A CASE STUDY IN TIME SERIES ANALYSIS

Prepared by: **MUHAMMET EDİZ ÖZDAŞ**
Student No: **090200323**
Submission Date: **June 22, 2025**
Course: **MAT 4901E**
Supervisor: **PROF. DR. ATABEY KAYGUN**

Contents

Introduction

A time series is a type of data that is indexed by regularly spaced time intervals and represents observations of a particular quantities over time. This quantities can be hourly tempeture, patient number for a hospital or daily stock prices.

In history, one of the first important researches about time series was written by Udny Yule in 1927, and this article defined the Autoregressive (AR) models . In the same year, the Moving Average (MA) model, which is another crucial method for time series, was introduced by Eugen Slutsky. In 1938, Herman Wold developed Autoregressive Moving Average (ARMA) models for stationary time series, and he stated that all time series can be decomposed into stochastic and deterministic parts. In 1970, *Time Series Analysis: Forecasting and Control* by G. E. P. Box and G. M. Jenkins was published. This book is the first and most widely used book to describe time series models in a detailed and systematic way. Moreover, the Autoregressive Integrated Moving Average (ARIMA) method was introduced for the first time in this book with Box-Jenkins Method. For variance-based models, the Autoregressive Conditional Heteroskedasticity (ARCH) model was first introduced by Robert F. Engle in 1982 to model time series with time-varying volatility for financial time series data and later, in 1986, Tim Bollerslev extended this study by proposing the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. These models are essential and they are commonly used to model volatility of time series.

In this study, we obtain statistical time series models for the prices of five different commodities, which are cocoa, coffee, sugar, copper and platinum, using data from 2020 to 2024. Different time series models such as AR, MA, ARMA, ARIMA, ARCH, and GARCH are applied to understand the behavior and volatility of each series. We aimed to determine the most suitable model for each commodity by evaluating model performance based on information criteria and residual analyisis. These models can be useful for forecasting and understanding the structure of commodity price movements.

Methodology

2.1 Autocovariance and Autocorrelation Function

Let $\{x_t;\ 1 \leq t \leq n\}$ be a time series sample of size n from $\{X_t\}$.

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_{t+k} - \bar{x})(x_t - \bar{x}) \tag{2.1}$$

where $\bar{x}$ is the sample mean of the time series and k stands for the lag value, which indicates the number of time series shifts.

Autocovariance shows how a time series is related to its past values. It measures how the current value and the past values move together. It is used to compute the autocorrelation function, which is an important function in time series analysis to determine the model parameters.

Autocorrelation function (ACF) defined as:

$$r_k = \frac{c_k}{c_0} \tag{2.2}$$

Similarly to autocovariance, autocorrelation shows the linear relationship between a time series and its lagged (shifted) values. It takes values between −1 and 1. Values close to 1 or −1 indicate strong positive or strong negative linear relationships, respectively. If the value is close to 0, it says that there is little or no linear relationship between the two time series. ACF will be important to determine the order of MA models.

2.2 Partial Autocorrelation

The partial autocorrelation function (PACF) in lag k, denoted ϕ_{kk}, is defined as the coefficient of X_{t-k} in the following linear regression model.

$$X_t = \phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \cdots + \phi_{kk}X_{t-k} + \varepsilon_t \tag{2.3}$$

Then:

$$\text{PACF}(k) = \phi_{kk} \tag{2.4}$$

By estimating this coefficient, the partial autocorrelation ϕ_{kk} can be calculated recursively.

The partial autocorrelation function (PACF) at lag k of a stationary time series $\{X_t\}$ with $E(X_t) = 0$ is

$$\phi_{11} = \text{Corr}(X_{t-1}, X_t) = \frac{\text{Cov}(X_{t-1}, X_t)}{[\text{Var}(X_{t-1})\,\text{Var}(X_t)]^{1/2}} = \rho_1 \tag{2.5}$$

$$\hat{Z}_t = X_t - \hat{X}_t \tag{2.6}$$

$$\phi_{kk} = \text{Corr}(\hat{Z}_{t-k}, \hat{Z}_t) = \frac{\text{Cov}(\hat{Z}_{t-k}, \hat{Z}_t)}{\left[\text{Var}(\hat{Z}_{t-k})\,\text{Var}(\hat{Z}_t)\right]^{1/2}}, \quad k \geq 2 \tag{2.7}$$

Where $\hat{X}_t$ denotes the regression (prediction) in 2.3 [1].

Similarly to the ACF, the PACF is a metric that shows the relationship and correlation with lagged values. However, unlike the ACF, the PACF measures the direct effect of past lags. For example, PACF(2) gives the correlation between the current value (lag 0) and the second lagged value, after removing the effects of the first lag. But ACF(2) shows the correlation between the current value and the values from both the first and second lags together.

2.3 Stationarity

A time series X_t is stationary if E(X_t) = μ is constant, E(X_t^2) ¡ ∞ and $\text{Cov}(X_t, X_{t+k}) = \gamma(k)$ is independent of t for each integer k. Most of the time series, especially financial time series, are nonstationary, and making them stationary is essential for time series models. If the time series is stationary, the probability distribution (or the data distribution) of the time series should be the same at any chosen point or window. Therefore, in many models, stationarity is the key assumption [2]. Differencing is a common method for making nonstationary time series stationary. By this method, seasonality and trend can be removed from time series.

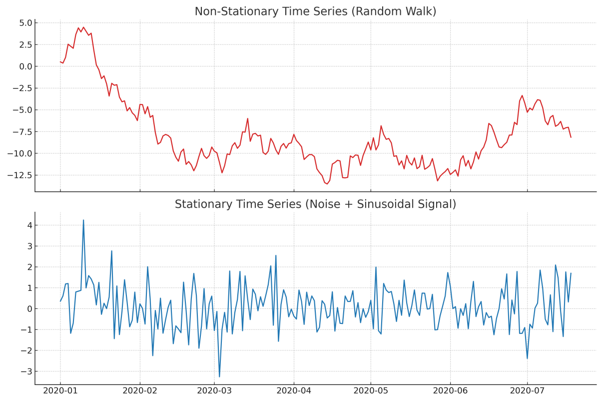

Figure 2.1: Stationary-nonstationary time series example

Looking at Figure 2.1, the first graph is an example of a nonstationary time series because it has no constant mean and constant variance for any interval. But the second plot stays around zero for each window. Also, it looks that it has constant variance.

2.3.1 Stationarity Tests

Stationarity can be observed from the plot of the time series, but a better way is to use a statistical test. KPSS, Ad-Fuller test and Phillips-Person test are most common tests for stationarity.

KPSS

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is commonly used to determine the stationarity in a time series. Basicly, it decomposes time series as deterministic trend, random walk and stationary error to test random walk variance is 0 or not [3]. This implies that:

Null hypothesis (H_0): The time series is stationary (level-stationary or trend-stationary).

Alternative hypothesis (H_1): The time series is nonstationary (has a unit root).

ADF

Another stationarity test is the Augmented Dickey–Fuller (ADF) which is proposed by Dickey and Fuller (1979). It checks whether the lagged levels have significantly relation with difference series or not. [4]. Shortly, it follows:

Null hypothesis (H_0):	The time series is nonstationary.
Alternative hypothesis (H_1):	The time series is stationary.

Phillips-Perron

The last stationarity test used in this study is the Phillips–Perron (PP) test, which is proposed by Phillips and Perron (1988). The PP test is similar to the ADF test but handles serial correlation and heteroskedasticity differently [5]. The hypotheses are written as follows:

Null hypothesis (H_0):	The time series is nonstationary.
Alternative hypothesis (H_1):	The time series is stationary.

2.4 Mean Models

2.4.1 Moving Average (MA) Models

The first models for stationary time series are Moving Average models. The Moving Average model of order q is denoted by MA(q) and defined as:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{2.8}$$

where $\{\varepsilon_t\} \sim$ WN$(0, \sigma_\varepsilon^2)$, that is, $\{\varepsilon_t\}$ is a white noise series, and $\mu, \theta_1, \ldots, \theta_q$ are real-valued parameters (coefficients) with $\theta_q \neq 0$.

The MA(q) model takes a weighted sum of the error terms obtained from WN$(0, \sigma_\varepsilon^2)$, so the series tends to stay around the average value. That is why it is called a ”moving average”. By using white noise terms in the calculation, MA models take past shocks or errors into account.

2.4.2 Autoregressive (AR) Models

Another model for stationary time series is Autoregressive models. The following equation is called the Autoregressive model of order p and is denoted by AR(p):

$$X_t = \psi_0 + \psi_1 X_{t-1} + \psi_2 X_{t-2} + \cdots + \psi_p X_{t-p} + \varepsilon_t \tag{2.9}$$

where $\{\varepsilon_t\} \sim$ WN$(0, \sigma^2)$, $E(X_s \varepsilon_t) = 0$ if $s < t$, and $\psi_0, \psi_1, \ldots, \psi_p$ are real-valued parameters (coefficients) with $\psi_p = 0$.

AR models use past lags to estimate the next value through regression, which is why they are called "autoregressive". AR models are constructed based on the regression coefficients obtained from this process.

2.4.3 Autoregressive Moving Average (ARMA) Models

The stationary model which uses the combination of AR and MA models is called Autoregressive Moving Average (ARMA) models. ARMA(p,q) model of order (p,q) defined as:

$$X_t = \psi_0 + \psi_1 X_{t-1} + \cdots + \psi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{2.10}$$

where $\{\varepsilon_t\}$ ~WN(0, σ^2), $E(X_s \varepsilon_t) = 0$ if $s < t$, and $\{\phi_k\}$ and $\{\theta_k\}$ are real-valued parameters (coefficients) with $\phi_p = 0$ and $\theta_q = 0$.

It follows from the equation, the ARMA model is an additive combination of AR and MA models. So it can take into account both past lagged values and error (shock) terms to model time series.

2.4.4 Autoregressive Integrated Moving Average (ARIMA) Models

Autoregressive Integrated Moving Average (ARIMA) models are very similar to the ARMA models. The main difference between these models ARIMA models are nonstationary models. ARIMA applies differencing to the time series until it becomes stationary, and then fits an ARMA model to the transformed data.

2.5 Order Determination

As we discuss in previous chapters, AR, MA, ARMA and ARIMA models can take different values for their orders. AR(p), MA(q) models has one, ARMA(p,q) models has two and ARIMA(p,d,q) models has three inputs indicate the order of models. For determine this orders there are several ways but most common methods are: Using the ACF/PACF function and using information criteria

2.5.1 ACF/PACF Plots

The first method to determine the orders of the model is to use the ACF and PACF functions. All of this models have an AR part and MA part. This parts has relation between ACF and PACF plots. The following table can be used to choose the suitable order for the model.

	MA(q)	AR(p)	ARMA(p, q), $p>0, q>0$
ACF Behavior	Cuts off after lag q	Tails off	Tails off
PACF Behavior	Tails off	Cuts off after lag p	Tails off

Table 2.1: ACF and PACF behavior of ARMA models

For the AR part, the PACF plot must be analyzed because AR models use past lags for estimation. This means that the AR model requires data that is directly related to the current observation (lag 0). For the MA part, the ACF plot should be considered because the MA models are based on the correlation of the error terms (shocks) at various lags. The ACF helps identify how many lagged error terms significantly affect the current value. According to Box et al., for more optimal results, the maximum number of lags considered in ACF and PACF functions should be n/4, where n is the sample size[6].

2.5.2 Information Criteria

Another way to find the order of the model is by using information criteria (IC). IC is a value for controlling the complexity of the model and evaluating the goodness of the model. To identify the best time series model using IC, all possible combinations of model orders must be tested one by one. The model with the lowest IC value is considered more suitable than the other models. The most common are AIC, BIC, and HQIC.

AIC

The Akaike Information Criteria (AIC) is proposed by Akaike (1974) and is used to find the most balanced model for time series analysis. In this study, the most of the optimal models were selected based on the lowest AIC scores [7]. AIC formula is given as follows:

$$-2\log(\text{maximum likelihood})+2\times(\text{number of independently adjusted parameters within the model})$$

BIC

Another information criterion is the Bayesian Information Criterion (BIC), which was introduced by Schwarz (1978). It is also known as the Schwarz Criterion. Like AIC, BIC is used to balance the model and select the order of time series models.[8]. BIC calculates by:

$$-2\times(\text{maximized log likelihood})+\log(n)\times(\text{number of estimated parameters})$$

HQIC

The third information criterion used in this study is the Hannan–Quinn Information Criterion (HQIC), which was proposed by Hannan and Quinn (1979). It is used in the same way as AIC and BIC to evaluate model performance. Because of its formulation, it produces values that fall between those of AIC and BIC [9]. HQIC evaluate with this:

$$-2 \times (\text{maximized log likelihood}) + \log\log(n) \times (\text{number of estimated parameters})$$

2.6 Variance Models

2.6.1 Autoregressive Conditional Heteroskedasticity (ARCH) Models

The Autoregressive Conditional Heteroskedasticity model (ARCH), which was introduced by Engle [10], is a variance-based model and attempts to model the volatility of time series. Generally, financial time series was preferred using log-return series.

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \alpha_2 X_{t-2}^2 + \cdots + \alpha_p X_{t-p}^2 \tag{2.11}$$

where $\omega \geq 0$, $\alpha_i \geq 0$, and $\alpha_p > 0$ are constants, $\varepsilon_t \sim \text{iid}(0,1)$, and ε_t is independent of $\{X_k;\, k \leq t-1\}$.

The ARCH model states that the conditional variance depends on the squared error terms in the past. Accordingly, when a shock (i.e. a large error) occurs, the conditional variance increases, leading to higher volatility in the days that follow. However, this higher volatility tends to decline relatively quickly and return to normal levels as the impact of past shocks vanish over time.

2.6.2 Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) is another commonly used model for financial time series to modeling volatility. Similar to the ARCH model, it aims to model the conditional variance of a series but differently GARCH uses the past variance values for calculation. It was introduced by Bollerslev (1986) [11] and it is defined as:

$$X_t = \sigma_t \varepsilon_t \tag{2.12}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{2.13}$$

where $\omega \geq 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\alpha_p > 0$, and $\beta_q > 0$ are constants, $\varepsilon_t \sim$ i.i.d.$(0, 1)$, and ε_t is independent of $\{X_k\ ;\ k \leq t-1\}$.

Similar to the ARCH model, the GARCH model also takes into account past error terms, but additionally includes past conditional variances. This allows the GARCH model to capture more persistent volatility over time. As a result, shocks have more persistent effects and the volatility returns to normal levels slower than ARCH.

In order to apply the ARCH/GARCH models, the time series should contain the ARCH effect. ARCH effect implies that the time series has no constant volatility, so its volatility can be modeled.

Engels ML test can be applied to determine the ARCH effect.

2.6.3 Engle LM Test

The Engle LM test is used to determine the presence of ARCH effects in time series data. The test, first proposed by Engle (1982), uses squared residuals to detect relationships among lagged terms, and then applies the Lagrange Multiplier (LM) method to perform the hypothesis test [10]. The hypotheses are stated as follows.

Null hypothesis (H_0): There is no ARCH effect.

Alternative hypothesis (H_1): There is an ARCH effect .

Experiments

3.1 Data preparation

All financial data used in this study were obtained using `yfinance`, Yahoo Finance's Python library. `yfinance` [12] allows stock prices on Yahoo Finance's website to be easily transferred to the Python environment and the financial data needed can be obtained as a dataframe [13]. Firstly, the abbreviation or ticker of the desired financial should be found from the website, then the data can be obtained by using the following piece of code.

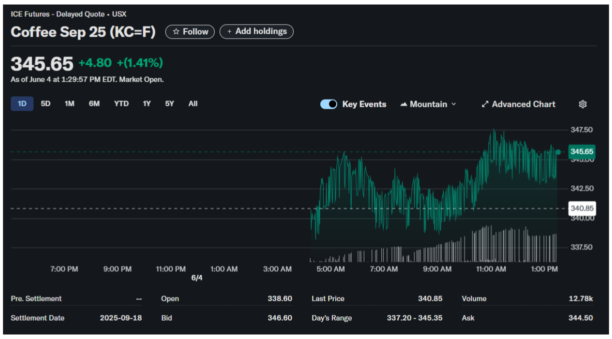

Figure 3.1: Coffee prices from https://finance.yahoo.com/ (Ticker is KC=F for coffee in yfinance)

Price	Close	High	Low	Open	Volume
Ticker	KC=F	KC=F	KC=F	KC=F	KC=F
Date					
2022-01-03	223.300003	226.949997	220.550003	226.149994	13645
2022-01-04	231.750000	235.750000	223.750000	224.750000	21914
2022-01-05	231.750000	233.750000	230.050003	232.050003	12325
2022-01-06	231.699997	232.000000	227.550003	230.600006	13695
2022-01-07	238.449997	240.500000	231.350006	232.000000	25955
...	...	...	...	...	...
2022-12-23	172.000000	172.649994	168.800003	168.899994	9136
2022-12-27	166.949997	171.000000	165.699997	171.000000	10951
2022-12-28	173.550003	174.949997	166.699997	166.750000	16432
2022-12-29	170.050003	173.399994	169.250000	172.949997	10617
2022-12-30	167.300003	172.449997	165.149994	169.100006	13417

251 rows × 5 columns

Figure 3.2: Coffee Dataframe from yfinance

This dataframe includes open, closed, the lowest and the highest prices for a day. For this study, it is only necessary to keep close prices.

Ticker	KC=F
Date	
2022-01-03	223.300003
2022-01-04	231.750000
2022-01-05	231.750000
2022-01-06	231.699997
2022-01-07	238.449997
...	...
2022-12-23	172.000000
2022-12-27	166.949997
2022-12-28	173.550003
2022-12-29	170.050003
2022-12-30	167.300003

251 rows × 1 columns

Figure 3.3: Close prices of coffee

In this study, the closing prices between 2020 and 2024 of coffee, cocoa, sugar, copper and platinum obtained by using this method and this series will be tried to model by using different methods.

3.1.1 ARIMA Models

First of all, it has to be checked that all data sets are stationary or not. For this, all time series can be checked with stationary tests as discussed in 2.3.1 but there is no need to do this because when the graphs are analyzed, it is clearly seen that time series are not stationary.

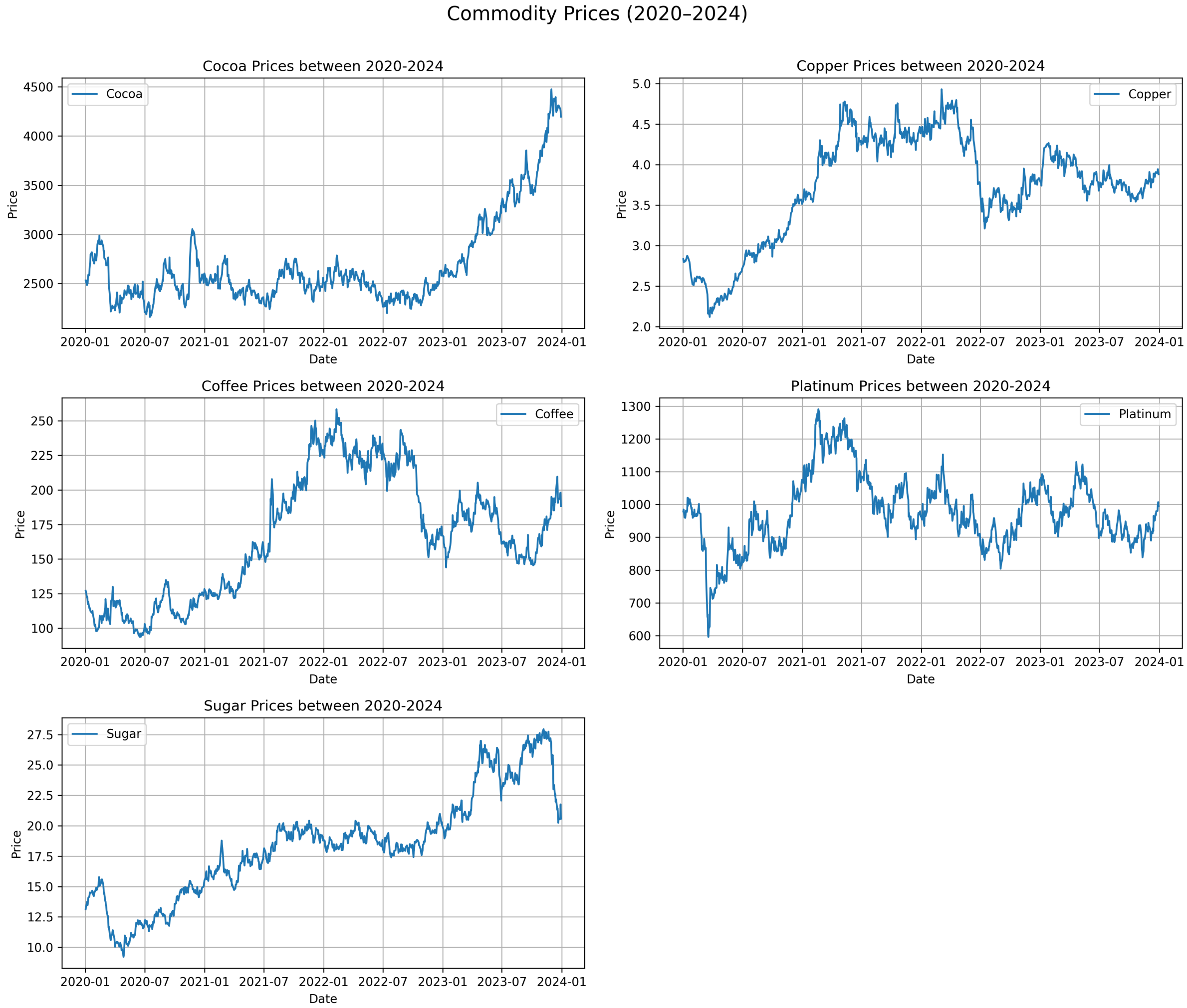

Figure 3.4: Daily prices of 5 commodities between 2020-2024

The results show that all time series are nonstationary, so we can use only ARIMA models for this series. ARIMA models use difference series to make them stationary and then it try to fit for stationary models. So, it can be taken difference by hand to determine parameters of the model.

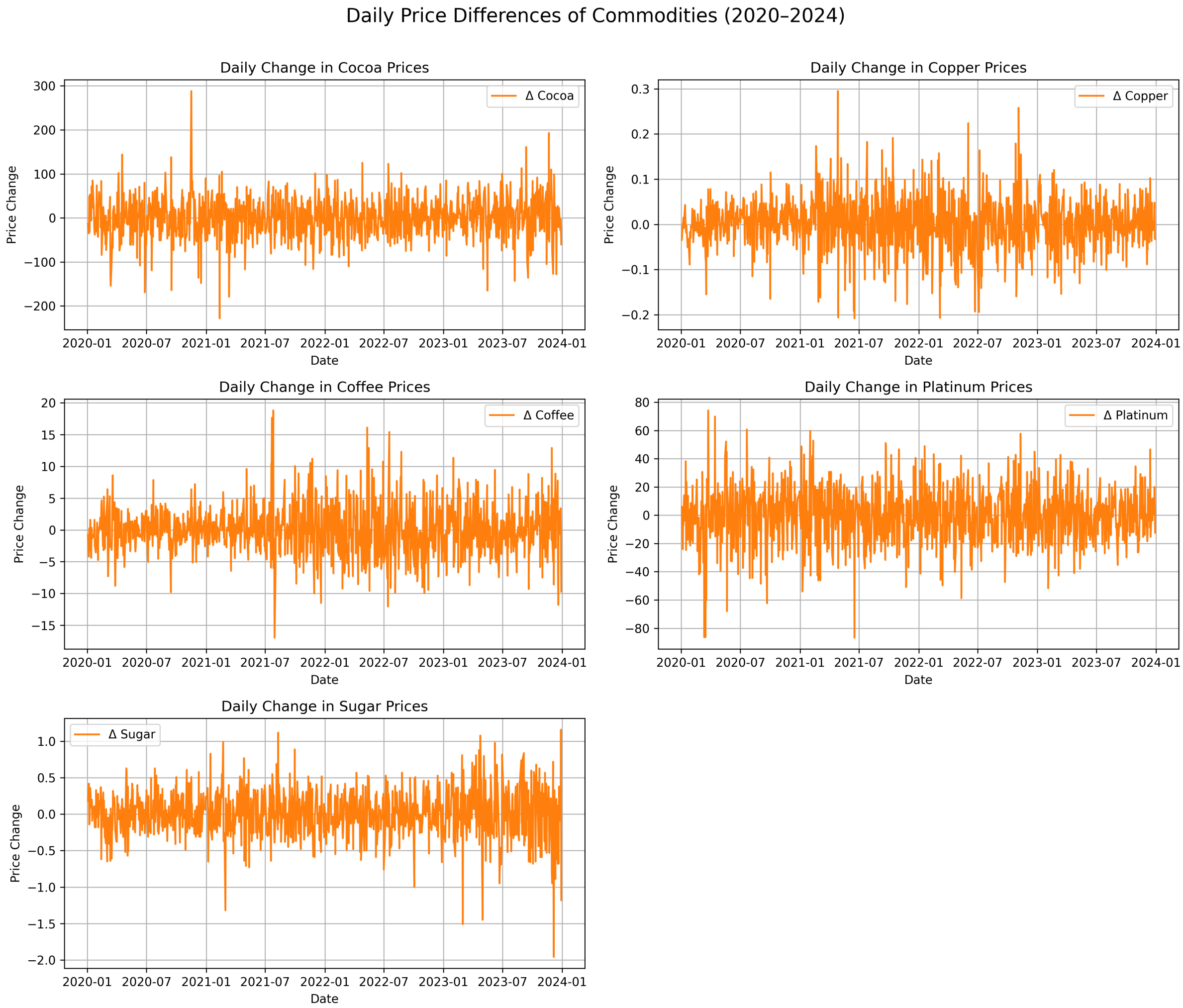

Figure 3.5: Difference series of 5 commodity

Figure 3.5 shows the difference series for commodities. It can be seen that all plots distributed in a constant variance around zero which indicate that the newly obtained series are stationary. But for certainty, this series has to be checked by using stationary tests mentioned in 2.3.1.

Table 3.1: Stationary Test Results for Difference Series

	Cocoa	Copper	Coffee	Platinum	Sugar
ADF Statistic	-20.63	-11.06	-31.47	-14.87	-8.98
ADF p-value	0.0000e+00	4.73e-20	0.0000e+00	1.69e-27	7.46e-15
ADF Result	Stationary	Stationary	Stationary	Stationary	Stationary
KPSS Statistic	0.2704	0.1583	0.1089	0.0371	0.0728
KPSS p-value	0.1	0.1	0.1	0.1	0.1
KPSS Result	Stationary	Stationary	Stationary	Stationary	Stationary
PP Statistic	-30.58	-31.81	-31.69	-31.42	-31.28
PP p-value	0.0	0.0	0.0	0.0	0.0
PP Result	Stationary	Stationary	Stationary	Stationary	Stationary

According to the Table 3.5, the order of models are (p,1,q) because time series become stationary after first difference. For determine p and q values, it should look at ACF/PACF plots.

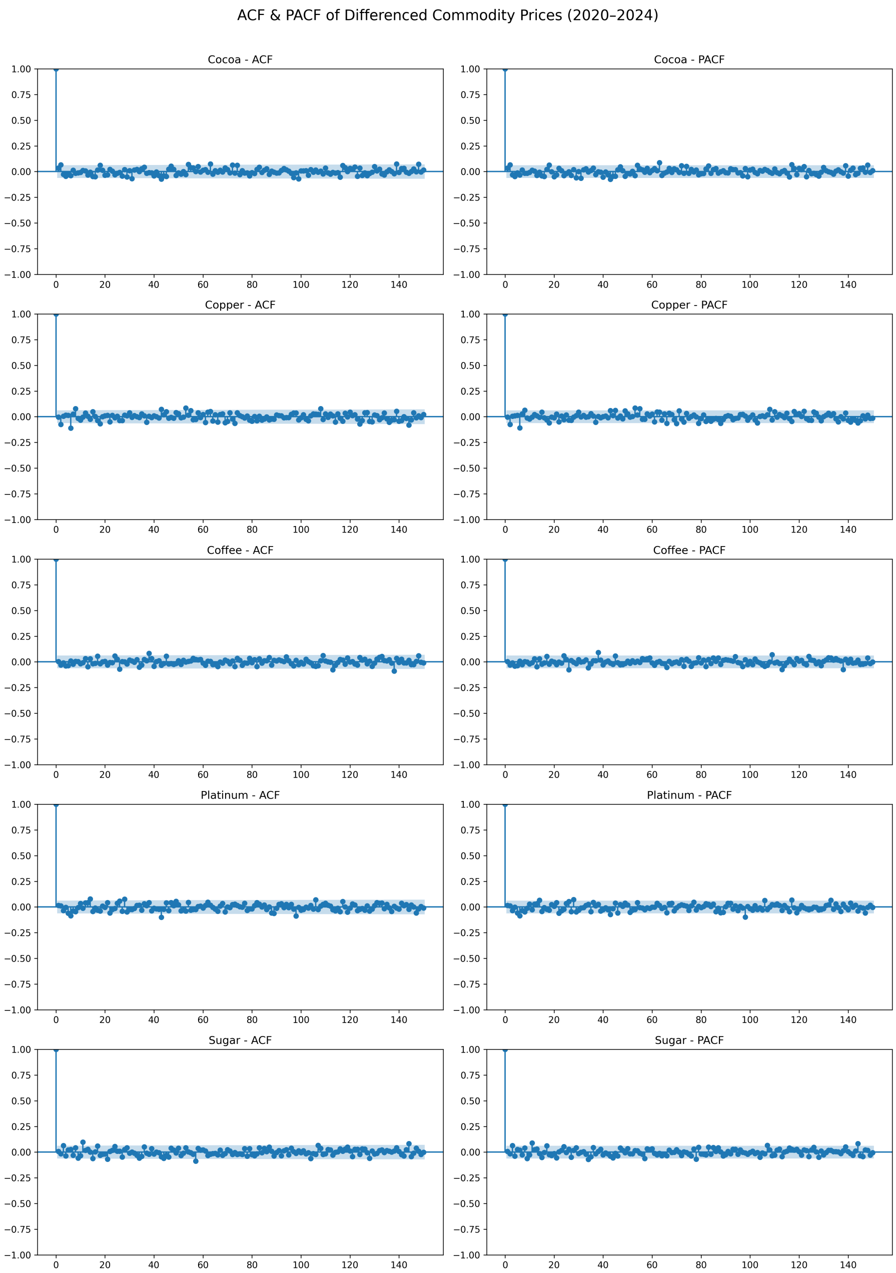

Figure 3.6: ACF/PACF Plots for differenced series of commodities

When graphics are analyzed in 3.6, it is hard to determine the order of the models. So orders can be determined by using information criteria. Table 3.2 shows the best ARIMA models for each commodity, based on the lowest AIC values among all combinations of p and q ranging from 0 to 5.

Table 3.2: Model selection results based on AIC

Series	Model	AIC	BIC	HQIC
Cocoa	ARIMA(0, 1, 0)	10569.23	10574.14	10571.10
Copper	ARIMA(2, 1, 2)	-2850.51	-2825.94	-2841.18
Coffee	ARIMA(0, 1, 1)	5618.62	5623.53	5620.48
Platinum	ARIMA(0, 1, 0)	8889.56	8894.48	8891.43
Sugar	ARIMA(2, 1, 2)	597.01	621.58	606.35

3.1.2 ARCH/GARCH Models

ARCH model is a variance-based model. In financial data, the variance is related to the return, so the log-return series should be used instead of the difference series. Log-return can be found with this formula.

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{3.1}$$

After finding the return series, all series must be checked for the ARCH effect. To this aim, the Engles LM test can be conducted.

Table 3.3: ARCH Effect Test Results for Commodity Log Returns

Commodity	LM Statistic	LM p-value	F Statistic	F p-value	ARCH Detected
Cocoa	10.4986	0.3979	1.0493	0.3994	No
Copper	45.6996	1.626e-06	4.7370	1.187e-06	Yes
Coffee	57.9508	8.831e-09	6.0854	4.947e-09	Yes
Platinum	127.9525	1.222e-22	14.5211	2.791e-24	Yes
Sugar	45.9493	1.465e-06	4.7641	1.065e-06	Yes

Looking at the Table 3.3, all the series have the ARCH effect except cocoa. So, copper, coffee, platinum and sugar can be modeled as ARCH/GARCH.

Similarly with the ARIMA models, order of ARCH models can be determined by using ACF/PACF plot or information criteria. However, in ARCH models, ACF/PACF plots should be considered for squared return series.

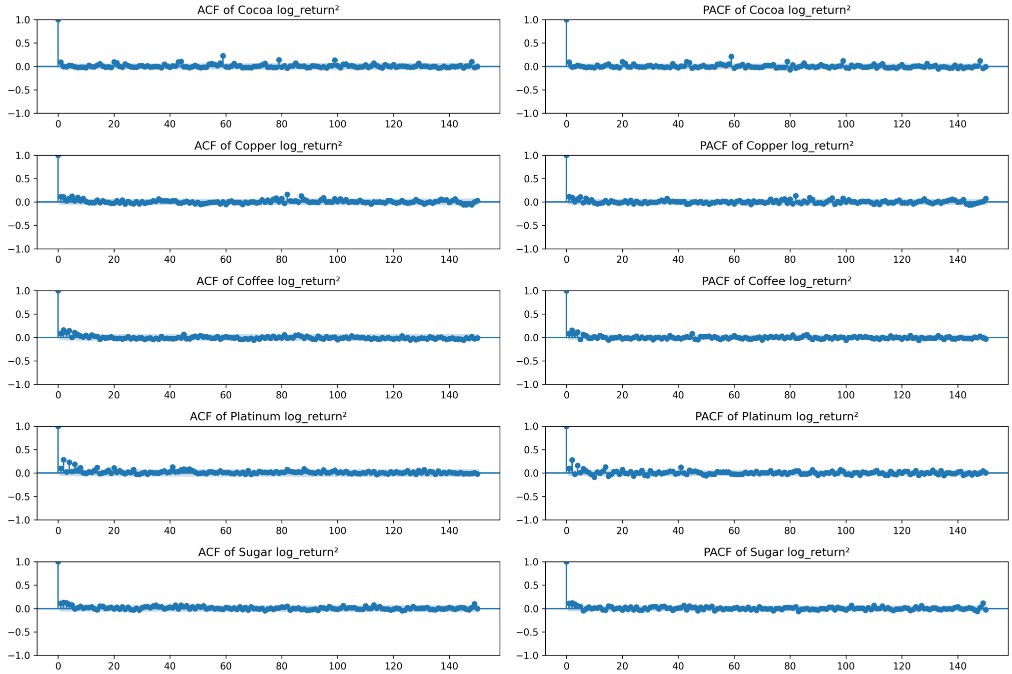

Figure 3.7: ACF/PACF plot of squared log-return series

Looking at the ACF/PACF plots in Table 3.7, it is not clear how to determine the order of the models. But there are little spikes around the first 5 lags. Using this, the lowest AIC values can be searched for each combination of p and q between 0 and 5.

After checking these combinations, the AIC values were obtained for these models:

Time Series	**Best GARCH Model (p,q)**	**AIC**
Copper	GARCH(1,1)	3681.42
Coffee	GARCH(2,1)	4489.23
Platinum	GARCH(1,1)	4315.18
Sugar	GARCH(1,1)	3968.25

Table 3.4: Best GARCH Model Orders and AIC Values for Each Commodity Time Series

3.1.3 ARMA/ARIMA + ARCH/GARCH

Another method of modeling a time series with ARCH/GARCH models to use the residuals of a mean model such as ARMA and ARIMA models. However, differently, these mean models should be constructed using log-return series instead of difference series. For ARMA part, the log-return series should be tested for stationarity by using standard stationary tests mentioned in 2.3.1.

Test	Cocoa	Copper	Coffee	Platinum	Sugar
ADF Statistic	-20.5482	-11.0458	-31.4916	-14.8713	-8.9575
ADF p-value	0.0000	5.23e-20	0.0000	1.66e-27	8.40e-15
ADF Result	Stationary	Stationary	Stationary	Stationary	Stationary
KPSS Statistic	0.2831	0.1541	0.1088	0.0371	0.0777
KPSS p-value	0.1	0.1	0.1	0.1	0.1
KPSS Result	Stationary	Stationary	Stationary	Stationary	Stationary
PP Statistic	-30.5608	-31.8609	-31.7068	-31.4660	-31.2513
PP p-value	0.0000	0.0000	0.0000	0.0000	0.0000
PP Result	Stationary	Stationary	Stationary	Stationary	Stationary

Table 3.5: Stationary Test Results of Log-Return Series

Table 3.5 shows that all series are stationary. After that, the best model for each time series can be found using the AIC values.

Commodity	ARMA Order (p,q)	AIC
Cocoa	(1, 3)	3958.1074
Copper	(2, 2)	3718.7228
Coffee	(0, 0)	4516.9793
Platinum	(4, 4)	4364.9192
Sugar	(2, 3)	4004.4055

Table 3.6: Best ARMA Model Orders with respect to AIC Values for Each Commodity

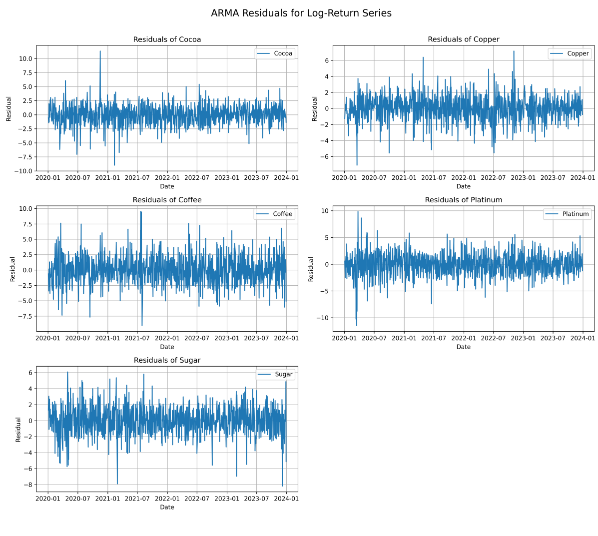

Figure 3.8: Residuals of models in Table 3.6

After finding the appropriate models, residuals need to tested for ARCH effect to model them ARCH/GARCH. Engle LM test can be applied to the residuals to check ARCH effect

Table 3.7: ARCH Effect Test (LM Test) Results

Commodity	LM Statistic	LM p-value	F Statistic	F p-value	ARCH Effect
Cocoa	10.8365	0.3704	1.0835	0.3718	No
Copper	37.6483	4.369e-05	3.8696	3.634e-05	Yes
Coffee	57.6361	1.012e-08	6.0504	5.714e-09	Yes
Platinum	115.6014	3.941e-20	12.9352	1.971e-21	Yes
Sugar	44.7289	2.434e-06	4.6316	1.809e-06	Yes

The table shows that residuals of copper, coffee, platinum and sugar have the ARCH effect. The best ARCH/GARCH models can be detected by using AIC values as same as the previous models.

Commodity	Best Model	Order (p,q)	AIC
Copper	GARCH	(2, 3)	3667.1081
Coffee	GARCH	(2, 1)	4483.6197
Platinum	GARCH	(1, 1)	4302.6299
Sugar	GARCH	(1, 1)	3957.6858

Table 3.8: Best GARCH Model Orders and AIC Values for Each Commodity (Based on ARMA Residuals)

Analysis

To evaluate the goodness and adequacy of the selected models, the residuals of the models will be analyzed in this chapter. Theoretically, the distributions of residuals need to converge to the normal distribution. That means that selected models are sufficient to modeling the deterministic part of the time series, the other part ,the residuals, represents a stochastic white noise process. To determine the distribution of the residuals, the `Fitter` library will be used. `Fitter` can evaluate approximately 80 different distributions to identify the most similar to the residuals distribution [14].

4.1 Cocoa

Table 4.1: Distribution fit results for Cocoa ARIMA(0,1,0) Residuals

Distribution	**SSE**	**AIC**	**BIC**	**KS Stat**	**KS p-value**
logistic	0.000040	1718.38	1728.21	0.0192	0.8448
genlogistic	0.000039	1730.55	1745.29	0.0203	0.7937
genhyperbolic	0.000039	1699.53	1724.09	0.0216	0.7300
johnsonsu	0.000039	1702.53	1722.18	0.0217	0.7229
norminvgauss	0.000040	1716.57	1736.22	0.0219	0.7132
vonmises_line	0.000041	1776.89	1791.63	0.0220	0.7051
t	0.000040	1674.42	1689.16	0.0222	0.6970
nct	0.000041	1676.34	1695.99	0.0234	0.6337
tukeylambda	0.000042	1678.45	1693.19	0.0240	0.6011
crystalball	0.000040	2009.71	2029.36	0.0248	0.5602

The best mean model for cocoa was ARIMA(0,1,0), as shown in Table 3.2. This implies that the difference series of cocoa converges to white noise. Based on Table 4.1, the best-fitting distribution according to the Kolmogorov-Smirnov (KS) p-value is the *logistic* distribution (0.8448). In contrast, when evaluated using the Akaike Information Criterion (AIC), t-distribution provided the best fit (1674.42). Regarding the Sum of Squared Errors (SSE), all of the distributions in the table yielded similar values about 0.0004.

Variance models cannot be applied to cocoa data because according to Table 3.3 and Table 3.7 cocoa does not have ARCH effect. So, the only model is ARIMA(0,1,0). Although the residual of the model converges to the t-distribution for some metrics, the model is not a good model because it only takes the difference series. Means that the difference series is white noise and it can not be modeled. No suitable model could be found for the cocoa data.

4.2 Copper

Table 4.2: Distribution Fit Results for Copper ARIMA(2,1,2)Model

Distribution	SSE	AIC	BIC	KS Stat	KS p-value
norminvgauss	21.6971	165.90	185.55	0.0161	0.9522
genhyperbolic	21.6948	167.70	192.26	0.0165	0.9442
t	21.4721	159.33	174.07	0.0167	0.9387
johnsonsu	21.8592	165.86	185.51	0.0175	0.9130
fisk	22.5698	187.51	202.25	0.0178	0.9007
logistic	22.5802	185.39	195.21	0.0179	0.9000
hypsecant	21.4842	151.79	161.61	0.0181	0.8899
nct	22.0606	167.53	187.18	0.0186	0.8719
burr12	22.9245	194.46	214.11	0.0192	0.8460
genlogistic	22.9642	193.73	208.47	0.0202	0.7977

According to Table 4.2, for the residuals of the ARIMA(2,1,2) model, the best-fitting distribution based on the KS p-value is the *normal inverse gaussian* distribution (0.9522). Furthermore, in terms of AIC and SSE, the t-distribution provides the best fit (159.33 and 21.4721 respectively) Notably, the t-distribution is the third-best distribution based on the KS p-value and it is very close to the highest value (0.9387).

Table 4.3: Distribution Fit Results for Copper GARCH(1,1) Model

Distribution	SSE	AIC	BIC	KS Stat	KS p-value
vonmises_line	0.0616	836.70	851.44	0.0166	0.9403
burr12	0.0623	855.58	875.23	0.0192	0.8471
nct	0.0622	845.61	865.27	0.0195	0.8312
genhyperbolic	0.0621	847.30	871.86	0.0199	0.8113
johnsonsu	0.0625	845.88	865.53	0.0202	0.7988
t	0.0618	840.18	854.92	0.0203	0.7929
logistic	0.0623	845.56	855.39	0.0206	0.7794
fisk	0.0623	847.62	862.36	0.0206	0.7774
genlogistic	0.0629	846.16	860.90	0.0208	0.7708
norminvgauss	0.0630	847.48	867.13	0.0209	0.7630

According to Table 4.3, the best-fitting distribution for the residuals of the Copper GARCH(1,1) model, based on the p-value, is the *von Mises* distribution. Similarly, in terms of AIC, the *von Mises* distribution again provides the lowest value (836.7). Regarding SSE, the differences between the top distributions are minimal, but the t-distribution and *von Mises* distribution both achieve very low SSE values (0.0618 and 0.0616, respectively).

Table 4.4: Distribution Fit Results for Copper ARMA(2,2)+GARCH(2,3) Model

Distribution	SSE	AIC	BIC	KS Stat	KS p-value
burr12	0.0561	867.14	886.79	0.0207	0.7766
t	0.0548	850.25	864.99	0.0219	0.7145
johnsonsu	0.0563	856.32	875.97	0.0221	0.7007
logistic	0.0549	857.00	866.82	0.0222	0.6957
norminvgauss	0.0565	857.94	877.59	0.0230	0.6511
mielke	0.0568	857.41	877.06	0.0242	0.5894
genlogistic	0.0551	858.95	873.69	0.0262	0.4849
gennorm	0.0555	867.18	881.91	0.0302	0.3115
hypsecant	0.0598	819.59	829.42	0.0319	0.2540
skewnorm	0.0675	1001.35	1016.09	0.0324	0.2376

According to table 4.4, which shows the best-fit distributions for the GARCH(2,3) model for residuals of ARMA(2,2), *Burr Type XII* distributions have the highest p-value for ks statistics (0.7766). Additionally, for SSE and AIC values, the best fitted distribution is t-distribution (0.0548 and 850.25 respectively).

For copper, all of the models give promising results. For instance, the residuals of the ARIMA(2,1,2) model, converge to the t-distribution in terms of SSE, AIC, and p-values. For the log-return series, the GARCH(1,1) model converges to the t-distribution only in terms of SSE. The GARCH(2,3) model, which uses the residuals of the ARMA(2,2) model, also converges to the t-distribution in terms of SSE, AIC and p-value. All models converges to t-distribution so it can be said that all models are valid, choice depend to the purpose of the model.

4.3 Coffee

Table 4.5: Distribution Fit Results for Coffee ARIMA(0,1,1) Model

Distribution	SSE	AIC	BIC	KS Stat	KS p-value
genhyperbolic	0.005572	1010.49	1035.05	0.0169	0.9318
gennorm	0.005545	1013.26	1027.99	0.0194	0.8377
hypsecant	0.006532	1001.59	1011.41	0.0195	0.8303
norminvgauss	0.006603	1007.06	1026.71	0.0229	0.6586
t	0.007573	1009.47	1024.21	0.0244	0.5792
johnsonsu	0.006957	1008.64	1028.29	0.0248	0.5580
laplace_asymmetric	0.007500	969.94	984.68	0.0254	0.5270
nct	0.007335	1012.09	1031.74	0.0266	0.4662
loglaplace	0.006948	971.72	986.46	0.0273	0.4333
logistic	0.008060	1033.82	1043.65	0.0275	0.4230

According to Table 4.5, based on the p-value, the best-fit distribution for the residuals of ARIMA(0,1,1) is the *generalized hyperbolic* distribution with a KS p-value of 0.9318. Similarly, in terms of AIC, the *hyperbolic secant* distribution provides the lowest AIC value (1001.59). Regarding SSE the *generalized normal* distribution achieves the lowest SSE values (0.005545).

Table 4.6: Distribution Fit Results for Coffee GARCH(2,1) Model

Distribution	**SSE**	**AIC**	**BIC**	**KS Stat**	**KS p-value**
mielke	0.0235	842.13	861.78	0.0135	0.9922
norminvgauss	0.0233	842.30	861.95	0.0139	0.9890
genlogistic	0.0234	840.60	855.34	0.0142	0.9856
fisk	0.0236	835.71	850.44	0.0143	0.9850
johnsonsu	0.0235	843.02	862.67	0.0148	0.9785
nct	0.0236	844.40	864.05	0.0158	0.9608
gennorm	0.0251	841.98	856.71	0.0194	0.8369
logistic	0.0254	830.93	840.75	0.0195	0.8302
t	0.0257	839.48	854.21	0.0224	0.6865
vonmises_line	0.0259	818.55	833.28	0.0235	0.6277

According to Table 4.6, the best-fitting distribution for the residuals of the GARCH(2,1) model based on the p-value is the *Mielke Beta-Kappa* distribution (0.9922). Additionally, in terms of SSE, the *Normal Inverse Gaussian* distribution provides the lowest error (0.0233) and *Logistic* distributions has lowest AIC values (830.93).

As seen in Table 3.6, the log-return series of coffee fits the ARMA(0,0) model, which indicates that the log-return series behaves like white noise. Therefore, the ARMA model does not make any changes to the log-return data. That is why the next step will give the same results as in Table 4.6 because they modeling the same data by using same methods.

In summary, for coffee prices, neither the GARCH(2,1) model nor the ARIMA(0,1,1) model converges to any acceptable distribution according to the corresponding tables. Thus, alternative methods may be needed to improve goodness of models.

4.4 Platinum

Table 4.7: Distribution Fit Results for Platinum ARIMA(0,1,0) Model

Distribution	SSE	AIC	BIC	KS Stat	KS p-value
nct	0.000375	1268.75	1288.40	0.0208	0.7710
genhyperbolic	0.000375	1270.51	1295.08	0.0209	0.7643
johnsonsu	0.000375	1268.12	1287.78	0.0209	0.7639
norminvgauss	0.000376	1267.78	1287.43	0.0210	0.7570
t	0.000381	1269.07	1283.81	0.0235	0.6285
dweibull	0.000402	1260.66	1275.40	0.0237	0.6167
genlogistic	0.000387	1254.94	1269.68	0.0238	0.6086
vonmises_line	0.000380	1252.56	1267.30	0.0239	0.6065
crystalball	0.000387	1288.61	1308.26	0.0243	0.5833
nakagami	0.000422	1332.38	1347.12	0.0251	0.5432

According to Table 4.7, the best-fitting distribution based on the KS p-value is the *Noncentral* t distribution (0.7710). Regarding SSE, the *Noncentral* t, *Generalized Hyperbolic*, and *Johnson's SU* distributions all achieve the lowest value (0.000375). In terms of AIC, the *von Mises* distribution provides the lowest score (1252.56).

Table 4.8: Distribution Fit Results for Platinum GARCH(1,1) Model

Distribution	SSE	AIC	BIC	KS Stat	KS p-value
burr12	0.0198	1090.65	1110.30	0.0211	0.7558
vonmises_line	0.0199	1127.21	1141.95	0.0211	0.7513
genlogistic	0.0201	1078.62	1093.36	0.0218	0.7166
burr	0.0202	1081.40	1101.05	0.0221	0.6995
norminvgauss	0.0205	1069.47	1089.12	0.0223	0.6914
johnsonsu	0.0202	1059.78	1079.43	0.0223	0.6909
nct	0.0200	1053.87	1073.52	0.0223	0.6905
genhyperbolic	0.0201	1055.77	1080.33	0.0224	0.6848
logistic	0.0204	1084.57	1094.39	0.0242	0.5898
fisk	0.0204	1086.62	1101.36	0.0242	0.5892

According to Table 4.8, the best-fitting distribution based on the KS p-value is the *Burr Type XII* distribution (0.7558). In terms of SSE, the *Burr Type XII* distribution also achieves the lowest error (0.0198) again. Regarding AIC, the lowest score is obtained by the *Noncentral* t distribution (1053.87).

According to Table 4.9,which shows the ARMA(4,4)+GARCH(1,1) model for log return series, the best-fitting distribution based on the KS p-value is the t distribution (0.9119) and t distribution also achieves one of the lowest SSE values (0.0212). In terms of AIC, again, the t distribution provides the best score (994.79).

[H]

Table 4.9: Distribution Fit Results for Sugar ARMA(4,4)+GARCH(1,1) Model

Distribution	SSE	AIC	BIC	KS stat.	KS p-value
t	0.0212	994.79	1009.52	0.0175	0.9119
johnsonsu	0.0211	998.97	1018.63	0.0184	0.8796
logistic	0.0217	999.04	1008.86	0.0185	0.8769
norminvgauss	0.0213	1005.14	1024.79	0.0186	0.8708
genlogistic	0.0216	996.46	1011.20	0.0194	0.8383
dweibull	0.0259	995.96	1010.70	0.0212	0.7482
crystalball	0.0210	1076.91	1096.56	0.0243	0.5845
gennorm	0.0234	1034.81	1049.55	0.0244	0.5769
nakagami	0.0237	1213.18	1227.92	0.0250	0.5461
exponnorm	0.0234	1208.17	1222.90	0.0252	0.5394

For platinum prices, the residuals of the ARMA(4,4) + GARCH(1,1) model clearly converge to the *t-distribution*, unlike the ARIMA(0,1,0) and GARCH(1,1) models. Therefore, it can be concluded that the ARMA(4,4) + GARCH(1,1) model is more suitable than other models for platinum prices.

4.5 Sugar

Table 4.10: Distribution Fit Results for Sugar ARIMA(2,1,2) Model

Distribution	SSE	AIC	BIC	KS Stat	KS p-value
nct	1.5858	591.38	611.03	0.0169	0.9318
t	1.5774	594.45	609.19	0.0171	0.9247
burr12	1.6022	606.21	625.86	0.0171	0.9245
genhyperbolic	1.5967	584.19	608.75	0.0173	0.9186
johnsonsu	1.5955	596.07	615.72	0.0175	0.9125
fisk	1.5911	612.48	627.22	0.0178	0.9005
logistic	1.5911	610.49	620.32	0.0178	0.9004
norminvgauss	1.6088	602.36	622.02	0.0180	0.8937
burr	1.6078	603.32	622.97	0.0182	0.8883
genlogistic	1.6088	601.18	615.91	0.0183	0.8847

According to Table 4.10, the best-fitting distribution based on the KS p-value is the *noncentral* t distribution (0.9318) and the second best is the t-distribution (0.9247). Also in terms of SSE, the t-distribution provides the lowest value (1.5774), closely followed by the *noncentral* t distribution (1.5858). Regarding AIC, the lowest score is achieved by the *generalized hyperbolic* distribution (584.19).

Table 4.11: Distribution Fit Results for Sugar GARCH(1,1) Model

Distribution	SSE	AIC	BIC	KS Stat	KS p-value
vonmises_line	0.0536	797.40	812.14	0.0207	0.7759
genhyperbolic	0.0556	803.01	827.58	0.0228	0.6644
nct	0.0556	802.46	822.11	0.0228	0.6623
johnsonsu	0.0559	802.88	822.53	0.0232	0.6416
t	0.0549	804.22	818.96	0.0235	0.6246
norminvgauss	0.0563	803.50	823.16	0.0236	0.6234
crystalball	0.0538	812.60	832.26	0.0239	0.6046
burr12	0.0573	796.37	816.03	0.0252	0.5369
tukeylambda	0.0561	802.18	816.92	0.0253	0.5333
fisk	0.0566	797.79	812.53	0.0260	0.4953

According to Table 4.11, the best-fitting distribution based on the KS p-value is the *von Mises* distribution (0.7759). This distribution also achieves the lowest SSE value (0.0536) and the best AIC score (797.40).

Table 4.12: Distribution Fit Results for Sugar ARMA(2,3)+GARCH(1,1) Model

Distribution	SSE	AIC	BIC	KS stat.	KS p-value
johnsonsu	0.0594	801.05	820.70	0.0213	0.7441
gennorm	0.0613	814.12	828.85	0.0226	0.6746
mielke	0.0655	821.82	841.47	0.0236	0.6203
logistic	0.0608	792.14	801.97	0.0237	0.6182
genlogistic	0.0608	794.15	808.89	0.0246	0.5707
crystalball	0.0573	806.34	825.99	0.0257	0.5131
pearson3	0.0618	864.15	878.88	0.0273	0.4333
skewnorm	0.0618	857.05	871.79	0.0279	0.4052
exponnorm	0.0608	886.24	900.98	0.0282	0.3935
norm	0.0608	884.24	894.06	0.0282	0.3935

According to Table 4.12, the best-fitting distribution based on the KS p-value and AIC value is the *Johnson's SU* distribution (0.7441 and 801.05,respectively). In terms of SSE, the lowest error obtained from the *crystal ball* distribution (0.0573)

For sugar prices, the ARMA(2,3)+GARCH(1,1) model and the GARCH(1,1) model do not converge to any acceptable distributions such as the t-distribution or the normal distribution. However, the residuals of the ARIMA(2,1,2) model converge to the t-distribution in terms of both KS p-value and SSE. In comparison, the ARIMA(2,1,2) model appears to be a more appropriate choice to model sugar prices than the ARMA(2,3)+GARCH(1,1) and GARCH(1,1) models.

Conclusion

In this project, we tried to model five different commodities with basic time series models and analyzed the residuals of the models to determine the goodness of the models. In the experiments, we examined how to choose the model, how to choose the order of the model, when to use statistical tests, and what should be taken into account when doing analysis. Although some models may look successful in the analysis, they may not be sufficient for real world problems that require forecasting such as trading or risk strategies. Using different methods may improve the results.

References

[1] Changquan Huang and Alla Petukhina. *Applied Time Series Analysis and Forecasting with Python*. Springer Nature Switzerland AG, Cham, Switzerland, 2022. Mathematics Subject Classification: 62-01, 62M10, 62-04, 62-07, 62M20, 62M45, 62F03, 62F10.

[2] Manu Joseph. *Modern Time Series Forecasting with Python*. Packt Publishing Ltd., Birmingham, UK, 2022. First published: November 2022.

[3] Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3):159–178, 1992.

[4] David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431, 1979.

[5] Peter C. B. Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.

[6] George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, NJ, 5 edition, 2016.

[7] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[8] Gideon E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

[9] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):190–195, 1979.

[10] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[11] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[12] Ran Aroussi. yfinance: Yahoo! finance market data downloader. `https://github.com/ranaroussi/yfinance`, 2015. Accessed: 2025-06-03.

[13] Yahoo Finance. Yahoo finance, 2025. Accessed: 2025-06-03.

[14] Thomas Cokelaer. Fitter documentation. `https://fitter.readthedocs.io/en/latest/`, 2021. Accessed: 2025-06-12.