APPLICATION OF ARIMA AND ARIMA-GARCH MODELS IN PREDICTING STOCK PRICES FOR BRITISH AMERICAN TOBACCO

Nakayo Adrine

142970

Submitted in partial fulfillment of the requirements for the Degree of Financial Economics at Strathmore University

Strathmore Institute of Mathematical Sciences

Strathmore University

Nairobi, Kenya

January, 2025

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Project contains no material previously published or written by another person except where due reference is made in the Research Project itself.

..……………...

……………..

……………..

This Research Project has been submitted for examination with my approval as the Supervisor.

Dr SEBASTIAN K KALOVWE

..……………...

SK ……………..

…………….................................. 01/29/2025

Strathmore Institute of Mathematical Sciences

Strathmore University

ABSTRACT

Forecasting stock prices is important for investors as it allows them to make the right decisions. Trends in historical data can be effectively captured by time series models such as ARIMA (Autoregressive Integrated Moving Average), but they are not able to capture the inherent volatility of financial markets. This volatility can lead to significant forecasting errors, impacting investment decisions. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) will also be used to address this limitation by modeling volatility. This study investigates the potential of a hybrid ARIMA-GARCH model for predicting stock prices using historical data. The hybrid model successfully captured volatility clustering, significantly enhancing forecasting accuracy. A comparison based on Theil's U-statistic showed a lower value for the ARIMA-GARCH model compared to ARIMA alone. This robust hybrid framework offers valuable insights into the price dynamics and volatility trends of BAT's stock, benefiting investors and financial analysts.

Table of Contents

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Background of the study

The tobacco industry, despite facing stringent regulations and ongoing public health concerns, remains a key sector for investors due to its capacity to generate substantial profits and its broader economic significance. British American Tobacco (BAT) stands out as a strong investment within this sector, leveraging its vast market presence and financial stability making it attractive to investors (Wright, 2024). However, British American Tobacco (BAT) faces significant negative speculation that could lead to volatility in its stock prices. These speculations include regulatory pressures and shifting consumer preferences toward e-cigarettes and reduced-risk products (Ghigini, 2024). These fluctuations present uncertainties to investors. Therefore, accurately forecasting these movements is crucial for helping investors navigate the complexities associated with the tobacco industry.

Many investors rely on brokerage recommendations. However, studies show they have little or no success in identifying stocks with significant price appreciation potential, as brokers may be subject to conflicts of interest and bias (finance, 2024). This limitation creates the need for a data-driven and objective approach to forecasting stock prices. The ARIMA-GARCH models, widely utilized in financial markets, offer a more reliable alternative by effectively predicting both the mean and variance of stock prices, providing a more informed basis for investment decisions.

This study focuses on leveraging ARIMA-GARCH models to predict BAT's stock price volatility. The ARIMA model, proposed by Box and Jenkins in 1970, has been used to evaluate and forecast time series data, including stock prices. ARIMA model forecasts both short- and long-term price movements by capturing patterns and trends in stock prices (Jin Liu, 2024). However, the market is prone to changing volatility (heteroskedasticity), where prices appear to follow a random walk process and are unpredictable. The GARCH model, proposed by Bollerslev (1986) and Taylor (1986), addresses the issue of heteroscedasticity. It explicitly models the volatility in the data, allowing for variations in the magnitude of price fluctuations. This makes GARCH particularly suitable for analyzing financial data with inherent volatility, like stock prices.

This study examines the combination of the existing models for predicting stock prices to achieve improved outcomes. This study will combine two models: ARIMA and GARCH. ARIMA will capture the underlying trends and seasonality in the data, while GARCH will model the volatility component, providing a more realistic picture of future price movements. Studies show that the integration of GARCH and ARIMA models has resulted in a 20% decrease in the volatility of a portfolio, thereby assisting institutions in efficiently mitigating and managing risks. (Jin Liu, 2024).

The ARIMA-GARCH model is particularly suitable for forecasting British American Tobacco (BAT) stock prices because it effectively captures the unique features often found in financial time series data. BAT's stock, like many other publicly traded securities, tends to exhibit non-stationary behavior, where prices follow trends and patterns that evolve over time, as well as periods of clustered volatility.

By combining them, this study may develop a more comprehensive approach to predicting stock prices. By understanding how stock prices can fluctuate, investors can make informed decisions and manage their risk exposure.

1.2 Statement of problem

Time series models like ARIMA have been effectively used to forecast stock prices by capturing historical trends and patterns. However, it does not capture the high volatility often present in financial markets (Jin Liu, 2024). The GARCH model is designed to model volatility but is not effective in capturing underlying trends. This is because GARCH models assume time series to have a conditional mean of zero, even though this assumption may not always hold true. The problem arises from the inadequacy of ARIMA and GARCH models when used independently to provide accurate stock price forecasts (Grachev, 2017). This can lead to significant errors in predicting stock prices which in turn can have substantial consequences for investors, financial institutions, and the overall market.

This problem is particularly relevant in the context of BAT, where stock price fluctuations are influenced by regulatory pressures and changing consumer preferences toward e-cigarettes and reduced-risk products. These uncertainties have created an unpredictable environment where traditional forecasting models struggle to provide accurate predictions for BAT's stock. Given the dynamic nature of the tobacco industry, this problem is likely to persist and possibly intensify in the future, making it more difficult to develop reliable forecasting models. This challenge emphasizes the need for innovative forecasting methods that can accommodate the dynamic nature of the tobacco industry.

Understanding and addressing this problem is crucial because improved stock price forecasts can lead to more informed investment decisions and better risk management strategies. For individual investors, better forecasts mean more informed decisions, potentially leading to higher returns and reduced risk. Policymakers, too, can benefit from better forecasting models as they seek to ensure market stability and protect investors. By offering more reliable tools for navigating the complexities of the stock market, this study can benefit various stakeholders, including individual investors, institutional investors, financial advisors, and policymakers.

1.3 Research Gap

Even though financial modeling and forecasting have advanced significantly, there is still a critical gap in the development of models that can simultaneously address stock price volatility and trend patterns particularly for British American Tobacco. The exact issue that this study addresses is the development of a hybrid forecasting model that incorporates the strengths of ARIMA and GARCH models. Specifically, the research seeks to determine whether integrating these models can enhance forecasting accuracy by capturing both the trends and the time-varying volatility of stock prices.

1.4 Research objectives

1.4.1 General objective

The main objective of this study is to develop and evaluate a hybrid ARIMA-GARCH model for improved forecasting of stock prices by capturing both trend and volatility dynamics.

1.4.2 Specific objectives

a. To develop a hybrid ARIMA-GARCH model and forecast stock prices.
b. Evaluate the performance of the hybrid model against a standalone ARIMA model using historical stock price data.
c. Investigate the impacts of the hybrid model on future research for British American Tobacco.

1.5 Significance of the study

By developing a hybrid model that integrates the strengths of both ARIMA and GARCH models, this study seeks to address the critical gap in the field of financial forecasting. This will improve the accuracy of stock price predictions. By capturing both trend patterns and volatility, the model addresses two aspects of stock price movements that are often analyzed separately. This integrated approach can lead to more reliable predictions, which are necessary for making informed investment decisions.

In practice, the hybrid model can be used by financial institutions, investment firms, and individual investors to enhance their forecasting capabilities. Improved accuracy in stock price predictions can lead to better investment strategies, risk management and portfolio optimization ultimately contributing to higher returns and reduced financial risks. This study opens avenues for further studies in financial modeling, encouraging the exploration of other hybrid models and their applications.

CHAPTER 2: LITERATURE REVIEW

2.1 ARIMA Model

Tamerlan's study (2022) investigated the effectiveness of the ARIMA model in forecasting Istanbul's stock market prices. This study uses monthly market price data from January 2009 to March 2021. The study found that ARIMA (3,1,5) is the most accurate model for predicting the increase in stock market prices. The ARIMA (3,1,5) model minimizes the forecast errors by closely matching the predicted values with the true values. The model accurately represented the fluctuations in the prices of Istanbul's stock market during that period.

Sidharth and Pramod (2022) in their study developed a model to forecast the closing stock prices of Tesla (TSLA) and NIO using stock price data from December 2018 to 2022. For Tesla (TSLA), the ARIMA (0,1,0) model was chosen as optimal. ARIMA (1,1,1) model was found to be the most optimal for NIO. Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) metrics were used to evaluate the performance of these models. The results showed a MAPE of 0.03% for Tesla and 0.18% for NIO, indicating better prediction for Tesla.

A study by Farooq Shahid (2023) used ARMA and ARIMA models to forecast prices in stock markets across Pakistan, India, Bangladesh, and Sri Lanka from 2012 to 2022. The model that had the smallest mean squared error (MSE), mean absolute error (MAE), and inequality coefficient was selected as the most accurate model. The study found that ARIMA (1,1,1), ARIMA (6,1,5) and ARIMA (4,1,1) were the most optimal models for Pakistan, India and Bangladesh respectively.

Padma, Suresh and Kumar (2022) in their study, used monthly share prices of the BSE Sensex from January 2000 to December 2020 to forecast the prices. It was found that the ARIMA (1,1,0) model was the most appropriate and successful model for predicting the BSE Sensex price index. The autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to determine the order of the autoregressive (AR) and moving average (MA) components in the ARIMA model. The portmanteau test was employed to verify the model's suitability.

Yutong (2023) conducted a study to assess the strength of the ARIMA model in forecasting short-term stock price patterns for Pinduoduo, a mobile online e-commerce platform in China. The analysis used historical stock prices from January 2020 to January 2021 that were obtained from Yahoo Finance. Logarithmic returns were applied to the closing prices in order to attain stationarity and stabilize variance. The choice was made using the Akaike Information Criterion (AIC) and the ARIMA (2,0,2) model was chosen. The model demonstrated a 61.54% accuracy rate in predicting the direction of log returns.

Shakir and Hela (2020) carried out a research to assess how well various ARIMA models performed in predicting Netflix stock prices. They compared Auto ARIMA and customized ARIMA (p, d, q) models. They used historical stock price data from April 2015 to April 2020. ARIMA (1,1,33) was found to be highly accurate because it had the lowest Mean Absolute Percentage Error compared to Auto ARIMA and ARIMA (1,2,33).

ARIMA model only might not capture market shocks well. By incorporating GARCH alongside ARIMA, the model can be enhanced to predict both the mean (average price) and volatility (fluctuations around the mean), therefore providing more accurate predictions.

2.2 ARCH / GARCH Model

Mbarek (2015) examined the fluctuation patterns of daily stock returns for the Tunisian index from 1984 to 2010 to determine the suitability of an ARCH model for modeling the volatility of these returns. An ARCH (3) model was identified as appropriate for modeling the fluctuations of the returns. The returns exhibited high volatility with clustering.

Awalludin (2018) used daily closing prices of Indonesian companies from July 2007 to September 2015 to analyze the behavior of stock returns. The analysis revealed non-normality characterized by fat tails signifying significant deviations from a normal distribution. The GARCH (1,1) model was employed to estimate volatility. It identified periods of heightened volatility, particularly notable during economic crises like the one experienced in late 2008. This study demonstrates how effective and reliable the GARCH (1,1) model is in identifying and quantifying volatility clusters in financial markets.

Incorporating the ARIMA model would capture the linear trends and give a detailed analysis of stock return behaviors.

2.3 Hybrid model ARIMA-GARCH

A study by Grachev (2017) explored the effectiveness of ARMA-GARCH models in predicting stock market returns. The research aimed to find an ARMA model suited for short-term return forecasts, investigate the presence of heteroskedasticity in the data, and develop an ARMA-GARCH model that captures both the average return and its volatility. The data used was weekly excess returns of the MSCI World Index from 2000 to 2017. ARMA (1,0) and GARCH (1,1) were found to be the best fits among various combinations of ARMA and GARCH models when compared using the Akaike Information Criterion (AIC). The findings emphasize the importance of modeling both the average return and its volatility.

Xiang (2022) conducted a study focused on improving the accuracy of WTI crude oil price forecasting. The study analyzed how well the ARIMA-GARCH model captured the complexities of oil price movements. The study utilized an ARIMA (1,1,0) model to address the trend and seasonality in WTI crude oil prices, while a GARCH (1,1) model was applied to capture the volatility. The accuracy of the forecasts of ARIMA and ARIMA-GARCH models was evaluated using the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). This hybrid approach leads to significant improvements in forecasting accuracy compared to the ARIMA model alone, as evidenced by the reduction in MAPE (from 1.549%) to (0.045%) and RMSE (from 1.032 to 0.071).

Gao (2021) conducted a study on forecasting the stock prices of companies trading on the Shanghai Stock Exchange (Ping) using the ARIMA-GARCH model. The most appropriate models were found to be GARCH (1,1) and ARIMA (3,1,4). Compared to using the ARIMA model alone, the study showed that the ARIMA-GARCH model is more successful at capturing both linear and nonlinear patterns in stock price data, enabling more accurate predictions of future prices.

Mustapa and Ismail (2019) study focused on developing an appropriate model to generate the short term forecasts of the monthly stock price of S&P500. They used the monthly adjusted closing prices from 2001 to 2018 and found that ARIMA (2,1,2) and GARCH (1,1) were the most effective for predicting the S&P 500 monthly stock prices in the short term. Trial and error approach was used to identify the best-fitting ARIMA model. Including GARCH (1,1) in their analysis significantly improved the performance of the ARIMA model by accounting for volatility clustering in the data. The hybrid ARIMA-GARCH model effectively captures the patterns in the S&P 500 stock prices and provides accurate short-term forecasts.

These studies highlight that ARIMA-GARCH models effectively capture volatility in financial markets. Ying Xiang (2022) specifically shows that combining ARIMA with GARCH leads to significant improvements in forecasting accuracy. ARIMA-GARCH models are robust tools capable of capturing the dual aspects of financial data.

The existing literature focuses on the application of ARIMA and GARCH models within broader financial markets and neglects industry specific contexts. This research aims to address that gap by applying ARIMA-GARCH model specifically to British American Tobacco (BAT) in the tobacco industry, which is characterized by significant volatility due to regulatory changes and evolving consumer preferences. By focusing on the tobacco industry, the study will offer tailored insights into the unique investment dynamics in this industry. Given the volatilities in this industry, ARIMA-GARCH provides a comprehensive framework that captures these dynamics more effectively than other models.

CHAPTER 3: METHODOLOGY

3.1 ARIMA model

The ARIMA model, proposed by Box and Jenkins (1976) is an extension of the ARMA model. It can be denoted as ARIMA (p, d, q). ARIMA can be explained by three model parameters p, q, and d, where p is the highest order of the AR (Autoregressive) model, q is the highest order of the MA (Moving average) model, and d is the order of differencing.

An initial step in creating the ARIMA model is converting a non-stationary time series into a stationary one. After this transformation, the model makes predictions about the dependent variable based on its past values as well as the random error term's current and historical values.

Diagnostic tests such as the Jarque-Bera normality test on residual error series, the Bayes information criterion (BIC), and the Akaike information criterion (AIC) are used to find the best model.

Through this model, we aim to capture the main trends, patterns, autocorrelations, and seasonalities present in the time series

3.1.1 Autoregressive models (AR)

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

Where:

$Zt \sim wn(0, \sigma_t^2)$

X_t is the value of the time series at time t.

ϕ_0 is a constant term.

ϕ_p's are the coefficients of the model representing how the current value relates to previous values.

The past values of the series, $X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p}$ determine its current value. (Shumway and Stoffer, 2016)

The order of AR model is denoted as p.

3.1.2 Differencing process

Differencing is crucial when dealing with non-stationary data. When the statistical properties such as its variance, mean and autocovariance of a time series change over time, it is considered to be non-stationary. (Brooks, 2008)

A non-stationary time series can be made stationary by differencing. This involves subtracting the current observation from the previous observation.

$$(Y_t - Y_{t-1})$$

Where:

Yt is the current value of the time series at time, t

Y_{t-1} is the previous value at the time, t−1

The order of differencing is denoted as d.

3.1.3 Moving average model (MA)

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q}$$

Where:

X_t represents the value of the time series at time t

Z_t denotes the white noise error term at time t

$\theta_q' s$ are the parameters of the model, showing the effect of previous error terms on the current value

$Z_{t-1}, Z_{t-2}, \ldots, Z_{t-q}$ are the previous error terms up to lag q.

It is linear combination of white noise processes, such that X_t depends on the current and previous values of a white noise disturbance term. (Brooks, 2008)

The order of an MA model is denoted as q.

3.1.4 ARIMA

The ARIMA model is expressed as:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

This model indicates that the current value of the series Y_t is influenced by a linear combination of its past values, as well as the current and past values of a white noise error component (Brooks, 2008).

3.2 Building ARIMA Model

Box and Jenkens, (1970) developed a practical approach to estimate the suitable ARIMA model. This process includes three key steps: identification, estimation, and diagnostic checking.

3.2.1 Model identification

First, we check the stationarity of the data series by plotting the data to get a general idea of the data and identify a trend in the plotted time series data. Then, the Augmented Dickey-Fuller (ADF) test is employed to determine if the data is stationary. If not, the data is transformed to achieve stationarity through differencing. The autocorrelation function, ACF is used to determine the AR and MA terms.

3.2.2 Estimation

The parameters of the identified ARIMA model are estimated. The optimal ARIMA model for the mean is selected, and a GARCH (1,1) model is added to account for volatility. The ACF is used to identify the AR and MA terms. The model with the lowest Akaike Information Criterion (AIC) value is the most preferred.

3.2.3 Diagnosis checking

This involves assessing the adequacy of the specified and estimated. Box and Jenkins proposed two approaches: over-fitting and residual diagnostics.

Over-fitting involves deliberately fitting a larger model than is necessary to capture the dynamics of the data. If the specified model is adequate, any additional terms added to the ARMA model would be insignificant.

Residual diagnostics involve checking the residuals for evidence of linear dependence, which, if present, it suggests that the originally specified model was inadequate to capture the features of the data. ACF or Ljung-Box tests can be used. If the fitted model is an adequate model for the data, then the residuals should satisfy the assumption of a non-autocorrelated random stationary process that is normally and independently distributed. (Mustapah and Ismail, 2019).

3.3 GARCH Model

Generalized autoregressive conditional heteroskedasticity, GARCH introduced by Bollerslev (1986) is an extension of Autoregressive conditional heteroskedasticity by Engle (1982). In a GARCH (p,q) model, the conditional variance is expressed as a linear function of the squared past squared error terms and conditional variances. The model is specified as follows:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

where:

σ_t^2 denotes the conditional variance at time t

ω is a constant term

α's are the coefficients associated with the past squared error terms indicating the impact of past shocks on the current volatility

β's are the coefficients associated with past conditional variances that indicate the persistence of volatility over time,

p is the order of the GARCH terms (lagged conditional variances), and

q is the order of the ARCH terms (lagged squared residuals).

The GARCH (1,1) model is often used due to its simplicity and effectiveness in capturing the dynamic nature of financial market volatility.

3.4 Testing the ARCH effect

We then test for the effect of the ARCH using ARCH Lagrange Multiplier (LM) test proposed by Engle (1982). If there exists ARCH effect, it means the variance is heteroskedastic.

The hypothesis tested is:

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \cdots = \alpha_n = 0$

$H_a : \alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \cdots \neq \alpha_n \neq 0$

3.5 Data

To carry out this study, time series British American Tobacco stock price data from Yahoo finance, is used to perform the task. Weekly stock price data from January 2010 to December 2020, is used for the development of models. Data from January 2021 to December 2021 is used for the validation of the developed models. The data analysis is conducted using the R Software. Missing values will be addressed through mean imputation to ensure the quality of the dataset. This approach will ensure accurate and reliable modeling and validation.

CHAPTER 4: DATA ANALYSIS

4.1 Introduction

This chapter presents the results of the ARIMA and ARIMA-GARCH models, as outlined in the methodology. It evaluates model performance, diagnostic tests, and forecast accuracy to determine the optimal approach for forecasting BAT stock prices.

4.2 Empirical data

The dataset for this study consists of weekly stock prices for British American Tobacco (BAT) obtained from Yahoo Finance. The data spans the period from January 2010 to December 2020, with additional data from January 2021 to December 2021 reserved for model validation. Key variables include the opening price, closing price, high price, low price, adjusted close price (which accounts for dividends and stock splits), and trading volume. To ensure data quality, missing values were addressed using mean imputation. Preprocessing steps, including transformations to achieve stationarity, were performed prior to model development.

4.3 Descriptive statistics

The summary of the adjusted close price, which is the primary variable of interest are presented in the table below:

	Time series data	Obs	Nobs	Mean	Median	Std dev	Skewness	Kurtosis
BAT Stock prices	Adjusted close price	W	574	492.98	571.50	12.71	-0.54	-1.07

Obs=observations, W=weekly, Nobs=number of observations

4.4 ARIMA Model

Stationarity test

The stationarity of BAT prices was tested using the Augmented Dickey-Fuller (ADF) test.

H_0: Unit root is present

H_a: Unit root is not present

The results were as follows:

Test statistic: −0.40919, lag order: 8 and p-value: 0.98590

With a p-value greater than the 0.05 significance level, we fail to reject the null hypothesis. This indicates that the BAT prices are non-stationary.

Differencing

Since the series is non-stationary, it cannot be directly used in ARIMA modeling. To address this, differencing was applied to the data to achieve stationarity, as required for ARIMA modeling. The stationarity of the differenced series was verified using a subsequent ADF test.

The results were as follows:

Test statistic: −12.962, lag order: 8 and p-value: 0.010

With a p-value of 0.010, which is less than the significance level of 0.05, the null hypothesis is rejected. This indicates that the differenced series is stationary, making it suitable for ARIMA modeling.

Identification of ARIMA parameters

Autocorrelation (ACF) and partial autocorrelation (PACF) plots were used to identify ARIMA parameters.

ACF Plot

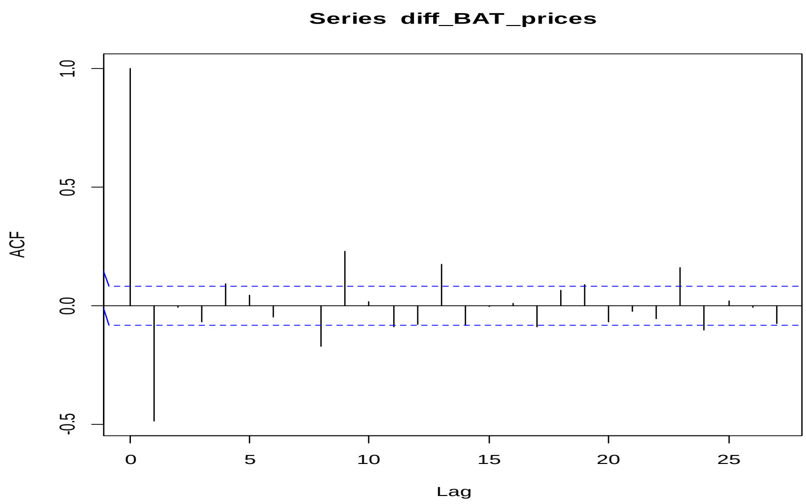


Figure 1:ACF plot

PACF Plot

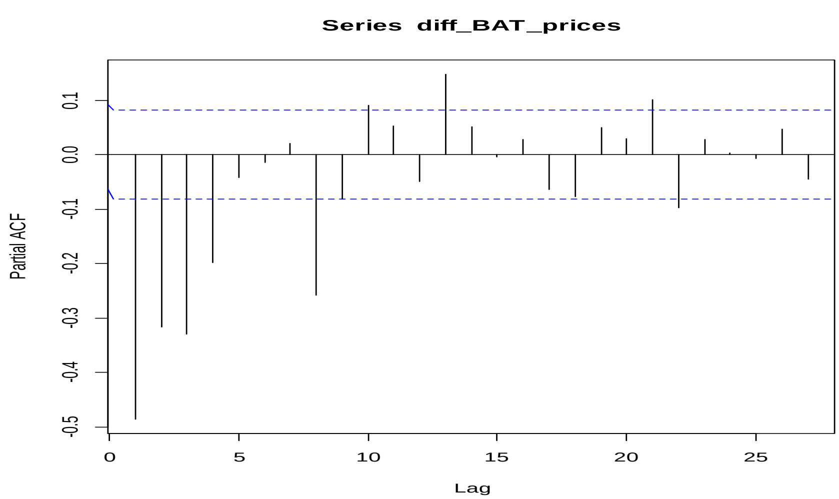


Figure 2:PACF plot

From these plots, the following ARIMA parameters were identified:

ARIMA (8,1,9):

The PACF plot shows a notable spike at lag 8, indicating a possible autoregressive component of order 8. The ACF plot's significant spike at lag 9 was retained as an MA component.

ARIMA (4,1,1):

The PACF plot has diminishing spikes around lag 4, suggesting a simpler autoregressive model of order 4 could be used. The ACF plot shows a quick decay after lag 1, indicating that a moving average term of order 1 might capture the short-term noise.

ARIMA (8,1,14):

The PACF plot has a spike at lag 8 and the ACF plot shows some autocorrelation persisting to higher lags (lag 14). Including this higher-order MA component could capture additional noise.

Model estimation

To identify the best-fitting model for forecasting stock prices, the Akaike Information Criterion (AIC) was used as the selection criterion. The AIC values for the estimated models were as follows:

Model	**AIC**
ARIMA (8, 1, 9)	6715.686
ARIMA (4, 1, 1)	6759.01
ARIMA (8, 1, 14)	6712.988

Based on the AIC values, the ARIMA model with the lowest AIC was selected as the best model. In this case, the ARIMA (8, 1, 14) was identified as the optimal model, as it minimizes the AIC, indicating better model fit.

Diagnosis testing

Diagnostic check was conducted on the residuals to evaluate the adequacy of the selected ARIMA model. The Ljung-Box test was used to test for autocorrelation in the residuals.

H_o: Residuals are uncorrelated

H_a: Residuals are correlated

The results were as follows:

Lag order: 25, p-value: 0.8557

The p-value from the Ljung-Box test is 0.8334, which is greater than the typical significance level of 0.05. This means that we fail to reject the null hypothesis that the residuals are uncorrelated. Therefore, the residuals do not exhibit significant autocorrelation, suggesting that the ARIMA model has adequately captured the temporal structure in the data. The model is therefore appropriate for forecasting stock prices.

Forecasting using ARIMA

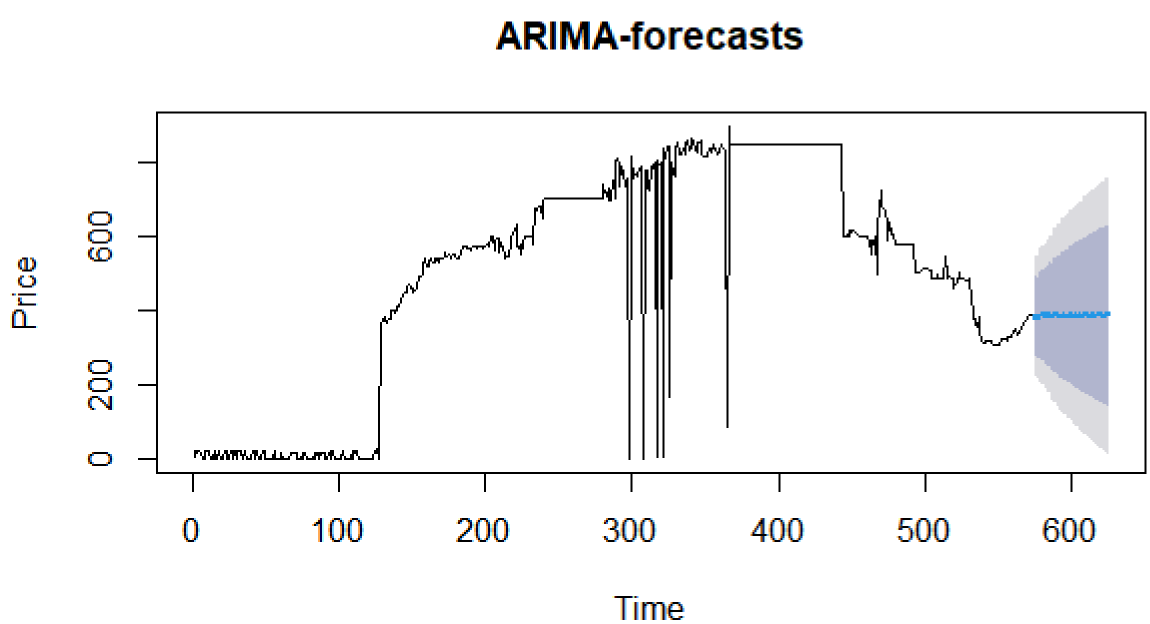


Figure 3:ARIMA forecasts

The forecast shows a rising trend in stock prices over the forecast period. This upward movement suggests an overall positive outlook based on historical data trends. However, the forecast is accompanied by some volatility, which is normal in stock price forecasting due to market fluctuations and other external factors.

4.5 Fitting GARCH (1,1) model

Model Specification

A GARCH (1,1) model was fitted to the residuals of the ARIMA(8, 1, 14) mean model to capture volatility clustering. The model assumes a standard GARCH process (sGARCH) with residuals following a normal distribution.

Residual diagnostics

Serial Correlation (Ljung-Box Test):

Squared residuals show no significant autocorrelation, indicating that the GARCH (1,1) model successfully captures volatility clustering.

ARCH Effects (ARCH LM Test):

The ARCH LM test yielded $p > 1$, confirming no remaining ARCH effects. This means that the GARCH (1,1) model adequately models conditional heteroskedasticity.

Model Interpretation

The GARCH (1,1) model effectively captures time-varying volatility in BAT stock prices, complementing the ARIMA mean model. By addressing both mean and variance dynamics, this combined approach provides a robust framework for analyzing and forecasting financial time series data.

Forecasting using ARIMA-GARCH

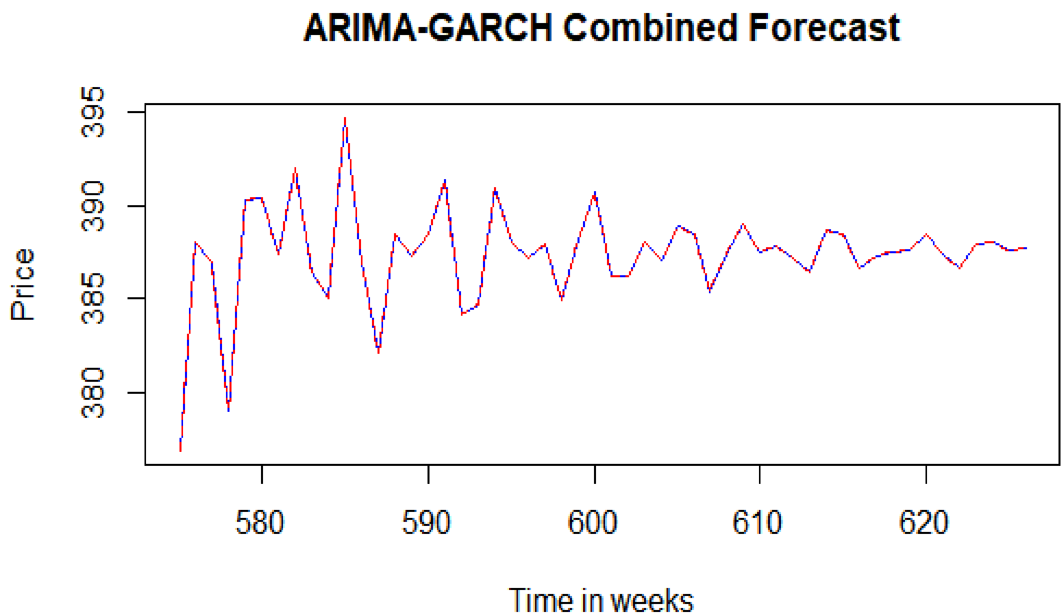


Figure 4:ARIMA-GARCH combined forecasts

The ARIMA-GARCH combined forecast illustrates fluctuations in predicted prices over time, reflecting the dynamics captured by the ARIMA model in the mean process and the GARCH model in the conditional variance. As the forecast progresses, the fluctuations appear to stabilize. This suggests a potential steadying trend in the stock prices over the forecast horizon. The GARCH component accounts for the conditional heteroskedasticity, providing a reliable prediction framework for financial data.

4.6 Model Comparison Using Theil's U-statistic

To evaluate the accuracy of the ARIMA and ARIMA-GARCH models, Theil's U-statistic was calculated for both models.

Theil's U-statistic is a measure of forecast accuracy, with lower values indicating better model performance. The calculated Theil's U-statistic for the ARIMA model was 0.1666. This value reflects the accuracy of the ARIMA model's forecast, comparing the squared errors of the forecast to the squared values of the forecasted means. The Theil's U-statistic for the ARIMA-GARCH model was 0.1628, which accounts for both the forecasted mean and volatility, capturing the conditional heteroskedasticity in the data.

Since the Theil's U-statistic for the ARIMA-GARCH model is lower than that for the ARIMA-only model, it suggests that the ARIMA-GARCH model provides a more accurate forecast. Therefore, the ARIMA-GARCH model is better than the ARIMA-only model based on Theil's U-statistic. This indicates that the ARIMA-GARCH model is more effective in capturing both the mean and volatility dynamics of the data, providing more reliable predictions compared to the ARIMA-only model.

CHAPTER 5: DISCUSSIONS AND CONCLUSIONS

5.1 Introduction

This chapter presents the conclusions drawn from the results of the study, limitations of the study and recommendations offering suggestions for further improvement in modeling and forecasting financial time series data.

5.2 Discussions

The primary aim of this study was to forecast stock prices for BAT using time series models, specifically ARIMA and ARIMA-GARCH models. The results of the study support the utility of combining ARIMA with a GARCH model to forecast financial data accurately. The key findings are summarized below.

The ARIMA (8,1,14) model demonstrated strong performance in analyzing the historical trends in British American Tobacco's (BAT) stock prices. The ARIMA model's forecast for BAT stock prices over the next 52 weeks predicts a gradual upward trend. This result aligns with BAT's historical stability, driven by its strong market presence and sustained consumer demand. However, as the forecast horizon extends, the widening confidence intervals emphasize the uncertainty associated with long-term predictions, reflecting potential impacts from regulatory shifts and evolving consumer behavior.

The ARIMA-GARCH model provided a more detailed and nuanced analysis of BAT's stock prices by accounting for both the underlying price trends and the observed volatility clustering, which is characteristic of financial market data. The GARCH (1,1) component effectively captured time-varying volatility, which is critical for understanding BAT's exposure to sector-specific risks such as regulatory interventions or changes in tobacco consumption patterns.

The ARIMA-GARCH model outperformed the standalone ARIMA model in forecasting accuracy, as demonstrated by its lower Theil's U-statistic. This highlights the importance of incorporating volatility dynamics, especially for a company like BAT that operates in a highly regulated and increasingly health-conscious market. The ARIMA-GARCH model's forecasts for BAT stock prices showed an overall upward trajectory, reflecting investor confidence in the company's ability to adapt to industry challenges. The model also identified periods of increased volatility, which were probably associated with significant events such as regulatory updates, financial reports, or changes in consumer preferences favoring reduced-risk products like e-cigarettes.

5.3 Limitations and Recommendations

This study has a few key limitations that offer valuable directions for future research. First, the ARIMA-GARCH model assumed that residuals followed a normal distribution, which can lead to biased results as financial data often exhibit non-normal characteristics like fat tails and skewness. Second, while the GARCH(1,1) model effectively captured volatility clustering, it failed to account for sudden volatility jumps commonly observed during periods of market crises or uncertainty. Lastly, the analysis did not include broader macroeconomic indicators, such as inflation or GDP growth, which may influence investor sentiment and market trends.

To address these limitations, future research should explore alternative distributions, such as the t-distribution, to better capture the non-normal features of financial data. Advanced GARCH models like Exponential GARCH (EGARCH) and Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) should also be considered to account for volatility asymmetries and sudden jumps. Additionally, integrating macroeconomic indicators into forecasting models could enhance their explanatory power and provide deeper insights into market dynamics.

Despite these limitations, the ARIMA-GARCH model proved effective in capturing time series dependencies and volatility clustering. Analysts and researchers are encouraged to use this model for forecasting in markets with volatility while adopting the suggested improvements for enhanced accuracy and robustness in future studies.

REFERENCES

Finance, Y. (2024, September 12). Yahoo finance. *Yahoo finance news*. https://finance.yahoo.com/news/invest-british-american-tobacco-bti-133012738.html

Ghigini, E. (2024, October 3). *Morgan Stanley downgrades British American Tobacco stock on weak US cigarette volumes*. Investing.com. https://www.investing.com/news/company-news/morgan-stanley-downgrades-british-american-tobacco-stock-on-weak-us-cigarette-volumes-93CH-3646527

Wright, L. (2024, May 10). *Understanding British American Tobacco's investment potential*. Investors' Hangout. https://investorshangout.com/understanding-british-american-tobaccos-investment-potential-59360-/

Mashadihasanli, T. (2022). Stock market price forecasting using the ARIMA model: An application to Istanbul, Turkiye. Journal of Economic Policy Researchers, 9(2), 439-454.

Tiwary, S., & Mishra, P. K. (2022). Time-series forecasting of stock prices using ARIMA: A case study of TESLA and NIO. In Proceedings of the 2nd Indian International Conference on Industrial Engineering and Operations Management (pp. 1517-1524).

Prabha, C. P., Kumar, K. S., & Kumar, S. S. (2022). ARIMA model in forecasting stock returns: A study with reference to BSE Sensex. International Journal of Novel Research and Development, 7(5), 809-816.

Ge, Y. (2023). PDD stock price prediction using ARIMA model, 83-87.

Khan, S., & Alghulaiakh, H. (2020). ARIMA model for accurate time series stocks forecasting. International Journal of Advanced Computer Science and Applications (IJACSA), 11(7), 524-528.

Grachev, O. Y. (2017). Application of time series models (ARIMA, GARCH, and ARMA-GARCH) for stock market forecasting (Honors Capstone No. 177). Northern Illinois University.

Shahid, M. A. F. (2023). Stock return prediction using ARIMA model: Evidence from South Asian markets (Master's thesis, Faculty of Management & Social Sciences, Department of Management Sciences).

Awalludin, S. A., et al. (2018). Modeling the stock price returns volatility using GARCH(1,1) in some Indonesia stock prices. *Journal of Physics: Conference Series, 948*, 012068.

Xiang, Y. (2022). Using ARIMA-GARCH model to analyze fluctuation law of international oil price. *Mathematical Problems in Engineering*.

Mbarek, B. K. (2015). Stock returns and forecast: Case of Tunisia with ARCH model. *Journal of Finance and Economics, 3*(3), 55-60.

Gao, J. (2022). *Research on stock price forecast based on ARIMA-GARCH model*. Institute of Economics, Shanghai University.

Liu, J. (2024). Navigating the financial landscape: The power and limitations of the ARIMA model. *Highlights in Science, Engineering and Technology, 88*, 747-752.

Mustapa, F., & Ismail, M. T. (2019). Modelling and forecasting S&P 500 stock prices using hybrid ARIMA-GARCH model. *Journal of Physics: Conference Series, 1366*, 012130.

Brooks, C. (2008). *Introductory econometrics for finance*. Cambridge University Press.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327.

Shumway, R., & Stoffer, D. (2011). *Time series analysis and its applications: With R examples*. Springer Science & Business Media.

APPENDIX

R codes

install.packages(“quantod”)

install.packages(“rugarch”)

install.packages(“forecast”)

install.packages(“FinTS”)

```
library(quantmod)

#Downloading the data
getSymbols("BAT", src = "yahoo",
        from = "2010-01-01",
        to = "2020-12-31",
        periodicity = "weekly")

## [1] "BAT"

#Viewing a sample of the data
head (BAT)

##            BAT.Open BAT.High  BAT.Low BAT.Close BAT.Volume BAT.Adjuste
d
## 2010-01-01 19.76820  20.7096 19.34460  20.42720     538673     20.42720
## 2010-01-08 20.14480  20.7096  2.64817   2.65363     310861      2.65363
## 2010-01-15  2.62087  20.1448  2.52258  19.86240      23507     19.86240
## 2010-01-22 20.09770  20.2389  2.33694  20.05060     513046     20.05060
## 2010-01-29 19.76820  20.0506  2.30141  19.76820      78185     19.76820
## 2010-02-05  2.05822  19.5329  2.02056  18.63860      25669     18.63860

tail(BAT)

##            BAT.Open BAT.High BAT.Low BAT.Close BAT.Volume BAT.Adjusted
## 2020-11-20      370   372.00     370    371.50     392200       371.50
## 2020-11-27      372   380.00     372    380.00      63200       380.00
## 2020-12-04      400   410.00     385    390.50      97800       390.50
## 2020-12-11      390   399.75     390    390.00      24400       390.00
## 2020-12-18      390   390.00     390    390.00      58900       390.00
## 2020-12-25      390   395.00     376    383.25       2900       383.25

#Saving the data
write.csv("BAT_stock_data.csv")

## "","x"
## "1","BAT_stock_data.csv"
```

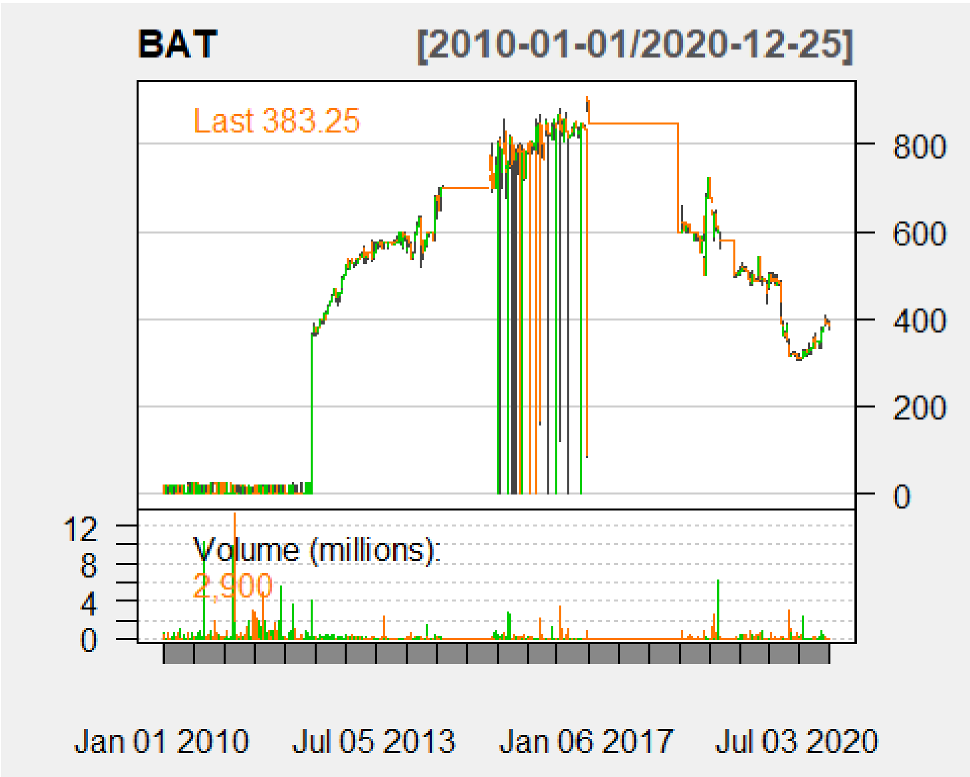
```
#Cleaning the data
#check for any missing values
sum(is.na(BAT))

## [1] 0

summary(BAT)

##      Index               BAT.Open         BAT.High          BAT.Low
##  Min.   :2010-01-01   Min.   :  0.117   Min.   :  0.4762   Min.   :  0.117
##  1st Qu.:2012-09-29   1st Qu.:318.250   1st Qu.:330.0000   1st Qu.: 21.164
##  Median :2015-06-29   Median :572.000   Median :580.0000   Median :541.500
##  Mean   :2015-06-29   Mean   :492.387   Mean   :507.7568   Mean   :465.143
##  3rd Qu.:2018-03-28   3rd Qu.:771.000   3rd Qu.:794.5000   3rd Qu.:702.000
##  Max.   :2020-12-25   Max.   :909.000   Max.   :910.0000   Max.   :878.000
##    BAT.Close         BAT.Volume        BAT.Adjusted
##  Min.   :  0.124   Min.   :       0   Min.   :  0.124
##  1st Qu.:319.250   1st Qu.:       0   1st Qu.:319.250
##  Median :571.500   Median :   32650   Median :571.500
##  Mean   :492.983   Mean   :  323374   Mean   :492.983
##  3rd Qu.:750.750   3rd Qu.:  211150   3rd Qu.:750.750
##  Max.   :900.000   Max.   :13241852   Max.   :900.000

#visualizing the data
chartSeries(BAT, theme="white")
```



```
#ARIMA model specification
library (forecast)
library(tseries)

# Checking the structure of the BATdata
str(BAT)
```

```
## An xts object on 2010-01-01 / 2020-12-25 containing:
##   Data:    double [574, 6]
##   Columns: BAT.Open, BAT.High, BAT.Low, BAT.Close, BAT.Volume ... with
1 more column
##   Index:   Date [574] (TZ: "UTC")
##   xts Attributes:
##     $ src    : chr "yahoo"
##     $ updated: POSIXct[1:1], format: "2024-12-06 09:15:24"
```

```
# Extracting the adjusted closing prices
BAT_prices <- BAT[, "BAT.Adjusted"]
BAT_prices
```

```
library(psych)
describe(BAT_prices)
```

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	574	492.98	304.6	571.5	509.42	328.95	0.12	900	899.88	-0.54	-1.07	12.71

```
##            BAT.Adjusted
## 2010-01-01     20.42720
## 2010-01-08      2. 65363
## 2010-01-15     19.86240
## 2010-01-22     20.05060
## 2010-01-29     19.76820
## 2010-02-05     18.63860
## 2010-02-12      2.06393
## 2010-02-19      1.99218
## 2010-02-26     18.91160
## 2010-03-05     19.01520
##        ...
## 2020-10-23    342.00000
## 2020-10-30    350.00000
## 2020-11-06    364.00000
## 2020-11-13    369.75000
## 2020-11-20    371.50000
## 2020-11-27    380.00000
## 2020-12-04    390.50000
## 2020-12-11    390.00000
## 2020-12-18    390.00000
## 2020-12-25    383.25000
```

```
# Making BAT_prices a numeric vector
BAT_prices <- as.numeric(BAT_prices)
head(BAT_prices)
```

```
## [1] 20.42720  2.65363 19.86240 20.05060 19.76820 18.63860
```

```
#checking for stationarity using ADF test
adf_BAT_prices <- adf.test(BAT_prices)
adf_BAT_prices
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  BAT_prices
## Dickey-Fuller = -0.40919, Lag order = 8, p-value = 0.9859
## alternative hypothesis: stationary
```

```
#data is non stationary so we apply differencing
#to transform the series into a stationary one
diff_BAT_prices <- diff(BAT_prices)
head(diff_BAT_prices)
```

```
## [1] -17.7735703  17.2087700   0.1882000  -0.2824001  -1.1296005 -16.5746694
```

```
tail(diff_BAT_prices)
```

```
## [1]  1.75  8.50 10.50 -0.50  0.00 -6.75
```

```
# Run the ADF test on the differenced series
adf_diff_BAT_prices <- adf.test(diff_BAT_prices)
```

```
## Warning in adf.test(diff_BAT_prices): p-value smaller than printed p-value
```

```
adf_diff_BAT_prices
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff_BAT_prices
## Dickey-Fuller = -12.962, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```

```
#Identifying the ARIMA parameters
dev.new()  # Opens a new graphics window
acf(diff_BAT_prices)
dev.new()  # Opens another new graphics window for PACF
pacf(diff_BAT_prices)
```

```
#fitting the ARIMA model
```

```
arima_model1 <- arima(BAT_prices, order = c(8, 1, 9))
arima_model2 <- arima(BAT_prices, order = c(4, 1, 1))
arima_model3 <- arima(BAT_prices, order = c(8, 1, 14))

AIC(arima_model1)

## [1] 6715.686

AIC(arima_model2)

## [1] 6759.01

AIC(arima_model3)

## [1] 6712.988

best_arima_model=arima(BAT_prices,order=c(8,1,14))
best_arima_model

##
## Call:
## arima(x = BAT_prices, order = c(8, 1, 14))
##
## Coefficients:
##          ar1      ar2      ar3      ar4      ar5     ar6      ar7      ar8
##      -0.3367  -0.8552  -0.4653  -0.5584  -0.1927  0.0537  -0.2944  -0.0294
## s.e.   0.3591   0.3078   0.4754   0.4285   0.4004  0.2906   0.1356   0.2113
##          ma1     ma2      ma3     ma4      ma5      ma6     ma7      ma8
##      -0.4906  0.6426  -0.2402  0.3339  -0.2315  -0.2222  0.3661  -0.2940
## s.e.   0.3565  0.2307   0.3093  0.1902   0.2047   0.1736  0.2209   0.2594
##         ma9     ma10   ma11     ma12    ma13     ma14
##      0.2292  -0.1099  0.125  -0.0086  0.1883  -0.1833
## s.e.  0.2013   0.1055  0.077   0.0759  0.0682   0.0807
##
## sigma^2 estimated as 6575:  log likelihood = -3333.49,  aic = 6712.99

summary(best_arima_model)

##
## Call:
## arima(x = BAT_prices, order = c(8, 1, 14))
##
## Coefficients:
##          ar1      ar2      ar3      ar4      ar5     ar6      ar7      ar8
##      -0.3367  -0.8552  -0.4653  -0.5584  -0.1927  0.0537  -0.2944  -0.0294
## s.e.   0.3591   0.3078   0.4754   0.4285   0.4004  0.2906   0.1356   0.2113
##          ma1     ma2      ma3     ma4      ma5      ma6     ma7      ma8
##      -0.4906  0.6426  -0.2402  0.3339  -0.2315  -0.2222  0.3661  -0.2940
```

```
## s.e.   0.3565  0.2307   0.3093  0.1902   0.2047   0.1736  0.2209   0.2594
##          ma9     ma10   ma11     ma12    ma13     ma14
##       0.2292  -0.1099  0.125  -0.0086  0.1883  -0.1833
## s.e.  0.2013   0.1055  0.077   0.0759  0.0682   0.0807
##
## sigma^2 estimated as 6575:  log likelihood = -3333.49,  aic = 6712.99
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE     MASE       ACF1
## Training set 2.152327 81.01769 33.03997 -2190.627 2210.338 1.260243 -0.0022
7606
```

```
AIC(best_arima_model)
```

```
## [1] 6712.988
```

```
#diagnostic checking
#ljung_box test
residuals= residuals(best_arima_model)
head(residuals)
```

```
## Time Series:
## Start = 1
## End = 6
## Frequency = 1
## [1]   0.02042718 -12.67337579   6.93277484   4.93178449   2.85519886
## [6]   3.17034758
```

```
tail(residuals)
```

```
## Time Series:
## Start = 569
## End = 574
## Frequency = 1
## [1] 20.811435 13.229716 18.718464 27.704774 10.400786  2.124751
```

```
Box.test(residuals,lag=25,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  residuals
## X-squared = 17.68, df = 25, p-value = 0.8557
```

```
#Forecast for the next 1 year
library(quantmod)
library(forecast)

# Get actual data (closing prices) for 2021
```

```
actualdata_2021 <- Cl(getSymbols("BAT", src = "yahoo",
                                 from = "2021-01-01",
                                 to = "2021-12-31",
                                 periodicity = "weekly",
                                 auto.assign = FALSE))

# Assuming 'best_arima_model' is the fitted ARIMA model
# Forecasting for the next 1 year (52 weeks)
forecasted_values <- forecast(best_arima_model, h = 52)
forecasted_values

# Plot forecasted values
plot(forecasted_values)

# Convert actual data to numeric (closing prices)
numeric_actualdata <- as.numeric(actualdata_2021)

# Checking if there are any NA values in the actual data
sum(is.na(actualdata_2021))

## [1] 0

#fitting a GARCH (1,1) model
library(rugarch)

#extract the residuals of arima
residuals_arima <- residuals(best_arima_model)
residuals_arima

#Specify and fit the GARCH(1,1) model to ARIMA residuals
garch_spec <- ugarchspec(variance.model = list(model = "sGARCH",
                         garchOrder = c(1, 1)), mean.model = list(armaOrder = c(8, 14),
                          include.mean = FALSE),
                         distribution.model = "norm")

garch_spec

##
## *---------------------------------*
## *       GARCH Model Spec          *
## *---------------------------------*
##
## Conditional Variance Dynamics
## ------------------------------------
## GARCH Model      : sGARCH(1,1)
## Variance Targeting   : FALSE
##
## Conditional Mean Dynamics
```

```
## -----------------------------------
## Mean Model      : ARFIMA(8,0,14)
## Include Mean    : FALSE
## GARCH-in-Mean       : FALSE
##
## Conditional Distribution
## -----------------------------------
## Distribution :  norm
## Includes Skew    :  FALSE
## Includes Shape   :  FALSE
## Includes Lambda  :  FALSE
```

```
garch_fit <- ugarchfit(spec = garch_spec, data = residuals_arima, solver = "hybrid")
summary(garch_fit)
```

```
##    Length     Class      Mode
##         1 uGARCHfit        S4
```

```
garch_fit
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(8,0,14)
## Distribution : norm
##
## Optimal Parameters
## -----------------------------------
##         Estimate  Std. Error   t value Pr(>|t|)
## ar1    -1.025201    0.002642 -388.0155 0.000000
## ar2    -0.115383    0.022981   -5.0209 0.000001
## ar3     0.058217    0.027084    2.1495 0.031592
## ar4    -0.787447    0.006876 -114.5246 0.000000
## ar5    -1.015586    0.007781 -130.5243 0.000000
## ar6    -0.435593    0.015311  -28.4488 0.000000
## ar7    -0.199947    0.022548   -8.8674 0.000000
## ar8     0.030969    0.005587    5.5429 0.000000
## ma1     1.551394    0.000746 2079.7328 0.000000
## ma2     1.032553    0.001206  855.9829 0.000000
## ma3     0.707771    0.000121 5847.5342 0.000000
## ma4     1.142409    0.001286  888.3407 0.000000
## ma5     1.875168    0.001819 1030.6085 0.000000
```

```
## ma6     2.029842    0.003979  510.0815 0.000000
## ma7     1.804103    0.001729 1043.5762 0.000000
## ma8     1.083873    0.000720 1505.3575 0.000000
## ma9     0.627700    0.001523  412.0461 0.000000
## ma10    0.673415    0.001019  660.5483 0.000000
## ma11    0.731745    0.001526  479.4400 0.000000
## ma12    0.746861    0.000652 1145.4648 0.000000
## ma13    0.312071    0.002072  150.6022 0.000000
## ma14   -0.022349    0.004322   -5.1713 0.000000
## omega  30.668959    9.286959    3.3024 0.000959
## alpha1  1.000000    0.157489    6.3497 0.000000
## beta1   0.485628    0.078236    6.2072 0.000000
##
## Robust Standard Errors:
##         Estimate  Std. Error   t value Pr(>|t|)
## ar1    -1.025201    0.023868 -42.95362  0.00000
## ar2    -0.115383    0.135213  -0.85334  0.39347
## ar3     0.058217    0.113033   0.51505  0.60652
## ar4    -0.787447    0.054033 -14.57343  0.00000
## ar5    -1.015586    0.027617 -36.77436  0.00000
## ar6    -0.435593    0.059964  -7.26428  0.00000
## ar7    -0.199947    0.159586  -1.25291  0.21024
## ar8     0.030969    0.100048   0.30954  0.75691
## ma1     1.551394    0.004670 332.23046  0.00000
## ma2     1.032553    0.009566 107.93584  0.00000
## ma3     0.707771    0.006619 106.92808  0.00000
## ma4     1.142409    0.004840 236.03755  0.00000
## ma5     1.875168    0.015612 120.11012  0.00000
## ma6     2.029842    0.035337  57.44275  0.00000
## ma7     1.804103    0.015035 119.99157  0.00000
## ma8     1.083873    0.010859  99.81503  0.00000
## ma9     0.627700    0.007116  88.20628  0.00000
## ma10    0.673415    0.019190  35.09110  0.00000
## ma11    0.731745    0.009270  78.93707  0.00000
## ma12    0.746861    0.003163 236.14493  0.00000
## ma13    0.312071    0.024768  12.59995  0.00000
## ma14   -0.022349    0.046245  -0.48328  0.62890
## omega  30.668959   43.804602   0.70013  0.48385
## alpha1  1.000000    1.411659   0.70839  0.47870
## beta1   0.485628    0.691860   0.70192  0.48273
##
## LogLikelihood : -2922.849
##
## Information Criteria
## ------------------------------------
##
```

```
## Akaike       10.271
## Bayes        10.461
## Shibata      10.268
## Hannan-Quinn 10.345
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                          statistic p-value
## Lag[1]                       1.589  0.2075
## Lag[2*(p+q)+(p+q)-1][65]   109.710  0.0000
## Lag[4*(p+q)+(p+q)-1][109]  147.854  0.0000
## d.o.f=22
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                  0.0004351  0.9834
## Lag[2*(p+q)+(p+q)-1][5] 0.4176460  0.9696
## Lag[4*(p+q)+(p+q)-1][9] 0.6880978  0.9957
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]    0.3265 0.500 2.000  0.5677
## ARCH Lag[5]    0.4902 1.440 1.667  0.8865
## ARCH Lag[7]    0.6226 2.315 1.543  0.9660
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  no.parameters>20 (not available)
## Individual Statistics:
## ar1    0.05873
## ar2    0.07531
## ar3    0.07396
## ar4    0.08766
## ar5    0.08415
## ar6    0.04730
## ar7    0.07810
## ar8    0.13338
## ma1    0.14429
## ma2    0.22575
## ma3    0.02427
## ma4    0.03714
## ma5    0.14820
```

```
## ma6    0.25009
## ma7    0.03101
## ma8    0.04633
## ma9    0.29288
## ma10   0.26267
## ma11   0.03370
## ma12   0.03593
## ma13   0.23115
## ma14   0.10785
## omega  0.10804
## alpha1 0.37992
## beta1  0.09973
##
## Asymptotic Critical Values (10% 5% 1%)
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias          0.41022 0.6818
## Negative Sign Bias 0.01506 0.9880
## Positive Sign Bias 0.01465 0.9883
## Joint Effect       0.18943 0.9793
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     91.23    2.010e-11
## 2    30     94.78    6.575e-09
## 3    40    104.19    7.544e-08
## 4    50    108.93    1.908e-06
##
##
## Elapsed time : 19.79966
```

```
garch_residuals <- as.numeric(residuals(garch_fit, standardize = TRUE))
garch_residuals
```

```
#Test for heteroskedasticity
library(FinTS)
```

```
length(garch_residuals)
```

```
## [1] 574
```

```
ArchTest(garch_residuals, lag = 24)
```

```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  garch_residuals
## Chi-squared = 4.6749, df = 24, p-value = 1
```

```
#Forecast using ARIMA and GARCH
# Forecast with ARIMA for 52 weeks
arima_forecast <- forecast(best_arima_model, h = 52)
forecasted_mean <- arima_forecast$mean

# Forecast volatility with the GARCH model
garch_forecast <- ugarchforecast(garch_fit, n.ahead = 52)
garch_forecast
```

```
##
## *------------------------------------*
## *       GARCH Model Forecast         *
## *------------------------------------*
## Model: sGARCH
## Horizon: 52
## Roll Steps: 0
## Out of Sample: 0
##
## 0-roll forecast [T0=0574-01-01]:
##         Series    Sigma
## T+1   16.93062    21.10
## T+2   27.80630    26.31
## T+3   17.85143    32.54
## T+4   22.08829    40.04
## T+5   -4.92502    49.12
## T+6  -22.94897    60.13
## T+7   -7.13451    73.50
## T+8  -13.81509    89.75
## T+9    6.21451   109.54
## T+10  16.91908   133.62
## T+11   0.57556   162.96
## T+12  16.67626   198.71
## T+13  -1.84663   242.26
## T+14 -12.70931   295.33
## T+15  -3.58815   360.02
## T+16 -17.71894   438.84
## T+17  -1.08363   534.92
## T+18   7.97463   652.02
## T+19   4.13858   794.74
## T+20  18.79245   968.70
## T+21   3.61607  1180.72
```

```
## T+22  -2.77157   1439.15
## T+23  -3.93586   1754.13
## T+24 -18.24154   2138.06
## T+25  -6.36957   2606.01
## T+26  -1.85064   3176.37
## T+27   2.27984   3871.56
## T+28  16.93310   4718.91
## T+29   8.19150   5751.71
## T+30   6.35415   7010.55
## T+31  -0.08942   8544.91
## T+32 -14.29910  10415.08
## T+33  -9.42800  12694.56
## T+34  -9.90156  15472.94
## T+35  -2.85957  18859.40
## T+36  10.99454  22987.03
## T+37   9.45003  28018.06
## T+38  12.69189  34150.19
## T+39   5.80865  41624.41
## T+40  -6.86747  50734.48
## T+41  -8.56467  61838.40
## T+42 -14.20447  75372.57
## T+43  -8.71611  91868.87
## T+44   2.50717 111975.61
## T+45   6.55186 136482.99
## T+46  14.59279 166354.13
## T+47  11.00250 202762.97
## T+48   1.98276 247140.38
## T+49  -3.78659 301230.38
## T+50 -13.64049 367158.70
## T+51 -12.57269 447516.33
## T+52  -6.08122 545461.31
```

```
forecasted_volatility= as.numeric(sigma(garch_forecast))
forecasted_volatility
```

```
##  [1]     21.09827     26.30545     32.53750     40.04357     49.12082
##  [6]     60.12715     73.49573     89.75227    109.53585    133.62406
## [11]    162.96361    198.70758    242.26071    295.33466    360.01523
## [16]    438.84435    534.92002    652.01799    794.74021    968.69564
## [21]   1180.72072   1439.14801   1754.13369   2138.05665   2606.00504
## [26]   3176.36918   3871.56433   4718.91160   5751.71165   7010.55342
## [31]   8544.90929  10415.07929  12694.56092  15472.93761  18859.39929
## [36]  22987.03359  28018.05614  34150.18567  41624.41429  50734.47847
## [41]  61838.40289  75372.57073  91868.87351 111975.61439 136482.98638
## [46] 166354.12691 202762.96898 247140.37668 301230.37794 367158.70473
## [51] 447516.33411 545461.31334
```

```
# Combine ARIMA forecasted mean and GARCH volatility
forecast_combined <- data.frame(
  Forecasted_Mean = arima_forecast$mean,
  Forecasted_Volatility = forecasted_volatility
)
forecast_combined
```

```
##    Forecasted_Mean Forecasted_Volatility
## 1         376.8396              21.09827
## 2         388.0556              26.30545
## 3         386.9058              32.53750
## 4         378.9894              40.04357
## 5         390.3328              49.12082
## 6         390.3720              60.12715
## 7         387.3434              73.49573
## 8         392.0434              89.75227
## 9         386.3707             109.53585
## 10        384.9740             133.62406
## 11        394.7279             162.96361
## 12        386.9928             198.70758
## 13        382.1524             242.26071
## 14        388.4857             295.33466
## 15        387.3150             360.01523
## 16        388.4415             438.84435
## 17        391.4121             534.92002
## 18        384.1433             652.01799
## 19        384.6897             794.74021
## 20        390.9295             968.69564
## 21        388.0822            1180.72072
## 22        387.1556            1439.14801
## 23        387.9573            1754.13369
## 24        384.9168            2138.05665
## 25        388.1556            2606.00504
## 26        390.7466            3176.36918
## 27        386.2437            3871.56433
## 28        386.1853            4718.91160
## 29        388.0275            5751.71165
## 30        387.1092            7010.55342
## 31        388.9308            8544.90929
## 32        388.4211           10415.07929
## 33        385.3447           12694.56092
## 34        387.3731           15472.93761
## 35        388.9667           18859.39929
## 36        387.4704           22987.03359
## 37        387.7976           28018.05614
## 38        387.1491           34150.18567
## 39        386.4343           41624.41429
```

```
## 40      388.6355      50734.47847
## 41      388.4919      61838.40289
## 42      386.6802      75372.57073
## 43      387.3242      91868.87351
## 44      387.5450     111975.61439
## 45      387.5618     136482.98638
## 46      388.4548     166354.12691
## 47      387.3918     202762.96898
## 48      386.6110     247140.37668
## 49      387.8878     301230.37794
## 50      387.9939     367158.70473
## 51      387.5679     447516.33411
## 52      387.7039     545461.31334
```

```
# Assuming you want a 95% confidence interval (z-score of 1.96 at 95% confidence)
# Calculate upper and lower bounds using GARCH volatility
forecast_combined$Upper_Bound <- forecast_combined$Forecasted_Mean
+ 1.96 * forecasted_volatility
```

```
##  [1] 4.135260e+01 5.155868e+01 6.377351e+01 7.848539e+01 9.627681e+01
##  [6] 1.178492e+02 1.440516e+02 1.759145e+02 2.146903e+02 2.619032e+02
## [11] 3.194087e+02 3.894668e+02 4.748310e+02 5.788559e+02 7.056298e+02
## [16] 8.601349e+02 1.048443e+03 1.277955e+03 1.557691e+03 1.898643e+03
## [21] 2.314213e+03 2.820730e+03 3.438102e+03 4.190591e+03 5.107770e+03
## [26] 6.225684e+03 7.588266e+03 9.249067e+03 1.127335e+04 1.374068e+04
## [31] 1.674802e+04 2.041356e+04 2.488134e+04 3.032696e+04 3.696442e+04
## [36] 4.505459e+04 5.491539e+04 6.693436e+04 8.158385e+04 9.943958e+04
## [41] 1.212033e+05 1.477302e+05 1.800630e+05 2.194722e+05 2.675067e+05
## [46] 3.260541e+05 3.974154e+05 4.843951e+05 5.904115e+05 7.196311e+05
## [51] 8.771320e+05 1.069104e+06
```

```
forecast_combined$Lower_Bound <- forecast_combined$Forecasted_Mean
- 1.96 * forecasted_volatility
```

```
##  [1] -4.135260e+01 -5.155868e+01 -6.377351e+01 -7.848539e+01 -9.627681e+0
1
##  [6] -1.178492e+02 -1.440516e+02 -1.759145e+02 -2.146903e+02 -2.619032e+0
2
## [11] -3.194087e+02 -3.894668e+02 -4.748310e+02 -5.788559e+02 -7.056298e+
02
## [16] -8.601349e+02 -1.048443e+03 -1.277955e+03 -1.557691e+03 -1.898643e+
03
## [21] -2.314213e+03 -2.820730e+03 -3.438102e+03 -4.190591e+03 -5.107770e+
03
## [26] -6.225684e+03 -7.588266e+03 -9.249067e+03 -1.127335e+04 -1.374068e+
04
## [31] -1.674802e+04 -2.041356e+04 -2.488134e+04 -3.032696e+04 -3.696442e+
```

```
04
## [36] -4.505459e+04 -5.491539e+04 -6.693436e+04 -8.158385e+04 -9.943958e+
04
## [41] -1.212033e+05 -1.477302e+05 -1.800630e+05 -2.194722e+05 -2.675067e+
05
## [46] -3.260541e+05 -3.974154e+05 -4.843951e+05 -5.904115e+05 -7.196311e+
05
## [51] -8.771320e+05 -1.069104e+06
```

```
# Plot the forecasted mean along with prediction intervals
plot(forecast_combined$Forecasted_Mean, type = "l", col = "blue",
    ylim = range(c(forecast_combined$Upper_Bound, forecast_combined$Lower_B
ound)),
    main = "ARIMA-GARCH Combined Forecast", xlab = "Time", ylab = "Price")
lines(forecast_combined$Upper_Bound, col = "red", lty = 2)
lines(forecast_combined$Lower_Bound, col = "red", lty = 2)
lines(numeric_actualdata, col = "black", lty = 1)

#Compute Theil's U-statistic for ARIMA model
arima_errors <- numeric_actualdata[(length(numeric_actualdata)-51):length(numeri
c_actualdata)] - arima_forecast$mean
arima_squared_errors <- arima_errors^2
arima_squared_forecasts <- arima_forecast$mean^2
U_arima <- sqrt(mean(arima_squared_errors / arima_squared_forecasts))
U_arima
```

```
## [1] 0.1666025
```

```
#Compute Theil's U-statistic for ARIMA-GARCH model
garch_errors <- numeric_actualdata[(length(numeric_actualdata)-51):length(numeri
c_actualdata)] - forecast_combined$Forecasted_Mean
garch_squared_errors <- garch_errors^2
garch_squared_forecasts <- forecast_combined$Forecasted_Mean^2 + forecast_com
bined$Forecasted_Volatility^2
U_arima_garch <- sqrt(mean(garch_squared_errors / garch_squared_forecasts))
U_arima_garch
```

```
## [1] 0.09248346
```

```
if (U_arima_garch < U_arima) {
  print("ARIMA-GARCH model is better than ARIMA-only model based on Theil's
U-statistic.")
} else {
  print("ARIMA-only model is better than ARIMA-GARCH model based on Theil's
U-statistic.")
}
```

```
## [1] "ARIMA-GARCH model is better than ARIMA-only model based on Theil's U-statistic."
```