA(0,1,2) models performed well, with high p-values confirming no significant misspecification and reduced RMSE/MAE. Despite the ARIMA(0,1,5) model's ability to generalise, the ARIMA(0,1,2) specification trained on the truncated data obtained the best overall results, making it the better choice for short to medium term cocoa price forecasting.

4.3 VARX Model

4.3.1 Model Equation

A VARX(p) model is specified as

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + B X_t + u_t$$

Where:

- Y_t is a $k \times 1$ vector of endogenous variables at time t which includes the differenced monthly average cocoa price.
- A_i are $k \times k$ coefficient matrices for the i-th lag of the endogenous variables, for $i = 1, \ldots, p$.
- p is the number of lags included in the model.

- X_t is a $m \times 1$ vector of exogenous variables at time t.
- B is a $k \times m$ matrix of coefficients for the exogenous variables.
- u_t is a $k \times 1$ vector of error terms (white noise), assumed to have zero mean and constant covariance matrix Σ_u.

4.3.2 Model Fitting

In order to satisfy VARX modelling assumptions, these models were only trained with stationary time series. For each training set, a VARX model was fitted with all endogenous variables (all transformed commodities price time series and the transformed Ghanaian CPI) and exogenous variables (transformed US Total and Food CPIs, transformed EU CPI, and transformed exchange rates). To ensure that the set of endogenous variables was causal, the Granger Causality Test was applied to test the effect of each training set's full VARX model's endogenous variables' causality with the stationary response. The results are shown in Table 5 and Table 6.

Transformed	AIC		HQ		SC		FPE	
Predictor	Lag (n)	P-value	Lag (n)	P-value	Lag (n)	P-value	Lag (n)	P-value
Oil	1	0.8620	1	0.8620	1	0.8620	1	0.8620
Sugar	9	< 0.01	2	0.8907	1	0.7274	9	1.6487e-5
Coffee	2	0.1227	1	0.0782	1	0.0782	2	0.1227
Ghana CPI	17	< 0.01	12	0.0023	1	0.9558	17	4.5939e-7

Table 5: Granger Causality Summary on Full Training Data

Transformed	AIC		HQ		SC		FPE	
Predictor	Lag (n)	P-value	Lag (n)	P-value	Lag (n)	P-value	Lag (n)	P-value
Oil	1	0.9955	1	0.9955	1	0.9955	1	0.9955
Sugar	5	0.0027	1	0.7452	1	0.7452	5	0.0027
Coffee	1	0.0164	1	0.0164	1	0.0164	1	0.0164
Ghana CPI	17	< 0.01	17	< 0.01	1	0.7629	17	< 0.01

Table 6: Granger Causality Summary on Truncated Training Data

For the truncated year training set, the transformed Ghanaian CPI, and transformed coffee, and transformed sugar commodities prices were chosen as endogenous variables (in addition to the transformed cocoa price) as they exhibited consistently statistically significant casualties with cocoa prices. For the full training set, by the same logic, the transformed Ghanaian CPI and the transformed sugar commodities price were selected as endogenous variables.

For each training set, the chosen set of endogenous data was used with R's varselect utility [14] at a range of maximum lags in order to determine potential AIC and FPE minimising lags with which a VARX model could be fit. Using maximum lags of 10, 12, 14, and 16, the optimal lags on the truncated training set were determined to be 9, 12, 14, and 16. Those of the full training set were determined to be 5, 12, and 13. Residual plots of these models are shown in Figure 12 and Figure 13. All residuals are mean-zero, though only the VARX(14) and VARX(16) models trained on the truncated training set can be argued to have homoskedastic residuals.

The best model by far was the VARX(14) trained on the truncated data. The model's forecasts demonstrated that it was able to capture the spike in average monthly prices observed in late 2024. This is

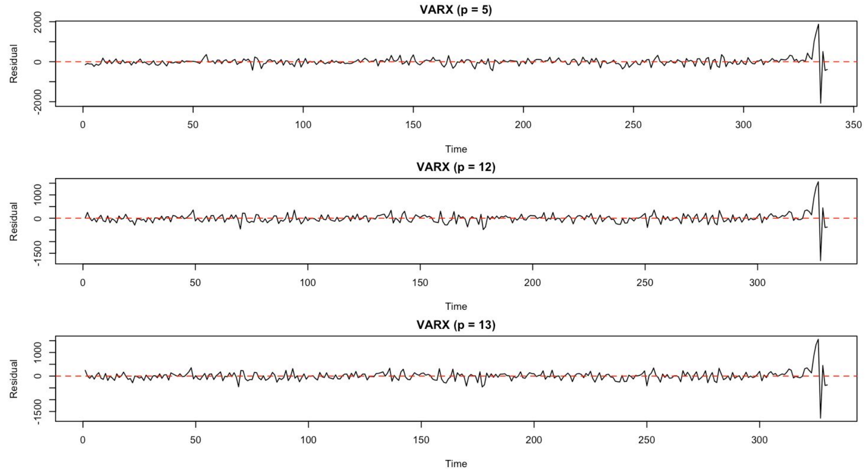

Figure 12: Residuals from VARX models at varying lags fitted to the full training set

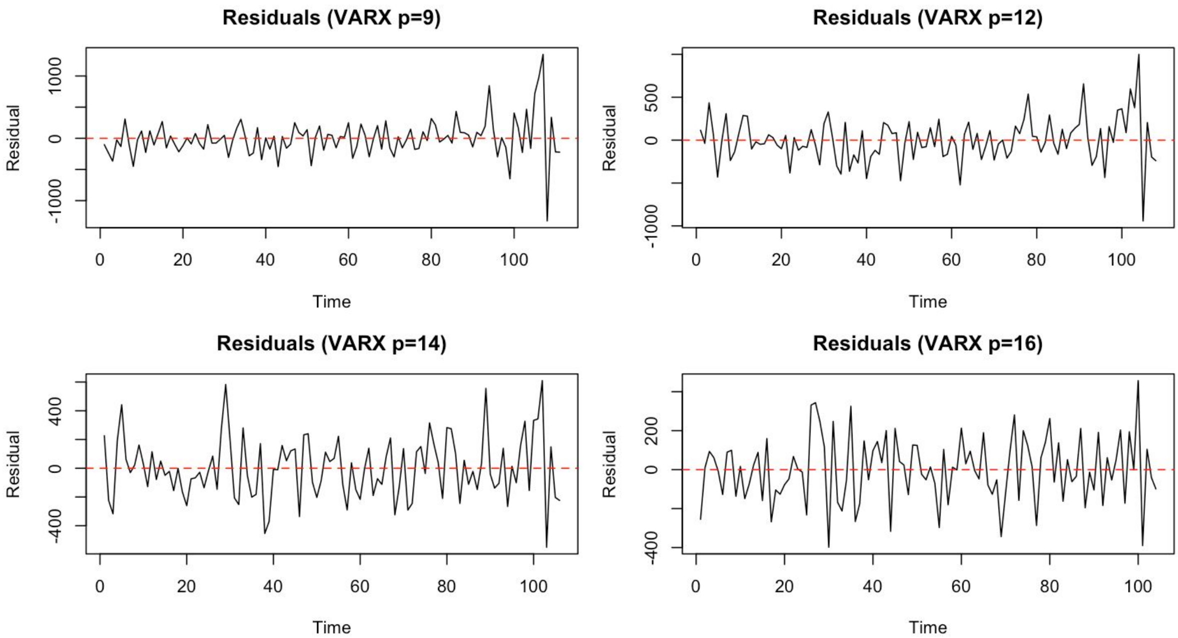

Figure 13: Residuals from VARX models at varying lags fitted to the truncated training set

Training Set	Truncated				Full		
Model	VARX(9)	VARX(12)	VARX(14)	VARX(16)	VARX(5)	VARX(12)	VARX(13)
MAE	1088.324	1694.380	569.849	2094.584	2672.741	1510.832	1574.340
RMSE	829.979	1236.952	543.601	1626.495	2223.015	1131.542	1177.869

Table 7: Forecast performance of different VARX model specifications (with Truncated and Full labels)

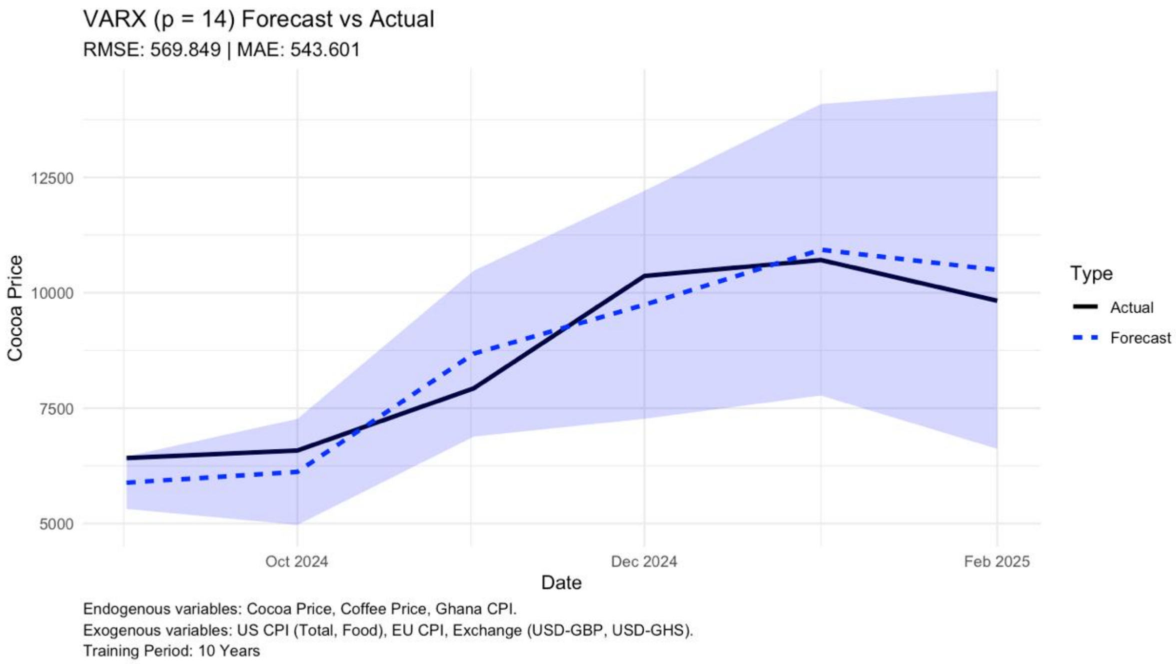

Figure 14: VARX(14) forecasts on the testing data

shown in Figure 14, which was obtained by undifferencing the model's predictions on the testing set.

4.4 Exponential Smoothing

Exponential smoothing methods produce forecasts by averaging past observations, assigning weights that decay exponentially as time between observations increases. Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. This means that more recent observations are assigned larger weights, which has a bigger impact on our forecasts. Equations of the pertinent types of ETS models are provided in Appendix 2.

4.4.1 Model Fitting

ETS models were trained solely on the stationary monthly average price data from both training sets. Figure 15 shows the trend, seasonal, and error components of these data. The identified seasonal component was negligible compared to the observed price time series. This suggested that there was no perceptible seasonal component present in the average monthly price of cocoa, and therefore that no ETS model fitted to the data should have a seasonal component. The trend component accounted for a substantial proportion of the observed variation, so both a multiplicative and additive trend specification were deemed appropriate models. Additionally, the preliminary decomposition revealed a heteroskedastic residual component. This observation supported the use of a multiplicative or additive error term within

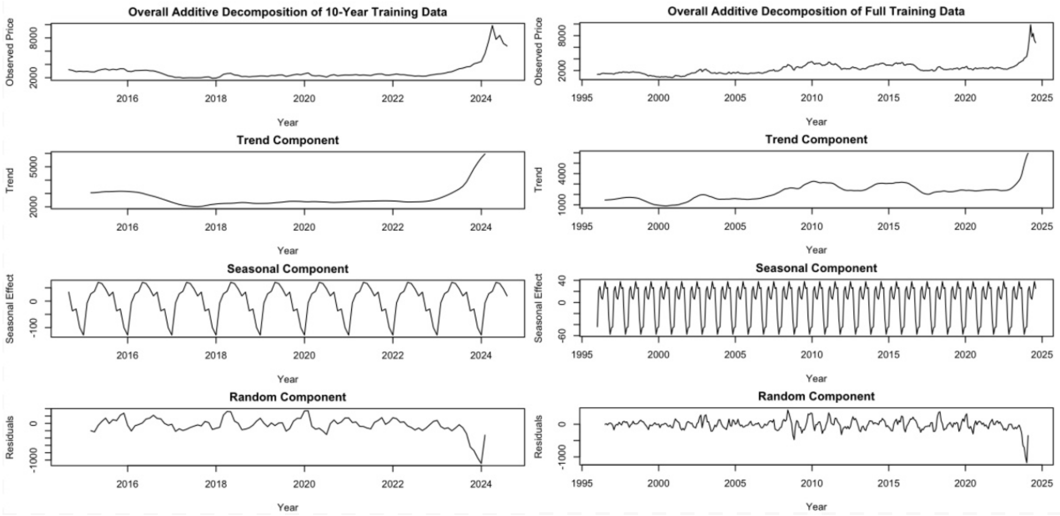

Figure 15: ETS decomposition of the truncated training data (left) and full training data (right)

our ETS framework.

In an attempt to capture these observed preliminary dynamics, an ETS(M,A,N) model, an ETS(A,A,N) model, and an ETS(M,M,N) model were fitted to the full monthly price dataset. Residual diagnostic plots shown in Figure 16 for the three models indicated that, aside from the period around the beginning of 2024, the residuals were centered around zero with well-behaved volatility resembling white noise. Additionally, the corresponding ACF and PACF plots confirmed the absence of significant autocorrelation among the residuals.

Our empirical findings and state-space modeling considerations suggest adopting an ETS specification, either ETS(M,A,N), ETS(M,M,N) or ETS(A,A,N). The validation period, consisting of the last six months of our data, was used to assess out-of-sample forecasting performance. The metrics in Table 8 summarise the performance of each ETS model specification within the testing dataset.

Model	Training Data	MAE	RMSE	AIC	BIC
ETS(M,A,N)	Truncated	2608.998	3280.061	1836.559	1850.496
ETS(M,M,N)	Truncated	2511.677	3147.560	1835.796	1852.521
ETS(A,A,N)	Truncated	1967.544	2438.239	2019.562	2033.499
ETS(M,A,N)	Full	2246.010	2848.863	5379.645	5402.689
ETS(M,M,N)	Full	2187.729	2762.945	5375.900	5398.944
ETS(A,A,N)	Full	1994.181	2476.721	5822.549	5841.752

Table 8: Model Performance Metrics

Across both the truncated and full training datasets, the ETS(A,A,N) trained model yielded the lowest MAE and RMSE values, suggesting superior forecast accuracy on the testing set relative to the other ETS specifications. However, its AIC and BIC were higher compared to the multiplicative error counterpart models, indicating a trade-off between forecast accuracy and model complexity.

The forecasted price levels derived from each ETS model exhibit distinct behaviors; the ETS(M,A,N) and ETS(M,M,N) models, with their multiplicative error specifications, capture the increasing volatility as prices rise, yet they tend to produce more conservative level forecasts. In contrast, the ETS(A,A,N) model, which uses an additive error structure, better tracks the observed price movements in the test period, as reflected by its lower MAE and RMSE.

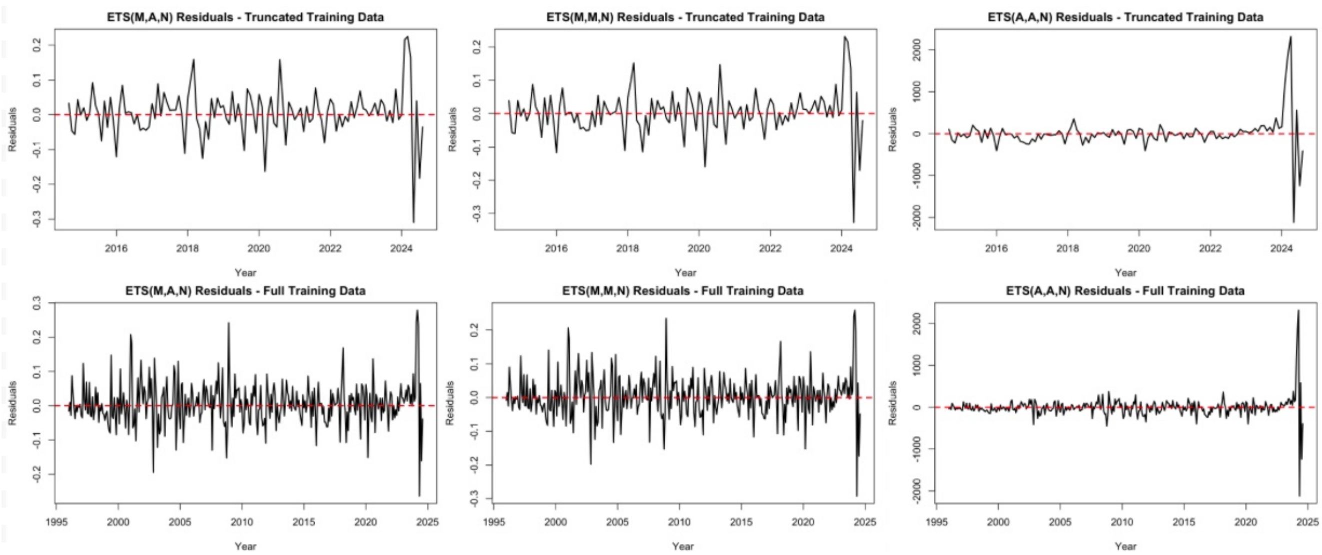

Figure 16: ETS residual plots from (left to right) ETS(M,A,N), ETS(M,M,N), and ETS(A,A,N)

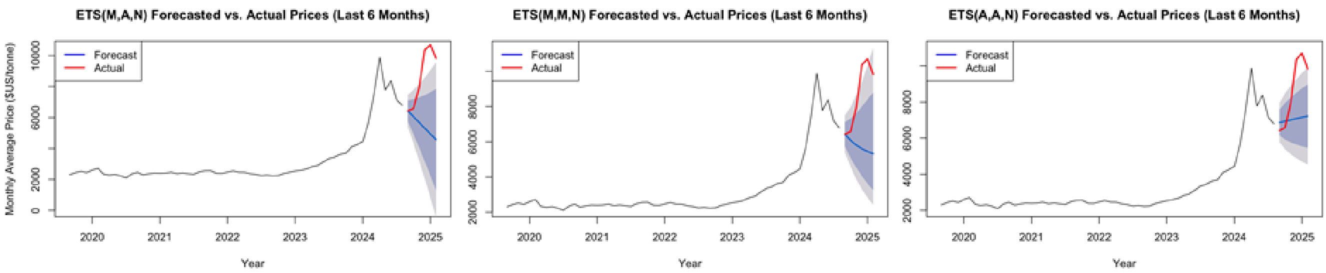

Figure 17: Forecasted ETS prices on testing data

Figure 17 provides visual comparisons of the 6 month forecasted versus actual prices, for the ETS models trained from the truncated training set. The forecast plot overlays the predictions on the actual monthly average cocoa price series. For the ETS(A,A,N) model, the forecast more closely follows the actual price path/movement.

Overall, while the ETS(M,A,N) and ETS(M,M,N) models provide parsimonious representations as indicated by lower AIC and BIC values, the ETS(A,A,N) model yields the most accurate forecasts in terms of MAE and RMSE, which can be observed in the plots. All models show that the actual monthly average cocoa prices still mostly lie within the 80% prediction interval, except for the ETS(M,A,N) specification. In summary, the ETS(A,A,N) specification fits the most assumptions while performing the best.

After fitting the model across the entire dataset, the following parameters were obtained, giving the model specification:

$$\begin{aligned}
\text{Observation:} \quad & \hat{y}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1} + \varepsilon_t, \\
\text{Level Equation:} \quad & \hat{\ell}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1} + 0.9999\,\varepsilon_t, \\
\text{Trend Equation:} \quad & \hat{b}_t = \hat{b}_{t-1} + 0.0311\,\varepsilon_t,
\end{aligned}$$

with $\sigma = 473.046$, and initial states $\hat{\ell}_0 = 3112.8648$ and $\hat{b}_0 = -2.021$.

5 Discussion and Conclusion

Model	Training Set	MAE	RMSE
ARIMA(0,1,2)-GARCH(1,1)	Truncated	2088.649	2661.740
VARX(14)	Truncated	569.849	543.601
ETS(A,N,N)	Truncated	1967.544	2438.239

Table 9: Comparison of best fitted models of each type

Table 9 provides a summary of the predictive power of the best fitted models of each investigated type. The estimated parameters of these models can be found in Appendix 3 and Appendix 4

The ARIMA(0,1,2)-GARCH(1,1) model demonstrated distinct trade-offs between capturing recent market dynamics and leveraging long-term historical information. While historical data is modelled appropriately, showing well-behaving residuals especially after implementing GARCH(1,1), the model does not accurately forecast the most recent innovations in cocoa prices, despite statistically significant coefficients. This suggests that there has been a recent structural change in cocoa pricing dynamics which cannot be modelled through analysing the cocoa price time series directly.

The ETS approach can quickly adapt to new innovations, resulting in our estimated ETS(A,A,N) model. Inspection of the model's coefficients reveals that the smoothing parameter α is approximately 1; this means that the ETS model's level forecast puts almost all the weight on the preceding observation, making it highly responsive to noisy data. The trend smoothing parameter β is only 0.0311, suggesting that the long-term trend component is only slightly influenced by the forecast error, leading to a very gradual change in the trend estimate. Both findings suggest that our best ETS model is designed to quickly adapt its level components to new data while keeping the long-term trend adjustments less sensitive.

These models suggest that there are underlying latent variables that cannot be captured accurately through price dynamics alone. Thus, our best performing model, the VARX(14) model, improves on these approaches by considering multiple lags of endogenous variables. The model captures how past values of cocoa prices and related factors such as commodity pricing and the CPI of cocoa-producing countries affect current cocoa prices. The inclusion of exogenous variables allows the model to account for external economic indicators such as exchange rates and the CPIs of cocoa-consuming countries that have a direct impact on cocoa pricing. This additional information improves the model's predictive accuracy, especially during periods of market volatility, which could not be captured with ETS and ARIMA-GARCH process.

In fitting the VARX(14) model, it was revealed that of the hypothesised endogenous predictors of price, those which exhibited statistically significant causality were the price of coffee and Ghana's CPI. On the other hand, over the past ten years, the prices of oil and sugar have not exhibited Granger-causal relationships with the price of cocoa.

The performance of the VARX(14) model comes with a trade-off of greater model complexity, since the longer lag structure impacts the endogenous variables, potentially leading to overfitting. However, due to the complex dynamics of cocoa pricing, the VARX(14) approach benefits from incorporating interdependencies with exogenous variables and therefore has an edge over the ETS and ARIMA-GARCH modelling approaches.

References

[1] Bellini Motovska, N., Gonzalez Gaviola, A., Koster, T., & Waarts, Y. (2024). *High Level analysis of recent and current yield in the West African cocoa sector: Trends, possible causes and recommendations for interventions.* (Report / Wageningen Economic Research; No. 2024-135). Wageningen Economic Research. https://doi.org/10.18174/674598.

[2] Bilgin, D., Ellwanger, R., & Bank of Canada. (2017). *A dynamic factor model for commodity prices.* In Bank of Canada Staff Analytical Note [Report]. Bank of Canada. https://www.bankofcanada.ca/wp-content/uploads/2017/09/san2017-12.pdf.

[3] Drachal, K. (2021). Forecasting selected energy commodities prices with Bayesian dynamic finite mixtures. *Energy Economics, 99*, 105283. https://doi.org/10.1016/j.eneco.2021.105283.

[4] Eurostat, Harmonized Index of Consumer Prices: All-Items HICP for Euro Area (19 Countries) [CP0000EZ19M086NEST], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CP0000EZ19M086NEST, March 21, 2025.

[5] Hyndman, R. J., & Khandakar, Y. (2008). *Automatic Time Series Forecasting: The forecast Package for R.* Journal of Statistical Software, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03.

[6] International Cocoa Organization. (2025, April 04). *Statistics - International Cocoa Organization.* https://www.icco.org/statistics/.

[7] *Investing.com - Brent Crude Oil Futures Price.* (n.d.). Investing.com. Retrieved March 21, 2025, from https://www.investing.com/commodities/brent-oil.

[8] *Investing.com - London Sugar Futures Price.* (n.d.). Investing.com. Retrieved March 21, 2025, from https://www.investing.com/commodities/london-sugar.

[9] *Investing.com - US Coffee C Futures Chart.* (n.d.). Investing.com. Retrieved March 21, 2025, from https://www.investing.com/commodities/us-coffee-c-streaming-chart.

[10] *Investing.com Canada - USD GHS historical data.* (n.d.). Investing.com Canada. https://ca.investing.com/currencies/usd-ghs-historical-data

[11] *Investing.com Canada - USD GBP historical data.* (n.d.). Investing.com Canada. https://ca.investing.com/currencies/usd-gbp-historical-data

[12] Jammeh, L. B., Guobadia, E. K., & Ugoh, C. (2023). Exponential Smoothing State Space Innovation Model for Forecasting export commodity Price Index in Nigeria. *African Journal of Economics and Sustainable Development, 6*(4), 74–84. https://doi.org/10.52589/ajesd-xqfamzmy.

[13] Kumar, K. A., Spulbar, C., Pinto, P., Hawaldar, I. T., Birau, R., & Joisa, J. (2022). Using Econometric Models to Manage the Price Risk of Cocoa Beans: A Case from India. *Risks, 10*(6), 115. https://doi.org/10.3390/risks10060115.

[14] Pfaff, B. (2008) Analysis of Integrated and Cointegrated Time Series with R. Second Edition. Springer, New York. ISBN 0-387-27960-1

[15] R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[16] Real Sector Economic Data – Bank of Ghana. (n.d.). https://www.bog.gov.gh/economic-data/real-sector/.

[17] Scarffe, C. (2024, July 11). *Price or quantity effect? The impacts of the pandemic on Canadian trade.* Global Affairs Canada [Report]. https://www.international.gc.ca/trade-commerce/assets/pdfs/economist-economiste/price-quantity_prix-volume-en.pdf.

[18] Shahzad, F., Bouri, E., Mokni, K., & Ajmi, A. N. (2021). Energy, agriculture, and precious metals: Evidence from time-varying Granger causal relationships for both return and volatility. *Resources Policy, 74*, 102298. https://doi.org/10.1016/j.resourpol.2021.102298.

[19] U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: Food in U.S. City Average [CPIUFDSL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPIUFDSL, March 21, 2025.

[20] U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPIAUCSL, April 4, 2025.

6 Appendix 1: R Code

Our R Code can be found on Github at https://github.com/isaachuahy/STA457_Project.

7 Appendix 2: ETS Model Equations

In the following ETS model equations, we define:

- y_t: The observed cocoa price at time t.
- ℓ_t: The level component of the cocoa price time series at time t.
- b_t: The trend component of the series at time t.
- ε_t: The error term (white noise) at time t, with zero mean and constant variance.
- α: The level smoothing parameter, with $0 \leq \alpha \leq 1$.
- β: The trend smoothing parameter, with $0 \leq \beta \leq 1$.

7.1 ETS(A,A,N) Model Equation

An ETS model with additive error, additive trend, and no seasonal component is specified as:

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$$

with state equations:

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\,\varepsilon_t$$
$$b_t = b_{t-1} + \beta\,\varepsilon_t$$

7.2 ETS(M,A,N) Model Equation

An ETS model with multiplicative error, additive trend, and no seasonal component is specified as:

$$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t),$$

with state equations:

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\,(\ell_{t-1} + b_{t-1})\,\varepsilon_t,$$

$$b_t = b_{t-1} + \beta\,(\ell_{t-1} + b_{t-1})\,\varepsilon_t.$$

7.3 ETS(M,M,N) Model Equation

An ETS model with multiplicative error, multiplicative trend, and no seasonal component is specified as:

$$y_t = \ell_{t-1}(1 + b_{t-1})(1 + \varepsilon_t),$$

with state equations:

$$\ell_t = \ell_{t-1}(1 + b_{t-1})(1 + \alpha\,\varepsilon_t),$$
$$b_t = b_{t-1}(1 + \beta\,\varepsilon_t).$$

8 Appendix 3: ARIMA(0,1,5)-GARCH(1,1) Estimated Parameters

Parameter	Estimate	Std. Error	t-Statistic	p-Value
μ	6.40613155	7.27790700	0.8802162	0.3787422
θ_1	0.25563459	0.05497776	4.6497816	3.322868e-06
θ_2	0.07520155	0.06461876	1.1637728	0.2445161
θ_3	0.02654404	0.05280600	0.5026710	0.6151956
θ_4	0.10390258	0.05929486	1.7523033	0.07972167
θ_5	-0.12802914	0.05464796	-2.3427983	0.01913973
ω	2714.0811138	1759.12300	1.5428600	0.1228647
α_1	0.33599597	0.12600880	2.6661596	0.007672326
β_1	0.66304043	0.11513500	5.7588073	8.471034e-09
shape	3.20054246	0.78666910	4.0684737	4.732209e-05

9 Appendix 4: VARX(14) Estimated Parameters

Estimated parameters for the exogenous variables of the VARX(14) model are provided in Table 9. Estimated parameters for the model's endogenous variables are provided in Table 9.

Transformed Variable	Estimate	Std. Error	t-Statistic	p-Value
Constant	-1.719e+01	6.107e+01	-0.282	0.779297
US CPI Total	2.469e+06	2.027e+06	1.218	0.228079
US CPI Food	3.035e+06	4.361e+06	0.696	0.489264
EU CPI	1.106e+02	9.955e+01	1.111	0.271064
USD/GBP Exchange	-2.579e+03	2.379e+03	-1.084	0.282855
USD/GHS Exchange	-1.015e+02	7.547e+01	-1.345	0.183848

Table 10: VARX(14) Estimated Parameters: Exogenous Variables

Transformed Variable	Lag	Estimate	Std. Error	t-Statistic	p-Value
Cocoa Price	1	-0.009	0.124	-0.074	0.941
Coffee Price	1	-6.520	4.008	-1.627	0.109
Ghana CPI	1	0.103	33.230	0.003	0.998
Cocoa Price	2	0.324	0.128	2.527	0.014
Coffee Price	2	8.733	4.103	2.129	0.038
Ghana CPI	2	23.130	34.200	0.676	0.502
Cocoa Price	3	-0.151	0.124	-1.213	0.230
Coffee Price	3	-1.241	4.173	-0.297	0.767
Ghana CPI	3	-21.780	29.970	-0.727	0.470
Cocoa Price	4	-0.105	0.204	-0.517	0.607
Coffee Price	4	-5.275	4.310	-1.224	0.226
Ghana CPI	4	-65.200	31.850	-2.047	0.045
Cocoa Price	5	-0.263	0.279	-0.942	0.350
Coffee Price	5	-6.223	4.642	-1.341	0.185
Ghana CPI	5	-24.760	34.270	-0.723	0.473
Cocoa Price	6	-0.124	0.296	-0.417	0.678
Coffee Price	6	-10.750	4.905	-2.191	0.032
Ghana CPI	6	112.400	33.940	3.311	0.002
Cocoa Price	7	0.410	0.287	1.428	0.159
Coffee Price	7	0.922	5.013	0.184	0.855
Ghana CPI	7	-78.260	35.970	-2.176	0.034
Cocoa Price	8	-0.004	0.287	-0.014	0.989
Coffee Price	8	-4.543	4.822	-0.942	0.350
Ghana CPI	8	-86.070	35.100	-2.452	0.017
Cocoa Price	9	0.316	0.293	1.077	0.286
Coffee Price	9	0.207	4.697	0.044	0.965
Ghana CPI	9	-2.946	34.530	-0.085	0.932
Cocoa Price	10	0.265	0.290	0.913	0.365
Coffee Price	10	1.686	4.525	0.373	0.711
Ghana CPI	10	134.100	33.240	4.034	0.000
Cocoa Price	11	0.289	0.293	0.986	0.328
Coffee Price	11	-5.682	4.464	-1.273	0.208
Ghana CPI	11	-31.810	33.910	-0.938	0.352
Cocoa Price	12	-0.061	0.295	-0.208	0.836
Coffee Price	12	5.200	4.371	1.190	0.239
Ghana CPI	12	-99.020	31.970	-3.097	0.003
Cocoa Price	13	0.036	0.287	0.126	0.900
Coffee Price	13	3.006	4.429	0.679	0.500
Ghana CPI	13	-40.800	36.460	-1.119	0.268
Cocoa Price	14	0.378	0.275	1.372	0.175
Coffee Price	14	13.150	4.280	3.073	0.003
Ghana CPI	14	141.700	34.840	4.069	0.000

Table 11: VARX(14) Estimated Parameters: Endogenous Variables