Time-Series Analysis of DJI, NASDAQ, and SPX: A Study on ETS and ARIMA Predictive Modelling (2012-2017)

Prepared for
Peter Kempthorne
7.4-PAFBPK-401007A

Prepared by
Group2
Group Members:
Eric - Rongze Gao
Mia – Xueyiting Wang
Cecilia – Sihan Wang
Rona - Hairong Zhang

Word Count: 13096

Contents

1. Executive Summary

1.1 Project Overview

This report embarks on an insightful journey into the world of stock indices through the application of time series analysis, particularly leveraging the ARIMA (Autoregressive Integrated Moving Average) and ETS (Exponential Smoothing) models.

The analysis of stock indices via time series methods provides a comprehensive understanding of market behaviors, highlighting trends, seasonal variations, and potential future movements. The goal of this report is to utilize these sophisticated statistical models to extract meaningful information from historical stock index data, enhancing predictive accuracy and informing strategic decision-making.

The ARIMA model, with its robust handling of autoregressive and moving average components, and the ETS model, with its ability to address error, trend, and seasonality components, together form a comprehensive approach. We will meticulously analyze the stock index data using these models, aiming to interpret the findings within the context of today’s dynamic financial landscape. The insights derived from this analysis are expected to empower investors, researchers, and market analysts with a deeper understanding of the stock index’s past behavior and its implications for future performance.

1.2 Relevant Theories/Models

For the research, we applied time series analysis to the study of the United States stock indices, including NASDAQ, DJIA, and SPX 500. We tried to unravel their underlying patterns, dependencies, and possibly forecast future movements.

We used both ARIMA model and ETS model to do the time series forecasting analysis for all these three stock indices. And after the analysis of each of them, we also compared the forecasting performance of ARIMA model and ETS model of each stock index and find out which one is better in each case.

In addition to the comparison of the models, we also analyzed the regularity, commonalities and differences behind these three stock indices, so as to explore the situation of the United States stock market and carry out further analysis.

Moreover, we also analyzed the influence of time window on the research, that is, the different research results brought by different time slices, and we got some summaries from that.

2. Introduction

2.1 Background of the Study

The global financial markets have evolved significantly in recent decades, becoming increasingly interconnected and complex. One of the world's most influential economy, the United States, has been at the forefront of this transformation. Its respective stock indices, including NASDAQ, DJIA, and SPX 500, have a profound impact on both domestic and global market dynamics.

Understanding these indices' behaviors is crucial for investors, policymakers, and researchers, as these indices often serve as a sign of the broader economic health of the US and even the whole market. Traditional analysis methods can be valuable, but they often fail to predict future trends effectively due to the inherently dynamic and non-linear nature of financial markets. This needs the adoption of more sophisticated analytical methods. One such method is time series analysis, which takes into account temporal dependencies and fluctuations in the data over time, providing a deeper understanding of patterns and predictive insights. This report aims to apply time series analysis to the study of the United States stock indices, with an aim to unravel their underlying patterns, dependencies, and possibly forecast future movements.

2.2 Motivation of the Research

In the ever-evolving world of finance, investors are always on the lookout for information that will give them an edge. As Warren Buffet famously said, "Risk comes from not knowing what you're doing." In this context, understanding the dynamics of stock indices becomes critical. A stock index reflects the performance of a group of shares and is often seen as a barometer of the economy's health. Analyzing stock indices gives a comprehensive overview of market trends, investor sentiment, and potential investment opportunities.

Given the dynamic and often volatile nature of financial markets, regular and detailed analysis of key stock indices is paramount. It helps investors make informed decisions, from choosing the right time to invest to selecting which sectors have the most promising outlooks. The more data we can harness, and the better our understanding of the market trends, the more equipped we are to anticipate future movements and take appropriate action. In an era where data is the new oil, we believe that the insights gleaned from this analysis could prove invaluable, ultimately contributing to robust financial strategies and, potentially, more stable and prosperous economies.

2.3 Scope of the Report

This report is focused on using time series analysis to examine the fluctuations and trends in the stock indices of the United States for 20 years, from 1999 to 2019. And to do the forecasting, we just use data for 5 years, from 2012 to 2017. Our study centers on stock market of the United States, represented by three main stock indices, the NASDAQ index and Dow Jones Industrial Average, and SPX 500. To conduct this analysis, we will use both primary data (derived from the actual historical performances of the indices) and secondary data (sourced from relevant financial and economic literature and reports). This study will specifically scrutinize broad market trends and patterns but will not delve into the performance of individual stocks or specific sectors within these markets. Potential limitations of this report include the unpredictable nature of the stock markets and the numerous external factors that can impact them, such as political events, policy changes, and global economic trends. Additionally, while our time series analysis can highlight patterns and suggest correlations, it is not intended to predict future performances of the indices.

3.Background

3.1. Background of the Report

Time series analysis is a significant branch of statistics, primarily focusing on data points collected in chronological order. This type of analysis can assist us in understanding hidden patterns within the data, predicting future trends, and evaluating variations at different time points (Box, Jenkins, Reinsel, & Ljung, 2015). Time series analysis is extensively applied in various fields, including economics, finance, biology, and engineering.

Despite the widespread recognition of the importance of time series analysis, challenges persist in its practical application, such as selecting appropriate models, handling missing and outlier values, and interpreting results (Hyndman & Athanasopoulos, 2018). Moreover, time series analysis plays a pivotal role in financial markets, but there are many issues with forecasting results, often leaving much to be desired. With the rapid development of big data and machine learning, the methods and techniques in time series analysis are continuously evolving, presenting new challenges and opportunities for researchers and practitioners.

The objective of this report is to explore these challenges through time series analysis and forecast results of datasets from three significant U.S. stock indices (NASDAQ, DJIA, SPX) in order to unearth future trends in the U.S. economy, and to introduce and explain the reasons behind the phenomena we discover. We will concentrate on two models (ARIMA and ETS) to analyze and discuss the forecasting results. We will first introduce the source and characteristics of the datasets, then describe the time series analysis methods we use, followed by reporting our results. Finally, we will discuss the implications and limitations of our findings, as well as directions for future research.

3.2. The concept definition

Time Series: A time series is a collection of data points gathered in time order, usually used for analyzing and predicting trends and patterns (Chatfield, 1984).

Autocorrelation: Autocorrelation is the correlation of a variable with its past values in a time series. This is an important concept in time series analysis, as many time series models assume some form of autocorrelation between data points (Box et al., 2015).

Stationarity: In time series analysis, if a time series' statistical properties (such as mean, variance, etc.) remain constant over time, we say that the time series is stationary. Many time series analysis methods require the data to be stationary or can be transformed into stationary in some way (Hyndman & Athanasopoulos, 2018).

Trend: A trend is the long-term direction or path of time series data as it changes over time. Identifying and dealing with trends is an important step in time series analysis (Chatfield, 1984).

Seasonality: Seasonality is the repeating pattern in time series data over a certain period, such as the seasonal changes each year or the weekdays and weekends each week. Dealing with seasonality is an important consideration in time series analysis (Hyndman & Athanasopoulos, 2018).

ARIMA Model: The ARIMA (Autoregressive Integrated Moving Average) model is a commonly used time series prediction model that combines the methods of autoregression (AR), differencing (I), and moving average (MA). The ARIMA model can handle non-stationary time series data (Box et al., 2015).

Autoregression (AR): Autoregression is a linear prediction model based on the past values of a time series. In the ARIMA model, the autoregressive part describes the dependency between the current value and past values (Box et al., 2015).

Differencing (I): Differencing is a method of making a non-stationary time series stationary. In the ARIMA model, the differencing part deals with the non-stationarity of the time series (Box et al., 2015).

Moving Average (MA): Moving average is a linear prediction model based on the past error terms of a time series. In the ARIMA model, the moving average part describes the dependency between the current value and past error terms (Box et al., 2015).

ETS Model: The ETS (Error, Trend, Seasonality) model is a commonly used time series prediction model that can handle time series data containing trends and seasonality. The main advantages of the ETS model are its simplicity and flexibility (Hyndman & Athanasopoulos, 2018).

4. Methodology

4.1 Research Design

Our research project is composed of two primary components. The first component involves a detailed investigation of three major United States stock indices: NASDAQ, DJIA, and SPX500. We selected this focus due to the significant influence these indices have on the global financial markets.

For this component, we gathered extensive data spanning a period of five years, from 2012 to 2017. This five-year period was chosen to provide a balance between having a substantial amount of data to analyze, but also limiting the scope to a timeframe where the economic conditions were relatively stable.

To analyse the data, we used two popular statistical forecasting models: the Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing State Space Model (ETS). We employed the Auto ARIMA function for the ARIMA model and a standard implementation for the ETS model, as both provide robust mechanisms for model selection and fitting.

In order to validate the quality of our model fits, we conducted a series of statistical tests. This included a residuals test to examine the difference between the observed and predicted values, and a Ljung-Box test to check for autocorrelation in our residuals. In addition, we examined the autocorrelation function (ACF) and plotted histograms to visually assess the distribution of our residuals. These tests helped us assess whether our models were adequately capturing the underlying patterns in the data, or if there was white noise indicating randomness in our residuals.

Based on the fit of these models, we then made predictions about future stock price trends. We also discussed the potential implications of these predictions, looking at how they could impact the economy and influence financial decision-making.

The second part of our research project involved examining the impact of different time windows on our forecasts. This was broken down into three separate investigations.

The first investigation involved comparing the effects of different lengths of time windows on our forecasts. We looked at two-time windows: a five-year window from 2012 to 2017 and a shorter three-year window from 2014 to 2017. This helped us understand how the choice of time window could impact the accuracy of our forecasts.

The second investigation explored the impact of different data extraction frequencies on our forecasts. We compared forecasts derived from daily data, weekly data, and monthly data. This analysis helped us understand how the granularity of data could influence the forecast.

The third investigation examined the impact of taking the mean of the data on the forecast results. We compared two scenarios: one where we did not take the mean and randomly extracted data for forecasting, and another where we first calculated the mean of the data and then made forecasts based on this mean.

4.2 Data Collection Methods

In our study, the choice of the time window for data analysis was a critical factor. An important aspect of this choice was the conscious decision to avoid periods containing Black Swan events. These unpredictable occurrences have a major impact on financial markets and can create outliers in the data that significantly distort the results of our analysis.

Examples of Black Swan events include financial crises, during which stock prices can plummet precipitously. Such extreme fluctuations can dramatically skew the understanding of regular market behavior and distort the modeling and prediction outcomes. Hence, our primary focus was to study the nature and characteristics of the data under normal conditions.

However, it is impossible to eliminate the risk of encountering a Black Swan event in the data. To mitigate this risk, we decided to limit the size of our time window to a maximum of ten years. We found that if the time window was larger than ten years, the chances of encountering a Black Swan event, and hence the presence of outliers in the data, increased significantly.

Furthermore, to ensure the reliability of our research, we performed an analysis to assess the quality of our chosen time window. It was critical to ensure that the data within our chosen time window exhibited good properties, in other words, that it was representative of normal market conditions and free of significant distortions.

After careful consideration, we decided to select the five-year period from 2012 to 2017. This decision was based on our assessment that this period provided a balance between having a sufficient amount of data for reliable analysis and limiting the risk of including a Black Swan event. This time frame also represented a period of relative stability in the financial markets, making it suitable for our investigation into the regular characteristics and behaviors of the stock indices.

4.3 Data Analysis Techniques

In this study, we mainly use the ARIMA model and ETS model. Before performing the model fitting, we need to confirm that the data are stationary, which is the assumed premise of the ARIMA model. We assume that a time series is stationary when its mean, variance, and autocorrelation structure are zero.

4.3.1 Taking logarithms

we need to find the white noise component before we can fit the model. However, the stock price is not white noise because there is no stable expectation, no constant variance, and it is very highly correlated with residuals. We have to take the logarithm of the stock price. If you take the natural logarithm of asset returns, the resulting distribution follows a normal distribution. At the same time the asset price return is very much like white noise. It follows both a normal distribution, residual expectation, and variance is basically zero. Occasionally, with a black swan incident, the fluctuations can be large, but the basic range is between plus and minus 0.1. So, we consider it a smooth observable variable.

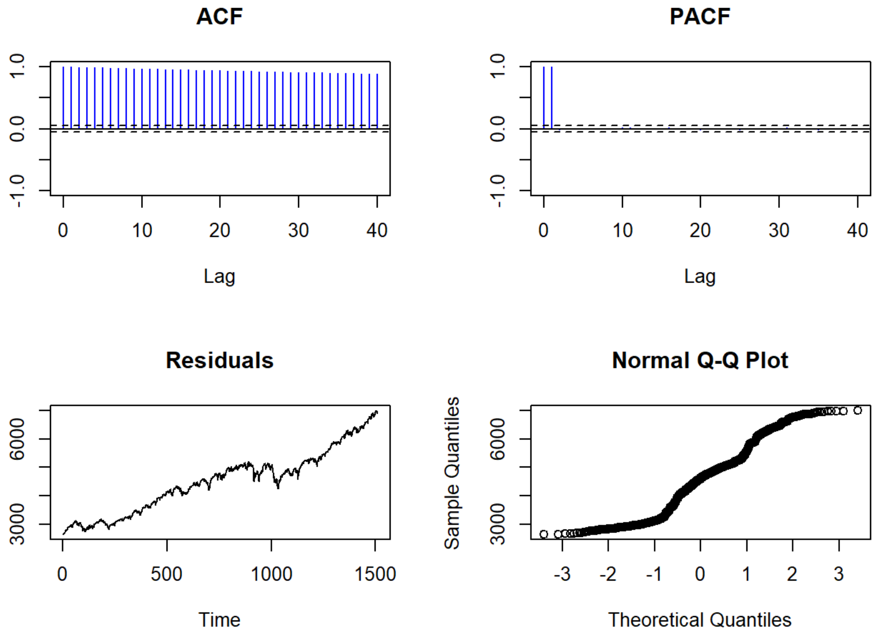
4.3.2 First Difference

We treat the data with a first order difference to try to make the time series stationary. The first order difference is equal to the difference between the closing price of the day and the opening price of the day. It can be viewed as the change in price over the course of the day. The purpose of this is to try to eliminate any trends in the data so that the new series has a constant mean and variance. In other words, the original price series is increasing over time. However, the difference in prices may not have a trend and we can obtain a constant variance.

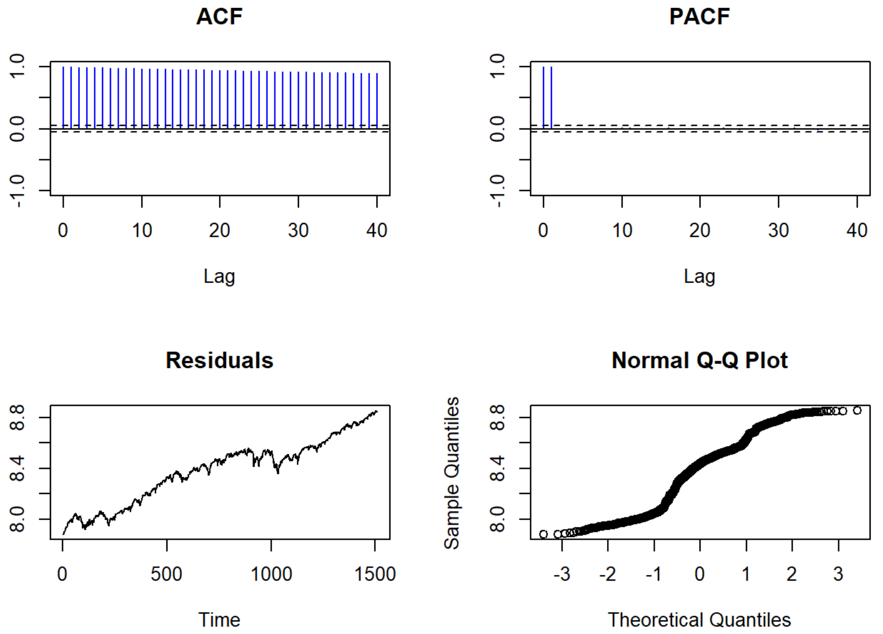
4.3.3. Stock index analysis

4.3.3.1. NASDAQ

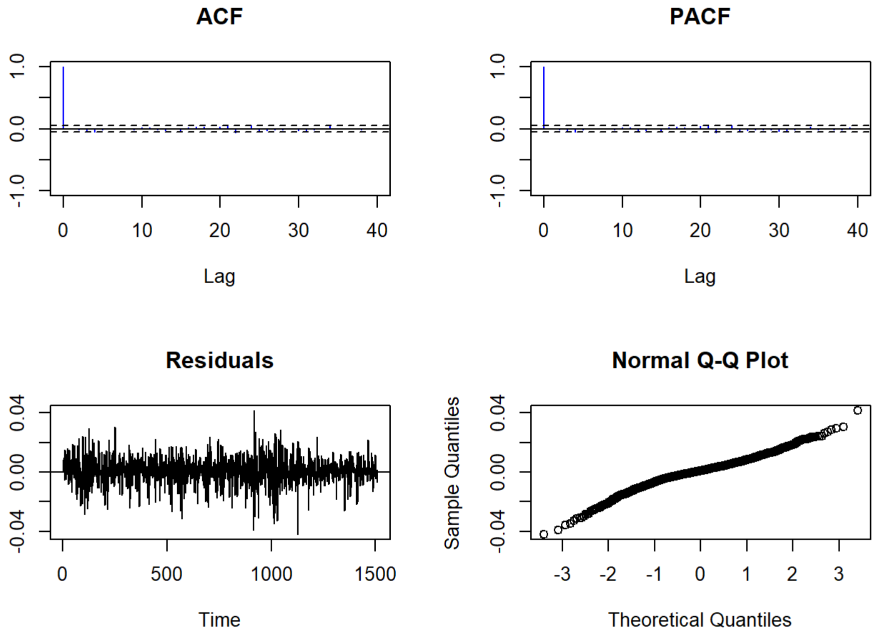
First, we examined the NASDAQ time series. Each lag is well beyond the critical value in ACF plot, and they are all with a slow steady downward trend. Moreover, it shows significant autocorrelations at higher lags, it could be an indication of seasonality. Autocorrelation at all lags is out of the critical value, and the PACF plot shows significant partial autocorrelations at the first few lags, thus, we do not think it is a white noise here. Furthermore, residuals show a significant upward trend. We infer that it is not white noise because its mean is not close to zero like a normal distribution, and it also has a clear pattern and trend. Last, we can find that the the points in Normal Q-Q are along the line, which means the data is normally distributed.



From the characteristics of the log data, we can find that the residuals fluctuated around zero and had its first, then it continued to rise again, and it has been maintaining an upward trend. Although it has some fluctuations with the increase of time, from which we do not see cyclical or regular fluctuations. The values of each lag in the ACF are very close to 1, they are well above the critical value, and these values are slowly decreasing. This indicates that they have strong autocorrelation and can be initially judged as non-white noise. From the PACF plot, lag1 and lag2 are both approximately equal to 1, also indicating that they have strong autocorrelation. From the results of the Normal Q-Q Plot, the points are all very close to the 45-degree line, and it can be assumed that it is normally distributed.

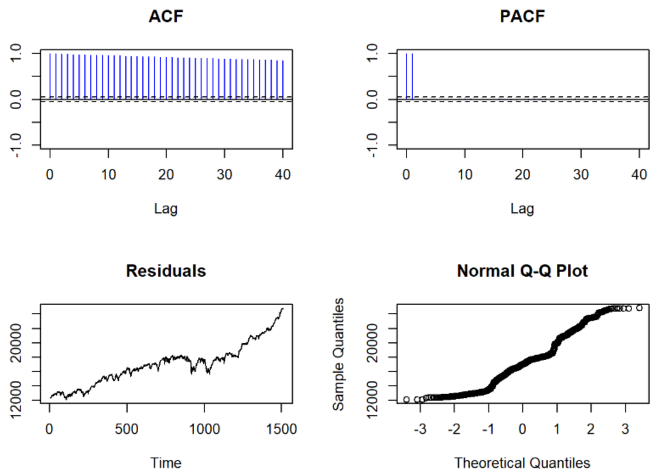


As a result, we chose to switch to another model and take the logarithm of this time series by performing a first-order difference on it. From the characteristics of the logdiff results: we find that the residuals fluctuated around between ±0.02. The mean of residuals is approximately equal to zero, but its variance closes to 7.961067e-05, this suggesting that while it fluctuates slightly around zero. ACF and PACF show that all lags are within the critical values, which indicates that they are not autocorrelated and perform well. From the results of the Normal Q-Q Plot, the points are all very close to the 45-degree line, and it can be assumed that it is normally distributed.

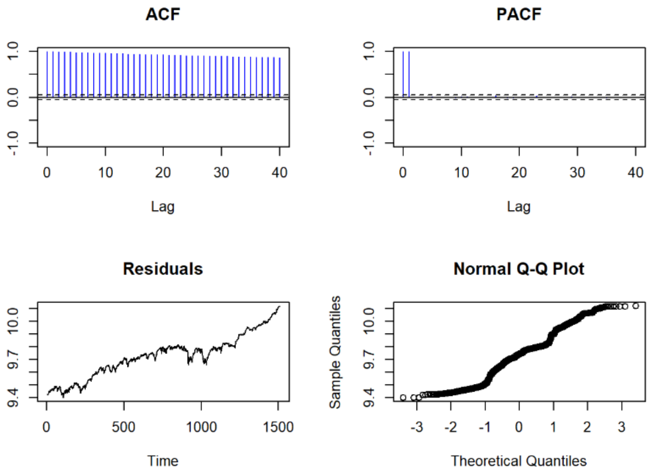


4.3.3.2 DJI

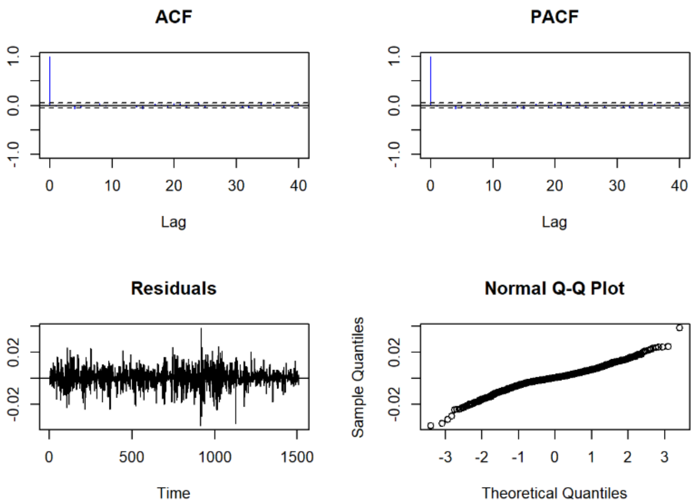
The time series analysis and forecasting process for DJIA index data is as follows:



Based on the original data, each lag surpasses the critical value of the Autocorrelation Function (ACF), and a slow, steady declining trend is clear. Moreover, significant autocorrelation is observed at higher lags, which may signify seasonality or periodicity. Given that all lags exhibit autocorrelation beyond the critical value, we conclude that this is not white noise. Furthermore, the residuals demonstrate a pronounced upward trend, reaching a significant peak around time 1500 and exceeding 22000. Additionally, the residuals fluctuate between approximately 800 and 1200 without a clear growth trend, which could be due to certain economic factors. We infer that it is not white noise because its mean does not approach zero as a normal distribution would, its variance is not constant, and it exhibits a certain upward trend with minor fluctuations.



Upon an in-depth examination of the log data's characteristics, a few key observations can be made. Firstly, a distinct upward trend is evident in the residuals, which is interspersed with random fluctuations. A remarkable large-scale fluctuation is noticeable between the time frame of 750 and 1250, where the trend of this section appears relatively flat. However, we do not discern any periodic or regular fluctuations that might suggest seasonality or cyclical patterns. Secondly, the autocorrelation function (ACF) values for each lag are remarkably close to 1, vastly exceeding the critical threshold. This suggests a strong autocorrelation, leading us to conclude that the data does not represent white noise.

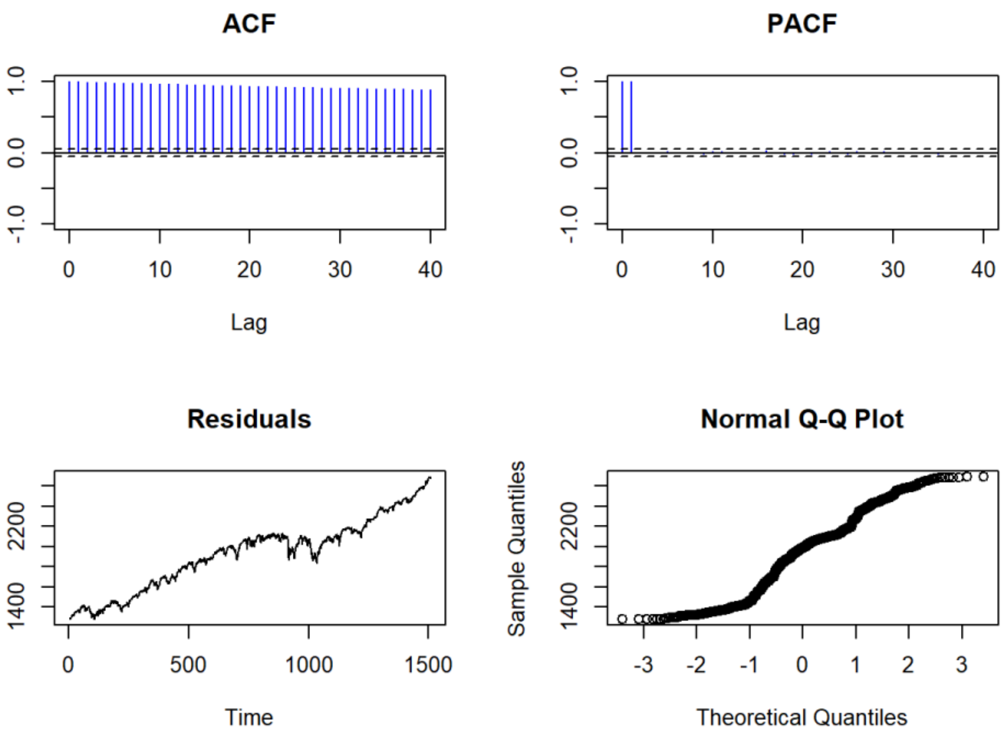


In light of these findings, we decided to further refine the data by applying a first-order differencing technique. Post this transformation, we observed that both the ACF and the partial autocorrelation function (PACF) exhibit significant cutoffs after the first lag. This is a clear indication that the first-order difference has successfully eliminated the strong autocorrelation and linear trend that was initially present in the data. Furthermore, the mean of the residuals stands at 0.0004576216, a value slightly above zero, and the variance remains relatively constant. This leads us to consider the data as white noise. Additionally, the residuals now align along a relatively smooth 45-degree line in the normal Q-Q plot, indicating an approximate normal distribution. Based on these comprehensive analyses, we can confidently assert that the differenced data has achieved stationarity. This transformation has effectively prepared the data for subsequent time series forecasting, thereby enabling more accurate and reliable predictions.

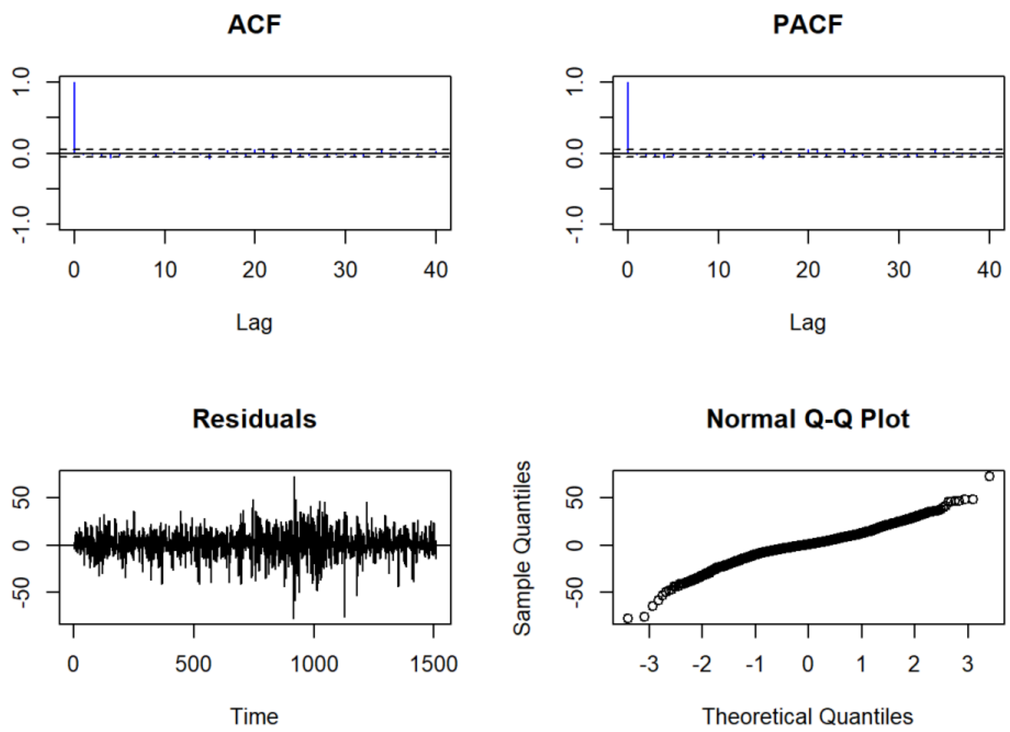
4.3.3.3 SPX500

From 1999 to 2019, the SPX500 index has experienced many unstable periods. And from 2012 to 2017, we can use these 5 years to do the forecasting. Through the analysis of past data, predicting the future trend and market conditions of the SPX500 index will provide an important judgment for the future development direction of the whole US stock market.

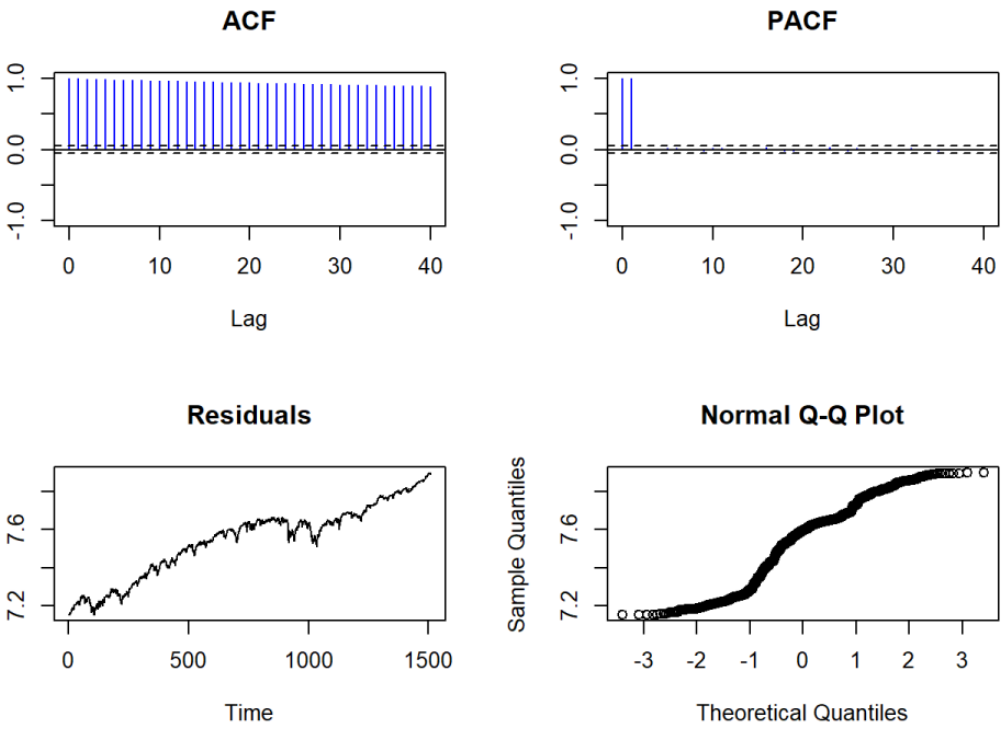
Firstly, we need to process the raw data efficiently. If we test the original data directly, we find that it is not a good object for analysis. As shown in the graphs below, since the ACF does not show a gradual decline and the PACF also does not have a sharp cutoff after lag, differencing may be necessary, and the data might not be stationary here. In addition, the residuals show a clear trend, so it indicates that the model has not fully captured the underlying structure, and differencing might be necessary. In the Normal Q-Q plot, the residuals also depart significantly from the reference line, which indicates the non-normality in the residuals. So differencing can help to improve the modeling process here.



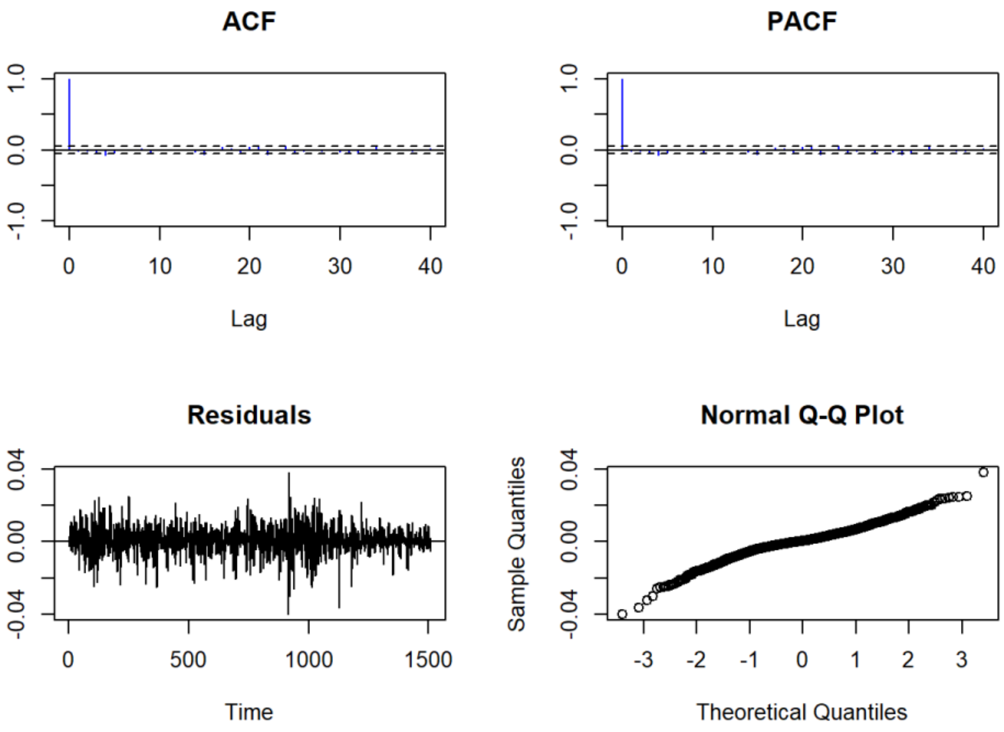
After applying one-time differencing operation here, from the graph below, we can see that the autocorrelation values in ACF drop quickly and the partial autocorrelation in PACF shows a sharp cutoff after lag 1, it suggests that one-time differencing has successfully removed any linear trends in the data. In the Normal Q-Q plot, the residuals follow a relatively straight line, indicating approximate normality. However, the residual plot still shows some unstable patterns, although it looks much more random than before.



To make the data more stable and feasible, we choose to do the logarithm transformation the original data. However, only operating log transformation of the original data is still not enough, because after the test, the same problems as the original data will appear, including ACF, PACF pattern does not meet the prediction assumptions requirements, and residuals problems exist. The test results are shown in the figures below.



Therefore, we applied one-time differencing operation again on the data with logrithm transformation. The results in the figures below show the feasibility of the processed data. To be more specific, both ACF and PACF show a sharp cutoff after lag 1, suggesting that the one-time differencing has successfully removed any linear trends in the data. Also, the residuals look almost random and relatively stable. And the residuals now follow a relatively straight line in the Normal Q-Q plot, indicating approximate normality of it. Based on this analysis, the differenced data has achieved stationarity, so the data has been already prepared to make subsequent time series predictions.



5. Forecast and Analysis

5.1 forecast for three stocks indexes

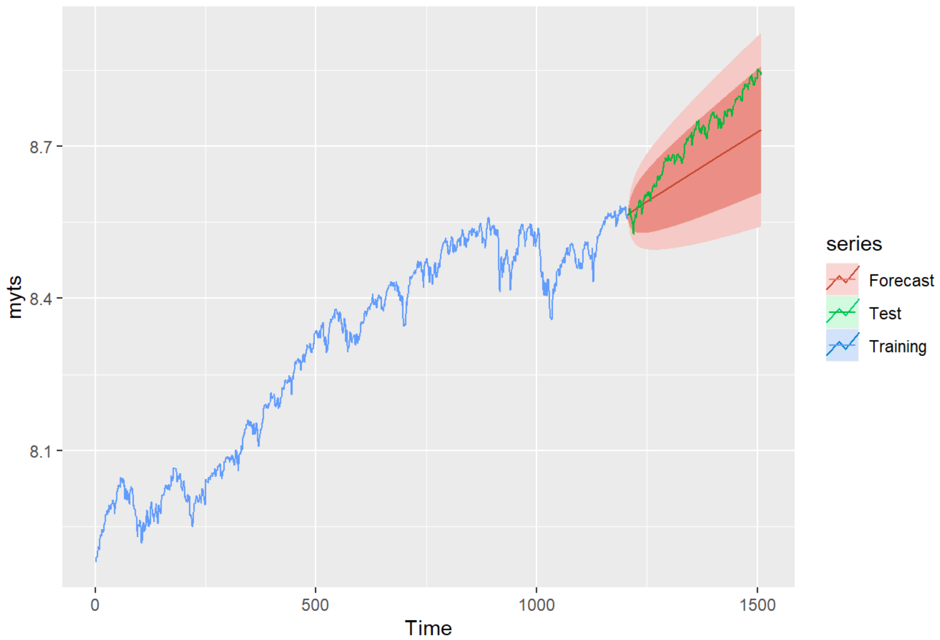
5.1.1. NASDAQ

We compare the fit of NASDAQ in ARIMA and ETS models and finally pick a fit model for it.

5.1.1.1. ARIMA Model fitting results and analyses:

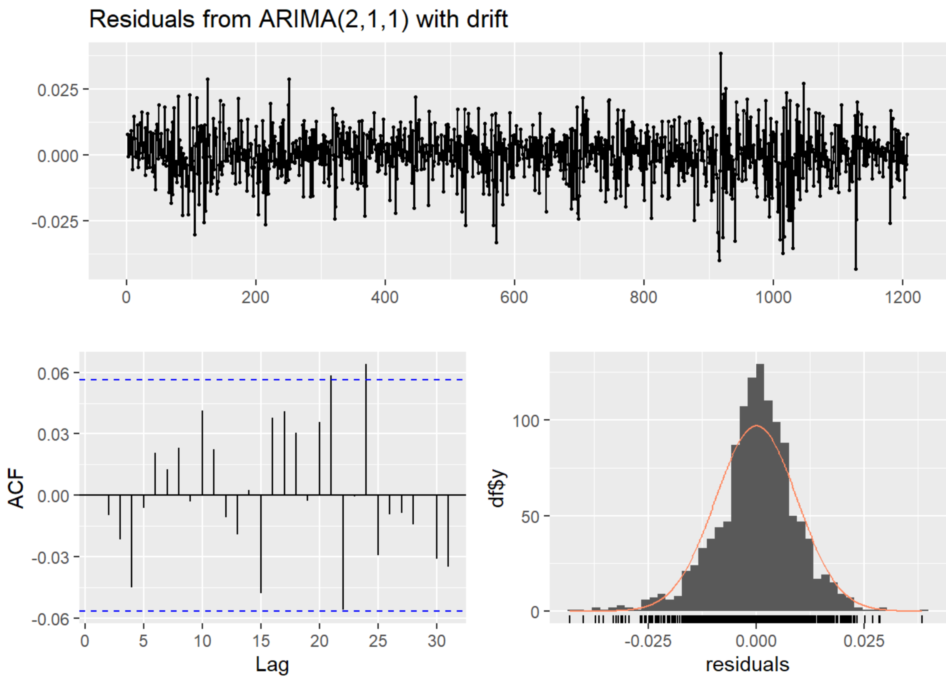
We plot the train, test and forecast curves in the same graph so that we can more clearly see the gap between the predicted and test values.

Overall, the time series shows a very clear upward trend from 2012 to 2015. Auto Arima helped us pick out ARIMA (2, 1, 1) as our fitted model. As you can see from the graph, after 2017, forecast results in a continuous increase in the stock price over time, and the angle of this diagonal line is roughly 45 degrees. When compared to the test results, both the predicted and test values are inside the confidence interval and the trend is consistent. Only the actual test value has a slight fluctuation and a steeper growth slope. To be precise, this prediction result is very reasonable and satisfactory to us.



From the results of residuals, the residuals are similar to the form of white noise. There are several reasons for this: firstly, from the histogram, it highly overlaps with the normal distribution curve, so we can understand that it looks to follow a normal distribution. Secondly, most of the lag values detected by ACF are inside the critical value, which is clearly no autocorrelation. In terms of the distribution of the overall residuals (ARIMA (2,1,1)), it tends to be around ±0.025, with one or two occasional deviations far from zero.

Statistically, the Ljung-Box test shows that the Q-value is 6.5888 and the p-value is 0.4729 which is greater than the commonly used significance level of 0.05. This means that we fail to reject the null hypothesis. The residuals are, as far as the test can tell, independently distributed.




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,1) with drift
## Q* = 6.5888, df = 7, p-value = 0.4729
##
## Model df: 3.   Total lags used: 10
```

The model fitting statistics indicated an adequate fit. The Ljung-Box test yielded a p-value of 0.4729, indicating that we cannot reject the null hypothesis of no autocorrelation among the residuals, which suggests that our model effectively captures the temporal structures within the data.

However, upon evaluating the model's forecasting performance, the results were less encouraging. For the training set, the Mean Error (ME), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) were all relatively low, indicating that the model did an excellent job in describing the historical data. The Mean Absolute

Scaled Error (MASE) for the training set was also around 1, signifying that the model performed comparably to a naïve benchmark model.

However, the model's performance deteriorated on the test set. The RMSE and MAE were significantly higher than those for the training set, and the Mean Absolute Percentage Error (MAPE) was 0.7049, suggesting that the average forecast was off by about 70% in absolute terms. Moreover, the MASE exceeded 8 for the test set, indicating the model's performance was much worse than a naïve model.

The autocorrelation of residuals at lag 1 (ACF1) was close to 1 for the test set, indicating that there could be remaining patterns in the test set that the model failed to capture. Additionally, the high value of Theil's U on the test set implies that the model may not be better than a simple random walk model at making forecasts.

In conclusion, despite the satisfactory model fitting on the training data, the ARIMA model demonstrated limited success in forecasting unseen data. This suggests the need to revisit our model selection process, consider additional model types, or explore potential structural changes within the time series data that might have been overlooked.

```
##                        ME        RMSE         MAE          MPE       MAPE
## Training set 3.720871e-05 0.009444754 0.006983597 0.0004688805 0.08416655
## Test set     5.928498e-02 0.069711460 0.061684200 0.6768479006 0.70490919
##                   MASE        ACF1 Theil's U
## Training set 0.9944159 0.000402455        NA
## Test set     8.7834032 0.978235597  10.95876
```

5.1.1.2. ETS Model fitting results and analyses:

We chose ETS (M, N, N) as the fitted model based on data characteristics. An ETS (M, N, N) model is a time series model with multiplicative error, no trend, and no seasonality. It is suitable for time series data where the variation is proportional to the level of the series, but there is no discernible trend or seasonal pattern.

The data from 2012 to 2015 shows a consistent upward trajectory, albeit with some minor fluctuations. This suggests a growing trend, which is typical of many economic and business series, indicating increasing market demand or productive capacity.

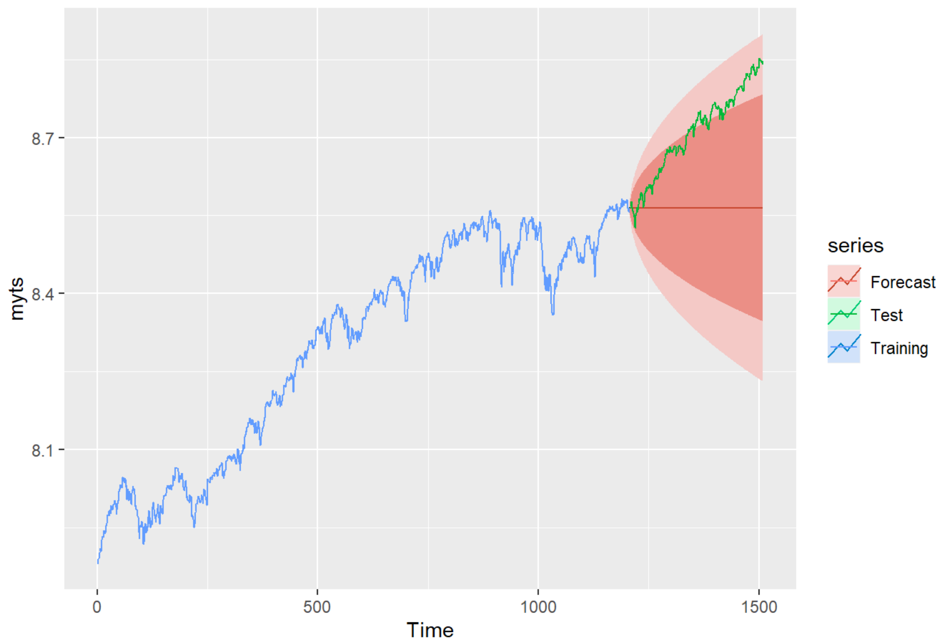
However, a significant shift in trend dynamics is observed from 2015 onwards. Here, we notice more substantial price fluctuations, indicative of market volatility. This could be due to various factors such as shifts in demand and supply, changes in market sentiment, or policy changes, among others.

Interestingly, the forecast made from 2017 onwards projects a steady, linear progression with no discernible incline, meaning the model doesn't anticipate any significant changes in the prices. The smooth, parallel price curve implies a forecast with a level trend, no seasonality, and constant variance over time. This may suggest an ETS (M, N, N) model or a similar non-trending model was used for the forecast, which expects prices to maintain an almost constant value.

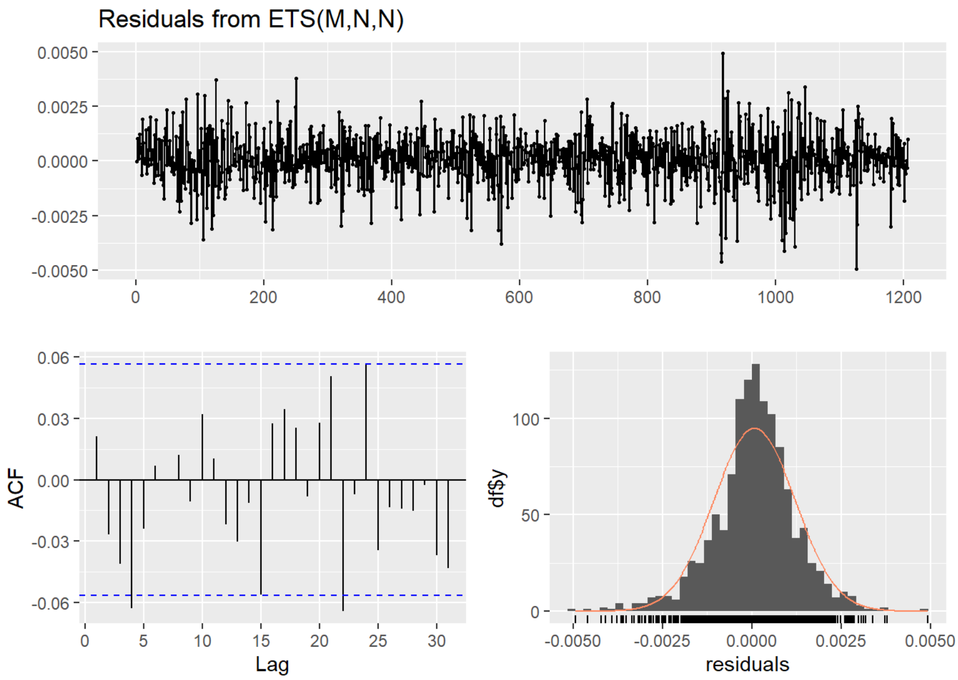
In contrast, the actual test results from this period show a sharp upward trend, suggesting the forecast significantly underestimated the future prices. The steep 45-degree angle indicates a strong positive trend, suggesting substantial growth or inflation in prices.

Despite this discrepancy between the predicted and actual trends, it's important to note that both are within the confidence interval. This means that the actual outcomes still fell within the range of uncertainty accounted for in the forecast model.

While this doesn't necessarily indicate a faulty model, it does imply that the model may have lacked some key information or failed to capture certain trend components. The difference between the forecast and actual results might indicate a need for a more dynamic model that can better adapt to shifts in trends, especially in a context where prices can be influenced by numerous fluctuating external factors.



From the plot of residuals of ETS (M, N, N), it can be seen that all residuals fluctuate between ±0.025, with a few outliers, where the most off-zero outlier is around 0.0050. There is no obvious seasonality or periodicity detected in the image of residuals, it fluctuates irregularly. If the data were not collected or recorded correctly, this could result in outliers. Sometimes, extreme events (shocks) can cause outliers. For example, a natural disaster might dramatically impact sales for a short period of time. The ACF results show that most of the lags are within the critical value, and only a few lags exceed the critical value by a small amount, such as lag4 and lag22. We can consider this to be a white-noise data because of the no autocorrelation in its lags. Furthermore, the histogram shows a very clear pattern of a normal distribution. residuals walk within the curve of the normal distribution and basically coincide with the normal distribution curve.



Q* = 10.604, and the null hypothesis of the Ljung-Box test is that the data are independently distributed. The larger the p-value, the more evidence we have to not reject the null hypothesis.

Since the p-value here is 0.3892 (above the typical significance level of 0.05), we would not reject the null hypothesis of independence among the residuals. This indicates that the residuals from the ETS (M, N, N) model are not significantly different from white noise, suggesting the model has appropriately captured the information in the series, and the residuals are just random fluctuations around zero.

```
##
##  Ljung-Box test
##
## data:  Residuals from ETS(M,N,N)
## Q* = 10.604, df = 10, p-value = 0.3892
##
## Model df: 0.   Total lags used: 10
```

The model performance metrics indicate contrasting results for the training and test data sets. When applied to the training data, the model shows an excellent fit. This is evidenced by the low values of the ME, RMSE, MAE, MPE, and MAPE. All these metrics are near zero, signifying minimal discrepancies between the actual and forecasted values. Furthermore, the MASE value is close to 1, suggesting our model's performance is on par with a benchmark naive model. The Autocorrelation Function at Lag 1 (ACF1) value is relatively low, pointing to a lack of significant autocorrelation in the residuals, a desirable trait indicating the model has adequately captured the information in the series.

However, the forecast model's performance deteriorates considerably when applied to the test set. The high values of ME, RMSE, MAE, MPE, and MAPE indicate a significant degree of forecast error, implying that the model failed to accurately predict unseen data points. The MASE value is substantially larger than 1, indicating that our model performs worse than a naive model. In particular, the ACF1 value near 1 indicates strong autocorrelation in the residuals, suggesting that there is a pattern in the test data that our model failed to capture. Additionally, Theil's U statistic, a measure of forecast quality, is high, further demonstrating the model's poor predictive performance.

These findings suggest that while the model has effectively captured the patterns in the training data, it failed to generalize to unseen data in the test set. This discrepancy may be due to overfitting, where the model has become too specialized to the training data and is unable to accurately predict new, unseen data. Alternatively, the underlying pattern in the data may have changed in the test period in a way that the model failed to capture.

```
##                        ME        RMSE         MAE         MPE       MAPE
## Training set 0.0005658478 0.009483681 0.007017078 0.006819715 0.08455626
## Test set     0.1446753505 0.166691575 0.145883177 1.652214585 1.66635679
##                    MASE       ACF1 Theil's U
## Training set  0.9991833 0.02159828        NA
## Test set     20.7727547 0.98797550  26.18957
```

5.1.1.3. Comparing the characteristics of the predicted data from both ARIMA and ETS models：

The comparison between the ARIMA (2, 1, 1) model and the ETS (M, N, N) model in terms of their performance on the given data set provides critical insights into the strengths and limitations of both models. Both models demonstrate a robust fit to the training data, effectively capturing the patterns within it, but exhibit poorer predictive performance on unseen test data. This could be indicative of overfitting or perhaps of a failure to capture some underlying structure in the data.

ARIMA models are popular in financial time series forecasting due to their ability to model data with non-constant variance and to capture a wide range of dynamic temporal structures. On the other hand, ETS models are simpler and offer more intuitive parameter interpretations. Yet, their restriction to multiplicative error, no trend, and no seasonality may limit their applicability in contexts where these factors are influential.

From a macroeconomic perspective, ARIMA models, by capturing autocorrelations and trends, may better handle economic variables that exhibit cyclical behavior or are impacted by policy changes. ETS models, meanwhile, might be more appropriate for stable, mature markets or sectors that do not show significant volatility or where changes are predominantly multiplicative.

Considering the stock market perspective, which is inherently volatile and influenced by numerous factors such as investor sentiment, macroeconomic indicators, and company-specific news, the ARIMA model's ability to model temporal dependencies can be beneficial. The observed trends and seasonality in financial markets make ARIMA a suitable choice for stock price predictions.

As per the economic situation's reality, the last decade has witnessed significant volatility due to events like the COVID-19 pandemic, various geopolitical tensions, and policy changes. Such events often lead to structural breaks in economic data, which could affect the performance of both ARIMA and ETS models. Therefore, it's essential to monitor such events and adjust the models accordingly.

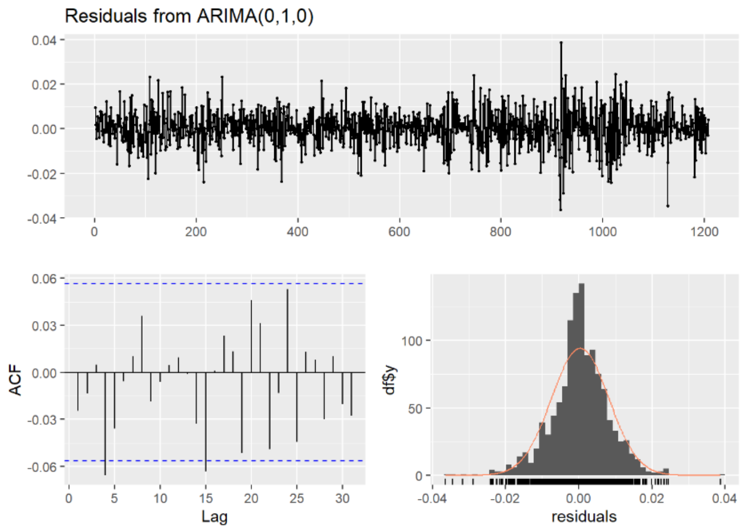
When forecasting future economic conditions, consider the models' past performance but also factor in the changes in economic structures and policies. For instance, significant government stimulus or changes in interest rates can affect future stock prices and overall economic conditions, potentially requiring adjustments to the model.

In summary, while both ARIMA and ETS models have their strengths, their applicability and performance may vary based on the data characteristics and the context. Therefore, model selection should be guided by a comprehensive understanding of the underlying data and the specific forecasting context. Periodic model evaluations and adjustments are necessary to maintain their predictive accuracy over time. Additionally, incorporating external factors such as policy changes and major events can further enhance the models' performance.

5.1.2.DJI

5.1.2.1. ARIMA

Next, we use the Auto. ARIMA model to compare different ARIMA models to fit data and choose the model with the lowest Akaike's Information Criterion (AIC) as the best fitting model. Then, based on the fitted ARIMA model, we generated forecasts for the next periods. The fitting result of the model is ARIMA (0,1,0). The forecast of ARIMA (0,1,0) shows an upward trend, but the predicted data are more within the large confidence interval of 95%, which indicates that the forecast result is not that good. According to the forecast, it can be seen that the data will reach above 10.2 before time 1500.



Next, we analyze the residuals: Firstly, by looking at the overall distribution of the residuals, the values occasionally deviate from zero, but they do not exceed ±0.04, which means the mean can be approximated to zero. Secondly, the lag values detected by the ACF are mostly within the confidence interval, only lag4 and lag15 exceed the critical value, indicating no correlation. The histogram shows that it has a high degree of overlap with the normal distribution, indicating that it appears to follow a normal distribution. According to our analysis, the residuals from fitting the ARIMA model approximate white noise, indicating that the model fits well.

```
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,0)
## Q* = 10.039, df = 10, p-value = 0.437
##
## Model df: 0.   Total lags used: 10
```

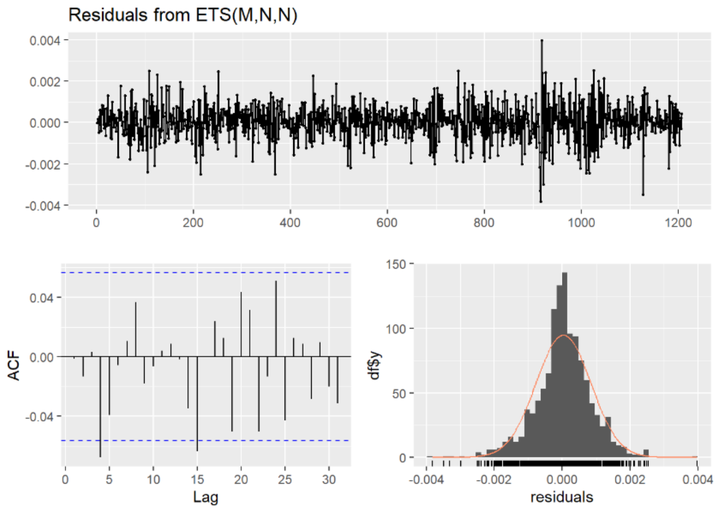
```
accuracy(future,myts.test)
```

```
##                        ME        RMSE       MAE         MPE      MAPE      MASE
## Training set 0.0003241628 0.007843256 0.0057627 0.003340238 0.0596112  1.000527
## Test set     0.1560696507 0.173229385 0.1563881 1.560865459 1.5641170 27.152308
##                     ACF1 Theil's U
## Training set -0.02461112        NA
## Test set      0.98452549  38.59677
```

Then, looking at the table, the p-value is 0.437, much greater than 0.05, which means that we do not have enough evidence to prove that our model is significantly different from the observed data, or that the model's performance is significantly better than random prediction. Analyzing from the RMSE, the value of the training set is 0.007843256, and the value of the test set is 0.173229385, both of which are relatively small, indicating a good degree of fit. In conclusion, although the RMSE value is small, the model does not perform well in future predictions, indicating that there may be room for improvement in the model.

5.1.2.2. ETS

Then we fit the data using the ETS model. Based on the characteristics of the data, ETS (M, N, N) was chosen as the fitting model, which represents an ETS model with a multiplicative error term, no trend, and no seasonality. From the prediction results of ETS (M, N, N), it can be observed that the prediction is very similar to that of the ARIMA model, predicting an upward trend, and more predicted data are within the larger confidence interval as well as 95%, which indicates that the prediction results are not that good.



Then we analyze the residual results, which are very similar to the analysis of the ARIMA model: First, looking at the overall distribution of residuals, the values occasionally deviate from zero, but do not exceed ±0.04, indicating an approximate mean of zero. Second, the lags detected by the ACF are mostly within the confidence interval, with only lag4 and lag15 exceeding the critical value, suggesting no autocorrelation. From the histogram, it can be seen that there is a high degree of overlap with the normal distribution, indicating that it seems to follow a normal distribution.

```
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,0)
## Q* = 10.039, df = 10, p-value = 0.437
##
## Model df: 0.   Total lags used: 10
```

```
accuracy(future,myts.test)
```

```
##                        ME        RMSE       MAE         MPE      MAPE      MASE
## Training set 0.0003241628 0.007843256 0.0057627 0.003340238 0.0596112  1.000527
## Test set     0.1560696507 0.173229385 0.1563881 1.560865459 1.5641170 27.152308
##                     ACF1 Theil's U
## Training set -0.02461112        NA
## Test set      0.98452549  38.59677
```

Looking at the table, the p-value is 0.445, which is much larger than 0.05, meaning that the model significantly outperforms random predictions. Analyzing from the RMSE, the value for the training set is 0.007836408, and for the test set it is 0.1733157485, both values are relatively small, indicating a good degree of fit. In summary, like ARIMA, although the RMSE value is small, the model does not perform well in future predictions, suggesting that there might be room for improvement in the model.

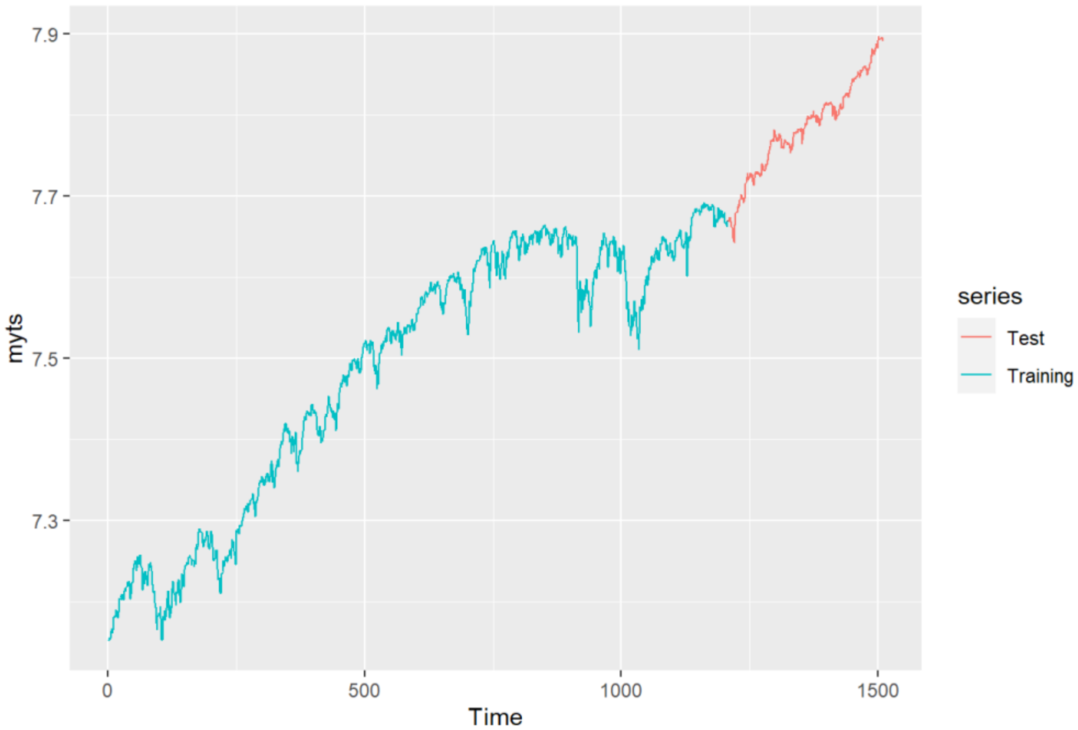
5.1.2.3. Compare

The most effective way to compare these two types of models is to compare the RMSE, MAE, and MAPE values of the ARIMA and ETS models. The lower the RMSE, MAE, and MAPE values, the higher the prediction accuracy. First, if only considering the RMSE on the training set, the RMSE of the ETS model is slightly lower than that of the ARIMA model. However, the difference between the two is very small and can almost be ignored. Therefore, the performance of these two models is very close. Although both models provide valuable insights, their accuracy is crucial and may require further optimization of the model.

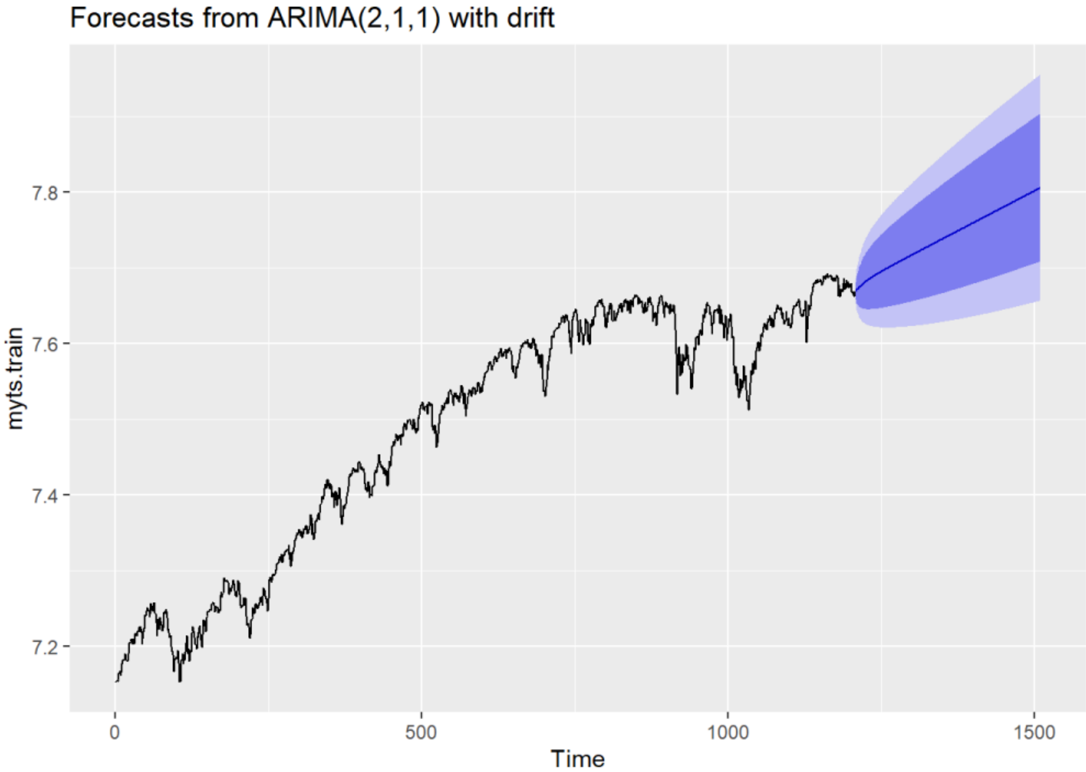
5.1.3 SPX500

5.1.3.1 ARIMA Model

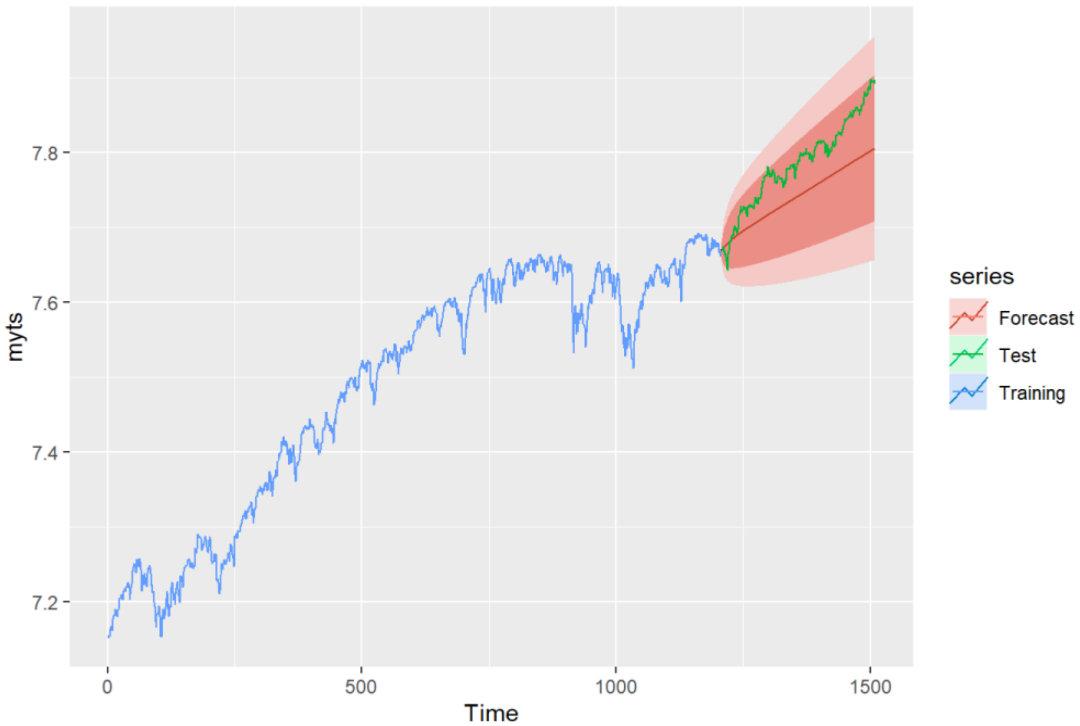
The first time series model we used here is ARIMA model. Firstly, in order to test the effect of the prediction, we divided the data into training set and test set. Here, we choose 80% of the data as the training set and the rest 20% to be the test set, as shown in the figure below.



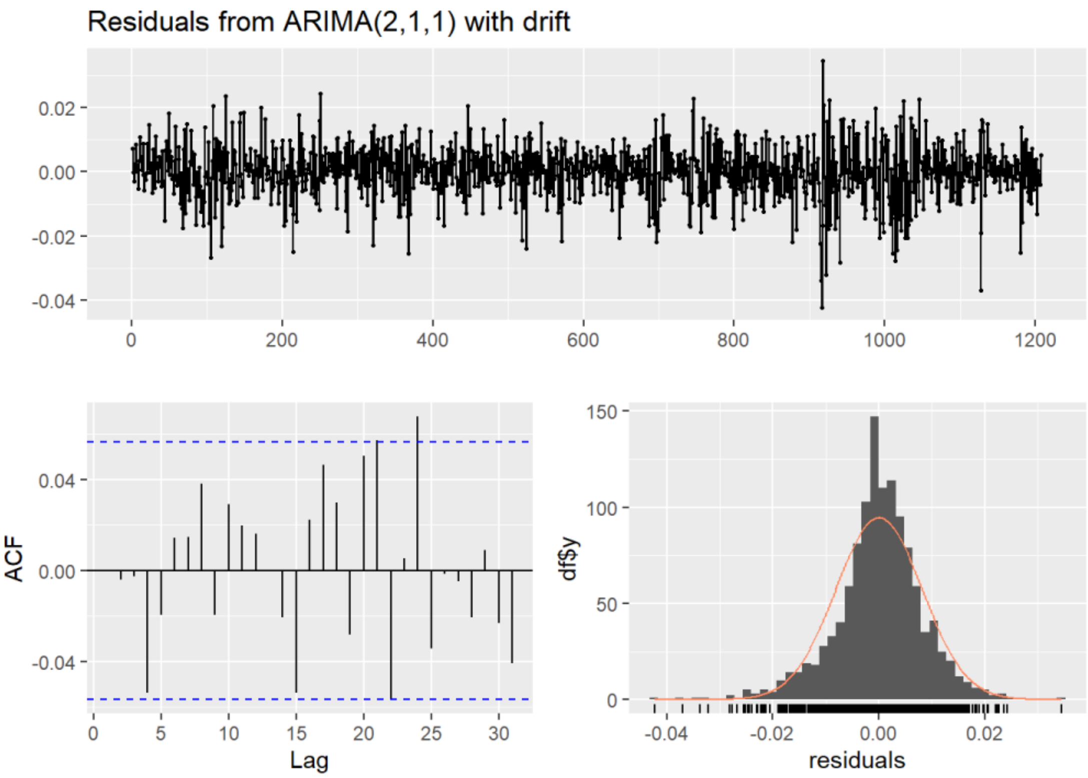
For the ARIMA model, we use the Auto. ARIMA function. As shown in the graph below, here is the ARIMA (2,1,1) model, which means that the parameters of the ARIMA model (P, D, Q) is (2,1,1) at this time. To be more specific, the potential autoregressive orders (p) is 2, the potential moving average orders (q) is 1, and the one-time differencing operation before means d is 1.



The graph below indicates that the forecast of the data by the ARIMA model is good with the test set, because the test set lies in the 80% confidence interval here.



To test the extent and effect of this prediction through ARIMA model, we can also check residuals here. As shown in the figures, although there are some obvious fluctuations in the residuals, it is relatively average and stable overall. In ACF, although there is one autocorrelation lag out of the critical line, it is still like a white noise here. And the residuals also show approximately normal pattern. So, for such a result, we can say that this is a relatively good prediction.



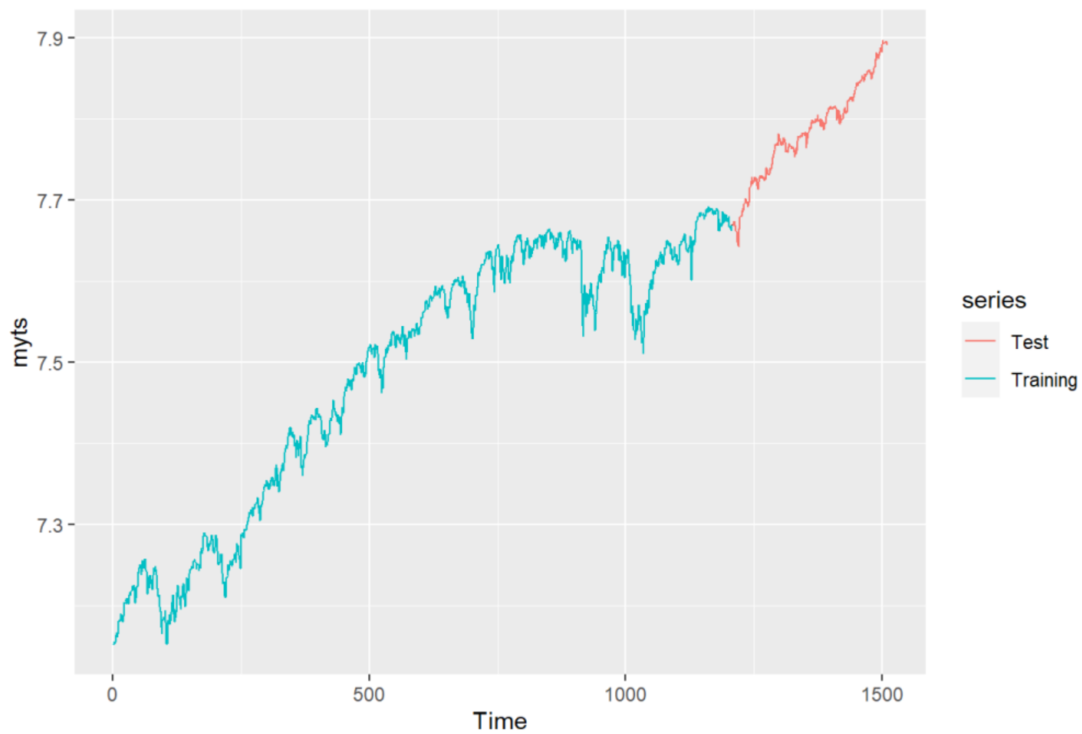
Then, we can also check its accuracy, that is, to evaluate it by some specific values. Here we can calculate several important values in evaluation metrics, in particular RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error). Since they all represent errors, the lower, the better. As we can see, the values of RMSE, MAE, and MAPE are very low for both the training set and the test set, indicating that the model is effectively minimizing the impact of large prediction errors. This shows that the prediction of this ARIMA model also still has certain feasibility and reliability from the analysis of RMSE and other values.

Therefore, in summary, the fitting degree of this ARIMA model to the prediction of the future trend of SPX500 is relatively high and appropriate.

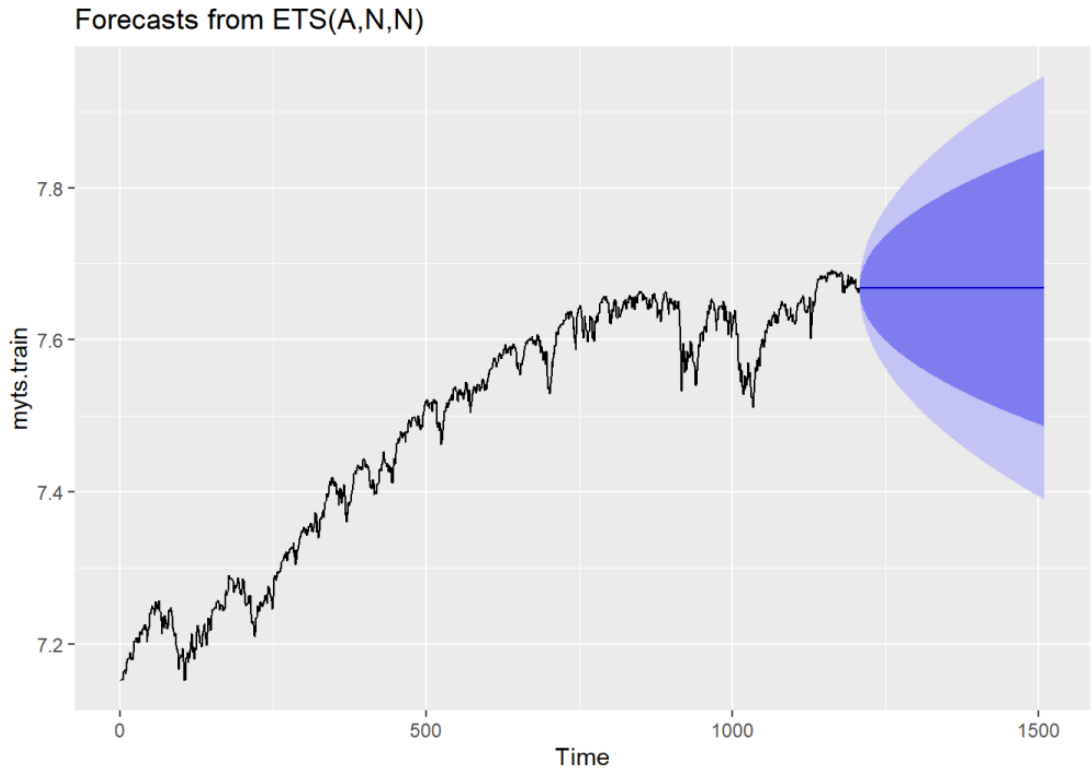
```
##                        ME        RMSE         MAE          MPE       MAPE
## Training set 2.505617e-05 0.008151126 0.005958462 0.0003575575 0.07964047
## Test set     4.282435e-02 0.048625278 0.044155729 0.5481007908 0.56549402
##                   MASE         ACF1 Theil's U
## Training set 0.9961046 3.288009e-05        NA
## Test set     7.3817245 9.702540e-01  10.75785
```

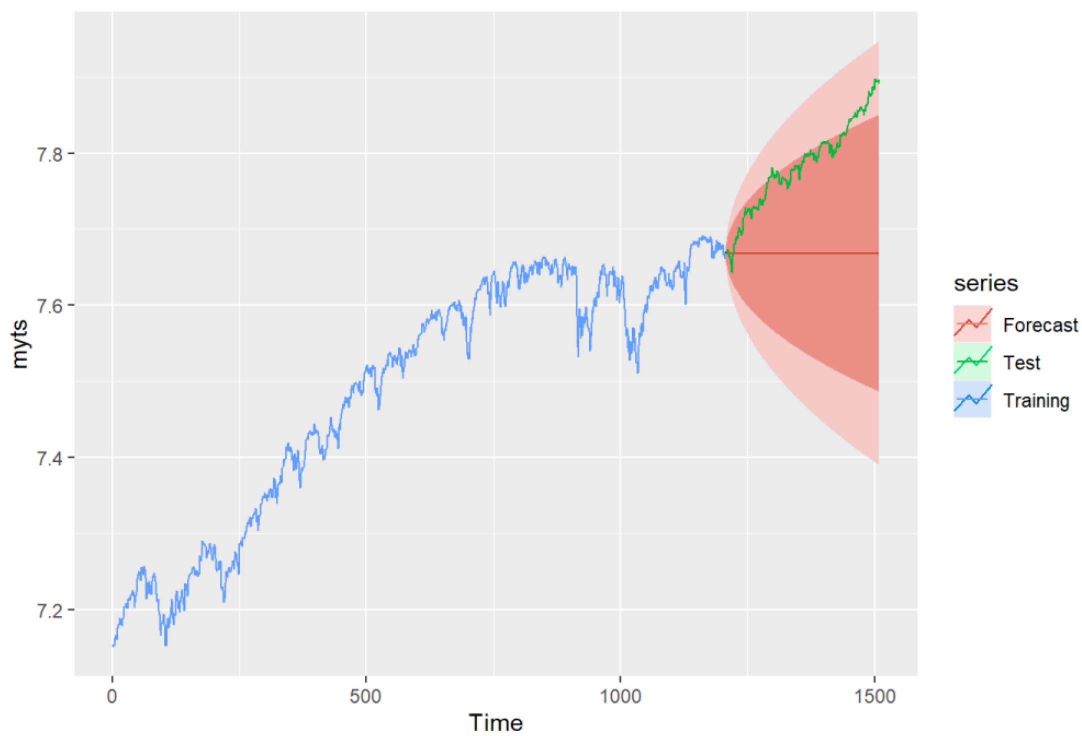
5.1.3.2 ETS Model

In addition to the ARIMA model, we can also use the ETS model to make a time series forecasting. Firstly, like ARIMA, we also use 80% and 20% to divide the data into training set and test set, as shown in the figure below.

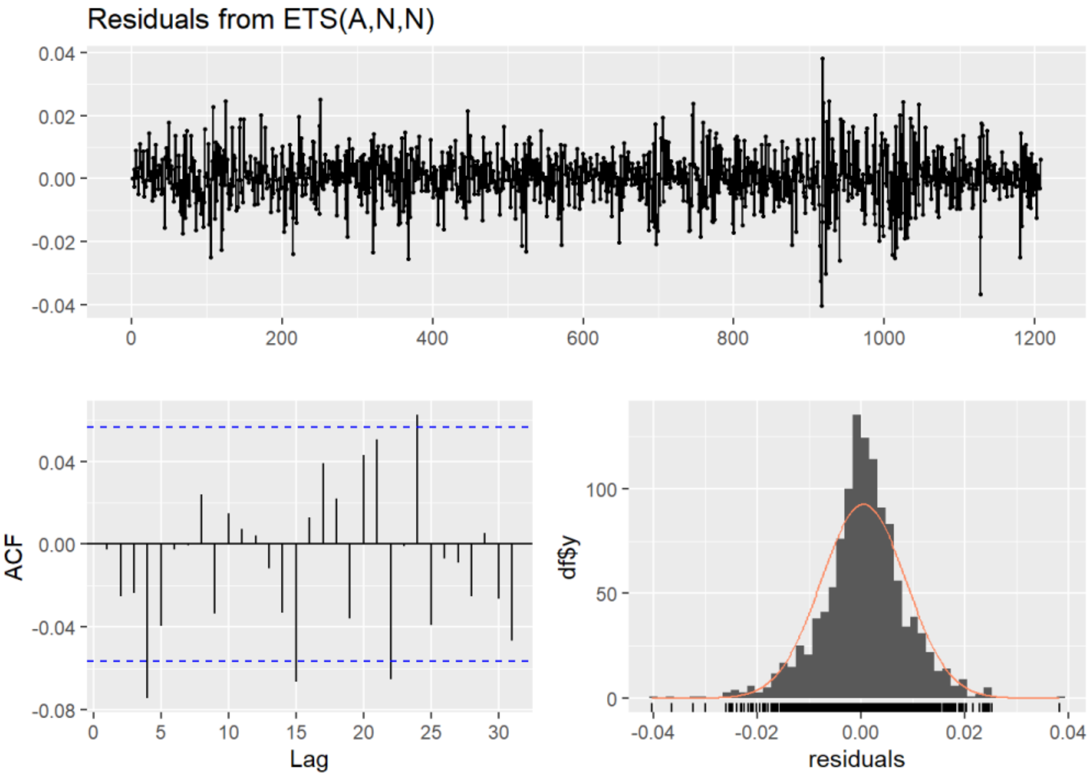


Then we use ETS model to analyze and predict the training set. From the graphs below, the forecast of the data by the ETS model may not so good, because some of the test set data are out of the main range (80% confidence interval) of the prediction.





Similarly, in order to demonstrate the effectiveness and accuracy of this model, we also adopted the methods of check residuals and accuracy here. In the following figures, the whole residuals basically show a uniform trend, with several prominent fluctuations. And residuals seem relatively normal. But ACF is not a white noise, since there are several lags out the the critical line. So ETS model's forecast of the data is feasible but not so good here.



Moreover, in evaluation metrics below, some of the RMSE, MAE, and MAPE values are low, but some are not low enough. For example, some are larger than 1. So, the effectiveness of the ETS model is not high enough here.

```
##                        ME        RMSE         MAE         MPE       MAPE
## Training set 0.0004295247 0.008188356 0.005975395 0.005738502 0.07985916
## Test set     0.1156707316 0.129583717 0.116344950 1.480446406 1.48926008
##                    MASE         ACF1 Theil's U
## Training set  0.9989354 -0.002532314        NA
## Test set     19.4499422  0.985232541  28.65392
```

5.1.3.3 Comparison of ARIMA model and ETS model

In conclusion, although both ARIMA model and ETS model can make reasonable and effective predictions on SPX500 data, we can also compare them to find the best model for forecasting SPX500 data. In order to compare the two models, the most efficient way is to compare the evaluation metrics, that is comparing the values of

RMSE, MAE, and MAPE between the ARIMA and ETS models. Lower values of RMSE, MAE, and MAPE indicate better forecasting accuracy. From the two-evaluation metrics here, all of the RMSE, MAE, and MAPE, and all other values of ARIMA model are lower than that of ETS model. And although the differences between some of them are very small and almost identical, if we have to choose an optimal one, based on most values, the ARIMA model is probably better than the ETS model here.

ARIMA model:

```
##                         ME        RMSE         MAE          MPE       MAPE
## Training set 2.505617e-05 0.008151126 0.005958462 0.0003575575 0.07964047
## Test set     4.282435e-02 0.048625278 0.044155729 0.5481007908 0.56549402
##                   MASE         ACF1 Theil's U
## Training set 0.9961046 3.288009e-05        NA
## Test set     7.3817245 9.702540e-01  10.75785
```

ETS model:

```
##                         ME        RMSE         MAE         MPE       MAPE
## Training set 0.0004295247 0.008188356 0.005975395 0.005738502 0.07985916
## Test set     0.1156707316 0.129583717 0.116344950 1.480446406 1.48926008
##                    MASE         ACF1 Theil's U
## Training set  0.9989354 -0.002532314        NA
## Test set     19.4499422  0.985232541  28.65392
```

5.1.4 Compare NASDAQ & DJIA & SPX

From 2012 to 2017, NASDAQ, DJIA, and SPX all trended upwards, showing a US bull market. But they had unique performances due to their different compositions. NASDAQ, packed with tech stocks, was more volatile, reflecting the high-risk, high-reward nature of the tech sector. DJIA, with 30 big blue-chip stocks, mirrors the US industrial scene. SPX, with a mix of 500 companies, provides a wider snapshot of the US economy.

5.2. The analysis of the Time Window

This report presents a comprehensive analysis on the impact of varying time window lengths, frequencies, and data averaging techniques on the performance of time series forecasting, specifically using the ARIMA model. The focus of our study is the S&P 500 index, which serves as a reliable indicator of the overall U.S. equity market performance.

The dataset used in our analysis spans from 2012 to 2017, a period rich in trends and fluctuations that lend themselves to time series analysis. To create a robust model and ensure a fair evaluation, the dataset has been partitioned into 80% for training and 20% for testing the ARIMA model.

The analysis is divided into three main aspects:

1.Impact of Time Window Length: The first part of our study concerns the influence of the length of the time window on the forecasting results. Two different time periods have been considered - the full period from 2012 to 2017, and a shorter one from 2014 to 2017. The ARIMA model's predictive power is examined for both these periods.

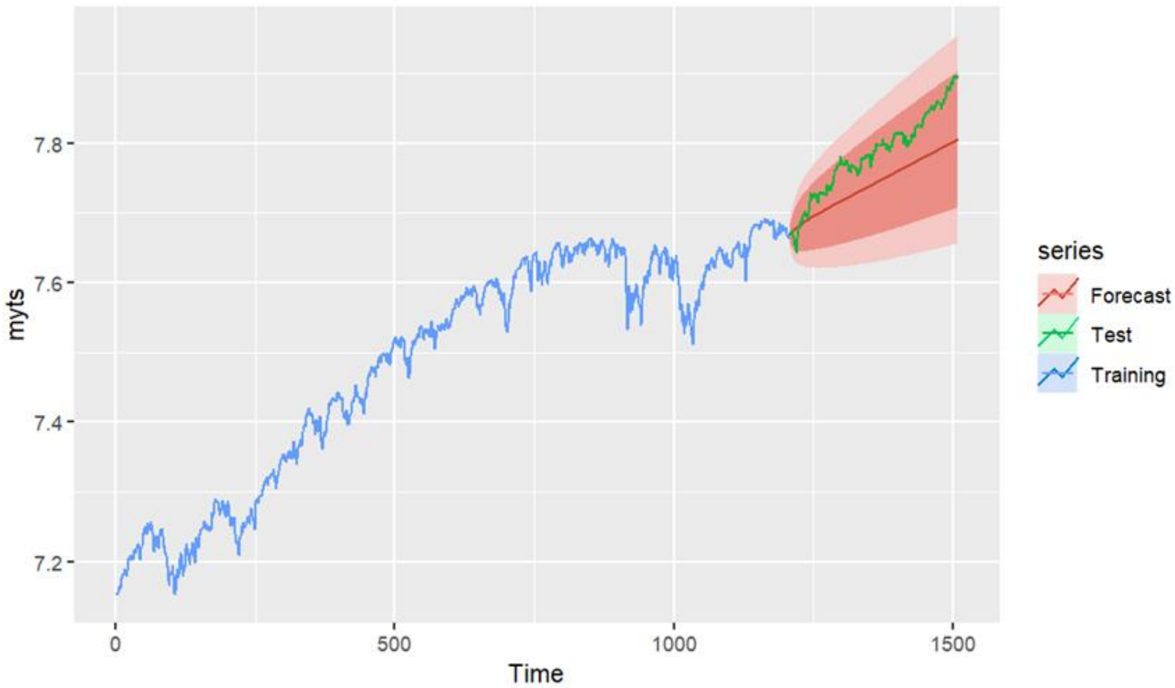
2.Impact of Time Window Frequency: The second part delves into the effects of data frequency. Three different frequencies - daily, weekly, and monthly - were analyzed using the full 2012-2017 dataset. The forecasts produced by the ARIMA model for each frequency were then compared.

3.Impact of Averaging: Lastly, the third part explores the influence of data averaging on the model's performance. Two weekly datasets from 2012 to 2017 were utilized - one created by selecting data every seven days and the other by averaging the data over each week.

By shedding light on these three aspects, this report seeks to enhance the understanding of how the choice of time window characteristics can influence the effectiveness of ARIMA-based time series forecasting in the context of equity market indices.

5.2.1. Impact of Time Window Length on Forecasting

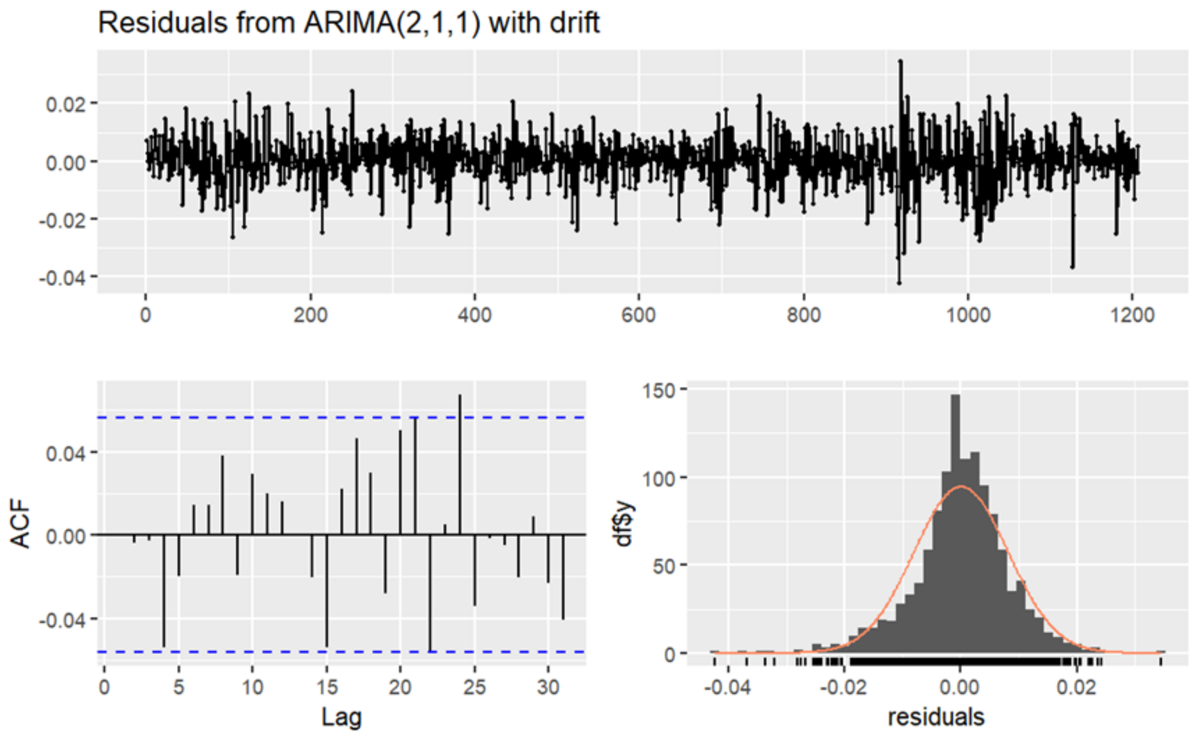
5.2.1.1Analysis of Five-Year Data (2012-2017)



Utilizing the data from 2012 to 2017, the ARIMA (2,1,1) model was applied after running the Auto.ARIMA function. The forecast graph successfully captured the trend of the test set. Remarkably, the test set data was entirely within the 80% confidence interval of the forecast, indicating a relatively precise prediction. And the area of confidence interval is small. The small size of the confidence interval area and the fact that the test set remained within the interval can be interpreted as low forecast uncertainty.

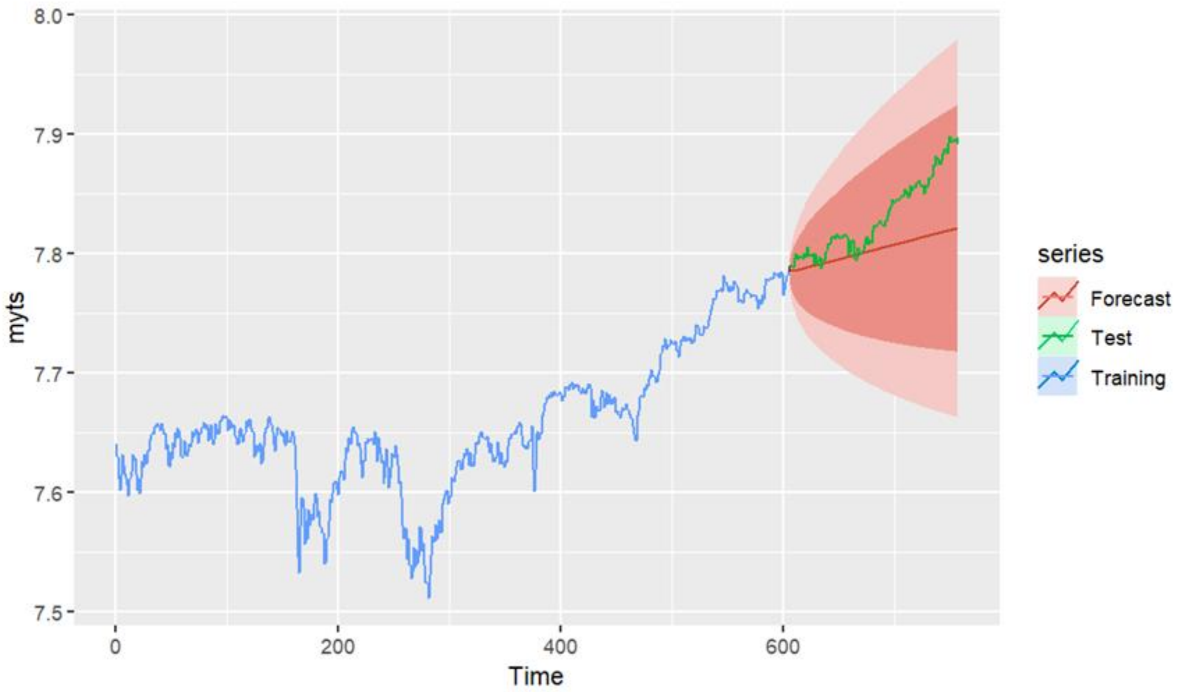
```
        Ljung-Box test

data:  Residuals from ARIMA(2,1,1) with drift
Q* = 7.7824, df = 7, p-value = 0.3522

Model df: 3.   Total lags used: 10

                       ME        RMSE         MAE          MPE       MAPE      MASE         ACF1 Theil's U
Training set 2.505617e-05 0.008151126 0.005958462 0.0003575575 0.07964047 0.9961046 3.288009e-05        NA
Test set     4.282435e-02 0.048625278 0.044155729 0.5481007908 0.56549402 7.3817245 9.702540e-01  10.75785
```

The Ljung-Box test resulted in a p-value of 0.3522, implying that the residuals are independent at all lags, and the model is well specified. In the Training set's RMSE is only 0.008151 and the test set's RMSE is 0.048625278



ACF plot analysis showed that apart from a minor exceedance at lag24, all other lags were within the critical value, which suggests that there is no autocorrelation in the residuals. Furthermore, residuals exhibited a near-normal distribution, which supports the suitability of the model.

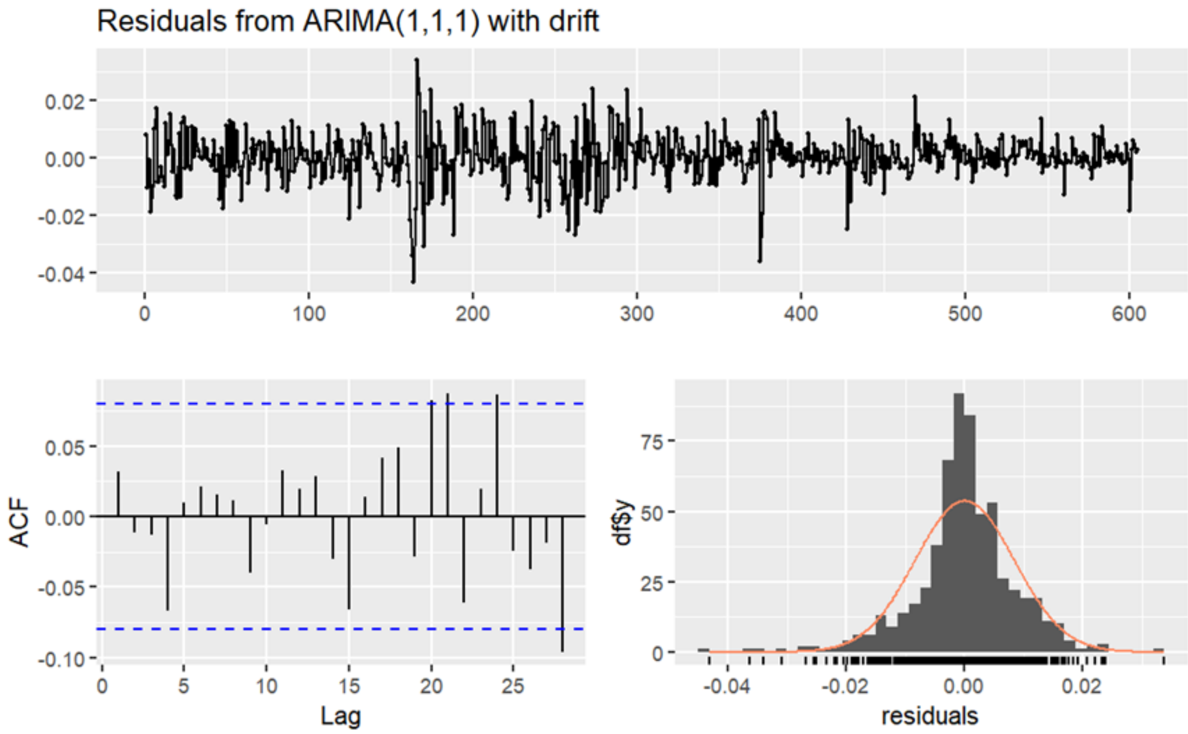
5.2.1.2. Analysis of Three-Year Data (2014-2017)



A different pattern was observed when analyzing the data from 2014 to 2017. This time, the ARIMA (1,1,1) model was applied. Although the 80% confidence interval still covered the entire test set, the forecast was less accurate compared to the five-year data, as indicated by the larger area of the confidence interval, suggesting higher forecast uncertainty.

```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,1) with drift
Q* = 5.1082, df = 8, p-value = 0.7459

Model df: 2.   Total lags used: 10

                       ME        RMSE         MAE           MPE      MAPE     MASE       ACF1 Theil's U
Training set 3.638195e-06 0.008429114 0.005928125 -0.0000557846 0.0776588 0.996595 0.03214627        NA
Test set     2.788398e-02 0.035589545 0.028078593  0.3549637326 0.3574610 4.720377 0.96752675  8.663283
```

The Ljung-Box test gave a p-value of 0.7459, indicating that the residuals are independent at all lags, and the model is adequately specified. The test set RMSE performance is even better than 5 years data, but the Training set’s RMSE is higher than 5 years which is 0.008429 indicating the underfitting.



The ACF plot showed that at lags 21, 24, and 28, the values exceeded the critical value, indicating some level of autocorrelation in the residuals, which can be a potential issue. Yet, the distribution of the residuals remained the shape of the normal distribution, but the lack of the enough data lead to the distribution doesn’t fit the normal distortion so well.

5.2.1.3. Comparison and Conclusion

In conclusion, our analysis demonstrates that the time window length significantly impacts the ARIMA model's predictive accuracy. In our specific case, a longer time window yielded better forecasting outcomes. This suggests that a longer historical record might provide more comprehensive information for the model to better understand the trend and cyclic patterns in the data.

However, it's crucial to note that in practical applications, extending the time window for your data can introduce more noise and potential outliers into your model, including unpredictable "black swan" events. These rare and severe occurrences can considerably affect market behavior, posing a challenge to forecasting models.

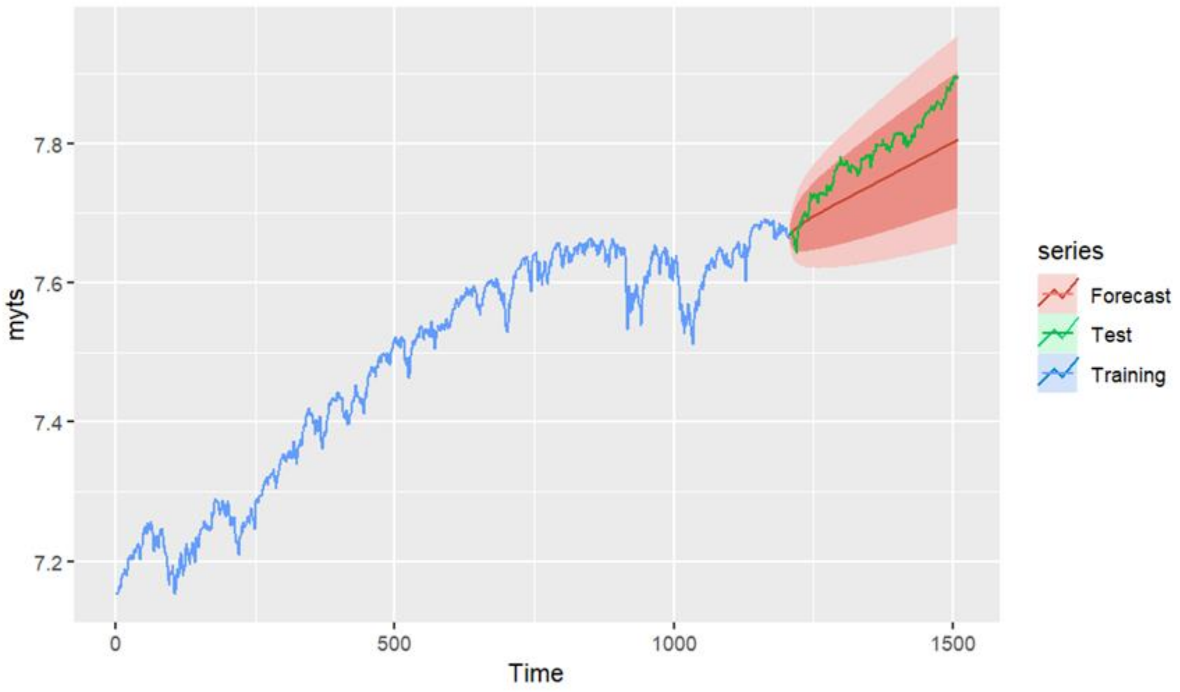
Moreover, an overly large dataset can also lead to prediction errors. It may incorporate more irrelevant or outdated information, thereby reducing the predictive power of the model. Furthermore, a larger dataset could span periods with different underlying market conditions. If market behavior changes over time, a model trained on an extensive historical dataset may not perform well in the current environment.

Therefore, while our analysis points to the benefits of a longer time window, it's essential to balance this against the potential risks and challenges. This includes the introduction of black swan events and shifts in market behavior that could impact the model's performance. Ultimately, the optimal time window length will depend on the specific characteristics of the dataset and the forecasting goals.

This underscores the importance of a comprehensive understanding of your data and the necessity of conducting robust sensitivity analyses. Such measures ensure the determination of the most suitable model and parameters for each specific forecasting task, offering more reliable and accurate predictions.

5.2.2. Impact of Time Window Frequency on Forecasting

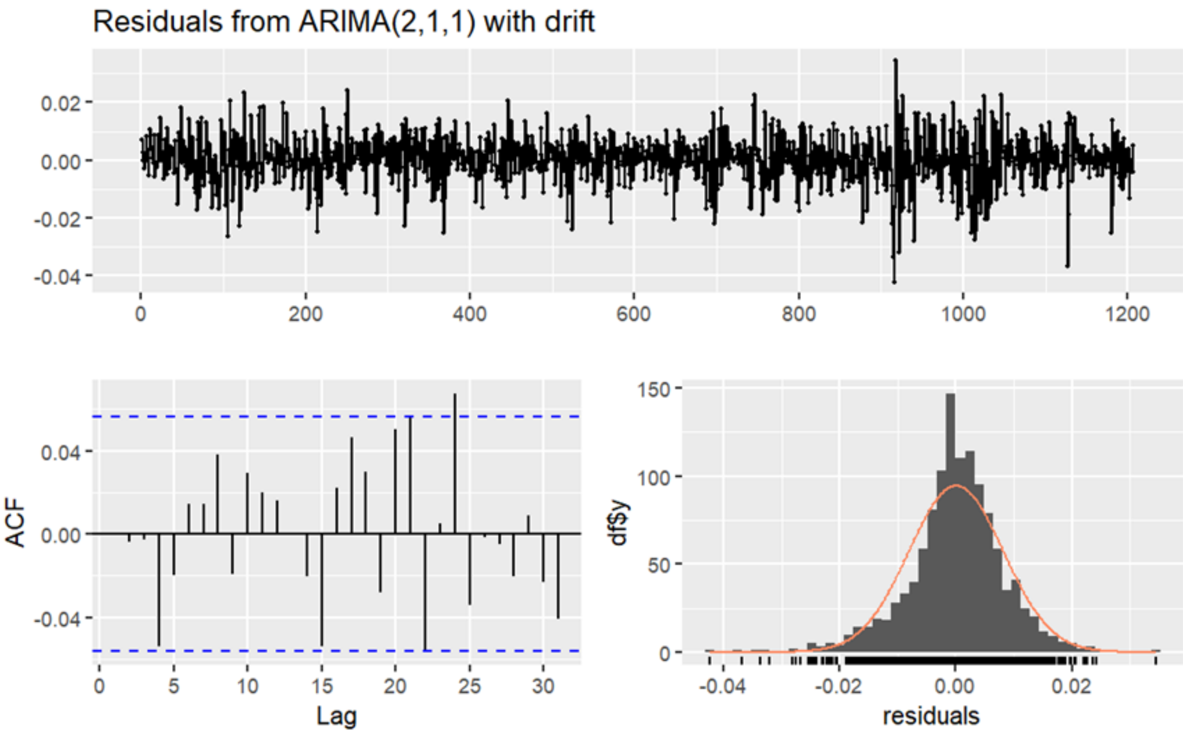
5.2.2.1Analysis of Daily Data



This daily data from 2012 to 2017 we have already analyzed it in 5 years data part, so that we will just simply repeat the analysis. The ARIMA (2,1,1) model was applied after running the Auto.Arima function. The forecast graph successfully captured the trend of the test set. Remarkably, the test set data was entirely within the 80% confidence interval of the forecast, indicating a relatively precise prediction. And the area of confidence interval is small. The small size of the confidence interval area and the fact that the test set remained within the interval can be interpreted as low forecast uncertainty.

```
        Ljung-Box test

data:  Residuals from ARIMA(2,1,1) with drift
Q* = 7.7824, df = 7, p-value = 0.3522

Model df: 3.   Total lags used: 10

                       ME        RMSE         MAE          MPE       MAPE      MASE         ACF1 Theil's U
Training set 2.505617e-05 0.008151126 0.005958462 0.0003575575 0.07964047 0.9961046 3.288009e-05        NA
Test set     4.282435e-02 0.048625278 0.044155729 0.5481007908 0.56549402 7.3817245 9.702540e-01  10.75785
```

The Ljung-Box test resulted in a p-value of 0.3522, implying that the residuals are independent at all lags, and the model is well specified. In the Training set's RMSE is only 0.008151 and the test set's RMSE is 0.048625278



ACF plot analysis showed that apart from a minor exceedance at lag24, all other lags were within the critical value, which suggests that there is no autocorrelation in the residuals. Furthermore, residuals exhibited a near-normal distribution, which supports the suitability of the model.

5.2.2.2. Analysis of Weekly Data

The forecast of the weekly data has lots of similarities. The predictive model begins by indicating values higher than the test parameter within this timeframe, adopting an approximate 30-degree slope. In subsequent forecasts, the predicted values drop below the test values. Nevertheless, the general trend remains consistent, and given that the projections fall within the confidence interval, these predictions can be deemed credible and reasonable.

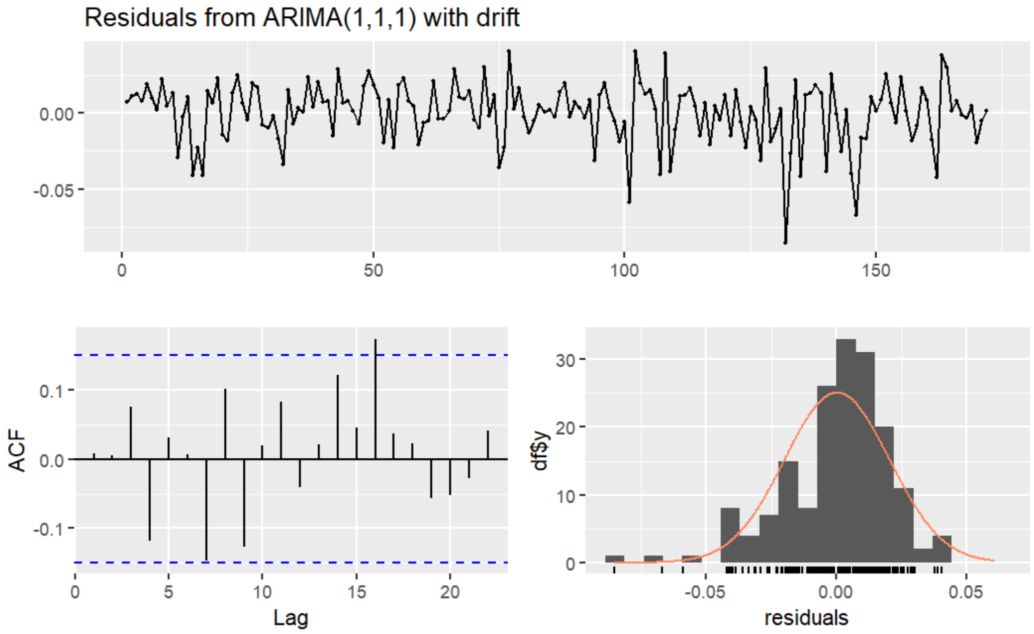
```
Ljung-Box test
data:  Residuals from ARIMA(1,1,1) with drift
Q* = 12.448, df = 8, p-value = 0.1323
Model df: 2.   Total lags used: 10
                      ME       RMSE        MAE         MPE      MAPE      MASE        ACF1 Theil's U
Training set 0.0001713824 0.02012570 0.01526111 0.002466202 0.2039130 0.9591801 0.008511103        NA
Test set     0.0319201239 0.04008905 0.03608085 0.407990798 0.4623216 2.2677271 0.796134121  3.165462
```

The test statistic Q* = 12.448, with degrees of freedom (df) = 8, yielding a p-value of 0.1323. The p-value is greater than the common significance level (0.05), suggesting that the null hypothesis cannot be rejected. This implies that the residuals are independently distributed, and there's no significant autocorrelation, indicating a good fit of the ARIMA model.

The 'Model df' is 2, which refers to the number of parameters in the ARIMA model, excluding the variance. The total lags used in this test is 10, which is typically chosen as a small multiple of the number of observations to the power 1/4.

Looking at the metrics for the training set: The ME is near zero, which implies that there is no persistent tendency for the forecasts to over- or under-forecast. The RMSE is quite small (0.02012570), indicating a model with good accuracy. The MAE is 0.01526111, which is low, showing that the model has minimal errors. The MPE and MAPE are also quite low, which suggests that the forecast errors, in percentage terms, are small. The MASE being less than 1 is also a good sign.

For the test set, the error metrics (ME, RMSE, MAE, MPE, MAPE, MASE) are higher than the training set, suggesting the model does not perform as well on unseen data as it does on the training set. This may be a sign of overfitting. A high Theil's U statistic (>1) indicates that the model's forecast is worse than a naive forecast, suggesting the model may need to be reassessed for its predictive power on future data.



The residual plot exhibits minimal fluctuations within a tight range of ±0.1 over time. This minimal dispersion indicates a good consistency in the model prediction errors.

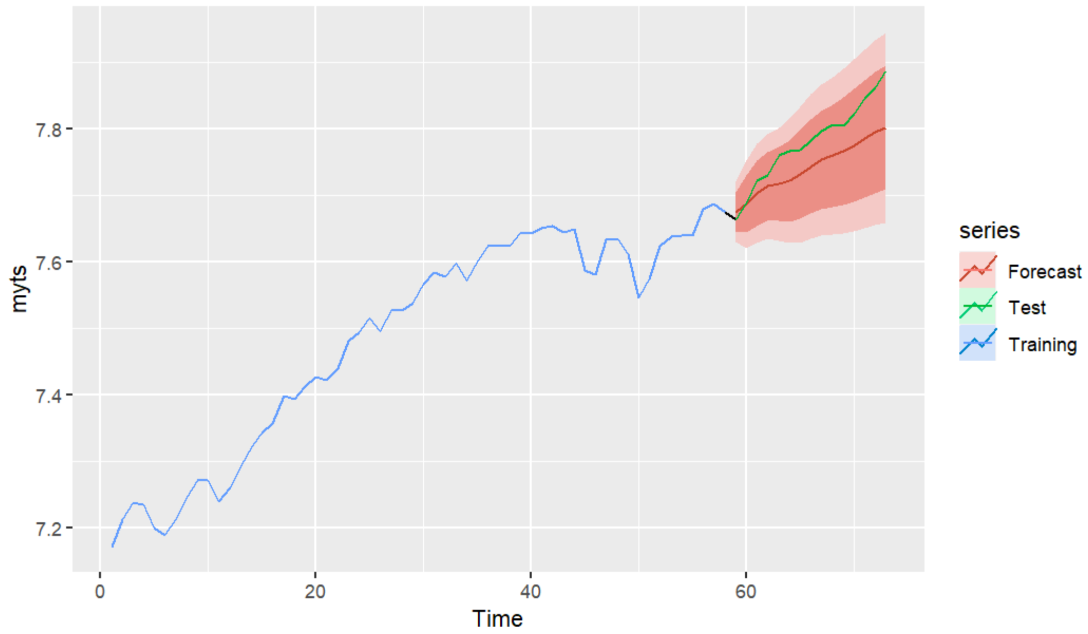
However, we note that the residuals show no apparent patterns of seasonality or periodicity, indicating the model has effectively captured underlying patterns in the dataset without overfitting to time-bound phenomena. The ACF provides a measure of the correlation between the residuals at different intervals or 'lags'. Our analysis indicates that the majority of lags fall within the critical value, signifying a desirable level of randomness in the errors. In a well-specified model, we would expect no correlation between errors at different time points, therefore residuals should not be significantly different from zero. This is mostly observed; however, the model does exhibit a slightly significant autocorrelation at lag 16. Despite the overall sound performance of the model, there are some areas of concern.

The histogram of the residuals indicates that the error distribution deviates somewhat from the normal distribution, skewing slightly to the right, and the lack of the enough data set lead to the bad fitting to the bell curve. This positive skewness suggests that there are more instances of overestimation than underestimation in our model’s predictions, which can be a source of systematic error in the model.

Overall, our model exhibits a relatively strong performance, as indicated by the minimal fluctuations in residuals, the lack of discernible seasonality or periodicity, and the largely negligible autocorrelations at most lags. Nonetheless, we recognize the minor concern of a slight significant autocorrelation at lag 16, and the positive skewness of the residual distribution. Although the model appears to be a generally good fit for the data, these identified areas suggest that further refinement may improve the model's prediction capabilities. Therefore, it might be beneficial to explore additional model specifications or transformations to address these residual characteristics and improve model performance."

Remember, while this analysis suggests that the model is generally well-specified, the final assessment should be based on whether the model meets the specific needs of the research question or application at hand. Even a well-fitted model might not be appropriate if it does not meet the specific assumptions and requirements of the study or application.

5.2.2.3. Analysis of Monthly Data



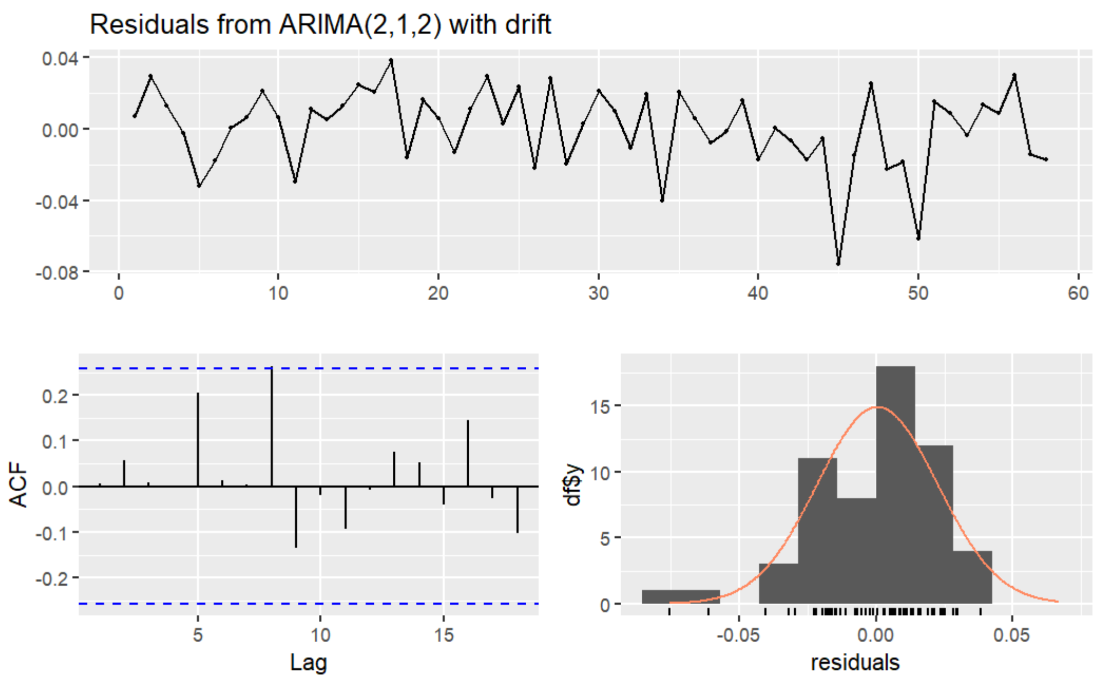
The forecast of the monthly data has shown notable effectiveness in capturing the growth trend of the asset, as well as a certain degree of fluctuation. The predictive model employed is ARIMA (2,1,2) chosen by the Auto Arima function, which is similar to the models used for daily and weekly data.

```
        Ljung-Box test

data:  Residuals from ARIMA(2,1,2) with drift
Q* = 9.1291, df = 6, p-value = 0.1664

Model df: 4.   Total lags used: 10

                       ME       RMSE        MAE         MPE      MAPE      MASE        ACF1 Theil's U
Training set 0.0004721178 0.02190386 0.01722273 0.007072524 0.2300238 0.8446693 0.006660298        NA
Test set     0.0388287037 0.04547456 0.04033658 0.496753370 0.5164284 1.9782615 0.604087687   2.45279
```

Ljung-Box test was performed on this model, yielding a Q* statistic of 9.1291 and degrees of freedom (df) = 6, resulting in a p-value of 0.1664. This p-value is higher than the common significance level of 0.05, suggesting that the residuals are independently distributed, and there's no significant autocorrelation. This indicates a good fit of the ARIMA model. The 'Model df' is 4, referring to the number of parameters in the ARIMA model, excluding the variance.

Looking at the metrics for the training set: The ME is near zero, which implies that there is no persistent tendency for the forecasts to over- or under-forecast. The RMSE is small (0.02190386), indicating a model with good accuracy. The MAE is 0.01722273, which is low, showing that the model has minimal errors. The MPE and MAPE are also relatively low, suggesting that the forecast errors, in percentage terms, are small. The MASE being less than 1 is a good sign.

The test set, however, showed a different story. The error metrics (ME, RMSE, MAE, MPE, MAPE, MASE) are higher than the training set, suggesting the model does not perform as well on unseen data as it does on the training set. This could be an indication of overfitting. A high Theil's U statistic (2.45279) suggests that the model's forecast is worse than a naive forecast, and the model may need reassessment for its predictive power on future data.



Regarding the residual plot and ACF analysis, it showed that most lags fall within the critical value, implying desirable randomness in the errors. However, due to the decreased frequency in the monthly data, the model shows some inadequacy in fitting the normal distribution, as evident from the residual histogram.

Despite these shortcomings, the overall performance of the model is relatively strong, capturing the general trend and fluctuations effectively. Nonetheless, the minor concern of the model potentially overfitting and not performing well with unseen data, and the slight deviation from the normal distribution in the residual histogram, suggest that there's room for improvement in the model's predictive capabilities. Further refinement of the model specifications or data transformations may improve its performance.

5.2.2.3. Comparison and Conclusion

In this analysis, we have evaluated the performance of ARIMA models on time series data at three different frequencies - daily, weekly, and monthly. This has provided us with valuable insights into the influence of data frequency on predictive modeling.

It's evident from our findings that reducing the frequency of the data has a discernible optimization effect on the forecast. Moving from daily to weekly and monthly data, the model complexity decreases, leading to a simpler model structure that is easier to interpret and computationally more efficient.

However, this reduction in frequency and resulting simplicity comes with its own set of challenges. As the frequency decreases from daily to weekly and then to monthly, the data becomes less dense. This decrease in data density leads to a diminished ability of the model to capture the more nuanced trends and fluctuations in the data. For the monthly data model in our study, this was observed as a notable deviation from normality in the residual distribution and worse performance on the unseen test data. These signs are indicative of potential underfitting, where the model, due to insufficient data, fails to capture the underlying structure of the data adequately.

Furthermore, despite the simplicity and interpretability advantages, the reduced frequency models exhibited minor exceedances at certain lags and slightly higher error metrics on the test set, suggesting the models could not fully capture the data patterns. This implies that while reducing data frequency simplifies the model and optimizes the forecasting process, it could lead to suboptimal performance if the data density becomes too low.

In conclusion, while reducing data frequency can offer certain benefits, it is crucial to strike a balance between simplicity and the ability to adequately capture the data's inherent patterns. A model's effectiveness is contingent on its capacity to capture the essential details without becoming overly complex or overly simplified. Future work should focus on optimizing this balance to ensure that any gains in computational efficiency or interpretability do not come at the cost of predictive accuracy.5.2.3. Impact of Averaging

5.2.3. Impact of Time Window Averaging on Forecasting

5.2.3.1. without average

The predictive model begins by indicating values higher than the test parameter within this timeframe, adopting an approximate 30-degree slope. In subsequent forecasts, the predicted values drop below the test values. Nevertheless, the general trend remains consistent, and given that the projections fall within the confidence interval, these predictions can be deemed credible and reasonable.

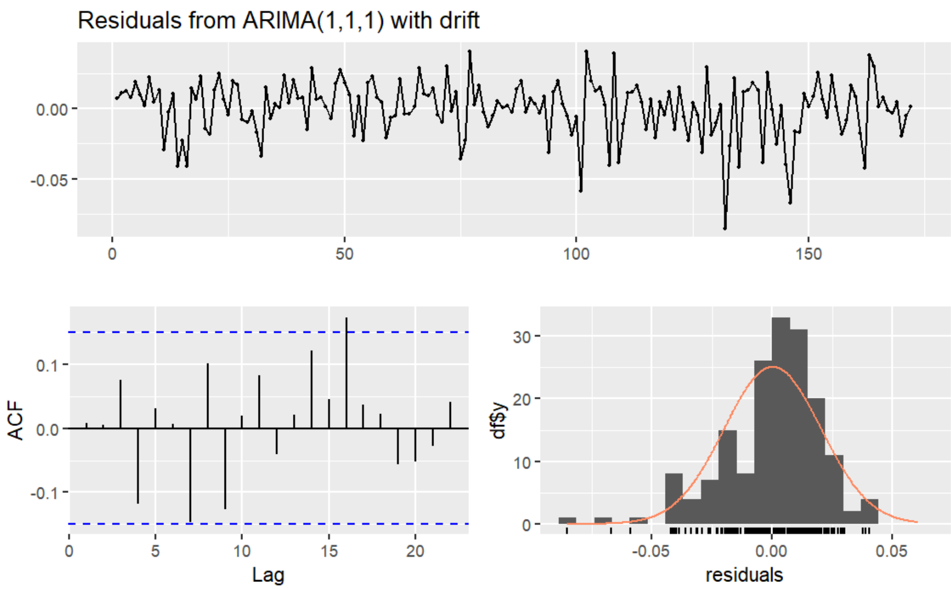
```
Ljung-Box test
data:  Residuals from ARIMA(1,1,1) with drift
Q* = 12.448, df = 8, p-value = 0.1323
Model df: 2.   Total lags used: 10
                       ME       RMSE        MAE         MPE      MAPE      MASE         ACF1 Theil's U
Training set 0.0001713824 0.02012570 0.01526111 0.002466202 0.2039130 0.9591801 0.008511103        NA
Test set     0.0319201239 0.04008905 0.03608085 0.407990798 0.4623216 2.2677271 0.796134121  3.165462
```

The test statistic Q* = 12.448, with degrees of freedom (df) = 8, yielding a p-value of 0.1323. The p-value is greater than the common significance level (0.05), suggesting that the null hypothesis cannot be rejected. This implies that the residuals are independently distributed, and there's no significant autocorrelation, indicating a good fit of the ARIMA model.

The 'Model df' is 2, which refers to the number of parameters in the ARIMA model, excluding the variance. The total lags used in this test is 10, which is typically chosen as a small multiple of the number of observations to the power 1/4.

Looking at the metrics for the training set: The ME is near zero, which implies that there is no persistent tendency for the forecasts to over- or under-forecast. The RMSE is quite small (0.02012570), indicating a model with good accuracy. The MAE is 0.01526111, which is low, showing that the model has minimal errors. The MPE and MAPE are also quite low, which suggests that the forecast errors, in percentage terms, are small. The MASE being less than 1 is also a good sign.

For the test set, the error metrics (ME, RMSE, MAE, MPE, MAPE, MASE) are higher than the training set, suggesting the model does not perform as well on unseen data as it does on the training set. This may be a sign of overfitting. A high Theil's U statistic (>1) indicates that the model's forecast is worse than a naive forecast, suggesting the model may need to be reassessed for its predictive power on future data.

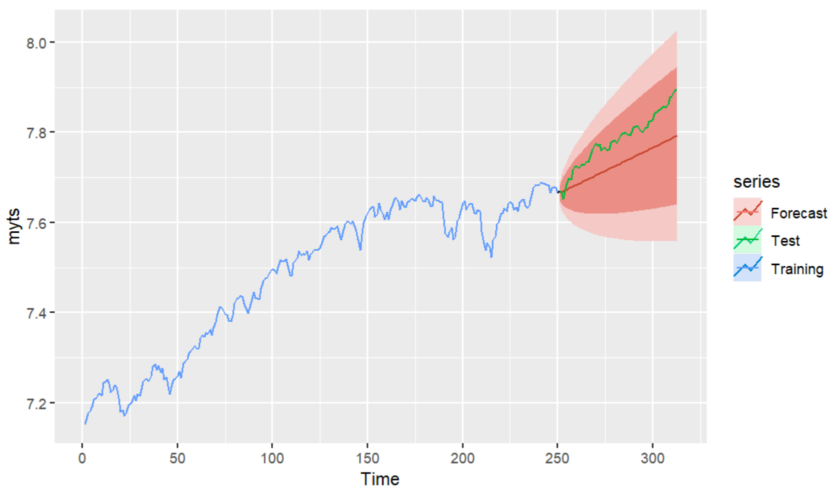


The residual plot exhibits minimal fluctuations within a tight range of ±0.1 over time. This minimal dispersion indicates a good consistency in the model prediction errors. However, we note that the residuals show no apparent patterns of seasonality or periodicity, indicating the model has effectively captured underlying patterns in the dataset without overfitting to time-bound phenomena. The ACF provides a measure of the correlation between the residuals at different intervals or 'lags'. Our analysis indicates that the majority of lags fall within the critical value, signifying a desirable level of randomness in the errors. In a well-specified model, we would expect no correlation between errors at different time points, therefore residuals should not be significantly different from zero. This is mostly observed; however, the model does exhibit a slightly significant autocorrelation at lag 16. Despite the overall sound performance of the model, there are some areas of concern. The histogram of the residuals indicates that the error distribution deviates somewhat from the normal distribution, skewing slightly to the right. This positive skewness suggests that there are more instances of overestimation than underestimation in our model’s predictions, which can be a source of systematic error in the model.

Overall, our model exhibits a relatively strong performance, as indicated by the minimal fluctuations in residuals, the lack of discernible seasonality or periodicity, and the largely negligible autocorrelations at most lags. Nonetheless, we recognize the minor concern of a slight significant autocorrelation at lag 16, and the positive skewness of the residual distribution. Although the model appears to be a generally good fit for the data, these identified areas suggest that further refinement may improve the model's prediction capabilities. Therefore, it might be beneficial to explore additional model specifications or transformations to address these residual characteristics and improve model performance."

Remember, while this analysis suggests that the model is generally well-specified, the final assessment should be based on whether the model meets the specific needs of the research question or application at hand. Even a well-fitted model might not be appropriate if it does not meet the specific assumptions and requirements of the study or application.

5.2.3.2 with average

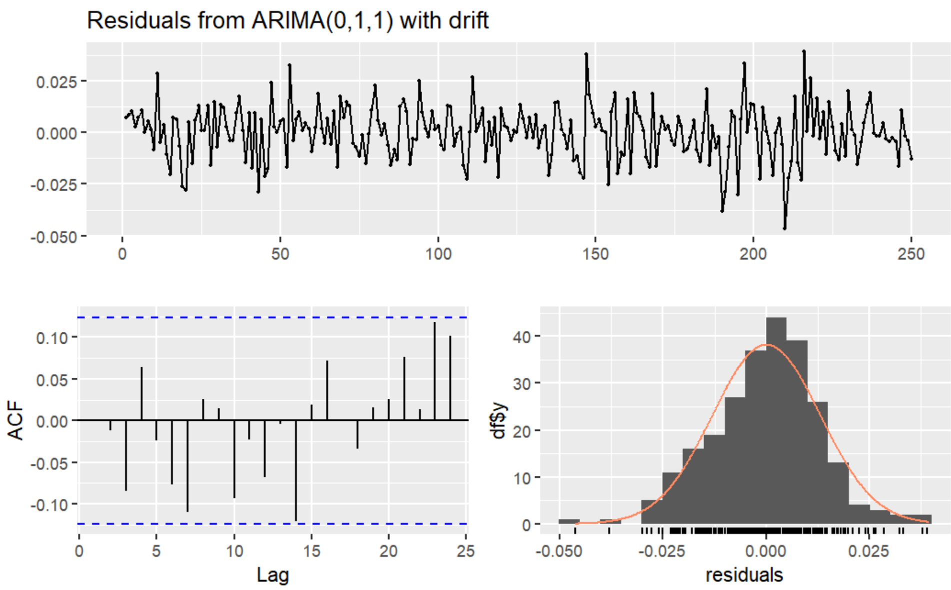


Upon inspection of the visual representation of the model's predictive performance, it's apparent that our predictions fall within the 80% confidence interval. This suggests that the model is successful in capturing the central tendency of the data with a reasonable level of accuracy. Moreover, the model's predictive trajectory aligns with the underlying trend in the data, showing an upward inclination similar to the actual test values.

However, there is a slight underestimation in the model's predictions relative to the actual test values. This implies that while the model has a good grasp of the general trend, it may be systematically under-predicting the response variable. Nonetheless, this underestimation is minor and the overall correlation between the predicted and actual values remains strong.

```
Ljung-Box test
data:  Residuals from ARIMA(0,1,1) with drift
Q* = 10.189, df = 9, p-value = 0.3354
Model df: 1.   Total lags used: 10
                       ME       RMSE        MAE          MPE      MAPE      MASE        ACF1 Theil's U
Training set 2.464104e-05 0.01310248 0.01010023 0.0004639407 0.1350363 0.9529594 0.001821072        NA
Test set     5.429537e-02 0.05926211 0.05496389 0.6954121511 0.7041479 5.1858576 0.883141483  7.104924
```

In analyzing the performance of the ARIMA (0,1,1) model with drift on the dataset, the Ljung-Box test yielded a p-value of 0.3354, suggesting that the residuals are independently distributed and showing no evidence of autocorrelation, a desirable characteristic of a well-performing time-series model. However, when examining model fit and accuracy, the situation is less straightforward. While the ME and RMSE for the training set suggest a good average performance and low dispersion respectively, these metrics increased noticeably when applied to the test set, indicating a reduction in predictive accuracy and increased error. This is further corroborated by larger MPE and MAPE values for the test set, which suggest an average prediction deviation of around 70% from the actual values. The MASE and Theil's U statistics are considerably greater than 1, implying that the model performs worse than a simple, naïve forecast on the test set. Moreover, the high ACF1 on the test set may hint at some unaccounted structure in the residuals. In conclusion, while the model exhibits satisfactory performance on the training set, its accuracy diminishes on the test set, indicating that further refinement and consideration of other models or additional feature engineering could be beneficial to improve its performance on unseen data.



The residuals in our analysis show minimal variance, fluctuating within the tight range of ±0.05, with most residing between ±0.025. This low variability indicates that our model provides consistent predictions with small error margins, which is a desirable property for any predictive model. Moreover, our residuals do not exhibit noticeable seasonal or cyclical fluctuations, suggesting that our model has adequately captured these patterns in the data. This is a positive attribute, as it indicates that the model has not overfitted to time-bound phenomena and will likely perform well on out-of-sample predictions. Evaluating the ACF of the residuals, we find that all lag values fall within the critical value, indicative of white noise distribution. In other words, the residuals show no autocorrelation, further strengthening the assertion that our model does not overfit and is capturing the inherent structure in the data accurately. We can conclude that after taking the average of the data, the ACF values fall within the critical value, indicating a reduced forecast error. This essentially means that our model is more efficient; the error between the predicted and actual values is smaller. This lower deviation makes our model more reliable, enhancing its predictive accuracy and stability over time. Upon examination of the histogram of residuals, we observe a near-normal distribution. This implies that the residuals are symmetrically distributed around zero, suggesting unbiased predictions from our model. However, a point of concern lies in the vertices of the histogram not aligning perfectly with those of the normal distribution curve, indicating a minor deviation from the ideal bell-curve pattern.

Overall, the analysis of residuals indicates that our model performs admirably, providing consistent predictions with minimal errors, adequately capturing seasonal and cyclical fluctuations, and ensuring negligible autocorrelation. The slight misalignment of the vertices in the histogram of residuals with the ideal normal distribution highlights the need for improvement. While this model fits the data well, this small discrepancy suggests potential refinement avenues to further enhance model performance.

5.2.3.3 comparison and conclusion

When introducing the process of averaging to our data analysis, we observed further enhancements in our model performance. Post-averaging, the fluctuations and errors of the numbers are noticeably reduced, resulting in a smoother series. This is closely aligned with the expectation of normal distribution, thereby leading to improved forecast predictions.

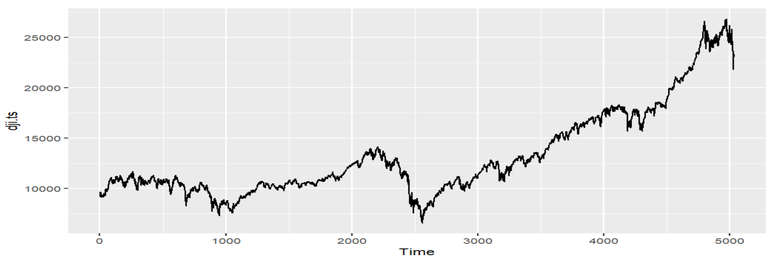
The reason behind the effective results generated by the average lies in its inherent nature as a data summary technique. Averaging is an act of distilling multiple data points into a singular piece of information, an 'average.' So, instead of dealing with seven days' worth of data separately, we combine them to yield a single averaged value.

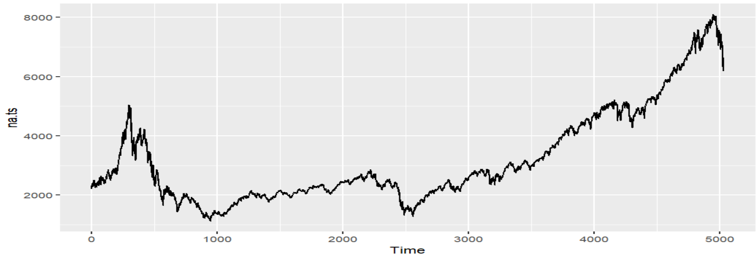
This practice translates into taking only one data point out of a week-long sample. This transformation of data naturally brings us to the concept of the Central Limit Theorem (CLT). The Central Limit Theorem is a fundamental principle in statistics, stating that the distribution of the sum (or average) of a large number of independent, identically distributed variables approaches a normal distribution, regardless of the shape of the original distribution. This explains why our averaged data, derived from multiple individual data points, better conforms to the normal distribution, increasing consistency and predictive accuracy of our model.

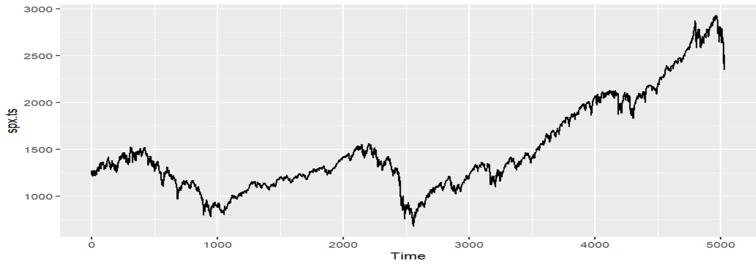
In blending the aforementioned analysis with the prior comparison of models based on random sampling and averaging, it becomes clear that both models present with commendable qualities. They exhibit consistent predictions, minimal autocorrelation, and satisfactory capture of seasonal and cyclical fluctuations. However, the process of averaging grants an edge in maintaining predictive accuracy across unseen data and better compliance with the normal distribution, thereby making it slightly superior. This does not negate the potential for improvement in both models, particularly concerning the minor deviation from the ideal bell-curve pattern observed in the residual histograms. Hence, future exploration may involve additional model specifications or transformations to further refine performance.

6. Discussion

6.1 The findings in the three Stock indexes and Comparison of the ETS Model and ARIMA Model







The observed trends across the board bear striking similarities. From the year 2012 through 2017, all three indices - NASDAQ, DJIA, and SPX - demonstrated an overarching upward trajectory. This pattern mirrors the bullish trend that pervaded the U.S. stock market during this period. However, the specific performances of these indices varied, largely influenced by their individual compositions.

The volatility exhibited by each index is distinct. NASDAQ, DJIA, and SPX each present different volatility profiles. Given that NASDAQ is primarily constituted of tech stocks, it tends to exhibit greater volatility compared to the DJIA and SPX. There are numerous factors contributing to this phenomenon, one of the most significant being the inherent growth nature of tech stocks. These stocks often carry the potential for high risks and high returns, which in turn may induce greater price volatility.

Moreover, each index represents different industry sectors. NASDAQ is predominantly composed of tech stocks, while the DJIA incorporates 30 substantial blue-chip stocks, thereby reflecting the U.S. industrial economy. On the other hand, the SPX encompasses 500 large and medium-sized companies that span across various industries in the U.S. As a result, it provides a more comprehensive reflection of the U.S. economic situation.

In conclusion, while the three indices share a common upward trend, they diverge in terms of volatility and industry representation. This variance underscores the importance of understanding the unique characteristics and implications of each index when interpreting their movements and making investment decisions.

6.2. The Time windows influences

Our analyses have indicated that the time window length exerts a significant influence on the predictive accuracy of ARIMA models, underscoring the pivotal role of time-series data's frequency, averaging, and duration in the performance of forecasting models.

The frequency of data emerged as a crucial factor in optimizing the model's performance. While reducing the frequency from daily to weekly or monthly simplifies the model and makes it more interpretable and computationally efficient, it may also lead to underfitting due to a decrease in data density. This decreased density limits the model's ability to capture nuanced trends and fluctuations in the data, potentially leading to inaccuracies in forecasts.

In considering the method of averaging, we compared models built on randomly sampled and averaged data. Averaging has the advantage of not only smoothing the data but also, due to the central limit theorem, ensuring that residuals are more normally distributed and can reduce the autocorrelation of residuals. Both the randomly sampled and averaged data models performed consistently, capturing the inherent patterns in the data effectively and ensuring minimal autocorrelation. However, the model based on averaged data showcased superior performance when dealing with unseen data, indicating its edge in predictive accuracy.

The length of the time window was another influential factor. Extending the time window for data can improve forecasting outcomes by providing the model with a more comprehensive historical record. However, care must be taken to balance the benefits of a longer time window against potential pitfalls, such as the introduction of more noise, outliers, and outdated information into the model.

In conclusion, the influence of the time window on ARIMA models is multifaceted, encompassing the frequency, the averaging method—particularly its role in smoothing data and reducing autocorrelation—and the duration of the time series data. Each of these factors has its own set of benefits and challenges, and finding a balance is essential for optimizing predictive accuracy. Future research should explore optimal practices in selecting the time window for specific forecasting tasks, keeping these factors in mind. The ultimate goal remains to ensure reliable and accurate predictions without compromising on the model's interpretability or computational efficiency.

7. Conclusion

In conclusion, this study has unveiled noteworthy findings regarding the behavior of the three primary stock indices - NASDAQ, DJIA, and SPX, and the predictive accuracy of ARIMA models under varying conditions. All three indices exhibited a consistent upward trend from 2012 to 2017, reflective of the overall bullish momentum in the U.S. stock market. However, each index presented unique volatility profiles and sectoral representations. The NASDAQ, heavily populated by tech stocks, demonstrated higher volatility, whereas the DJIA and SPX, which mirror the U.S. industrial economy and a broader range of industries, respectively, showed differing degrees of fluctuations. These findings underline the importance of understanding the distinct attributes of each index before making investment decisions.

Furthermore, our exploration into the impact of the time window on ARIMA models revealed its significance in affecting the model's predictive accuracy. The frequency of data, averaging method, and duration of the time series data each play pivotal roles in the performance of the model. While reducing data frequency can increase model efficiency and interpretability, it might also lead to underfitting due to lower data density. The averaging method, on the other hand, smoothens the data and reduces autocorrelation, with models built on averaged data showing superior predictive capabilities on unseen data.

Finally, while longer time windows can enhance forecast outcomes by offering a more comprehensive historical perspective, potential issues such as noise, outliers, and outdated information must be considered. Thus, a careful balance among these factors is essential to optimize predictive accuracy while preserving computational efficiency and interpretability. Future research can extend this work by establishing optimal practices for selecting time windows specific to various forecasting tasks, ultimately aiming for reliable and accurate predictions that can aid investors and market analysts in their decision-making process.

Reference

Yahoo Finance. (2023, July 17). S&P 500 Index (^GSPC) Stock Price, News, Quote & History. Retrieved July 17, 2023, from https://finance.yahoo.com/quote/%5EGSPC/

Yahoo Finance. (2023, July 17). Dow Jones Industrial Average (^DJI) Stock Price, News, Quote & History. Retrieved July 17, 2023, from https://finance.yahoo.com/quote/%5EDJI?p=%5EDJI

Yahoo Finance. (2023, July 17). NASDAQ Composite (^IXIC) Stock Price, News, Quote & History. Retrieved July 17, 2023, from https://finance.yahoo.com/quote/%5EIXIC

Chatfield, C. (1984) 'Introduction', The Analysis of Time Series: An Introduction, pp. 1–11. doi:10.1007/978-1-4899-2921-1_1.

Wilson, G.T. (2016) 'Time Series Analysis: Forecasting and control, 5th edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1', Journal of Time Series Analysis, 37(5), pp. 709–711. doi:10.1111/jtsa.12194.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts. https://otexts.com/fpp2/

Appendices

Code_1

```
#following code use python to grab the stock index data from Yahoo Finance
import pandas as pd
import yfinance as yf
start_date = "2012-01-01"
end_date = "2017-12-31"
spx=yf.download('^GSPC',start=start_date, end=end_date)
data=spx.reset_index()
three_years_data = data[data['Date'] > data['Date'].max() - pd.DateOffset(years=3)]
two_years_data = data[data['Date'] > data['Date'].max() - pd.DateOffset(years=2)]
freq_seven_data = data.iloc[::7, :]
freq_seven_mean_data = data.resample('7D', on='Date').mean()
freq_30_data = data.resample('30D', on='Date').mean()
five_years_data = data
three_years_data.to_csv('process/SPX_3years.csv')
two_years_data.to_csv('process/SPX_2years.csv')
freq_seven_data.to_csv('process/SPX_freq7.csv')
freq_seven_mean_data.to_csv('process/SPX_freq7_mean.csv')
five_years_data.to_csv('process/SPX_5years.csv')
freq_30_data.to_csv('process/SPX_freq30.csv')
NASDAQ=yf.download('^IXIC', start=start_date, end=end_date)
DJI=yf.download('DJI', start=start_date, end=end_date)
NASDAQ.to_csv('data/NASDAQ.csv')
DJI.to_csv('data/DJIA.csv')
spx.to_csv('data/SPX.csv')
```

Code_2

```
#following code use R and the data from code_1 to generate plot
 library(fpp2)
library(rmarkdown)
library(itsmr)
library(readr)
spx.y5<-read_csv("process/SPX_5years.csv")
spx.y2<-read_csv("process/SPX_2years.csv")
spx.y3<-read_csv("process/SPX_3years.csv")
spx.f7<-read_csv("process/SPX_freq7.csv")
spx.f7m<-read_csv("process/SPX_freq7_mean.csv")
```

```
spx.f30<-read_csv("process/SPX_freq30.csv")
na<-read_csv("data/NASDAQ.csv")
dji<-read_csv("data/DJIA.csv")
spx<-read_csv("data/SPX.csv")
#following code use spx as example
data<-spx

data.ts<-ts(data[,"Close"])
data.log<-ts(log(data[,"Close"]))
data.diff<-diff(data.ts)
data.logd<-diff(data.log)
test(data.ts)
test(data.diff)
test(data.log)
test(data.logd)
data_mean<-mean(data.logd)
data_var<-var(data.logd)

data_mean
data_var

#following code use ARIMA model as example
len<-length(data.log)
t<-floor(len*0.8)
myts <- ts(data.log)
myts.train <- window(myts,end=t)
myts.test <- window(myts,start=t+1)
fc <- auto.arima(myts.train)
future <- forecast::forecast(fc, h=(len-t))
autoplot(myts) +
  autolayer(myts.train, series="Training") +
  autolayer(myts.test, series="Test")
autoplot(future)
p <- autoplot(myts) +
  autolayer(myts.train, series="Training") +
  autolayer(future, series="Forecast")
p + autolayer(myts.test, series="Test")
checkresiduals(fc)
accuracy(future,myts.test)
```

Table of work distribution

Eric - Rongze Gao:
In charge of
1. coding and programing
2. Conceive analysis and work distribution
3. 5.2 The analysis of the Time Window
4. 6. Discussion and 7.conclusion

Mia – Xueyiting Wang:
In charge of
1. 4. Methodology
2. 4.3.3.1 the analysis of the NASDAQ
3. 5.1.1 the forecast of the NASDAQ
4. 5.2.3 the analysis of the averaging

Cecilia – Sihan Wang:
In charge of
1. 3. Background
2. Responsible for the summary of 5.1
3. 4.3.3.2. the analysis of DJI
4. 5.1.2 the forecast of DJI

Rona - Hairong Zhang:
In charge of
1. 1. Executive Summary and 2. Introduction
2. 4.3.3.3 the analysis of the SPX
3. 5.1.3 the forecast of the SPX