NANYANG TECHNOLOGICAL UNIVERSITY
AI6123-TIME SERIES ANALYSIS
Assignment 3 Individual
NTAMBARA Etienne G2304253K
May 3, 2024

NTAMBARA Etienne G2304253K [1 2 3] May 3, 2024

1. Introduction

In this project, we aim to analyze a dataset consisting of daily historical stock prices of Apple Inc., sourced from Yahoo Finance, covering the period from **2002-02-01** to **2017-01-31**. The dataset encapsulates critical trading metrics such as open, high, low, close, and adjusted close prices recorded daily over fifteen years. With a total of 3775 attributes, the primary objective is to identify significant trends in Apple stock prices during this period. Additionally, the project seeks to develop and optimize a predictive model capable of forecasting future stock price movements based on historical data. This comprehensive analysis will not only enhance understanding of past stock behaviours but also improve financial decision-making processes based on robust model predictions (yah).

2. Project Implementation

2.1. Original Dataset Plot

Figure 1. Original Dataset Plot Results

This (Figure 1) represents the adjusted closing price of Apple Inc. (AAPL) stock over 15 years from February 2002 to January 2017. The price is plotted on the vertical axis in U.S. dollars, with time progressing on the horizontal axis, showcasing the stock's overall upward trend with some volatility clustering and significant price increases, particularly noticeable after 2009. The time series data points demonstrate the market's changing valuation of Apple's stock, reflecting investor reactions to company performance, market conditions, and economic events over the years.

2.2. Statistical Summary of Original Data set

Table 1. Summary Statistics of AAPL Closing Prices

Statistic	AAPL.Close
Min. Date	2002-02-01
Min. Close Price (USD)	0.2343
1st Qu. Date	2005-10-29
1st Qu. Close Price (USD)	1.9632
Median Date	2009-07-31
Median Close Price (USD)	6.6204
Mean Date	2009-07-31
Mean Close Price (USD)	10.7746
3rd Qu. Date	2013-05-01
3rd Qu. Close Price (USD)	19.1407
Max. Date	2017-01-30
Max. Close Price (USD)	33.2500

An overview of the closing prices of Apple Inc.'s shares from 2002-02-01 to 2017-01-31 is shown statistically in Table 1. The closing price of the stock varied from $0.2343 at the lowest to $33.2500 at the highest. 25% of the closing prices, according to the first quartile ($1.9632), were below this value, and 75% of the prices, according to the third quartile ($19.1407), were below this higher figure. The dataset's mean (average) price was slightly higher at $10.7746 than the median closing price of $6.6204, which could indicate a bias in the distribution toward higher values.

2.3. Histogram of original data

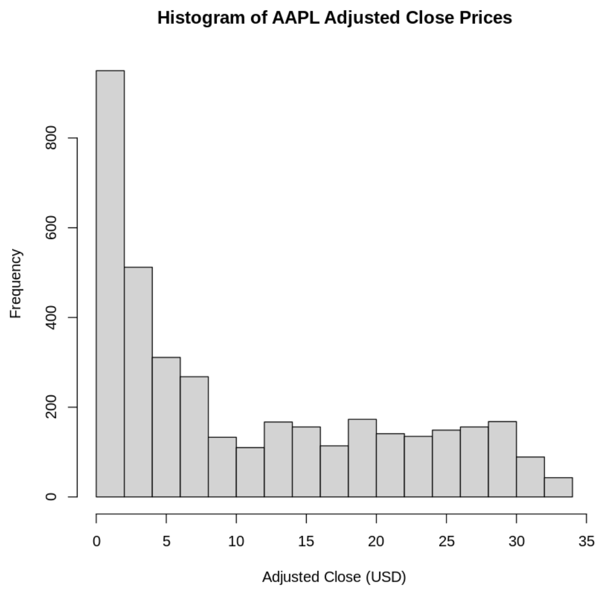

Figure 2. Histogram of original Data

The distribution of Apple Inc.'s adjusted closing stock prices, in U.S. dollars, throughout the given period is shown by the histogram from (Figure 2). It displays the most common price range, which is between \$0 and \$5, meaning that the stock closed many trading days inside this range. Higher closing prices may not have been as prevalent during the observed period as the frequency declines as the price rises.

2.4. Plotting ACF and PACF From Original Dataset

Augmented Dickey-Fuller Test

data: adj_close Dickey-Fuller = -2.3884, Lag order = 15, p-value = 0.4139 alternative hypothesis: stationary

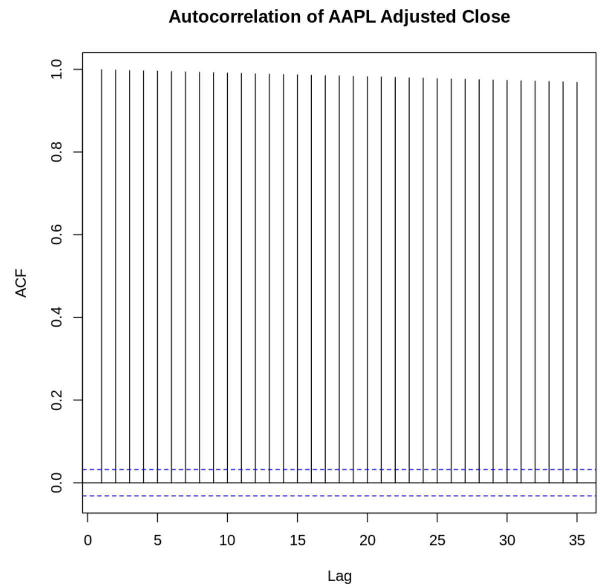

Figure 3. ACF

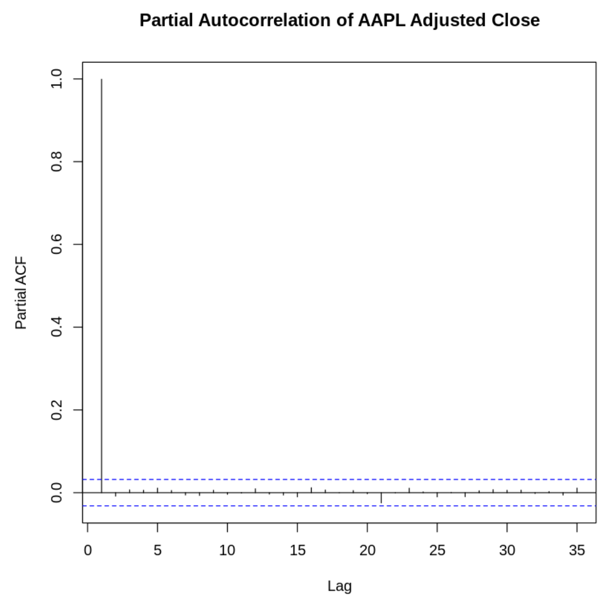

Figure 4. PACF

The adjusted close prices of AAPL show a progressive reduction in correlation on the autocorrelation function (ACF) plot shown in (Figure 3), indicating non-stationarity with possible trend or seasonality aspects. Beyond the first few lags, the partial autocorrelation function (PACF) plot shown in (Figure 4) reveals no significant correlations, suggesting that previous values have minimal bearing on subsequent values. The non-stationary nature of the series is confirmed by the Augmented Dickey-Fuller (ADF) test, which returns a p-value of 0.4139, which is not low enough to reject the null hypothesis of a unit root. The combination of the ACF plot and the high p-value indicates that the stock prices most likely follow a random walk.

2.5. Seasonal Decomposition

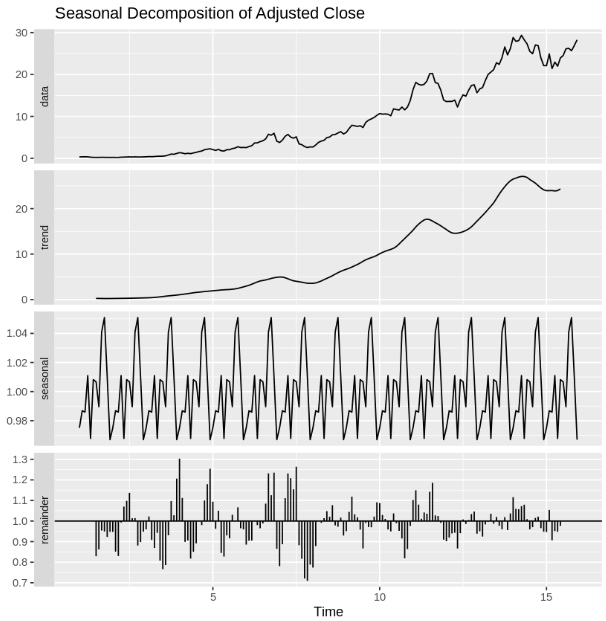

Figure 5. Seasonal Decomposition

Figure 5 displays the results of a Classical Decomposition on Apple's monthly stock data. To capture annual seasonality, the original daily data is transformed into a monthly time series at a predetermined frequency. This algorithm displays the underlying patterns and growth trends by multiplying the time series into trend, seasonal, and residual components. The trend component highlights growth that is not steady but fluctuates over time by showing the stock price of Apple as it gradually rises. Regular patterns that occur throughout the year are captured by the seasonal component, which shows predictable swings that might be balanced by logging or using a Box-Cox correction, among other treatments.

2.6. Dataset Transformation

2.7. Dataset Splitting (Train-Test Sets)

In total, we have 3775 observations where the train-test split method, the transformed time series data is divided into two parts: the 'train_set', which consists of the initial segment minus the last 30 observations which remain with 3745 observations, and the 'test_set', which comprises the final 30 observations. The model is trained on the 'train_set' to learn the patterns and relationships within the historical data. The 'test_set' is then used to evaluate the model's forecasting accuracy, effectively testing how well the model can predict future data points. This split is crucial for validating the model's performance and its ability to generalize to new, unseen data.

2.7.1. Log Returns

The logarithm of the modified closing prices is divided to determine the log returns. When prices show exponential development, this transformation stabilizes volatility and standardizes returns for comparison across various investments.

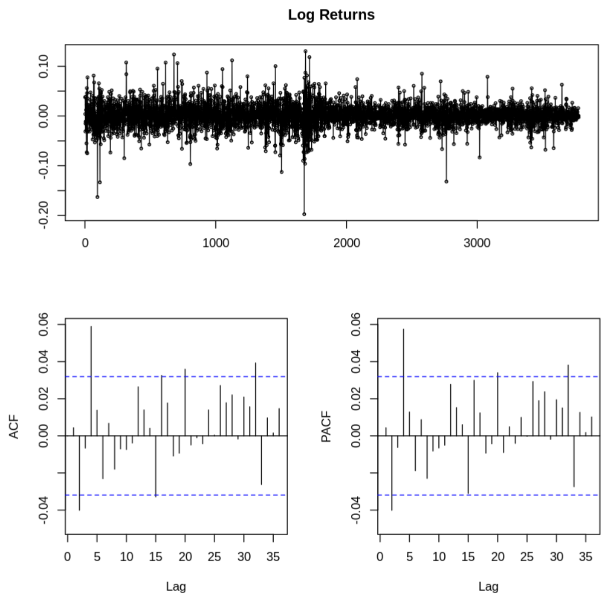

Figure 6. Log Returns

The log returns of adjusted closing prices are displayed in the "Log Returns" in (Figure 6), which shows how stock returns vary over time. The comparable ACF and PACF plots that follow show no evidence of autocorrelation because the majority of values fall inside the confidence intervals, demonstrating the relative independence of the past and current returns. The random walk theory, which is frequently seen in effective financial markets, is consistent with this pattern of weak autocorrelation.

2.7.2. Simple Returns

The percentage change in the adjusted closing prices from one period to the next is measured by simple returns. This computation, which eliminates NA values for cleanliness, is especially simple and displays the amount that has changed over time.

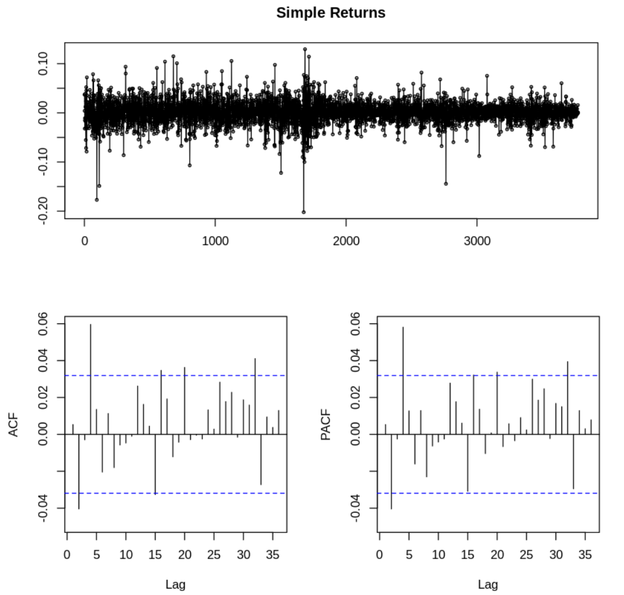

Figure 7. Log Returns

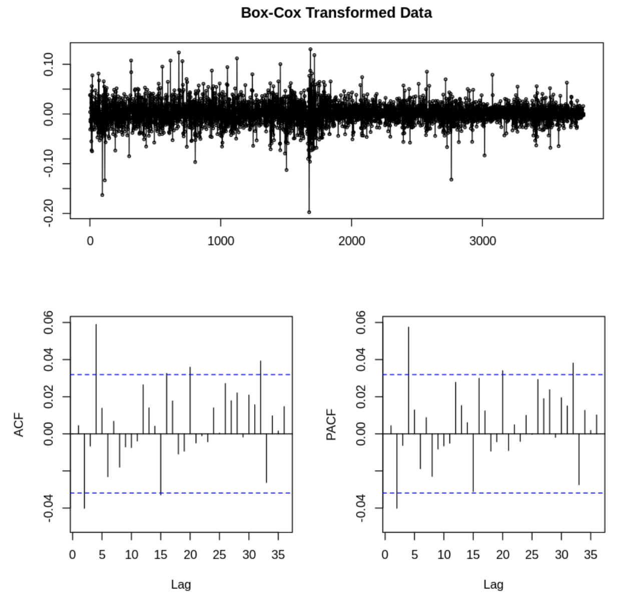

Figure 8. Box-Cox Transformation

The (Figure 7) displays the percentage change in stock prices over time, and the plot indicates how variable those returns are. The majority of the bars in the ACF plot below show that there is little autocorrelation between the simple returns. Additionally, the PACF plot shows low autocorrelation, a sign of market efficiency that implies previous price movements have little bearing on upcoming price changes.

2.7.3. Box-Cox Transformation

Using an appropriately selected lambda, the Box-Cox transformation is used to normalize the data and address non-constant variance or heteroscedasticity. Because of the variance stabilization and increased symmetry in the distribution, the data are more suited for linear modelling.

The "Box-Cox Transformed Data" (Figure 8) shows variations in stock prices that have undergone a Box-Cox transformation to reduce variance and improve analytical suitability. Since the correlations lie within the confidence ranges, the ACF plot suggests that there is no substantial autocorrelation and that the converted data does not exhibit significant linear dependency over time lags. The absence of substantial correlations in the PACF plot further supports the notion that the Box-Cox transformation improved the time series' consistency with the presumptions of numerous statistical models.

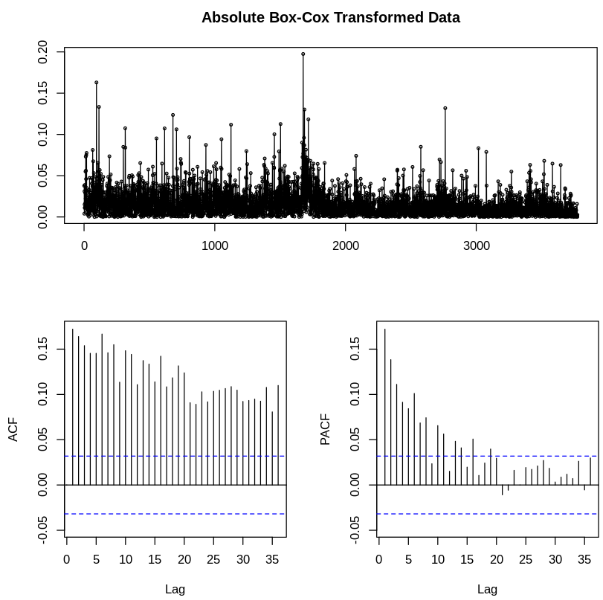

Figure 9. Box-Cox Transformation Absolute Value

The "Absolute Box-Cox Transformed Data" (Figure 9) illustrates the absolute values of the stock returns that have been Box-Cox transformed, emphasizing the size of changes regardless of their direction. More noticeable autocorrelation may be seen in the ACF figure below in the early delays, indicating some degree of linear predictability in the number of returns. However, the PACF figure shows that this predictability disappears after a few delays, with the majority of data falling inside the confidence intervals and indicating no meaningful link at longer lags.

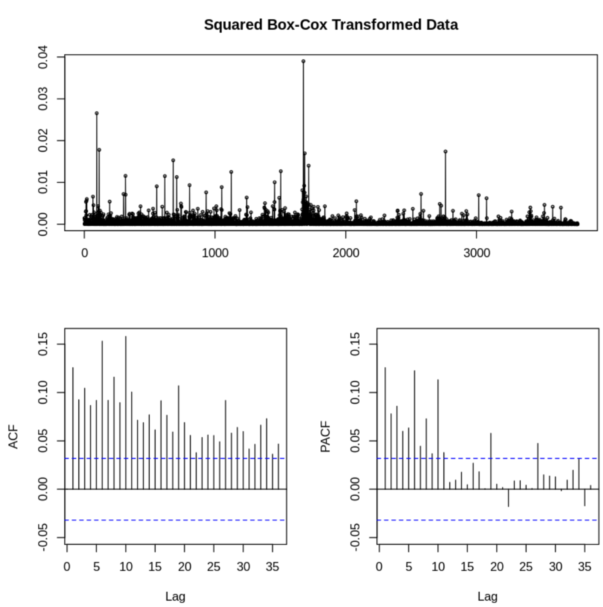

Figure 10. Box-Cox Transformation Square

The "Squared Box-Cox Transformed Data" plot from Figure 10 shows the squared values of Box-Cox transformed returns, emphasizing the variability and volatility clustering in the data. The ACF plot reveals significant autocorrelation at initial lags, indicating that volatility tends to cluster over time. The PACF plot, conversely, shows fewer significant correlations, suggesting that the squared transformations reveal a pattern where past volatility influences near-future volatility but diminishes as the lag increases.

2.7.4. Log Transformation and Differencing

By applying a special case of the Box-Cox transformation (lambda = 0), a logarithmic transformation is achieved. To render this data stationary—which is necessary for some forms of time series modelling and analysis—it is then differences.

2.7.5. Data Visualization and Stationarity Testing

To see patterns and dependencies in the modified data series, time series plots are used in conjunction with the autocorrelation (ACF) and partial autocorrelation (PACF) functions. To further ensure that the data are appropriate for forecasting and other statistical analysis, the stationarity of the Box-Cox transformed data is tested using the Augmented Dickey-Fuller (ADF) test.

According to the results of the Augmented Dickey-Fuller (ADF) test, the time series 'data_transformed' is stationary. The null hypothesis of a unit root (showing non-stationarity) can be rejected at a common significance level with a test statistic of -14.928 and a p-value of 0.01. The test's selection of 15 lags takes data autocorrelation into account, guaranteeing the validity of the test statistic.

2.7.6. QQ Plot

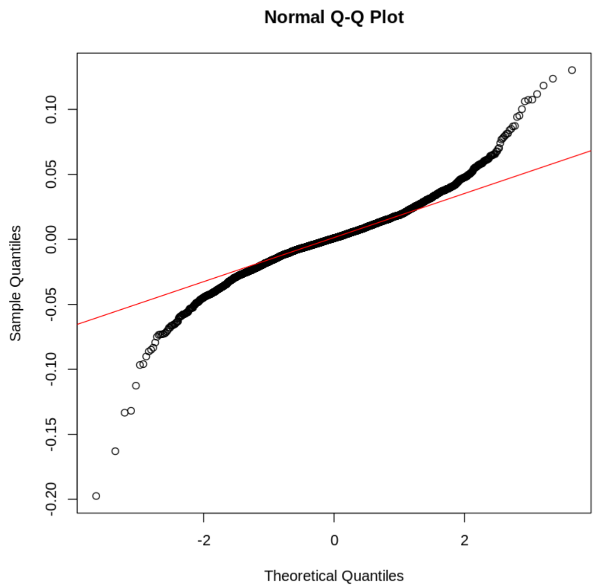

Figure 11. Normal QQ Plot

The data is seen to stray from normality in the Normal Q-Q Plot, where an obvious S-shaped curve points to heavy tails and distribution skewness. The data has a minor leftward asymmetry, as indicated by the skewness value of roughly -0.190150869898795. The data has more prominent peaks and fatter tails, as indicated by the kurtosis value of roughly 5.43351876083983, which is larger than the kurtosis of a normal distribution, which is 3 (Lavine, 2008), (wik).

2.8. GARCH Models

2.9. Extended Auto-correlation Function (EACF)

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	x	o	x	o	o	o	o	o	o	o	o	o	o
1	x	x	o	x	o	o	o	o	o	o	o	o	o	o
2	x	x	o	x	o	x	o	o	o	o	o	o	o	o
3	x	x	o	o	o	x	o	o	o	o	o	o	o	o
4	x	x	x	x	o	x	o	o	o	o	o	o	o	o
5	x	x	x	x	x	x	o	o	o	o	o	o	o	o
6	x	x	o	x	x	x	o	o	o	o	o	o	o	o
7	x	x	x	x	x	x	x	o	o	o	o	o	o	o

Figure 12. EACF Parameters of Original Transformed Data Should be p=4 and q= 0

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	o	o	o	o	o	o	x	x	o	o	x	o	o
2	x	x	o	o	o	o	o	o	x	o	o	x	o	o
3	x	x	x	o	o	o	o	o	o	o	o	o	o	o
4	x	x	o	o	o	o	o	o	o	o	o	o	o	o
5	x	x	x	x	x	o	o	o	o	o	o	o	o	o
6	x	x	x	x	x	x	o	o	o	o	o	o	o	o
7	x	x	x	x	x	x	x	o	o	o	o	o	o	o

Figure 13. EACF parameters of Absolute Values of the Transformed Data should be p=1, q=1, On the other hand, p = 2,q=2 again p=3,q=3 are the possible p and q values

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	o	o	o	o	x	x	o	o	x	o	o	o	o
2	x	x	o	o	o	x	o	o	o	x	o	o	o	o
3	x	x	o	o	o	x	o	o	o	x	o	o	o	o
4	x	x	x	x	o	x	o	o	o	x	o	o	x	o
5	x	x	x	x	x	o	o	o	o	x	o	x	o	o
6	x	x	x	x	x	x	o	o	o	x	o	o	x	o
7	x	x	x	x	x	x	x	o	o	x	o	o	x	o

Figure 14. EACF parameters of Squared Values of the Transformed Data should be: p=1 and q=1.

Figure 15. Overall analysis of Extended Autocorrelation Function (EACF) plots demonstrating various model parameters.

The decision from (Figures 15) to select the GARCH(1,1) model is rooted in analyzing the Extended Autocorrelation Function (EACF) across various data transformations. While different EACF analyses suggested several model parameters, the choice of (1,1) emerges as a consistent and thus preferred option. This reflects the GARCH model's capability to capture a common characteristic in financial time series: volatility clustering. The subsequent step involves validating the GARCH(1,1) model's fit by examining its residuals, ensuring that the model adequately captures the time series data's patterns without leftover structure in the variance.

2.10. Fitted Model

The GARCH(1,1) gives the AIC of: -18547.17. The below tables show more information optimal GARCH(1,1) model.

Table 2. Initial Parameters

I	Initial X(I)	D(I)
1	4.522856e-04	1.000
2	5.000000e-02	1.000
3	5.000000e-02	1.000

Table 3. Relative Function Convergence

Description	Value
FUNCTION	-1.274374e+04
RELDX	2.007e-06
FUNC. EVALS	65
GRAD. EVALS	25
PRELDF	9.052e-11
NPRELDF	9.052e-11

Table 4. Final Parameters

I	FINAL X(I)	D(I)	G(I)
1	4.447546e-06	1.000	-3.326e+03
2	4.769441e-02	1.000	-6.972e-01
3	9.441194e-01	1.000	-9.961e-01

The GARCH(1,1) model fit findings for Apple's daily historical stock prices from 2002 to 2017 are summarized in the (tables 2, 3, 4, 5, 6). The extremely low p-values suggest that the model's parameters are highly significant. Diagnostic tests indicate that while the fitted model's residuals show non-normality (because of the low Jarque-Bera test p-value), the squared residuals do not show autocorrelations (because the Box-Ljung test p-value is not significant), suggesting that the model does a good job of capturing volatility clustering.

Table 5. GARCH Model Summary

Coefficient	Estimate	Std. Error	t value	Pr($\geq \|t\|$)
a0	4.448e-06	6.825e-07	6.517	7.17e-11
a1	4.769e-02	3.472e-03	13.735	¡ 2e-16
b1	9.441e-01	4.227e-03	223.335	¡ 2e-16

Table 6. Diagnostic Tests

Test	Data	X-squared	df	p-value
Jarque Bera Test	Residuals	1968.1	2	$< 2.2e-16$
Box-Ljung Test	Squared.Residuals	1.1213	1	0.2896

2.11. Diagnostic Analysis of GARCH(1,1) Model Residuals

Ljung-Box test

data: Residuals Q* = 14.518, df = 10, p-value = 0.1507

Model df: 0. Total lags used: 10

Ljung-Box Test: With a p-value above the typical significance threshold, this test does not detect autocorrelation in the residuals at the first 10 lags, supporting the model's adequacy.

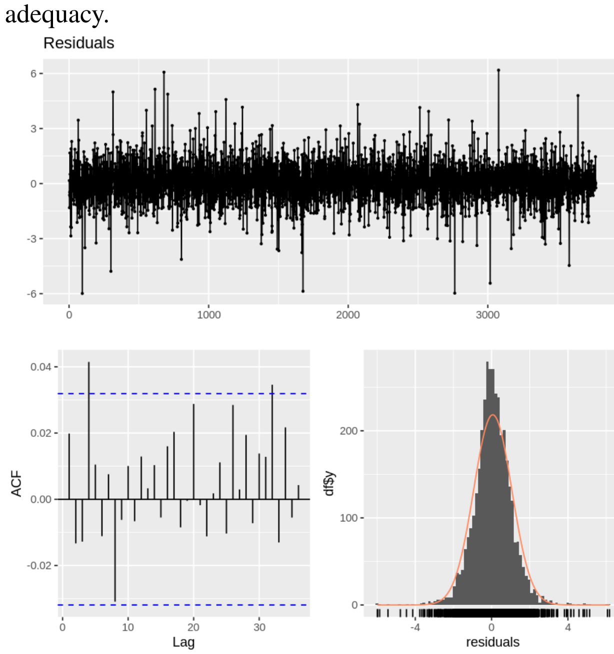

Figure 16. Diagnostic Analysis of GARCH(1,1) Model Residuals

Residuals Plot from Figure 16: The residuals fluctuate around zero without a pattern, indicating a model that captures the data's volatility well. ACF Plot: The ACF indicates no significant autocorrelation among residuals, suggesting that the model accounts for the time-series autocorrelation appropriately. Histogram vs. Normal Curve: The residuals mostly conform to the normal distribution, with some divergence at the tails hinting at potential heavy-tailed behaviour.

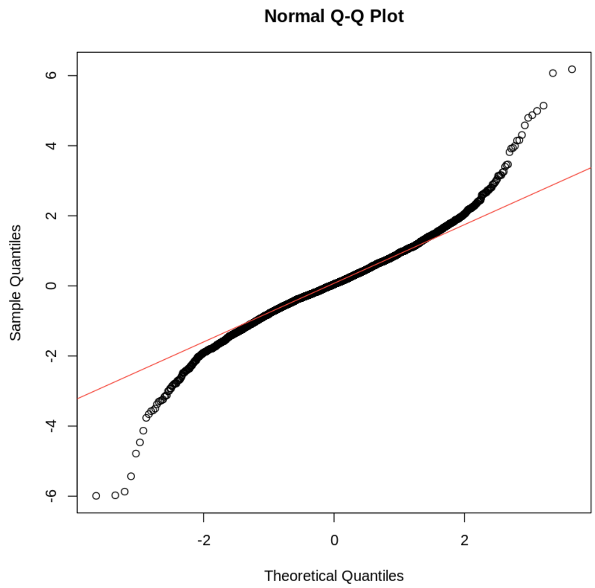

Figure 17. QQ-Plot of GARCH(1,1) model fitted

Q-Q Plot from (Figure 17): There's an S-shape deviation from the normal line, implying that the residuals have heavier tails than those of a normal distribution.

In summary, these diagnostics collectively imply that the GARCH(1,1) model is a good fit for the data, with the caveat of non-normality in the distribution of residuals, which is not uncommon in financial time series.

2.12. Diagnostic Plots for Absolute vs Square Residuals of GARCH(1,1) Model

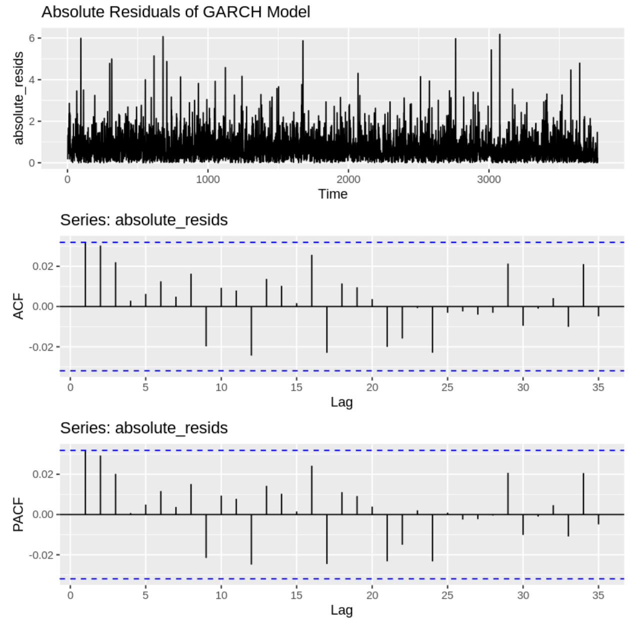

Figure 18. Diagnostic Plots for Absolute Residuals of GARCH Model

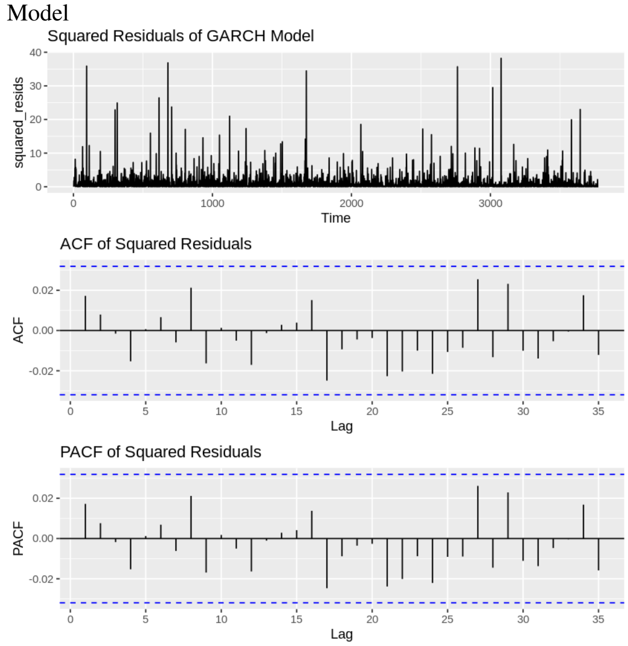

Figure 19. Diagnostic Plots for Square Residuals of GARCH Model

For the Absolute Residuals of the GARCH(1,1) Model plot shown in Figure 18: When comparing the absolute residuals plot to the squared residuals plot, the former shows less dramatic spikes but still shows volatility.

Bars within the confidence boundaries can be seen in the ACF plot for absolute residuals, indicating that the volatility has been sufficiently taken into account by the model.

Similar to the ACF plot, the PACF plot is devoid of bars that surpass the confidence limits, indicating that there are no substantial autocorrelations in the residuals of the model.

For the Squared Residuals of the GARCH(1,1) Model plot from Figure 19:

A feature that GARCH models frequently capture is periods of significant volatility clustering, as suggested by the plot of residuals, which has big spikes. From the ACF of squared residuals, it is clear that no meaningful volatility patterns remain unmodeled. The autocorrelation is minimal. Additionally confirming the model's suitability is the PACF of squared residuals, which shows no discernible autocorrelation.

2.13. eGARCH Volatility Forecasting for Apple Stock

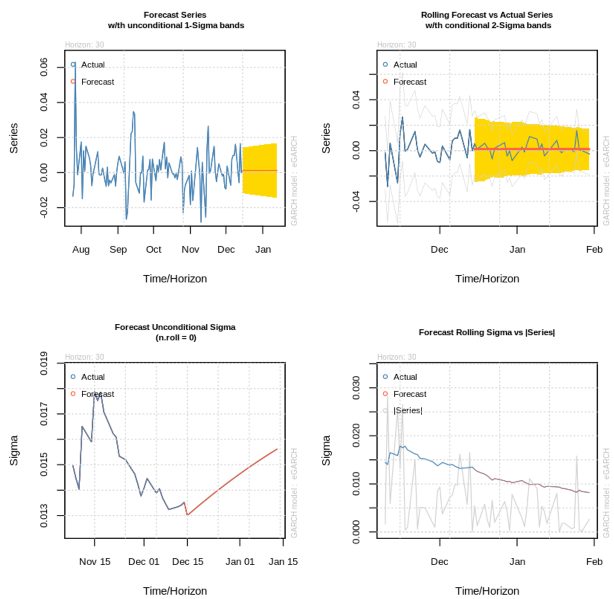

Figure 20. eGARCH(1,1) Volatility Forecasting for Apple Stock

The forecasts of the eGARCH model for Apple's stock prices over a thirty-observation horizon are displayed in (Figure 21). There are times when the first plot's actual price fluctuations exceed the predicted 1-standard deviation confidence intervals, suggesting that the volatility was either overestimated or underestimated.

While capturing more volatility, the second plot depicts occurrences across the 30-day horizon where actual volatility exceeds expectations by extending the confidence intervals to two standard deviations. The model's projected long-term volatility over the 30 observations is depicted in the third plot; the adaptive rolling forecasts, which update with each new data point, are shown in the fourth plot, which contrasts the predictions with the actual volatility experienced during the 30-day forecasting period. All of these plots evaluate how well the model predicts the degree of stock price volatility clustering over a given period.

2.14. Simulated Returns from eGARCH Model

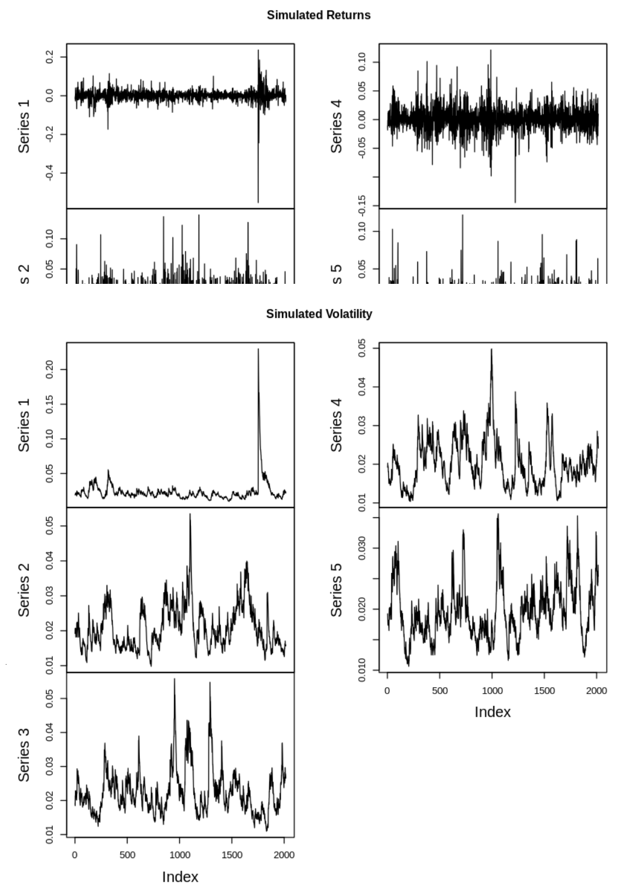

Figure 22. The eGARCH model forecasts variability and risk over time, which is visually represented by the volatility trajectories for each simulated returns series.

2.16. Simulated Stock Prices from eGARCH Model

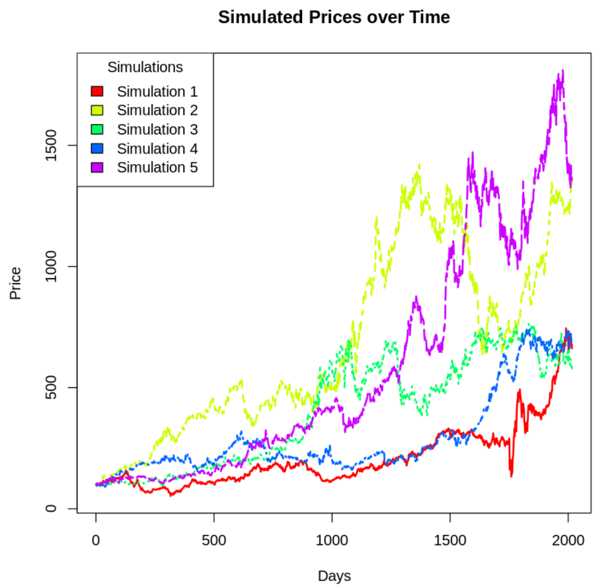

Figure 23. The plot displays the evolution of the original stock price under several simulated scenarios, demonstrating the range in possible future stock prices. It does this by converting simulated returns into price pathways.

3. Conclusion

Key volatility patterns and trends have been successfully detected by the eGARCH(1,1) model when applied to 15 years of stock data from Apple Inc. The model captures the dynamic aspect of the financial time series, as shown by simulations of future stock returns and volatility, albeit the residual distribution shows signs of non-normality. The Ljung-Box test and Q-Q plots, as well as the model fitting and diagnostic checks, indicate that the model is robust and captures the important aspects of the data without leaving a sizable amount of autocorrelation unaccounted for in the residuals.

All things considered, the model offers a strong basis for predicting short-term changes in stock prices while acknowledging the inherent uncertainty in such projections. It's a useful tool for predicting possible future behaviour and for deciding on investment or financial planning strategies with knowledge. Users should be mindful of the model's limits and the possibility of uncommon, extreme occurrences that might not be fully captured by the historical data that was used to train the model.

4. PART 2 Technical Questions

4.1. Question 1

The process X_t defined as

$$X_t = \begin{cases} Y_t & \text{if } t \text{ is even} \\ Y_t + 1 & \text{if } t \text{ is odd} \end{cases}$$

where Y_t is a stationary time series, is not stationary. This can be seen by examining the statistical properties of X_t.

For a time series to be stationary, its mean, variance, and autocorrelation must be constant over time. If Y_t is stationary, its expected value $E[Y_t]$ is some constant, say μ. However, the expected value of X_t is

$$E[X_t] = \begin{cases} \mu & \text{if } t \text{ is even} \\ \mu + 1 & \text{if } t \text{ is odd} \end{cases}$$

which is not constant over time as it alternates between μ and $\mu + 1$.

Moreover, while the variance of Y_t remains unchanged for even t, the variance of X_t would not change for odd t as adding a constant to a random variable does not affect its variance. However, the autocorrelation function of X_t, denoted as $\gamma_X(\tau)$, will be the same as that of Y_t, denoted as $\gamma_Y(\tau)$, for lags τ that are multiples of 2. For other lags, $\gamma_X(\tau)$ will differ due to the introduced constant on the odd terms. This pattern violates the assumption of the constant autocorrelation function required for stationarity.

In conclusion, the time series X_t is not stationary as its mean is not constant over time, and the autocorrelation function does not remain constant for all lags.

4.2. Question 2

Proving for Stationarizing a Time Series

Given the time series

$$X_t = (1 + 2t)S_t + Z_t$$

where $S_t = S_{t-12}$, we seek to transform X_t so that the transformed series is stationary.

A time series is stationary if its mean, variance, and autocovariance function are invariant over time. That is, for all t and τ, the following must hold:

1. $E[X_t] = \mu$ is constant (mean stationarity).
2. $Var[X_t] = \sigma^2$ is constant (variance stationarity).
3. $Cov(X_t, X_{t+\tau}) = \gamma(\tau)$ depends only on τ (covariance stationarity).

Steps for Transformation

Detrending

The term $(1 + 2t)$ indicates a linear trend. By taking the first difference, we attempt to eliminate this trend:

$$\Delta X_t = X_t - X_{t-1} = 2S_t + Z_t - Z_{t-1}$$

Seasonal Adjustment

In order to overcome the seasonal component, we further difference the series with a lag of 12:

$$\Delta_{12} X_t = X_t - X_{t-12}$$

Since $S_t = S_{t-12}$, the S_t terms cancel out, and we are left with:

$$\Delta_{12} X_t = Z_t - Z_{t-12}$$

Combining Transformations

Combining the detrending and seasonal differencing operations, we define the transformation as:

$$Y_t = \Delta(\Delta_{12} X_t)$$

$$Y_t = \Delta(Z_t - Z_{t-12})$$

$$Y_t = (Z_t - Z_{t-1}) - (Z_{t-12} - Z_{t-13})$$

If Z_t is stationary, then the differenced series Y_t will also be stationary. This combined transformation should render X_t stationary by removing both the deterministic trend and the seasonal component. The stationarity of the transformed series Y_t can then be confirmed using statistical tests such as the Augmented Dickey-Fuller test.

4.3. Question 3

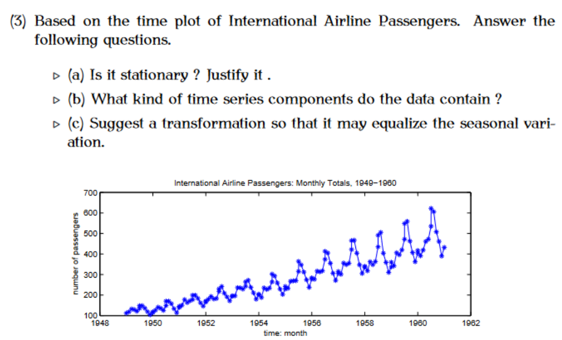

Figure 24. Question 3

(a) Is it stationary? Justify it.

The given time series is **not stationary**. The justifications are as follows:

- There is a noticeable upward **trend** over time, which suggests an increase in the number of passengers as time advances.
- Regular and repeating fluctuations within each year suggest a **seasonal** pattern in the data.
- The amplitude of the seasonal fluctuations increases with time, indicating that the **variance** is not constant.

Stationarity requires that the mean, variance, and autocorrelation structure of the series do over time. This series violates those conditions.

(b) What kind of time series components do the data contain?

The time series components present in the data include:

- A **trend component** characterized by a long-term increase in the number of passengers.
- A **seasonal component** that shows regular patterns within each year, corresponding to seasonal variations in airline travel.
- A **cyclical component** might be present, although further analysis would be required to identify any longer-term cycles beyond the seasonal fluctuations.

(c) Suggest a transformation so that it may equalize the seasonal variation.

To transform the series and stabilize the seasonal variations, the following steps could be considered:

1. Apply **seasonal differencing** to remove the seasonal component with the transformation $Y_t = X_t - X_{t-12}$, where X_t is the original series.
2. Use a **log transformation** to stabilize the increasing variance with the transformation $Y_t = \log(X_t)$.
3. Combine both by first applying the log transformation and then the seasonal differencing to the logged data.

These transformations aim to reduce the impact of both trend and seasonality, making the series more likely to satisfy the conditions of stationarity. Statistical tests, such as the Augmented Dickey-Fuller test, should be employed post-transformation to confirm stationarity.

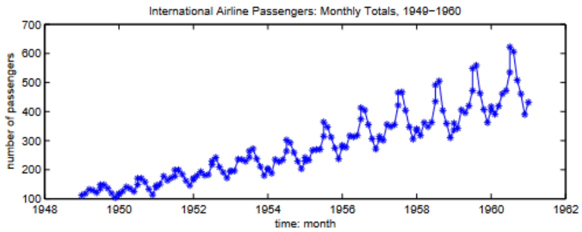

Figure 25. International Airline Passengers: Monthly Totals, 1949-1960

References

Kurtosis. https://en.wikipedia.org/wiki/Kurtosis. Accessed on: 2024-04-27.

Apple inc. (aapl) stock historical prices & data. https://finance.yahoo.com/quote/AAPL/history?p=AAPL&period1=1012521600&period2=1485820800. Accessed: 2024-04-26.

Lavine, M. *Introduction to Statistical Thought*. 2008. Free textbook available at https://people.math.umass.edu/~lavine/Book/book.html.

5. Appendix

You can access the codebase used to implement this project from this link: https://rb.gy/l9coh3.

[1]Nanyang Technological University (NTU), Singapore [2]School of Computer Science and Engineering (SCSE) [3]Master of Science in Artificial Intelligence. Correspondence to: Ntambara Etienne <ntam0001@e.ntu.edu.sg>.

Proceedings of the 38^{th} *International Conference on Machine Learning*, PMLR 139, 2021.