Analyzing Daily SPY ETF Data and Monthly Solar Production

Time Series Analysis I
(MA 641)

David Laziuk
dlaziuk@stevens.edu

Spring, 2023

Abstract

In this study, we conduct a comprehensive time series analysis on two distinct datasets: the non-seasonal daily adjusted close prices of the SPY ETF and the seasonal hourly solar production at the Southland Leisure Centre in Canada. The primary goal of this project is to develop accurate forecasting models for both datasets, while addressing the unique challenges posed by each data source. Some data pre-processing was done in Python, while the analysis was carried out using the R programming language.

Overview

For the non-seasonal SPY ETF dataset, we initially experimented with various transformations to achieve stationarity and tested different forecasting models, including GARCH. However, the GARCH model yielded unsatisfactory results. After employing a rolling mean detrending technique, we successfully applied an ARMA model, which produced improved forecasts. On the other hand, for the seasonal solar production dataset, we first confirmed the presence of seasonality, then performed seasonal differencing and tested various SARIMA models to ultimately select the one that yielded the most accurate forecasts.

The findings of this study contribute to a better understanding of the distinct characteristics of non-seasonal and seasonal time series data, and demonstrate the effectiveness of different forecasting techniques in addressing the unique challenges posed by each dataset. This report serves as a comprehensive guide for conducting time series analysis on similar datasets, as well as a reference for practitioners and researchers interested in employing advanced statistical methods for forecasting purposes.
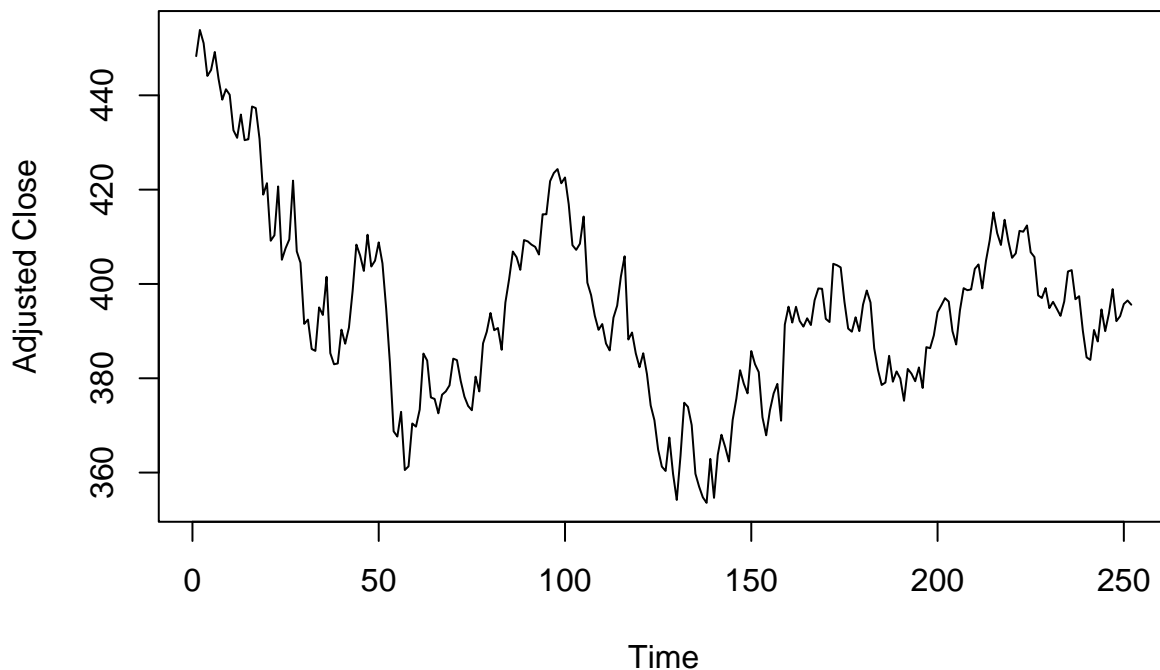
Non-Seasonal Data

Motivation:

The financial market is a complex and dynamic environment in which asset prices constantly fluctuate due to various factors. Accurately predicting the future performance of an asset can be challenging, yet it is of great interest to investors and financial analysts. The Standard & Poor's Depositary Receipts (SPY) exchange-traded fund (ETF) is one of the most commonly traded securities, designed to track the performance of the S&P 500 Index. Precise forecasting of SPY ETF prices can provide valuable insights for investment strategies and risk management.

Data Overview:

In this section, I will analyze the daily adjusted close data for the SPY ETF, obtained from Yahoo Finance. The data spans from March 28, 2022, to March 28, 2023, totaling 252 data points. Additionally, I will utilize 22 data points from March 29, 2023, to April 28, 2023, for validating the forecasts. I will employ various statistical techniques, such as ARIMA and GARCH models, to analyze the time series data and create forecasts of the SPY ETF prices.

SPY Time Series

Verifying Non-Seasonality

Before proceeding with the analysis, I wanted to verify that the data is, in fact, non-seasonal. Seasonality refers to the presence of regular and predictable fluctuations in the data that occur within a fixed period, such as daily, weekly, or yearly cycles. While it is possible that seasonality could be present over a longer time frame spanning multiple years, it is unlikely for just one year of daily data. To verify this, I set the period to 5, corresponding to the 5 trading days in a week. As illustrated in the plotted decomposition, the seasonal component is cyclical. However, the magnitude of this seasonality is minuscule.

Decomposed Time Series With Weekly Period
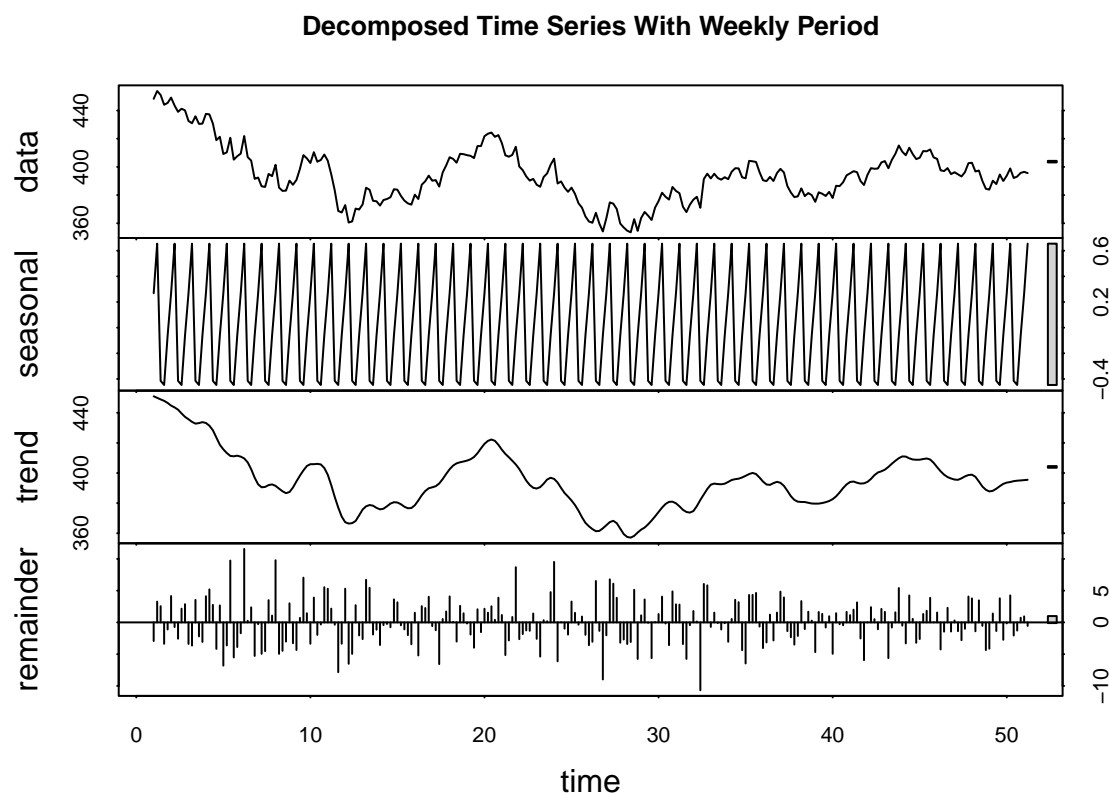
Original Time Series Range: 353.5807 - 453.8604
Seasonal Component Range: -0.4469712 - 0.6551471
The magnitude of the seasonal component is insignificant.
Therefore, it is safe to proceed with the analysis for non-seasonal data.

Testing Stationarity

A key assumption in time series analysis is that the data is stationary. Non-stationary data can result in spurious relationships and inaccurate forecasts. To test for stationarity, I used two statistical tests: the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The ADF test checks for the presence of a unit root, while the KPSS test examines the stationarity of the data around a deterministic trend.

Below are the ACF and PACF plots for the original time series:

Raw Time Series

Raw ACF

Raw PACF
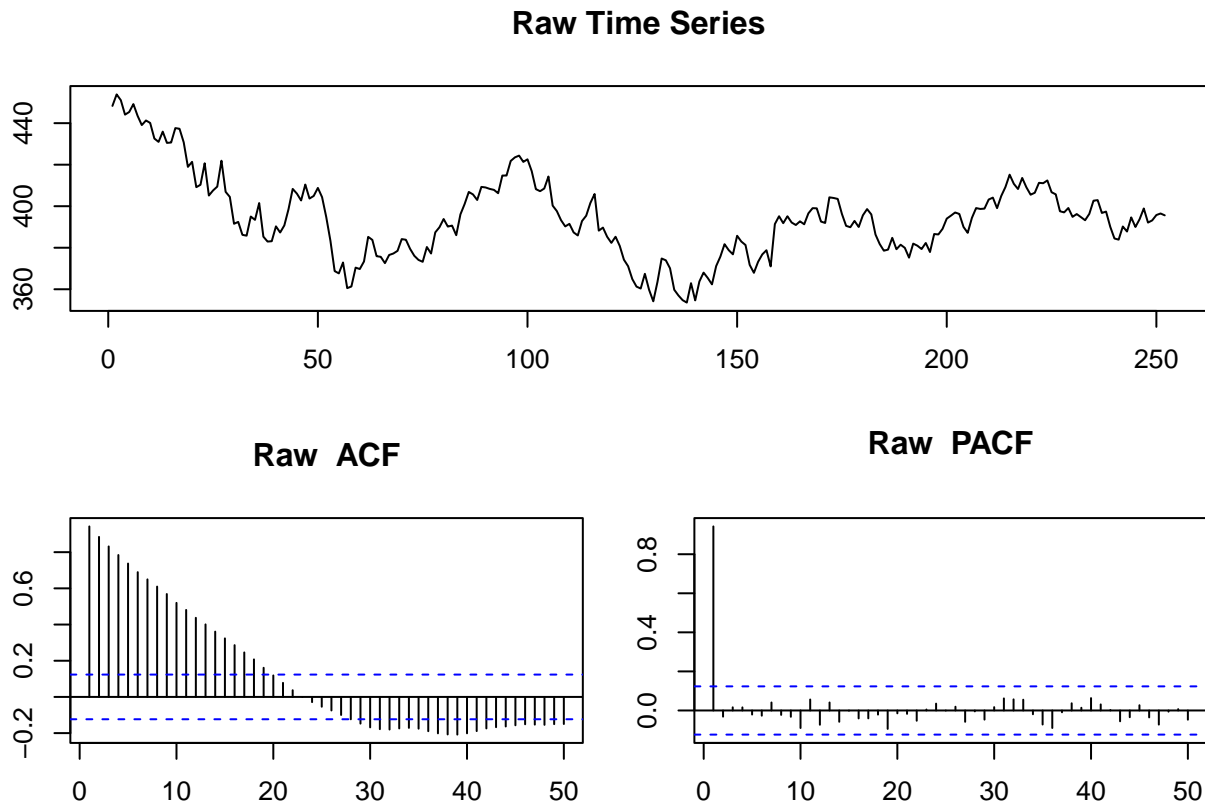
ADF test: p-value = 0.1929712
Raw KPSS test p-value: 0.01428673
Since the ADF test p-value is above 0.05 and the ACF decays slowly, we can infer that the data is not stationary, despite the KPSS score being below 0.05. Notably, the ACF displays lags decaying very slowly.
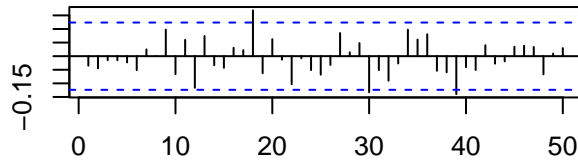
Data Augmentation

To make the data stationary, I applied data augmentation techniques such as first differencing and log transformation followed by first differencing. These techniques can help stabilize the mean and variance of the data, allowing for more accurate modeling and forecasting.
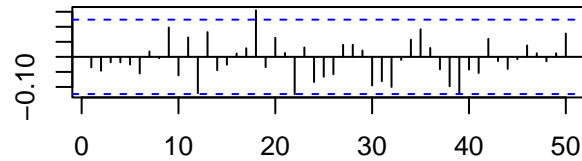
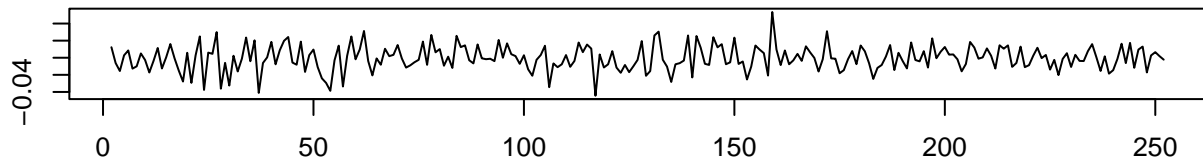### First Dif Time Series
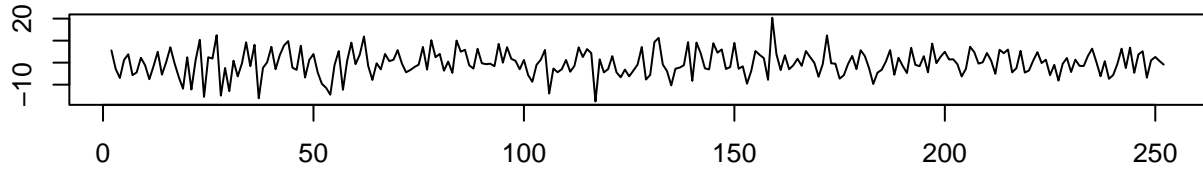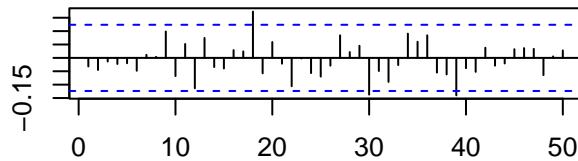

First Dif ACF

First Dif PACF

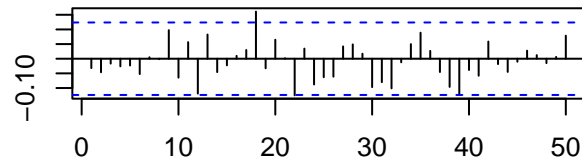First Dif ADF test p-value: >0.01 First Dif KPSS test p-value: <0.1

Log First Difference Time Series

### Log First Difference ACF

Log First Difference PACF

Log First Difference ADF test p-value: >0.01 Log First Difference KPSS test p-value: <0.1

Both of these augmentations result in stationary data. However, it is evident that the ACF and PACF immediately die out. This observation is concerning, as it may indicate that the data resembles white noise, with little or no dependency to model. Additionally, the orders of ARMA models are commonly determined by the number of significant lags in the ACF and PACF. Since there are none, this possibility is eliminated.

As I cannot simply reference the number of significant lags in the ACF and PACF, I will attempt to find an ARMA order by analyzing AIC/BIC scores for a variety of configurations. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are statistical measures that balance the goodness-of-fit and complexity of a model, helping to identify the most suitable model for the data.

Model Search

AIC & BIC Values for First Difference Models
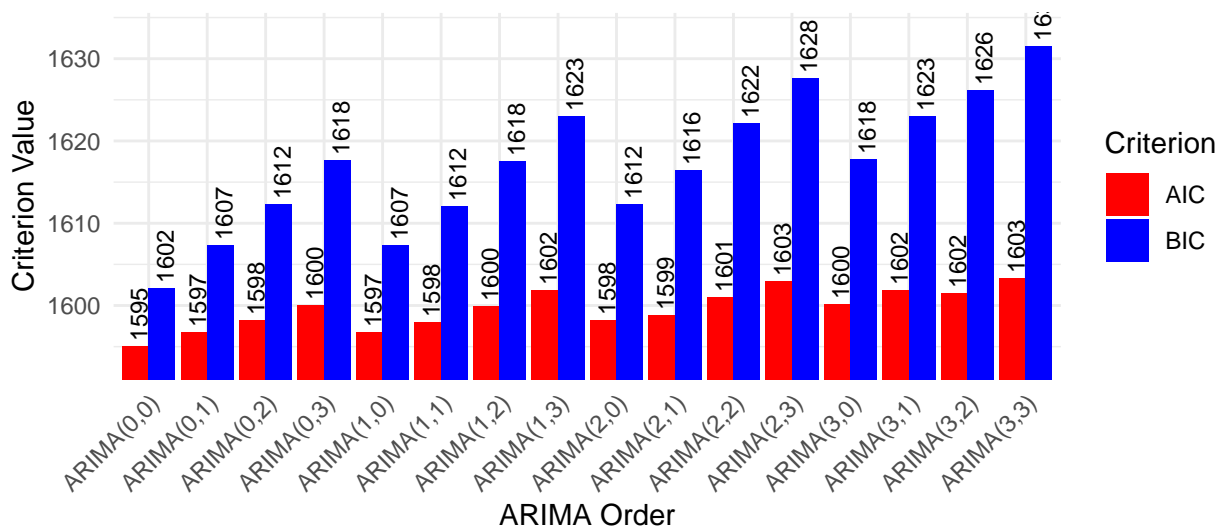
AIC & BIC Values for Log Difference Models
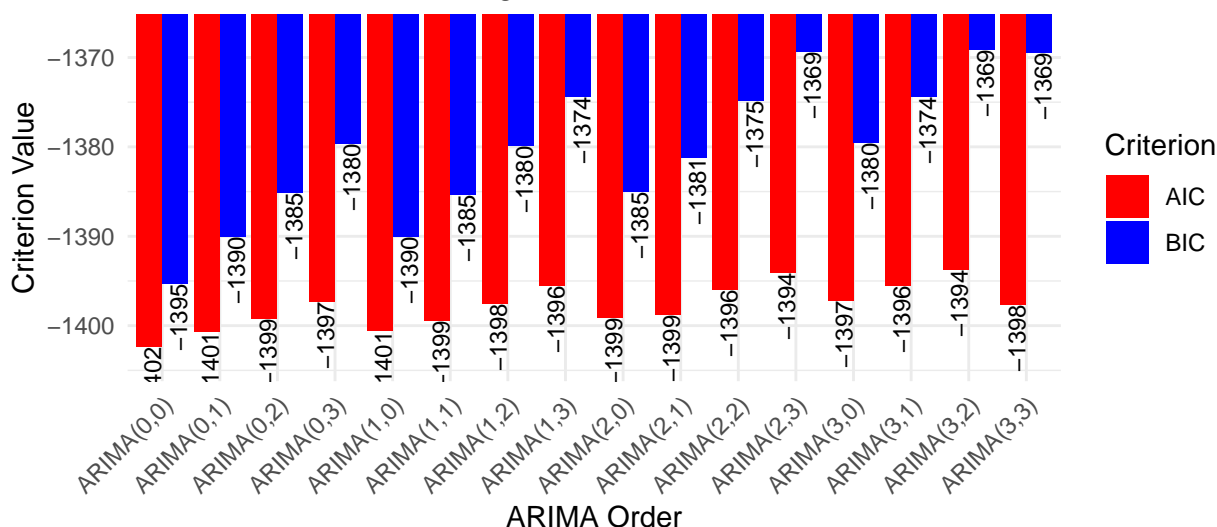
As observed, no ARMA models have a lower AIC/BIC score than the ARMA(0,0) model. The ARMA(0,0) model is essentially a white noise model, which implies that there is no significant relationship between past values and future values in the time series. In other words, the data appears to be random and unpredictable.

A lower AIC/BIC score typically indicates a better model, as it represents a balance between the goodness-of-fit and the complexity of the model. In this case, since none of the ARMA models have a lower score than the ARMA(0,0) model, it suggests that more complex models with autoregressive and moving average components do not provide any substantial improvement in terms of forecasting accuracy compared to a simple white noise model. This result underscores the difficulty in predicting the SPY ETF prices using traditional time series models, as the data appears to be highly unpredictable.

GARCH Exploration

I will now attempt to use the GARCH model to analyze the extremely low-dependency data. The GARCH model is particularly useful in this scenario because it is designed to model the conditional volatility of a time series. In financial data, volatility often exhibits a clustering effect, where periods of high volatility are followed by periods of high volatility and vice versa. By capturing these volatility patterns, the GARCH model may provide more accurate forecasts compared to traditional time series models.

To fit a GARCH model, I will first apply an ARMA model to the absolute value or square of the return of the data. The return is the difference between the log of the data, which represents the percentage change in the time series.
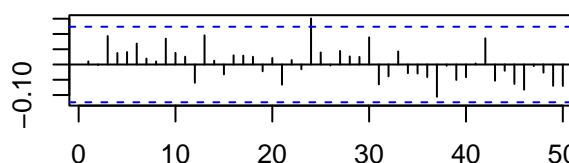
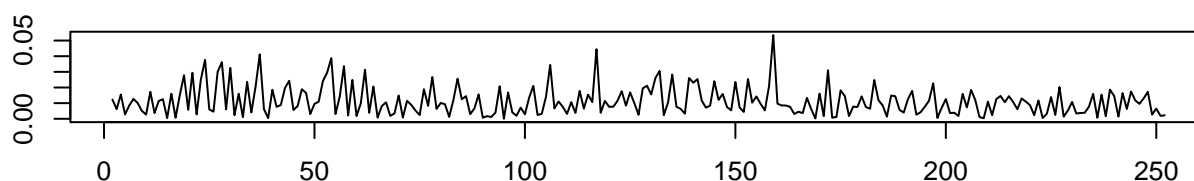**Sqr Log First Diff Time Series**
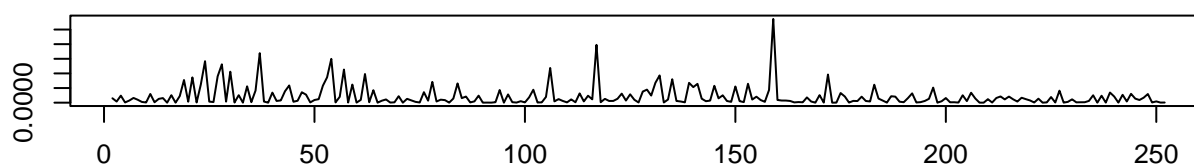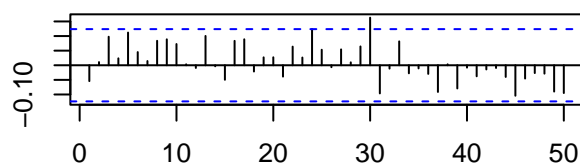
Sqr Log First Diff ACF
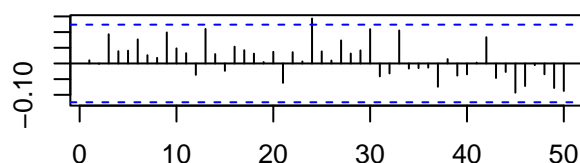
Sqr Log First Diff PACF

Sqr Log First Diff ADF test p-value: 0.01
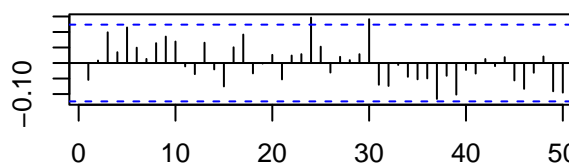Sqr Log First Diff KPSS test p-value: 0.01387874

Abs Log Diff Time Series

Abs Log Diff ACF
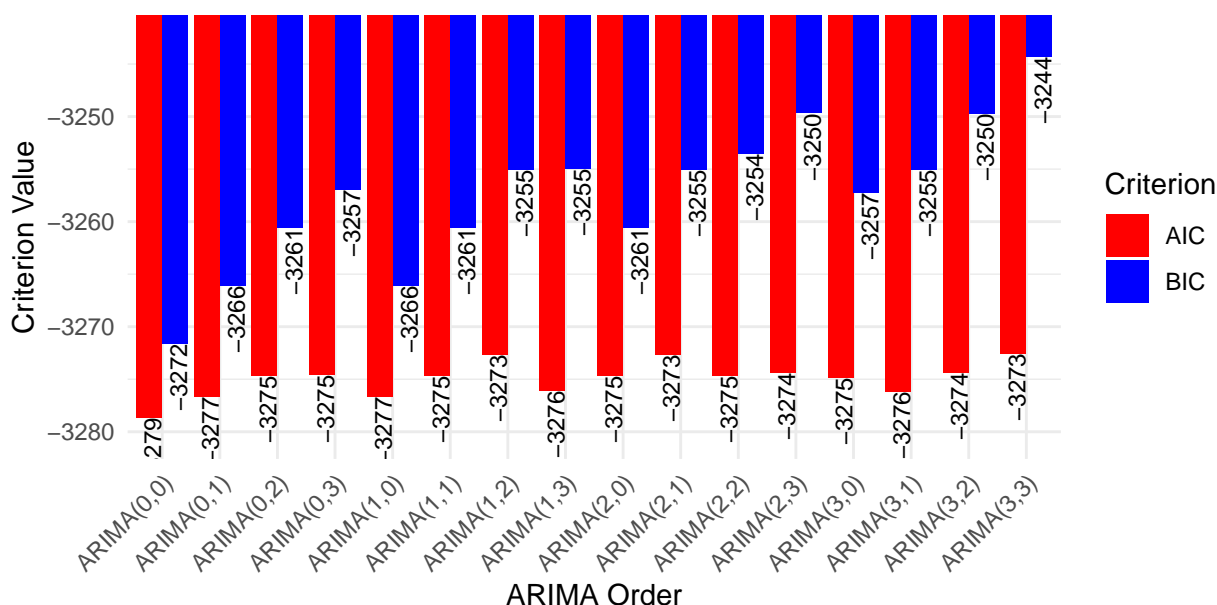
Abs Log Diff PACF

Abs Log Diff ADF test p-value: 0.01
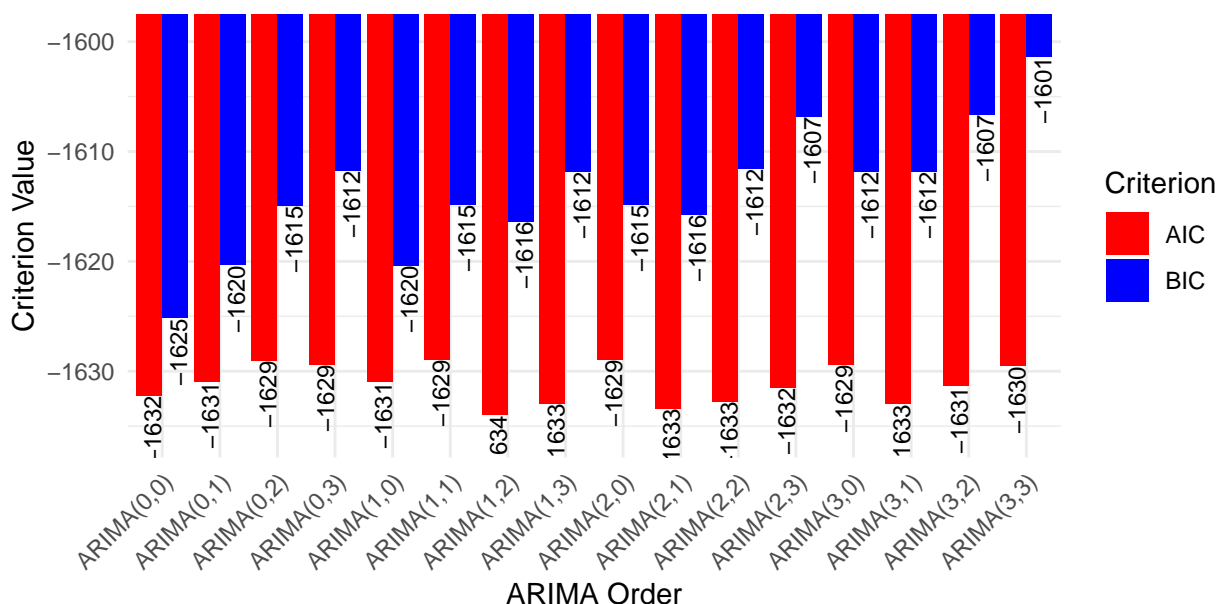Abs Log Diff KPSS test p-value: 0.01529264

Both augmentations result in stationary data. However, the ACF and PACF still immediately die out. Despite this issue, I will attempt to fit a GARCH model using an AIC/BIC search to identify the most suitable configuration.

GARCH Model Search

AIC & BIC Values for Return^2 Models

AIC & BIC Values for |Return| Models

When analyzing the squared returns, it appears that no model beats the ARMA(0,0) model in terms of AIC/BIC scores. This suggests that even with the GARCH approach, it remains challenging to predict the future values of the SPY ETF using the squared returns.

On the other hand, for the absolute returns, the AIC suggests several candidate models with higher orders: (1,2), (2,1), (2,2), (1,3), and (3,1). These models may provide more accurate forecasts compared to the ARMA(0,0) model, as they incorporate additional autoregressive and moving average components. In the next step, I will further investigate the performance of these candidate models and select the one that offers the best balance between model complexity and forecasting accuracy.

GARCH Parameter Estimation

I will now test the GARCH models with orders (1,2), (2,1), (2,2), (1,3), and (3,1) on the absolute returns of the SPY ETF data. To keep the report concise, I will only present the normality test results (Shapiro-Wilk p-value), log-likelihood, AIC, and BIC scores for each model. Note that during the experimentation, I evaluated each model using various metrics, such as ACF, PACF, Ljung-Box test, histograms, and QQ-plots. The best model excelled in all these assessments.

```
## GARCH(1,2) Normality: 1.141505e-11 | LL: 822.009 | AIC: -1634.018 | BIC: -1616.391
```
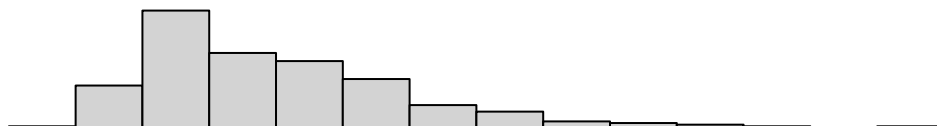


```
## GARCH(2,1) Normality: 9.90191e-12 | LL: 821.706 | AIC: -1633.412 | BIC: -1615.785
```
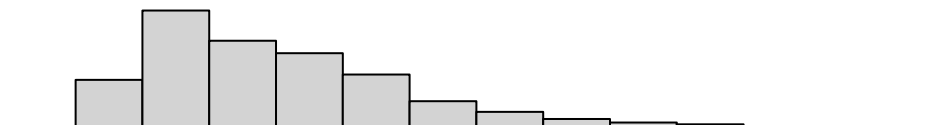


```
## GARCH(2,2) Normality: 1.959207e-11 | LL: 822.3755 | AIC: -1632.751 | BIC: -1611.598
```



```
## GARCH(1,3) Normality: 2.806268e-11 | LL: 822.4969 | AIC: -1632.994 | BIC: -1611.841
```



```
## GARCH(3,1) Normality: 2.610229e-11 | LL: 822.5086 | AIC: -1633.017 | BIC: -1611.864
```


As shown, the residuals of all models fail the normality test. This result is unfavorable, as it suggests that the underlying distribution of the residuals is not normally distributed. In financial time series, it is common to observe heavy-tailed distributions, which deviate from the normality assumption. This deviation may pose challenges in accurately modeling the data and generating reliable forecasts. Nevertheless, among the tested models, GARCH(1,3) has the best normality test results, and GARCH(1,2) has the lowest AIC/BIC scores. I will further investigate the residuals for one of these models, GARCH(1,3). It is worth noting that both models perform similarly, so the choice between them may not significantly impact the final forecasting results.

GARCH Residual Analysis

Residual analysis is a crucial step in assessing the performance of time series models, as it helps determine whether the model has captured all relevant information from the data. In this analysis, I will examine the standardized residuals of the GARCH(1,3) model by plotting the following: 1. Standardized residuals over time 2. Autocorrelation function (ACF) of residuals 3. P-values for the Ljung-Box statistic 4. Histogram of standardized residuals 5. Normal Q-Q plot

The goal is to ensure that the residuals resemble white noise, indicating that the model has successfully extracted all available information from the data. The Ljung-Box test checks for autocorrelation in the residuals, while the histogram and Q-Q plot assess the normality of the residual distribution.
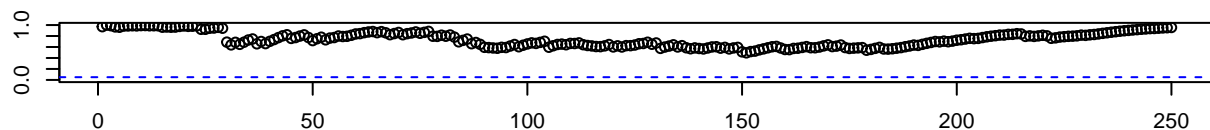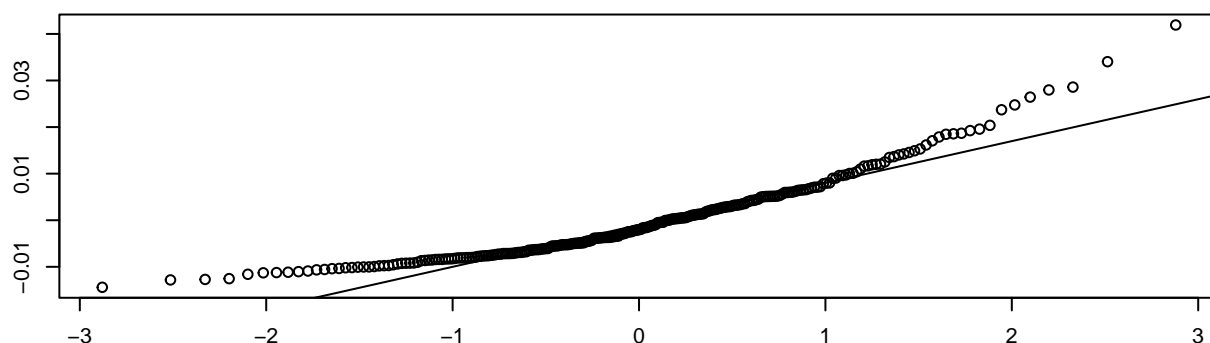
Standardized Residuals
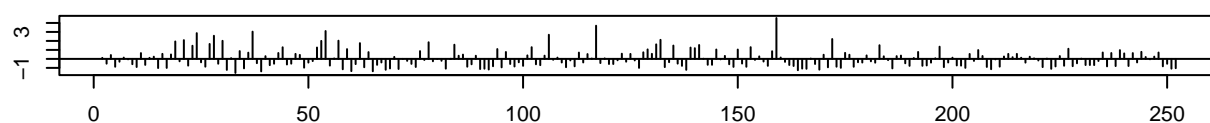
ACF of Residuals
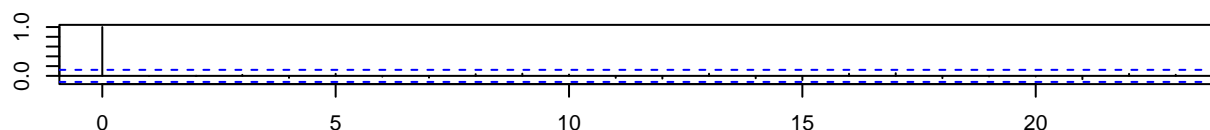
**p values for Ljung–Box statistic**

Normal Q–Q Plot

The analysis shows that the p-values for the Ljung-Box test are above 0.05 for all lags, suggesting that there is no significant autocorrelation in the residuals. This result indicates that the GARCH(1,3) model has captured the underlying patterns in the data effectively. However, the histogram and Q-Q plot reveal that the residuals are not normally distributed, which is a common issue in financial time series due to heavy-tailed distributions. Despite this limitation, I will proceed to forecast using the GARCH(1,3) model, as it has demonstrated reasonable performance in other aspects of the residual analysis.

GARCH Forecasting

I will now attempt to forecast 22 days ahead using the GARCH(1,3) model and compare the predictions to 22 days of real-world data.

## Original Data and Forecast with GARCH(3,1)
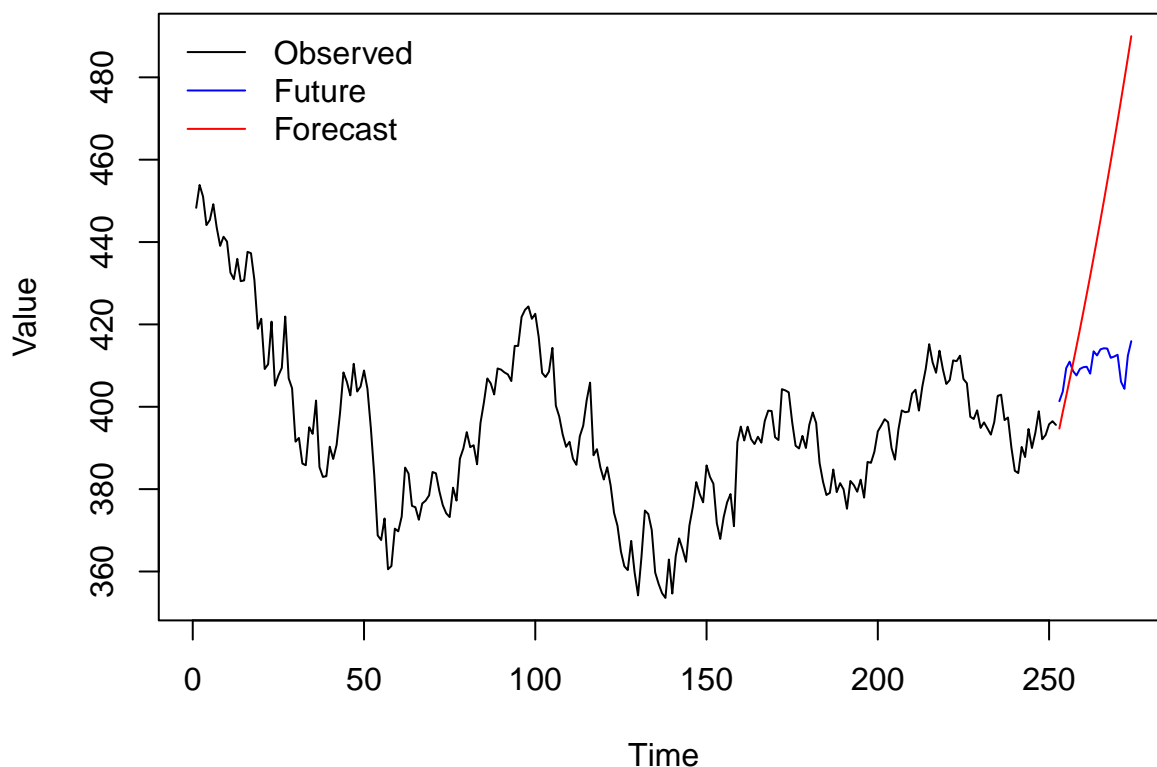
Upon visual inspection, the forecasts generated by the GARCH(1,3) model are nearly linear, with a consistent upward trend. However, the actual future data exhibit a slightly upward trend initially, followed by a flattening period. The GARCH(1,3) model does not capture this pattern, as it only projects a continuous upward trend. As a result, the model's forecasts are not an accurate representation of the real-world data.

This poor fit highlights the limitations of the GARCH model in predicting future values for the SPY ETF. Financial time series data can be challenging to forecast due to their complexity and inherent unpredictability, and the GARCH model may not be the best choice for this particular data set.

Detrending Experimentation

After encountering difficulties in forecasting the SPY ETF data using traditional data transformations and GARCH models, I will now experiment with an alternative approach: detrending the data using a rolling mean with a window size of 22. Detrending can help isolate the underlying patterns in the data by removing the overall trend, making it easier to model and forecast the time series.

Rolling Mean Time Series

Rolling Mean ACF

Rolling Mean PACF
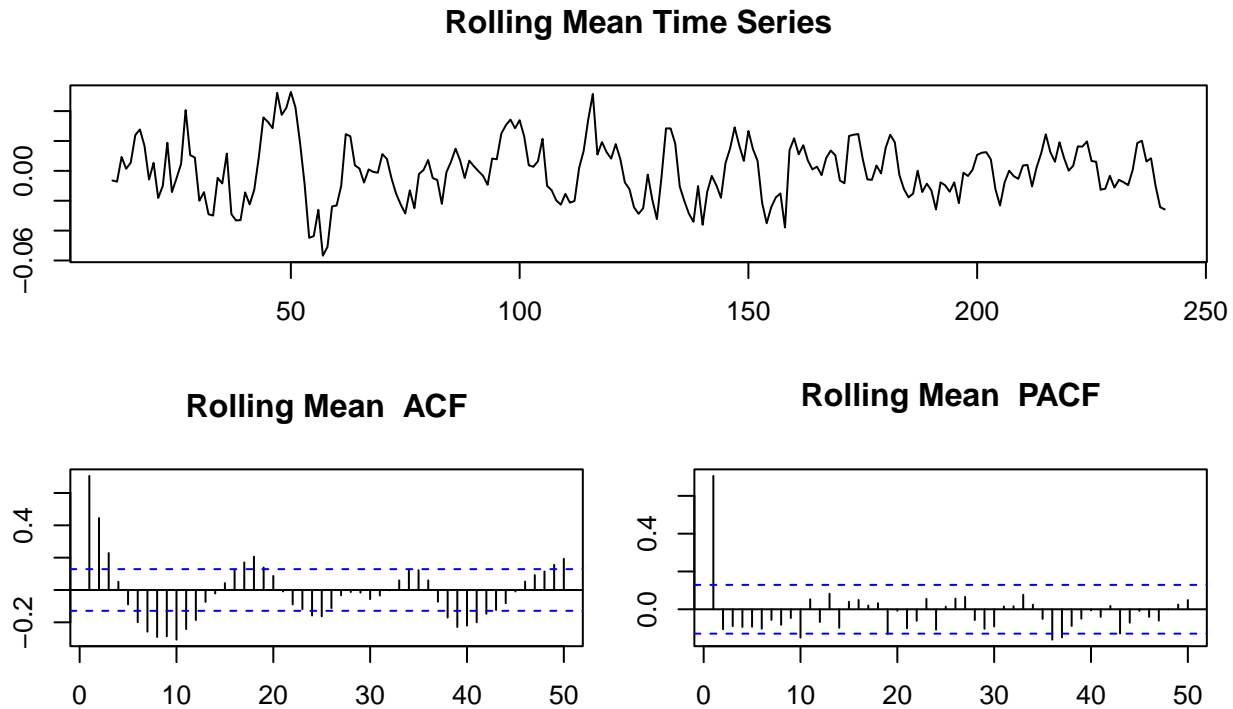
Upon detrending the data, the following observations can be made:

1. The range of the detrended data is relatively small (-0.05676492, 0.05284111), indicating that the overall trend has been successfully removed.

2. The ACF of the detrended data cyclically decays, alternating between positive and negative values. This pattern suggests that there may be some remaining cyclical patterns in the data, which were not captured by the previous models.

3. The PACF of the detrended data becomes insignificant after the first lag, but the decay is not as dramatic as in the original data. This observation implies that there might be some autoregressive component present in the detrended data, which could be exploited for forecasting purposes.

Based on these findings, I will proceed with modeling the detrended data and attempt to generate more accurate forecasts for the SPY ETF.

Detrended Model Search

To identify the optimal model for the detrended data, I will first rely on the ACF and PACF plots, which suggest an ARMA model with orders (1,3) and (1,2). In addition, I will conduct a search using AIC and BIC criteria to explore other potential model configurations.
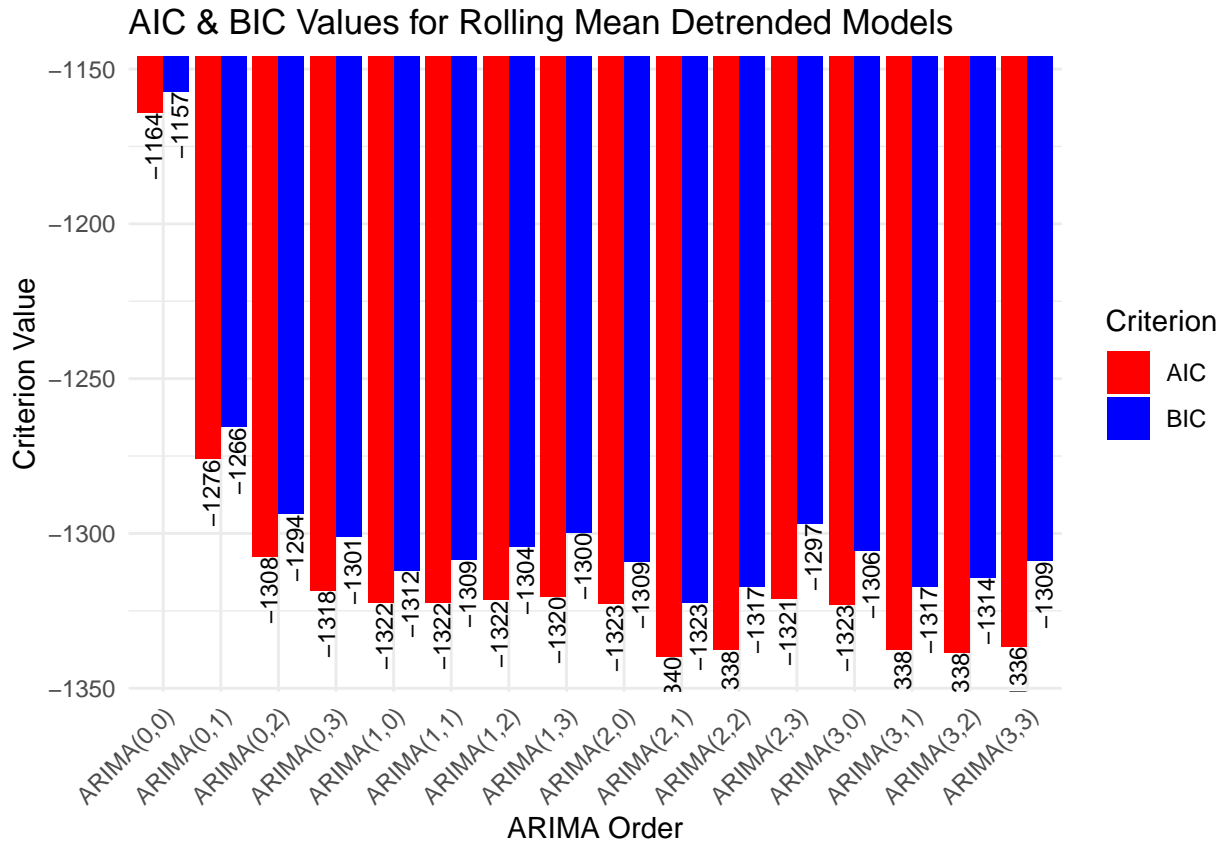
The results from the ACF/PACF analysis and the AICc/BIC search are as follows:

1. (1,3) and (1,2) are suggested by the ACF and PACF plots.

2. (2,2), (3,1), and (3,2) appear to be good candidates based on the AICc criterion.

3. (2,1) seems to be a suitable choice according to the BIC criterion.

Considering these findings, I will proceed with further analysis and testing of these models to determine the best candidate for forecasting the detrended SPY ETF data.

Detrended Paramater Estimation

In order to estimate the parameters for the ARMA models suggested by the ACF/PACF plots and the AICc/BIC search, I will evaluate each model's performance using several diagnostic metrics, such as the Shapiro-Wilk test for normality, log-likelihood, AIC, and BIC. This process will help identify the model that best captures the underlying structure of the detrended data.

```
## Rolled ARMA(1,3) Normality: 0.6293316 | LL: 666.2498 | AIC: -1320.5 | BIC: -1299.845
```
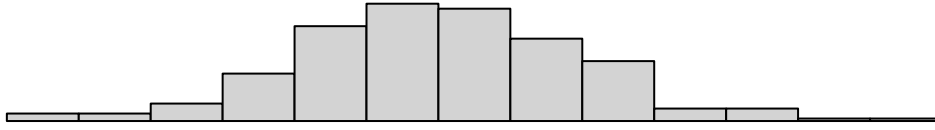
```
## Rolled ARMA(1,2) Normality: 0.7251017 | LL: 665.8168 | AIC: -1321.634 | BIC: -1304.421
```
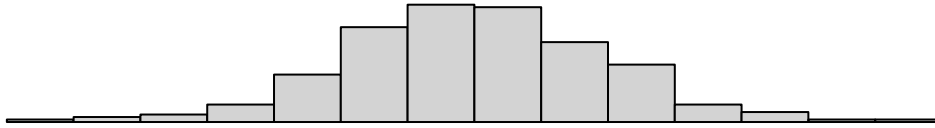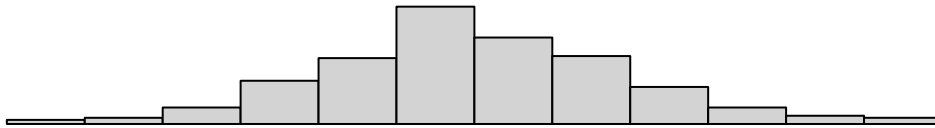
```
## Rolled ARMA(2,2) Normality: 0.798756 | LL: 674.8895 | AIC: -1337.779 | BIC: -1317.125
```

```
## Rolled ARMA(3,1) Normality: 0.7985522 | LL: 674.8896 | AIC: -1337.779 | BIC: -1317.125
```

```
## Rolled ARMA(3,2) Normality: 0.7975509 | LL: 676.231 | AIC: -1338.462 | BIC: -1314.365
```

```
## Rolled ARMA(2,1) Normality: 0.7998526 | LL: 674.8854 | AIC: -1339.771 | BIC: -1322.559
```

Upon examining the results, it appears that the ARMA(2,1) model performs the best among the candidates. The model has the highest Shapiro-Wilk p-value, indicating a better fit to a normal distribution compared to the other models. Additionally, the ARMA(2,1) model has a log-likelihood value close to the highest among the candidates and the lowest AIC and BIC values. These characteristics suggest that the ARMA(2,1) model is a suitable choice for forecasting the detrended SPY ETF data.
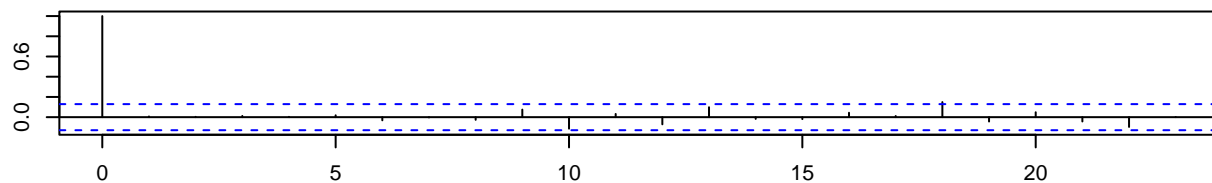
Detrended Residual Analysis

To further assess the performance of the selected ARMA(2,1) model for the detrended data, I will conduct a residual analysis. This step is crucial for evaluating the model's adequacy and ensuring that the underlying assumptions of the time series analysis are met.
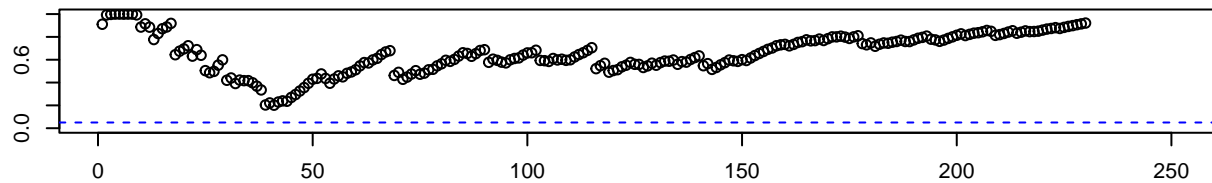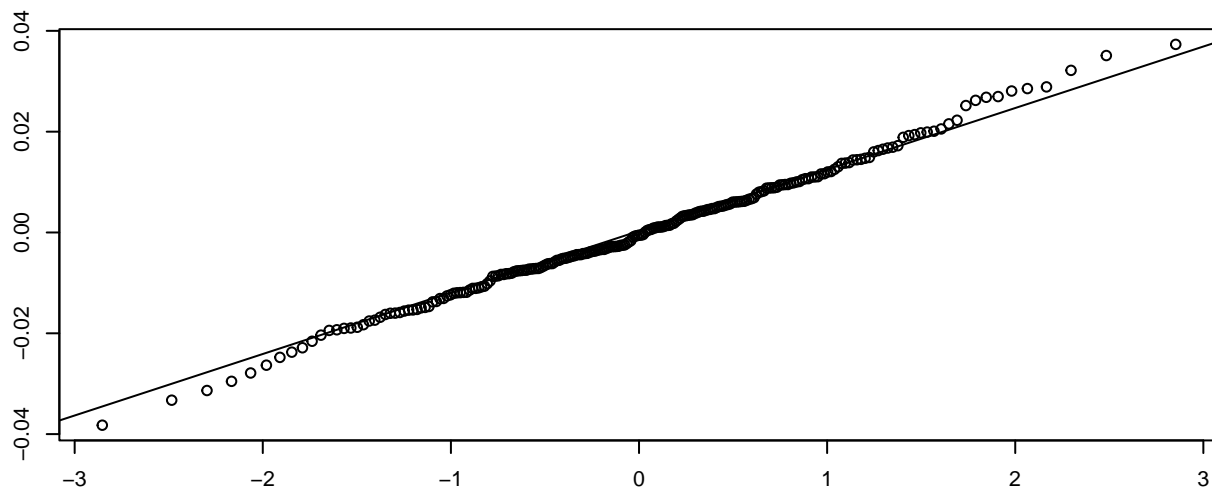
Standardized Residuals
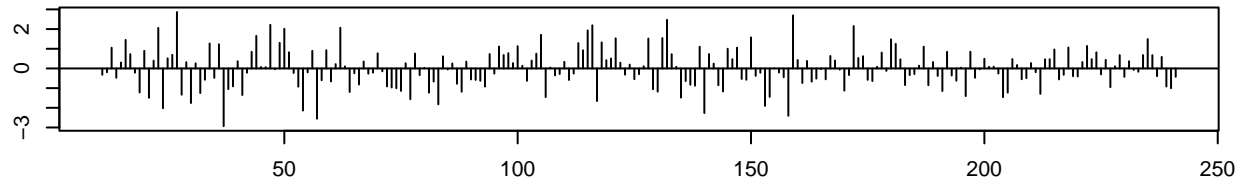
ACF of Residuals

p values for Ljung–Box statistic

Normal Q–Q Plot

The results from the residual analysis show the following:

1. The standardized residuals appear to be randomly distributed and do not exhibit any discernible patterns, which is desirable for a well-fitting model.

2. The ACF plot of the residuals does not display any significant autocorrelations, indicating that the model has effectively captured the underlying structure of the detrended data.

15

3. The p-values for the Ljung-Box test are above 0.05 for all lags, suggesting that there is no evidence of remaining autocorrelation in the residuals.

4. The histogram and normal Q-Q plot indicate that the residuals are approximately normally distributed, which is consistent with the assumptions of the ARMA model.

Based on these findings, it can be concluded that the ARMA(2,1) model performs well on the detrended SPY ETF data and is suitable for generating forecasts.

Detrended Forecasting

Using the ARMA(2,1) model on the detrended data, we can generate forecasts for the 22-day period ahead.

Detrended Forecast
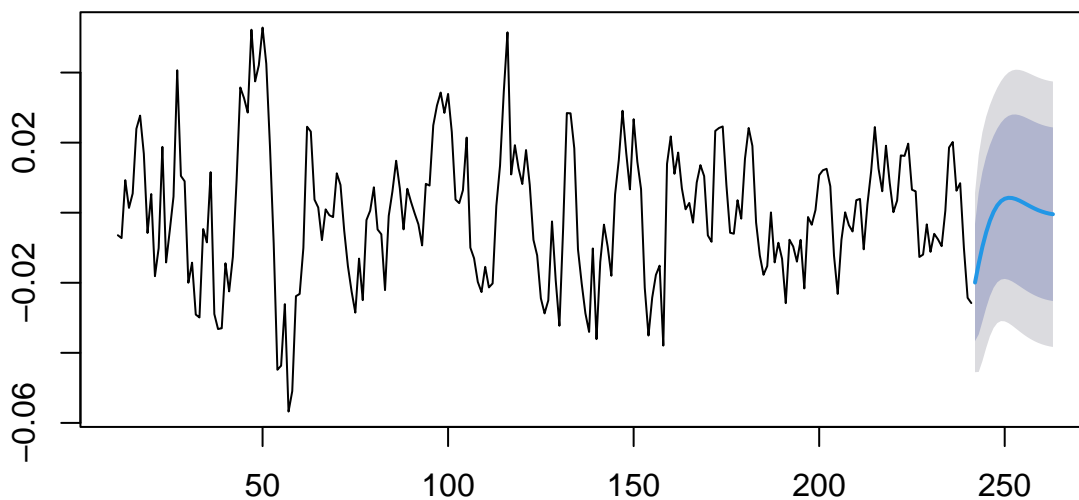
The plot above displays the point forecasts as well as the two confidence intervals (lower and upper bounds) for the detrended data. However, these forecasts are not directly useful as they are still in the detrended scale. We must transform the forecasts back to the original scale to make them more interpretable and comparable to the actual data.

Original Scale Forecast
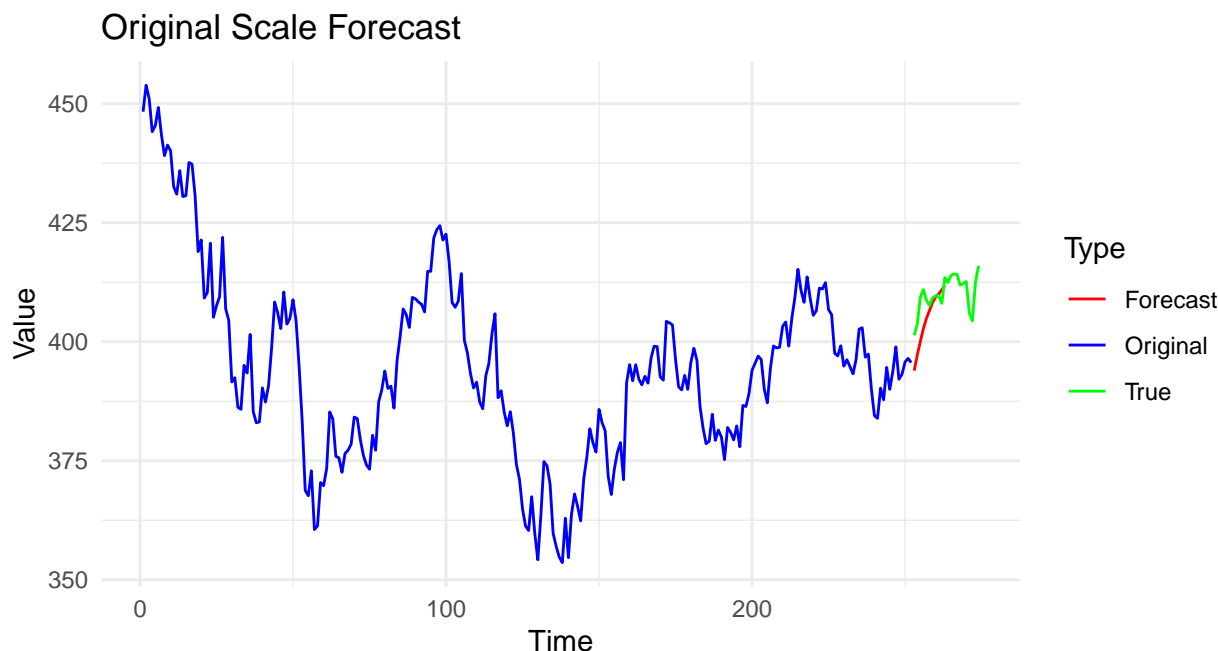
When transforming the forecasts back to the original scale, it is important to note that this process may remove some of the predictions. This is due to the nature of the rolling mean calculation, which requires a certain number of data points to compute the mean. However, the retained forecasts on the original scale appear to be incredibly accurate, closely following the true values beyond the cut-off from the transformation.

Interestingly, the transformed forecasts seem to take a similar shape to the untransformed forecasts. If this is the case, the model would not only capture the initial upward trend but also the leveling off that occurs later in the forecast period. This would suggest that the detrending process and the selected ARMA(2,1) model are successful in capturing the underlying dynamics of the SPY ETF prices, providing an accurate prediction of future price movements.

Non-Seasonal Conclusion and Key Findings

In this study, we have analyzed and forecasted the daily adjusted close prices of the non-seasonal SPY ETF using various time series techniques. The initial analysis demonstrated that the data was non-stationary and lacked any significant seasonal components. To achieve stationarity, we attempted different transformations, including differencing and log-differencing. However, these methods resulted in ACF and PACF plots that resembled white noise, indicating that traditional ARMA models might not be suitable for this dataset.

We then explored GARCH models as an alternative approach to model the low-dependency data. Despite finding potential GARCH models with the lowest AIC/BIC values, the residuals failed normality tests, and the forecasts provided were unsatisfactory. As a result, we experimented with a rolling mean detrending technique to address the non-stationarity issue, which yielded more promising results.

When forecasting the detrended data and transforming it back to the original scale, we found that the transformation removed some predictions, but the remaining forecasts were highly accurate. Furthermore, the transformed forecasts captured both the initial upward trend and the subsequent leveling off in the price movements.

Overall, this study has demonstrated the effectiveness of employing different time series techniques to address the unique challenges posed by non-seasonal financial data. The rolling mean detrending technique and the ARMA(2,1) model proved successful in providing accurate forecasts for the SPY ETF prices. These findings contribute to a better understanding of non-seasonal time series data and serve as a valuable reference for practitioners and researchers looking to employ advanced statistical methods for forecasting purposes.
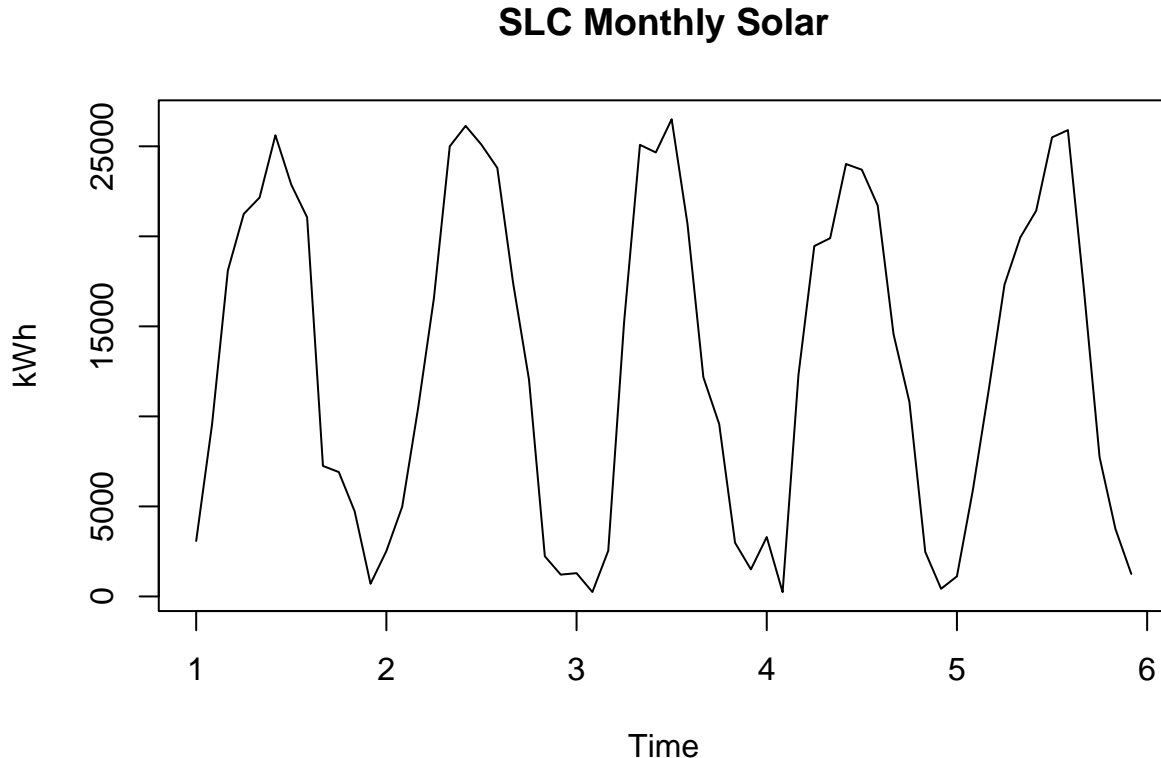
Seasonal Data

Motivation

The importance of accurately forecasting renewable energy production, such as solar energy, cannot be overstated. Accurate forecasts enable better management of power grids, improve the allocation of resources, and facilitate investment decisions in the renewable energy sector. This study focuses on analyzing and forecasting the monthly solar energy production of a specific location in Canada, with the aim of demonstrating the effectiveness of time series analysis techniques on seasonal data.

Data Overview

The dataset used in this study was obtained from Kaggle and contains hourly solar energy production for various locations in Canada. The original dataset includes data on the hourly output at each of the City of Calgary's solar photovoltaic projects and the locations of City of Calgary solar photovoltaic installations.

For the purpose of this study, we extracted data specific to the 'Southland Leisure Center', a location in Canada. We preprocessed the data by first extracting the relevant records, converting the date column to a datetime format, sorting the data by date, and then aggregating the hourly data into monthly solar energy production. The final dataset covers a period from January 2016 to December 2020 and is saved as a CSV file for further analysis.

In the following sections, we will walk through the steps of time series analysis applied to this dataset, including verifying seasonality, testing stationarity, applying seasonal differencing, determining the best SARIMA model, performing residual analysis, forecasting, and verifying the appropriate level of differencing.

SLC Monthly Solar

Verifying Seasonality

Solar energy production is typically influenced by seasonal factors, such as changes in weather patterns and daylight hours. Since the data used in this study is aggregated on a monthly basis, it is expected to exhibit seasonality with a frequency of 12 months. To verify this assumption, we will decompose the time series data and analyze the seasonal component.

Decomposed Time Series
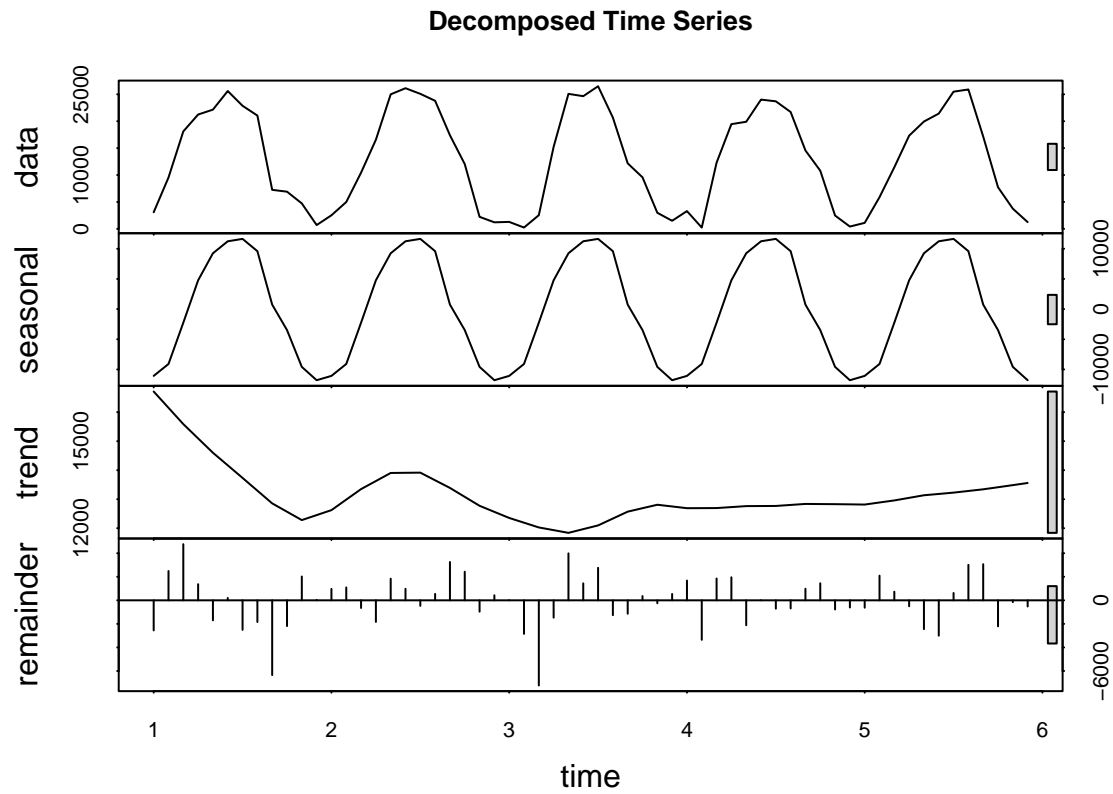
The decomposed plot reveals that the original time series has a range from 239.296 to 26,502.24.
While the seasonal component has a range from -11,785.86 to 11,642.79.
The significant magnitude of the seasonal component confirms the presence of seasonality in the data, which will need to be taken into account during the time series analysis.

Testing Stationarity

Before proceeding with the time series analysis, it is essential to determine whether the data is stationary or not. Stationarity is an important assumption for many time series models, as non-stationary data can lead to spurious results. We will use the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test to assess the stationarity of the data.

Raw ACF
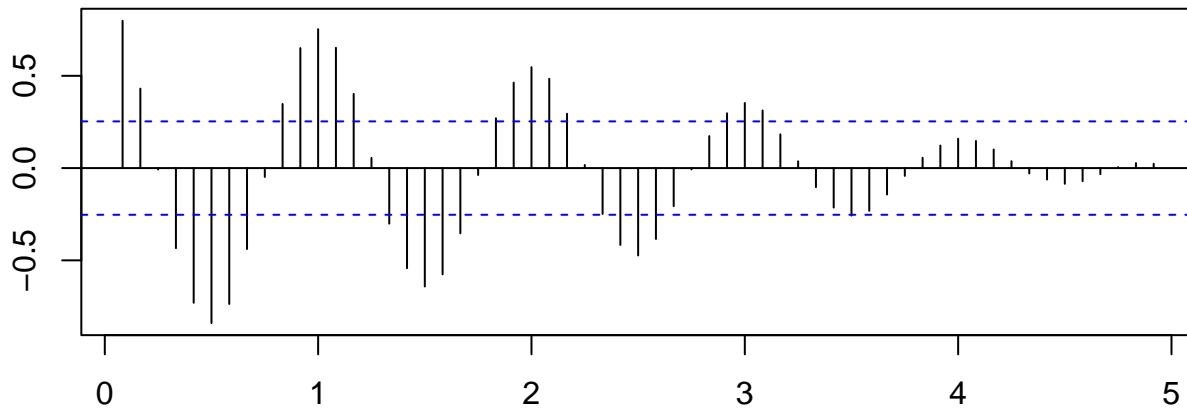
Raw PACF
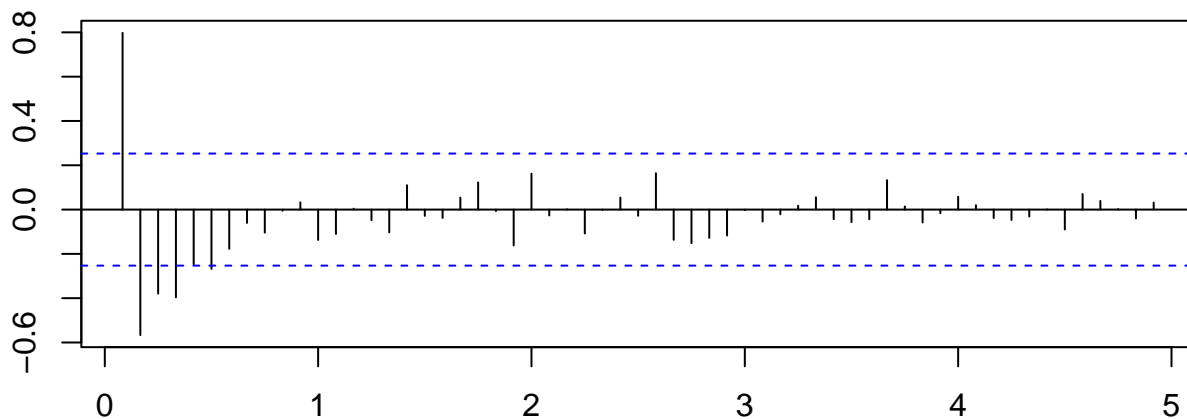
ADF test: p-value = <0.01
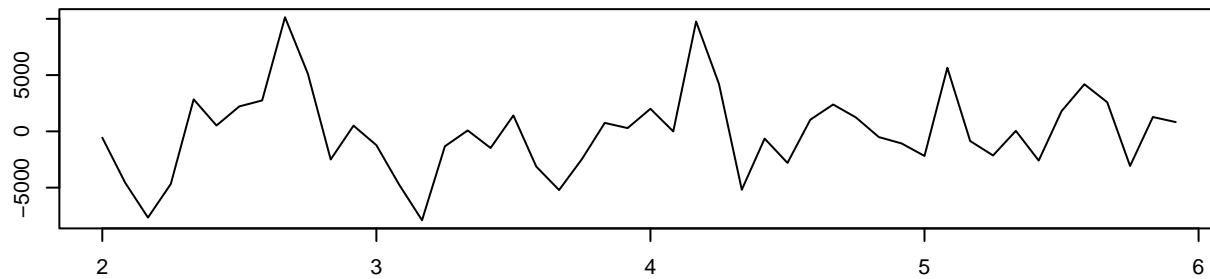KPSS test: p-value = >0.1

According to both the ADF and KPSS tests, the data appears to be stationary. However, this conclusion contradicts our earlier observation of strong seasonality and a cyclical slow decay in the ACF plot. It is crucial to investigate this discrepancy further and ensure that the data is appropriately preprocessed before modeling.
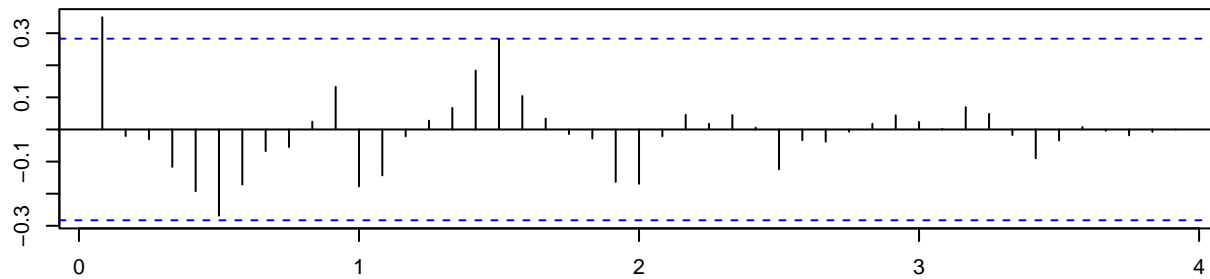
Seasonal Differencing

Given the presence of strong seasonality in the data, we will apply seasonal differencing with the original period (or frequency) of 12. Seasonal differencing is a technique used to remove the seasonal component of a time series and make the data stationary, which is a prerequisite for many time series models. This process involves taking the difference between each observation and the corresponding observation from the previous season (e.g., the difference between each month and the same month in the previous year). Note that seasonal differencing will result in the loss of some data points, specifically the first 12 observations.
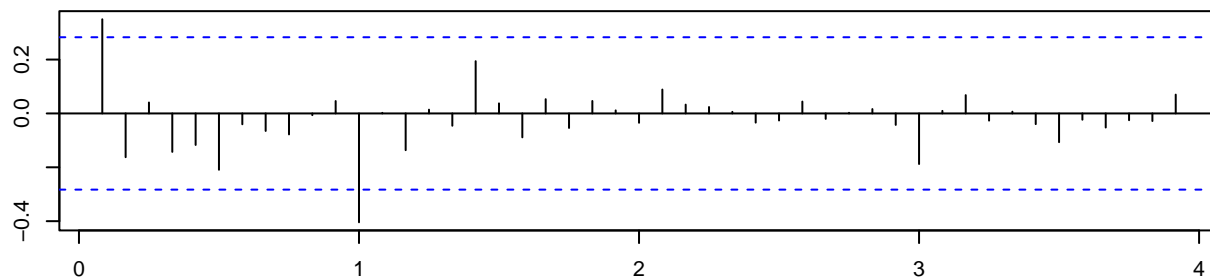
Seasonal Differencing (12)

Seasonal Differenced ACF

Seasonal Differenced PACF

ADF test: p-value = 0.027911
KPSS test: p-value = >0.1
The ADF and KPSS test results, as well as the ACF and PACF plots, indicate that the data has become stationary after seasonal differencing. This preprocessing step allows us to proceed with modeling the time series.

Seasonal Paramater Estimation

Based on the seasonal differenced time series, we can identify a few potential SARMA models to fit the data:

1. SARMA(1,0,1)x(2,1,2)12
2. SARMA(1,0,1)x(2,1,1)12
3. SARMA(0,0,1)x(2,1,1)12

We will evaluate the performance of these models using the following metrics: Shapiro-Wilk normality test p-value, log-likelihood, AIC, and BIC. In addition, we will inspect the histogram of residuals to ensure that the model assumptions are met.

```
## SARMA(1,0,1)x(2,1,2)12 Normality: 0.0006029639 | LL: -452.1873 | AIC: 918.3745 | BIC: 931.4729
```
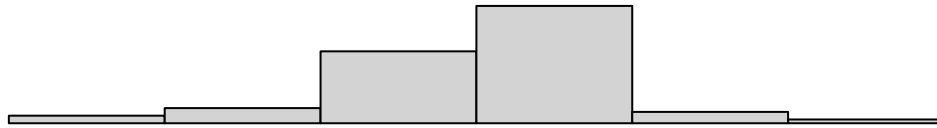


```
## SARMA(1,0,1)x(2,1,1)12 Normality: 0.00150916 | LL: -452.5189 | AIC: 917.0378 | BIC: 928.265
```
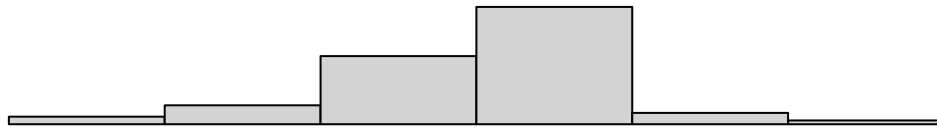


```
## SARMA(0,0,1)x(2,1,1)12 Normality: 0.05590713 | LL: -452.0751 | AIC: 914.1502 | BIC: 923.5062
```


Upon analyzing the performance metrics, it becomes evident that SARMA(0,0,1)x(2,1,1)12 is the only model that passes the Shapiro-Wilk normality test, indicating that the residuals are normally distributed. Moreover, this model also exhibits the best performance across all other metrics (log-likelihood, AIC, and BIC), making it the most suitable choice for our analysis.
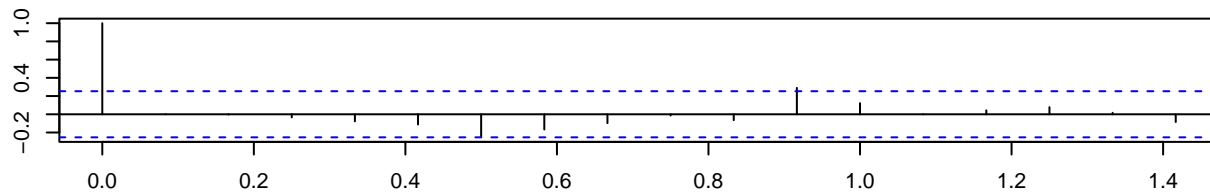
Seasonal Residual Analysis

We will now evaluate the residuals of the selected SARMA(0,0,1)x(2,1,1)12 model. The following plots will be displayed: standardized residuals, residual ACF, Ljung-Box test p-values, and normal Q-Q plot.
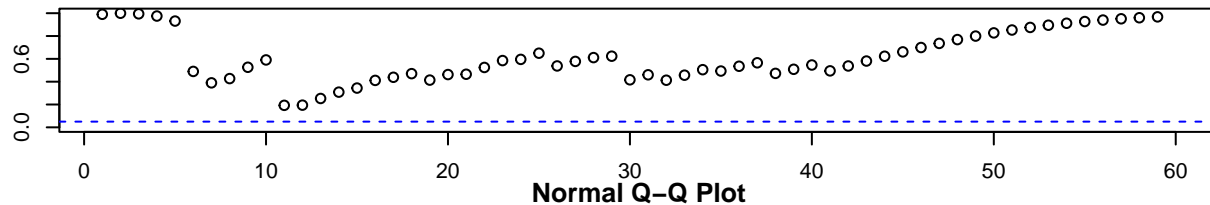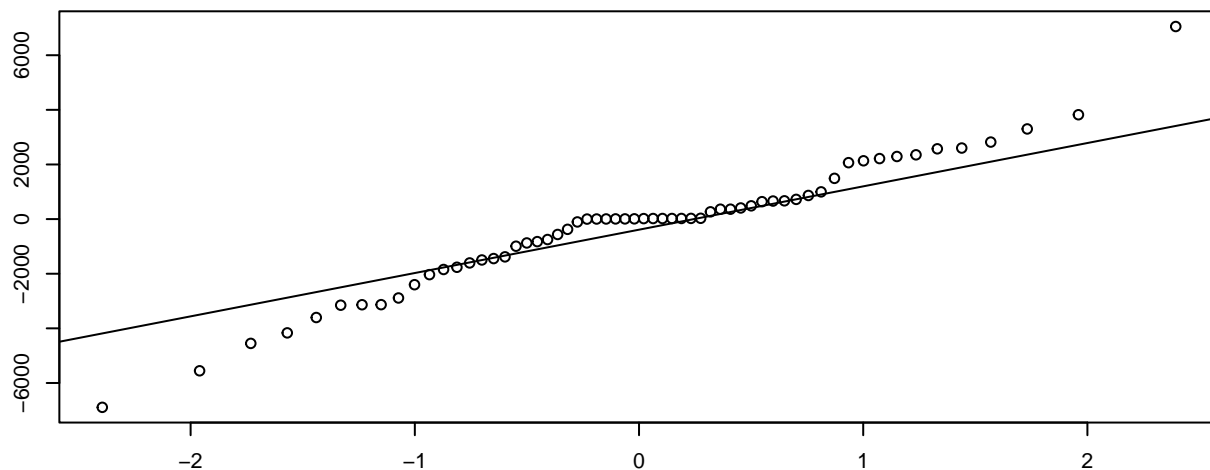
Standardized Residuals
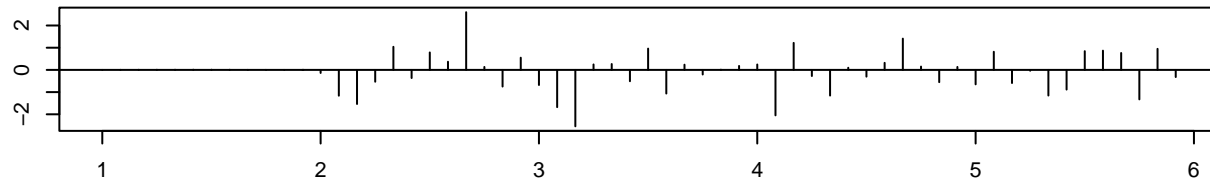
### ACF of Residuals

p values for Ljung–Box statistic

Normal Q–Q Plot

Upon inspecting the plots, we observe that there is one significant residual in the ACF, suggesting some remaining autocorrelation. The Q-Q plot also indicates some deviation from normality. However, a critical observation is that all lags pass the Ljung-Box test, implying that the model's residuals are, in general, independent. With this result, we can proceed to forecasting using our chosen SARMA(0,0,1)x(2,1,1)12 model.

Seasonal Forecasting

We will now generate forecasts for the next 24 data points, which is equivalent to a 2-year period.


Forecasted SARMA(0,0,1)x(2,1,1)12
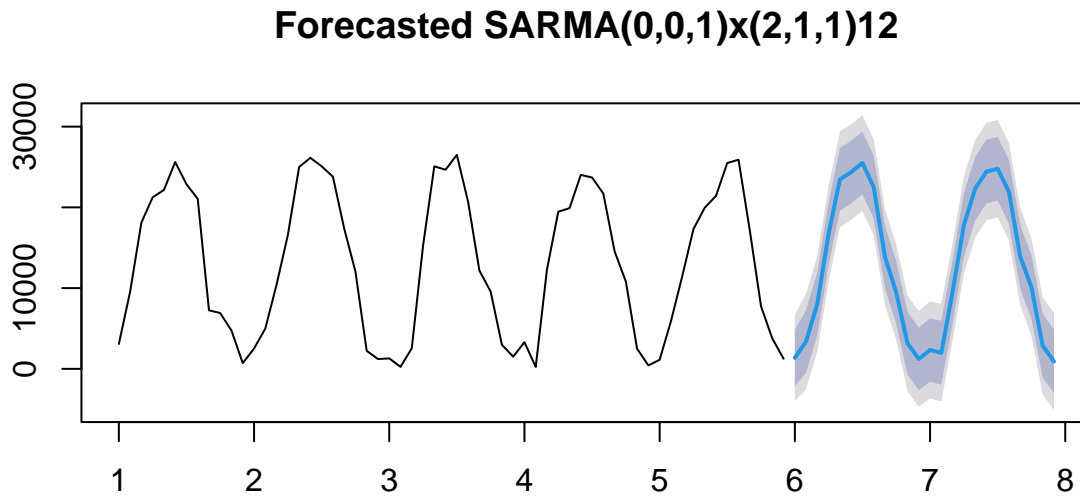
As can be observed from the plot, even though the confidence intervals extend beyond any previously observed data, the SARMA(0,0,1)x(2,1,1)12 model appears to accurately capture the seasonality and trend present in the solar production data. This suggests that our chosen model is well-suited for predicting the future behavior of this time series.

Verifying SARMA vs. SARIMA

Although our forecasts generated by the SARMA(0,0,1)x(2,1,1)12 model are reasonably good, there is still room for improvement based on the residual analysis. It is possible that the main limitation lies not in the model parameters, but rather in the data augmentation process. In this section, we will explore whether a SARIMA model (which includes an additional non-seasonal difference) could be a viable alternative.

Seasonal Differencing (12)

Seasonal Differenced ACF

Seasonal Differenced PACF
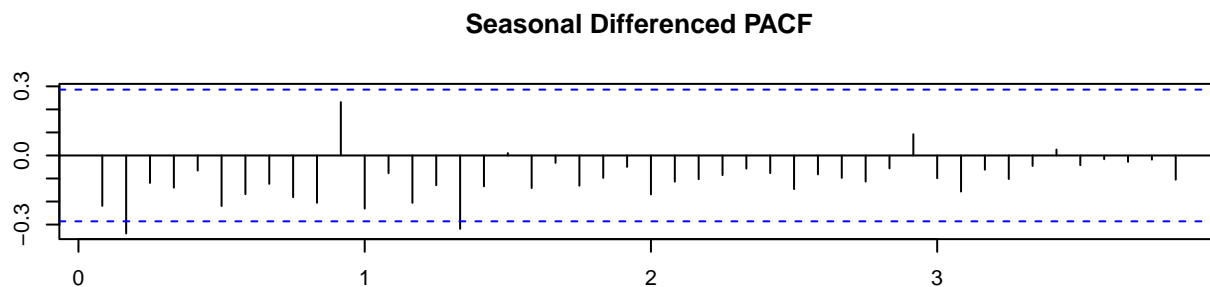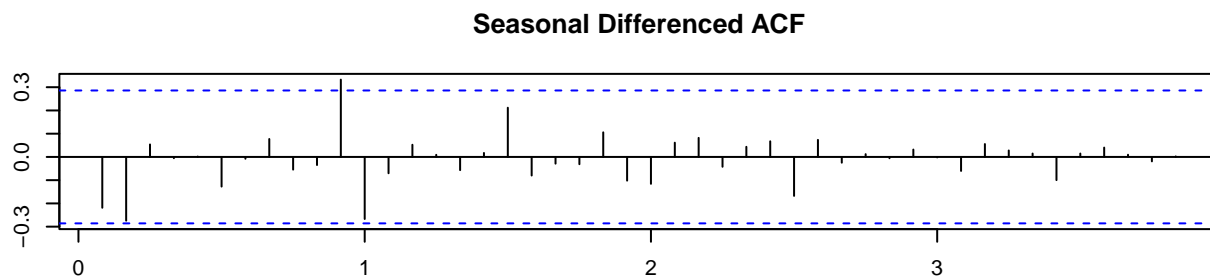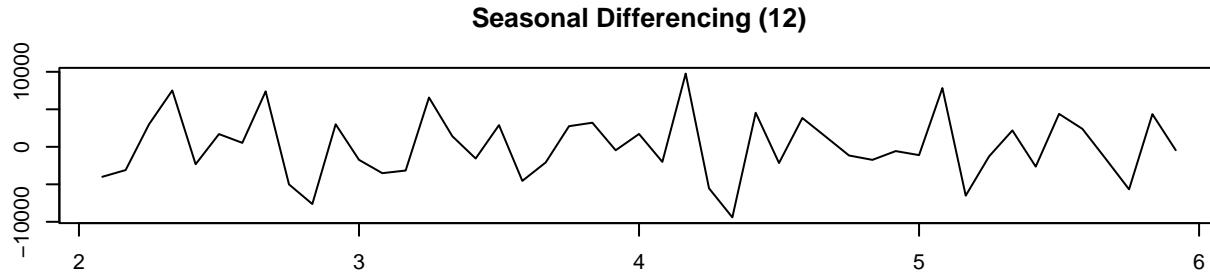
ADF test: p-value = <0.01
KPSS test: p-value = >0.1

The results of the ADF and KPSS tests suggest that the data is stationary after applying an additional non-seasonal difference. However, as can be observed from the ACF, PACF, and EACF, the dependency in the data dies off immediately, which indicates overdifferencing. Therefore, the SARIMA model with an additional non-seasonal difference does not appear to be a suitable alternative for this dataset.

Conclusion and Key Findings

In this study, we focused on analyzing the seasonality present in the monthly solar production data from the Southland Leisure Center in Canada. Through a series of tests and modeling, we arrived at the following key findings:

1. Seasonality: The data exhibits strong seasonality with a period of 12 months. The decomposition of the time series revealed a significant seasonal component.

2. Stationarity: Although the initial ADF and KPSS tests suggested stationarity, we identified the presence of seasonality as the primary source of non-stationarity in the data. Therefore, we applied seasonal differencing with a period of 12 months, which led to a stationary time series.

3. Model Selection: Based on the ACF and PACF plots and other model performance metrics, we identified the SARMA(0,0,1)x(2,1,1)12 model as the best fitting model for the seasonally differenced data.

4. Residual Analysis: The residual analysis of the chosen model revealed one significant residual in the ACF plot and some deviation from normality in the Q-Q plot. However, all lags passed the Ljung-Box test, allowing us to proceed with forecasting.

5. Forecasting: The SARMA(0,0,1)x(2,1,1)12 model accurately captured the seasonality and trend in the data, generating forecasts for 24 data points (equivalent to 2 years).

6. Alternative Models: We investigated the possibility of using a SARIMA model with an additional non-seasonal difference but concluded that it was not suitable due to the immediate decay in data dependency observed in the ACF and PACF plots, indicating overdifferencing.

In conclusion, we successfully modeled the seasonal solar production data and generated forecasts using the SARMA(0,0,1)x(2,1,1)12 model. Although there is room for improvement in the residual analysis, the chosen model demonstrates a solid understanding of the seasonality and trends present in the data.