MATH 4425 Introductory Time Series: Project

Chan Tsz Ngong, Lau Ying Yee Ava

Table of Contents

[1 Introduction 1](#_Toc103559108)

[2 CSOP Hang Seng Index Daily (-2x) Inverse [7500.HK] 2](#_Toc103559109)

[2.1 Source of Data 2](#_Toc103559110)

[2.2 Reason of selecting 7500.HK for analysis 2](#_Toc103559111)

[2.3 ARIMA-GARCH model fitting 2](#_Toc103559112)

[2.3.1 Transform to Near Normality 3](#_Toc103559113)

[2.3.2 Order Selection: Finding d 3](#_Toc103559114)

[2.3.3 Order Selection: Finding p and q 4](#_Toc103559115)

[2.3.4 Parameter Estimation for ARIMA(p,d,q) model 7](#_Toc103559116)

[2.3.5 ARIMA Forecasting 7](#_Toc103559117)

[2.3.6 Testing for ARCH effect 8](#_Toc103559118)

[2.3.7 Parameter Estimation for GARCH(1,1) model 9](#_Toc103559119)

[2.3.8 ARIMA-GARCH Forecasting 10](#_Toc103559120)

[3 Monthly Average of Global Solar Radiation (MGSR), measured in King’s Park 11](#_Toc103559121)

[3.1 Source of Data 11](#_Toc103559122)

[3.2 Reason of selecting MGSR 11](#_Toc103559123)

[3.3 Model Selection 12](#_Toc103559124)

[3.3.1 Visualizing the data 12](#_Toc103559125)

[3.3.2 Data Transformation 12](#_Toc103559126)

[3.3.3 Differencing the Data (Obtaining and ) 12](#_Toc103559127)

[3.3.4 Finding and 14](#_Toc103559128)

[3.3.5 Prediction 16](#_Toc103559129)

[3.3.6 Conclusion 17](#_Toc103559130)

[4 Reference 17](#_Toc103559131)

1 Introduction

In this project, we examined two Time Series Data Sets, which include:

1. The Historical Adjusted Close Price for CSOP Hang Seng Index Daily (-2x) Inverse
2. Monthly Average of Global Solar Radiation, measured at King’s Park

We examined the two Time Series Data using ARIMA-GARCH model and Seasonal ARIMA model respectively, and below are the findings of the Time Series Analysis.

2 CSOP Hang Seng Index Daily (-2x) Inverse [7500.HK]

2.1 Source of Data

The Stock Data of 7500.HK are obtained from Yahoo Finance [1]:

<https://finance.yahoo.com/quote/7500.HK/history?>

We examined a total of 725 rows of data ranging from May 27, 2019, to Apr 29, 2022. For illustration purposes, we took the first 715 rows of data to fit the ARIMA-GARCH model and use the last 10 rows of data to illustrate the accuracy of ARIMA-GARCH Forecasting.

2.2 Reason of selecting 7500.HK for analysis

7500.HK is an inverse ETF with leverage provided by CSOP Asset Management. CSOP Asset Management used Futures to construct a portfolio which enables them to provide an ETF that follows the Hang Seng Index inversely, with a multiplicative factor of 2, and minimize the daily tracking error. Different from another inverse ETF provided by CSOP Asset Management 7300.HK, which does not have leverage and is thus usually used by investors for hedging the bearish market, 7500.HK aims to provide investors with a way to gain from the bearish market.

In Hong Kong, individual investors could not short sell stocks easily. Before the launch of 7500.HK, individual investors could only buy 7300.HK or derivatives that is risky, such as Callable Bear Contract and Warrants, to make a profit from the bearish market. As a result, 7500.HK provides a choice for individual investors to take advantage of the bearish market, especially from mid of 2019 till now, by taking a medium level of risk.

2.3 ARIMA-GARCH model fitting

To start with, we first plot the Adjusted Close of 7500.HK

Chart, histogram

Description automatically generated

Fig. 2.1. Visualizing the data pattern of 7500.HK from May 27, 2019, to Apr 29, 2022.

2.3.1 Transform to Near Normality

Then we create a QQ-plot for our training data to observe if the data is Normal or not.

Chart, histogram

Description automatically generated

Fig. 2.2. Visualizing the QQ plot of the first 715 row of data.

We can observe that the data is not close to normal, so (natural) log transform is attempted to transform the to Normal Data.

2.3.2 Order Selection: Finding d

After transforming the data by taking natural log, a plot is created to observe if the time series of the transformed data is stationary.

Chart

Description automatically generated

Fig. 2.2. Visualizing the training data pattern after log transformation.

As the data is not stationary, first-order difference is taken to make the data stationary.

Chart

Description automatically generated

Fig. 2.3. Visualizing the training data pattern after log transformation and differencing

As the data become stationary after taking first-order difference, the order d is chosen to be 1, so the log return of the Adjusted Close Prices follows an ARIMA(p,1,q) model.

2.3.3 Order Selection: Finding p and q

First, Ljung-Box Test, with the null hypothesis of having the first 20 ACF values equal to 0, is used to test if the residuals of the log return are of a white noise process.

Text

Description automatically generated

Table 2.1. Result of Ljung-Box Test of Time Series

As the p-value is smaller than 0.05, we concluded that the null hypothesis - the residual is a white noise process - should be rejected, and there are autocorrelations between data.

To obtain the order of AR and MA coefficient, plots of ACF and PACF are used.

Chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Fig. 2.4. ACF and PACF plot of .

According to Fig. 2.4, the ACF plot cuts off at the 10th ACF, while the PACF plot cuts off after the 6th PACF. Therefore, we chose p = 6 and q = 10 as the order.

However, for higher accuracy, ARIMA(p,1,q) model for p and q from 0 to 10 are fitted to the log-transformed Adjusted Close Price and performed Ljung-Box Test with the adjusted degree of freedom 20-p-q to test the randomness of the residuals of the ARIMA model. The p values are as follows.

Table

Description automatically generated

Table 2.2. Testing results of possible combinations of p and q with its respective p-value obtained in Ljung-Box Test .

Then AIC of the fitted ARIMA(p,1,q) models are recorded in the table below. For entries with the p-values smaller than 0.05, which means there is significant autocorrelation in the residual of the ARIMA fit, is adjusted to ‘Inf’. This is to facilitate the use of min() function when finding the minimum AIC with p-value > 0.05 in the AIC Matrix in Table 2.3.

Table

Description automatically generated

Table 2.3. Testing result of possible combinations of p and q with its repective AIC of fitted model. Combinations with p-value < 0.05 are adjusted to ‘Inf’ for grid search purpose.

As the best model in terms of AIC is the model with the minimum AIC, p = 4 and q = 5 are selected for the remaining analysis.

2.3.4 Parameter Estimation for ARIMA(p,d,q) model

AR and MA coefficients are obtained from the arima() function in R and the results are as follows

Text

Description automatically generated

Table 2.4. Coefficients of ARIMA.

Again, the result of Ljung-Box Test with adjusted degree of freedom 20-4-5 = 11 are as follows

Text

Description automatically generated with medium confidence

Table 2.4. Result of Ljung-Box Test of residuals of ARIMA model.

With the above model fitting procedures, the model with lowest AIC is selected.

2.3.5 ARIMA Forecasting

In the following, 10-step ahead Forecasting is performed and the mean forecast and the forecasting interval (95%, Z value is taken to be 1.96) are obtained by using the predict() function in R. Exponential Transformation is taken to revert the natural log of Adjusted Close Price to Adjusted Close Price. The following table compares the mean forecast and forecasting interval of ARIMA model with the actual Adjusted Close Price.

Table

Description automatically generated

Table 2.5. 10 days ahead forecast and forecasting interval of ARIMA model comparing with actual adjusted close.

For illustration purpose, we plot the result with the recent 100 rows of data.

Chart, histogram

Description automatically generated

Fig 2.5. plot of 10 days ahead forecast and forecasting interval comparing with actual adjusted close.

2.3.6 Testing for ARCH effect

Ljung-Box test for the squared of the first 12 ACF of the residuals of the ARIMA model obtained above is used to test for the ARCH effect, and the results are as follows.

Text

Description automatically generated with low confidence

Table 2.6. Result of Ljung-Box Test of squared residuals of ARIMA model.

As the p-value is much smaller than 0.05, we reject the null hypothesis that there is no ARCH effect. To model the stochastic volatility of the model, GARCH(1,1) model is used to fit the residuals of the ARIMA model.

2.3.7 Parameter Estimation for GARCH(1,1) model

GARCH coefficients are obtained from the garchFit() function in R and the results are as follows.

Text

Description automatically generated with medium confidence

Table 2.7. Coefficients of GARCH model of ARIMA residuals with several test statistics.

We can observe from Table 2.7 that both alpha and beta coefficients of the GARCH(1,1) model are statistically significant, while the mean of residuals and the intercept term are not statistically significant.

2.3.8 ARIMA-GARCH Forecasting

In the following, 10-step ahead forecasting is performed to the GARCH model is obtained by using the predict() function in R.

The MLE of h is taken to be the squared of the mean error/ standard error of the estimate.

Table

Description automatically generated

Table 2.8. 10-days ahead error forecast of GARCH model.

As the exact formula for the n-step ahead forecasting interval is not provided in the lecture notes, the forecast variance of each step is assumed to be independent. As a result, the variance of the n-step ahead variance is calculated by the sum of the first h to the current h.

The mean forecast used comes from the ARIMA model, and the 95% forecasting interval is obtained by using adding or subtracting 1.96 times the square root of n-step ahead forecast variance from the GARCH model. Exponential transformation is taken to revert the natural log of Adjusted Close Price to Adjusted Close Price. Table 2.9 compares the mean forecast and forecasting interval of ARIMA-GARCH model with the actual Adjusted Close Price and the ARIMA forecasting Interval.

Table

Description automatically generated

Table 2.9. 10 days ahead forecast and forecasting interval of ARIMA-GARCH model comparing with actual adjusted close and forecasting interval of ARIMA model.

For illustration purpose, we plot the result of ARIMA-GARCH forecasting interval with the recent 100 rows of data.

Chart, histogram

Description automatically generated

Fig. 2.6. 10-days ahead forecast and forecasting interval of ARIMA-GARCH model comparing with actual adjusted close.

From Fig 2.6, we observe that the forecasting interval of ARIMA-GARCH model is wider than the ARIMA model in this data set. This result shows that the 95% forecasting interval of the ARIMA-GARCH model includes a wider interval to forecast the unexpected fluctuation of the market.

3 Monthly Average of Global Solar Radiation (MGSR), measured in King’s Park

3.1 Source of Data

The dataset is publicly available on the data platform hosted by the Hong Kong Government [2]. It records a total of 10866 daily global solar radiation values (in ) from July 1992 to March 2022. In this project, we grouped the values of the same month to obtain the monthly average, which is a total of 357 entries.

Here we split the 80% of the data as the training dataset, and 20% of the data as the testing dataset. This results in having 286 entries for training, while 71 for testing.

3.2 Reason of selecting MGSR

Global solar radiation (GSR) accounts for both diffuse and direct solar radiation that reaches the Earth’s surface. The task of GSR prediction is important in two ways. Firstly, it is the major energy source for life on Earth. It has huge influence on the climate of our habitat and natural processes such as evaporation and glacier melt. Prediction of GSR would be beneficial in analyzing our living environment. Secondly, it has practical implications on solar energy technologies. Hong Kong has been actively developing infrastructure for the generation of renewable energy. The study of the GSR pattern is useful for determining the location of a large-scaled solar energy infrastructure as well as accelerating energy harvesting of solar panels.

3.3 Model Selection

3.3.1 Visualizing the data

From Fig 3.1, the MGSR values follow a yearly cycle: it first decreases from July, then increases from January. Therefore, the model may follow the seasonal ARIMA model, which is

.

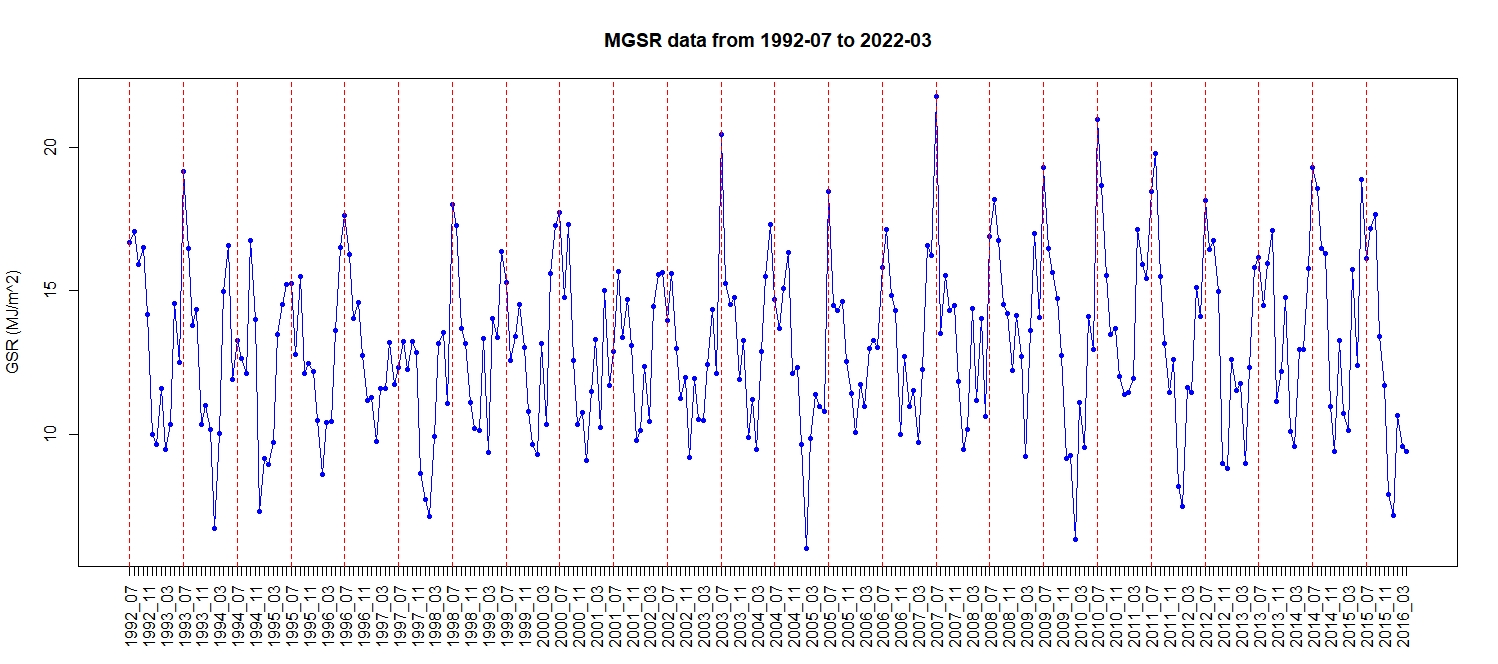


Fig. 3.1. Visualizing the data pattern of MGSR from July 1992 to March 2022. The vertical red lines partition the data into one-year intervals.

3.3.2 Data Transformation

Since the MGSR values spread across a large interval, we first take the log transformation, i.e. where is the MGSR data. The transformed values are shown in Fig. 3.2. The range of the values are reduced.

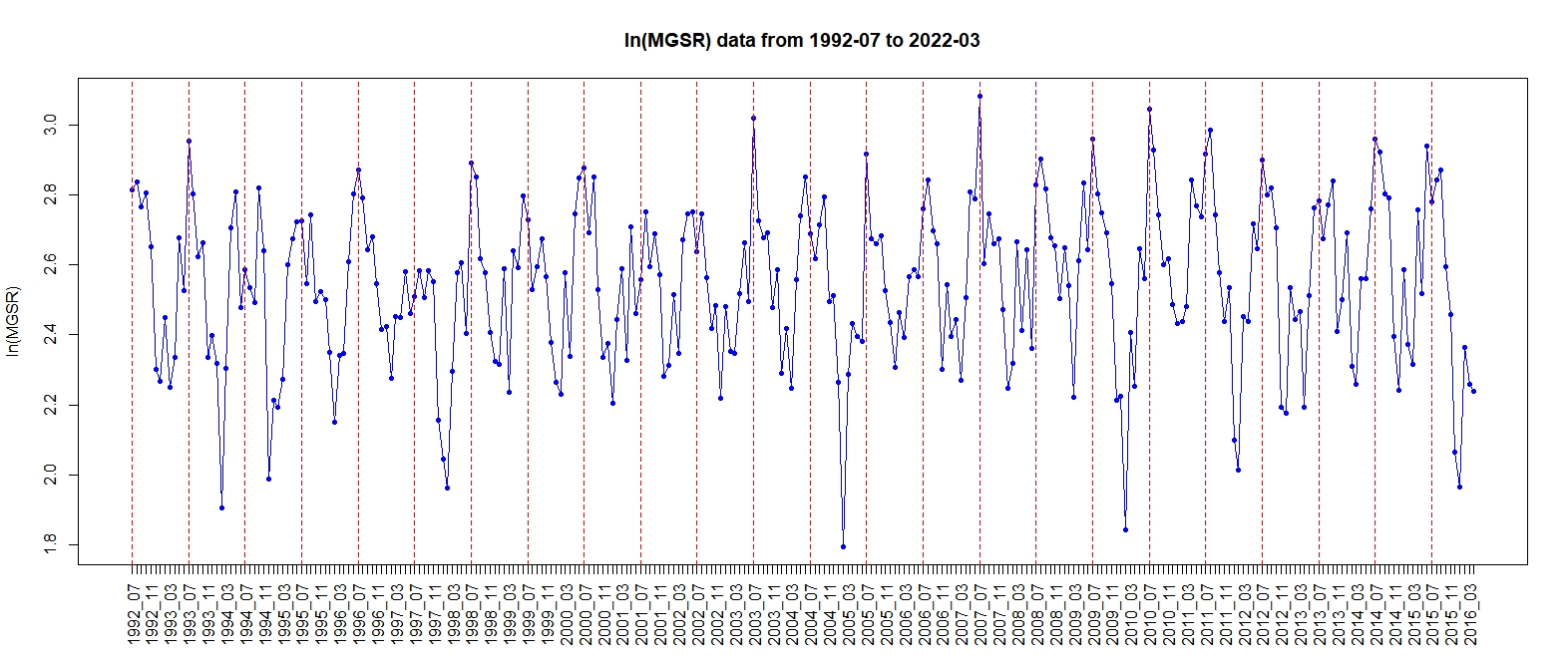
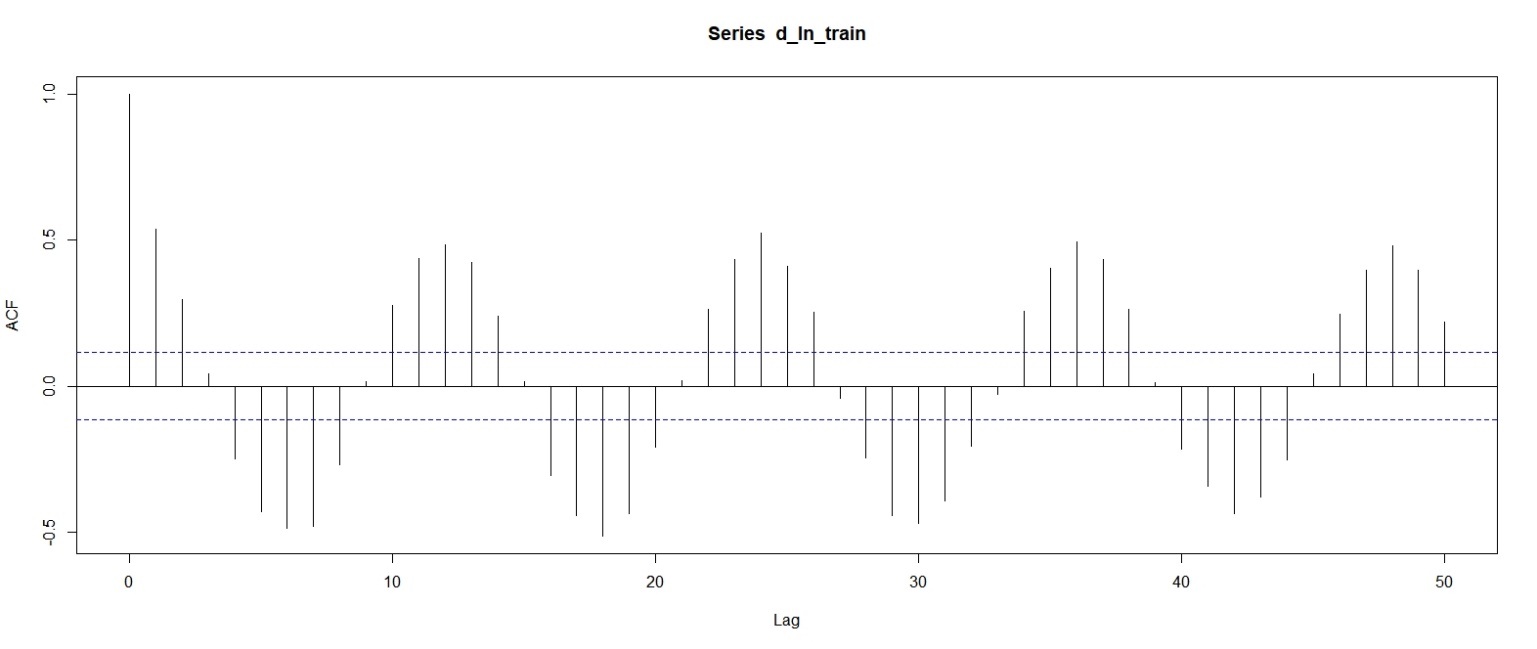


Fig. 3.2. Visualizing the data pattern after log transformation. Denoted as .

3.3.3 Differencing the Data (Obtaining and )

From the ACF plot of in Fig. 3.3, we can observe that the ACF values decay very slowly, which may indicate that the data is not stationary. As the plot has shown a periodic cycle of 12 lags, we try to take the difference, so that .



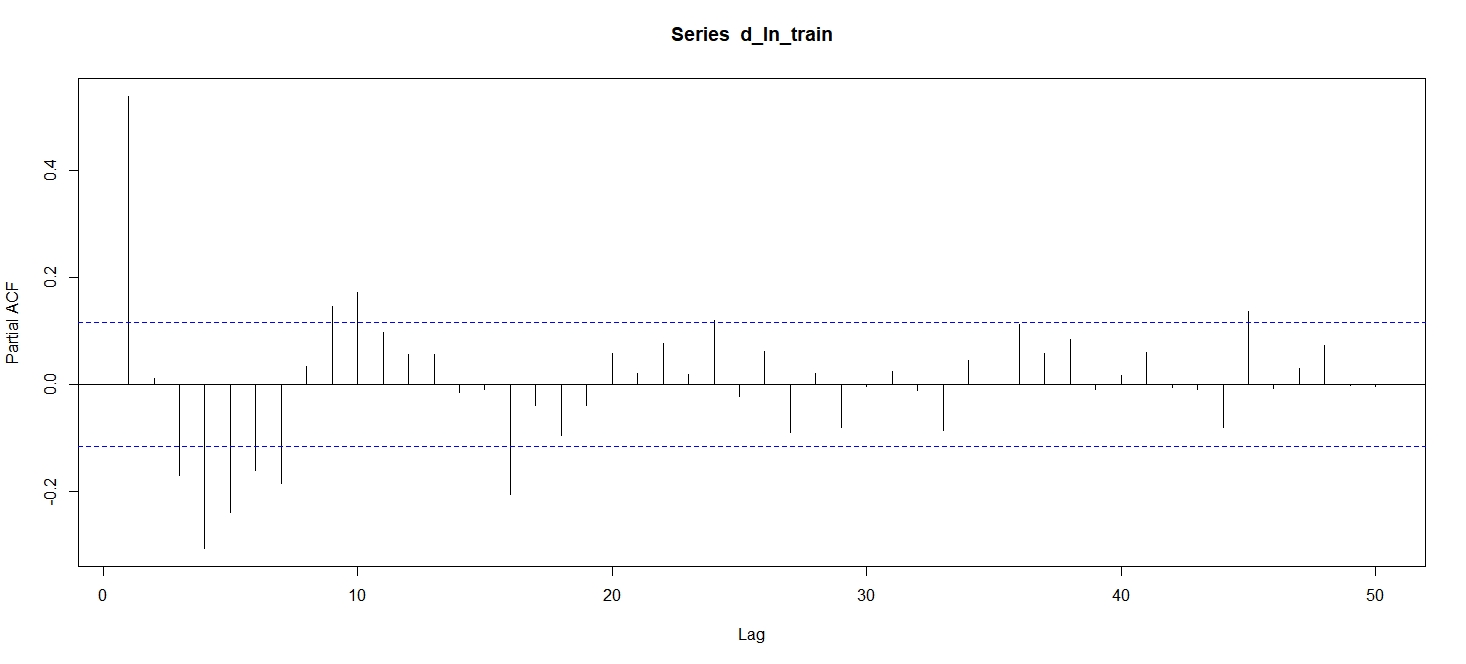


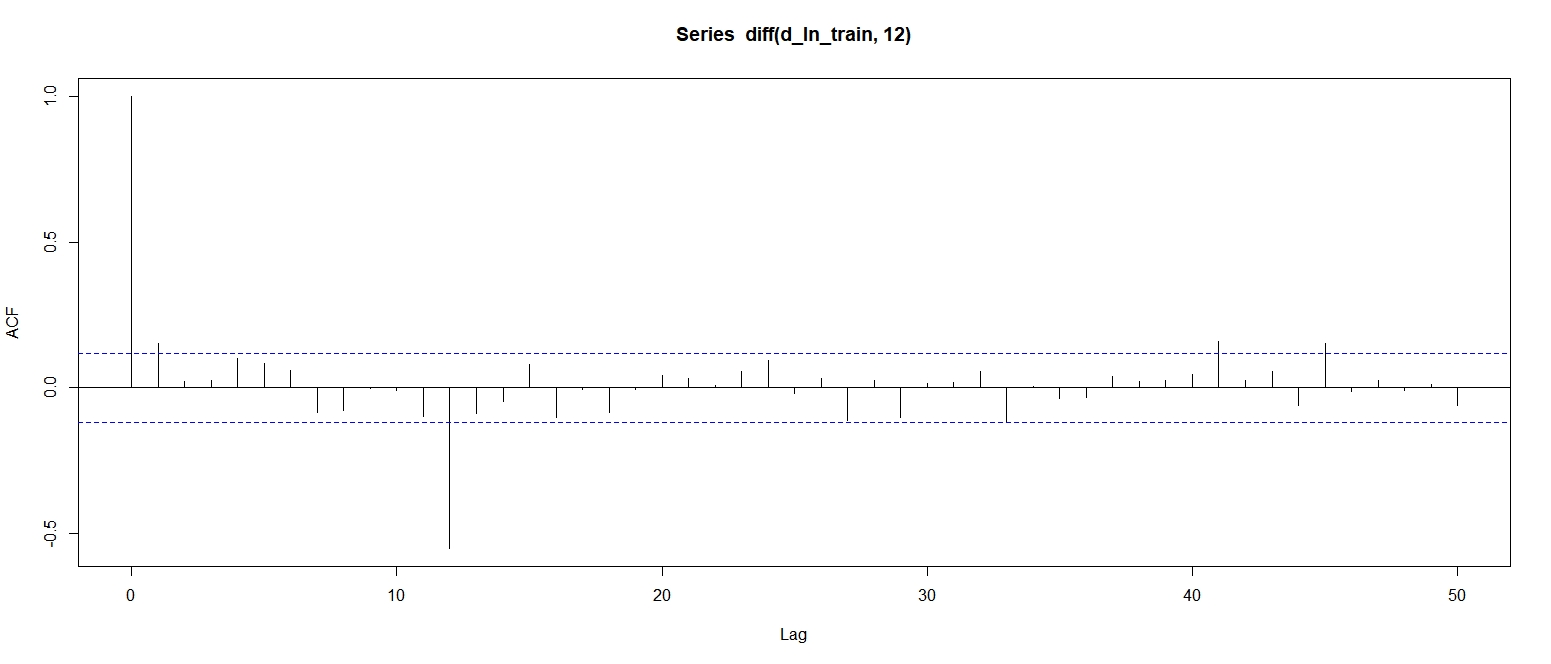
Fig. 3.3. ACF and PACF plot of .

After differencing , we can see that in Fig. 3.4, the data is stationary, and only few significant ACF and PACF values are shown in Fig 3.5.

A picture containing timeline

Description automatically generated

Fig. 3.4. Visualizing the data pattern after log transformation and differencing, i.e. .



A picture containing diagram

Description automatically generated

Fig. 3.5. ACF and PACF of .

Then from the ACF plot of Fig. 3.5, we can observe that is significant, which may imply that . Similarly, from the PACF plot of Fig. 3.5, values like and shows that or , while a spike in may imply that .

3.3.4 Finding and

Drawing conclusions from Fig. 3.5, we tested with values , , and , i.e., a total of combinations. The results are shown in Table 3.1, ranked in ascending order of AIC. All the p-values of L-Jung Box test and AIC are calculated after the removal of parameters that are not statistically significant to be non-zero with confidence interval of 95%.

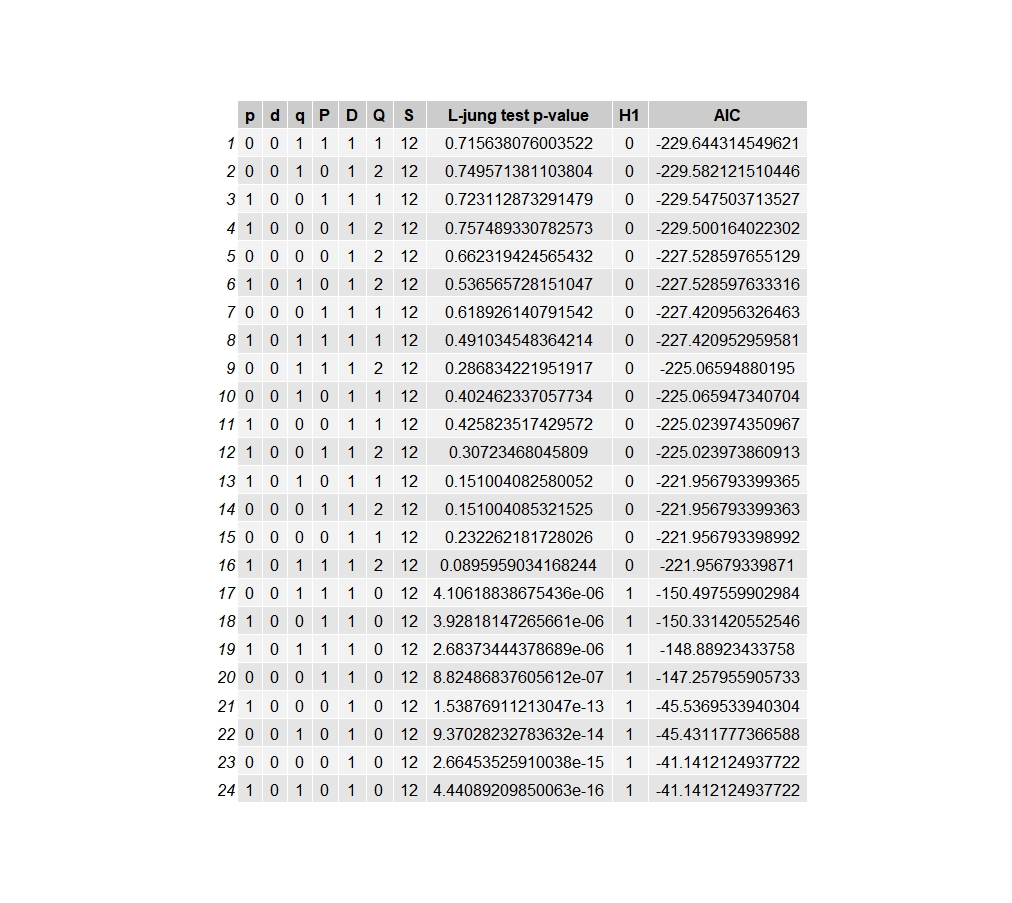


Table 3.1. Testing result of possible combinations, sorted in ascending order of AIC. The left 7 columns represent the parameters of the seasonal ARIMA model. Then it shows the p-value of the L-Jung box test, tested with degree of freedom , as suggested in [3]. The column “H1” indicates whether the null hypothesis of the L-Jung box test – the residuals of the model are not independently distributed - is rejected, equivalent to having the p-value of L-Jung box test smaller than 0.05. Having a value 0 means that the null hypothesis is not rejected.

From table 3.1, we select model 1 which has the lowest AIC for further analysis. Particularly, the coefficients of model 1 SARIMA is shown in Table 3.2. Its distribution of residuals of model 1 are shown in Fig. 3.6.

Table

Description automatically generated

Table 3.2. Coefficients of model 1: SARIMA.

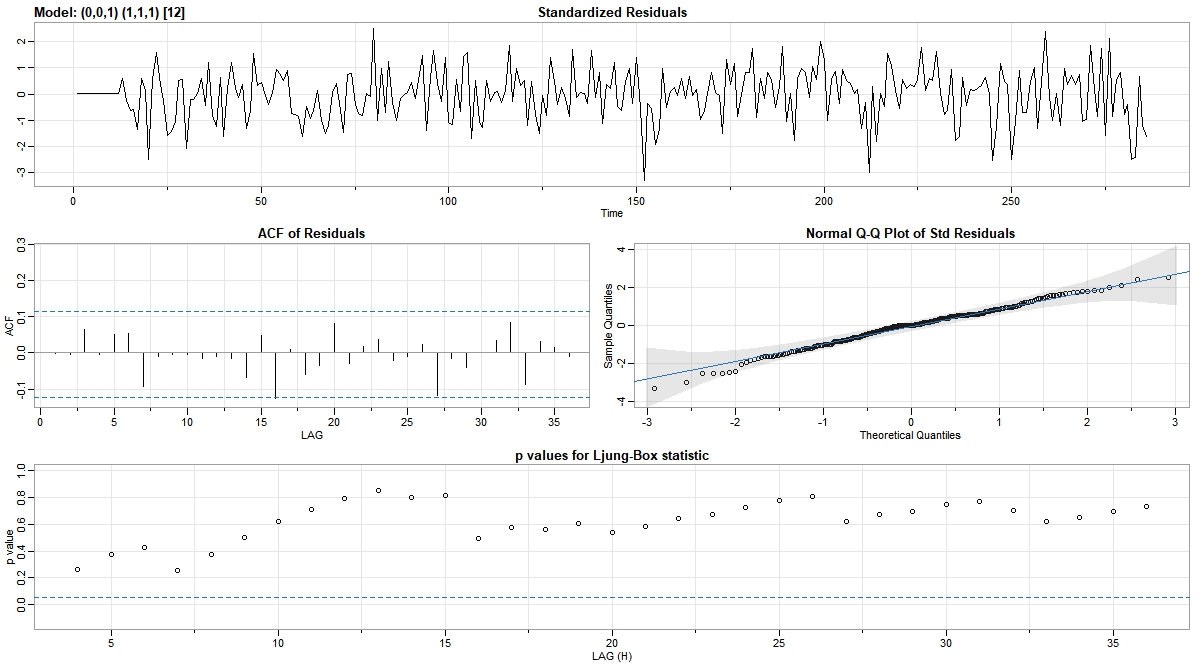


Fig. 3.6. Residual plot of model 1: SARIMA.

3.3.5 Prediction

Here we only conduct prediction for the model with smallest AIC, which is model 1: SARIMA.

Predictions are generated and compared with all the values in the test dataset, i.e., 71 steps ahead. We first generate the predictions based on , which is shown in Fig. 3.7. Then the prediction values are transformed into its original form as shown in Fig. 3.8.

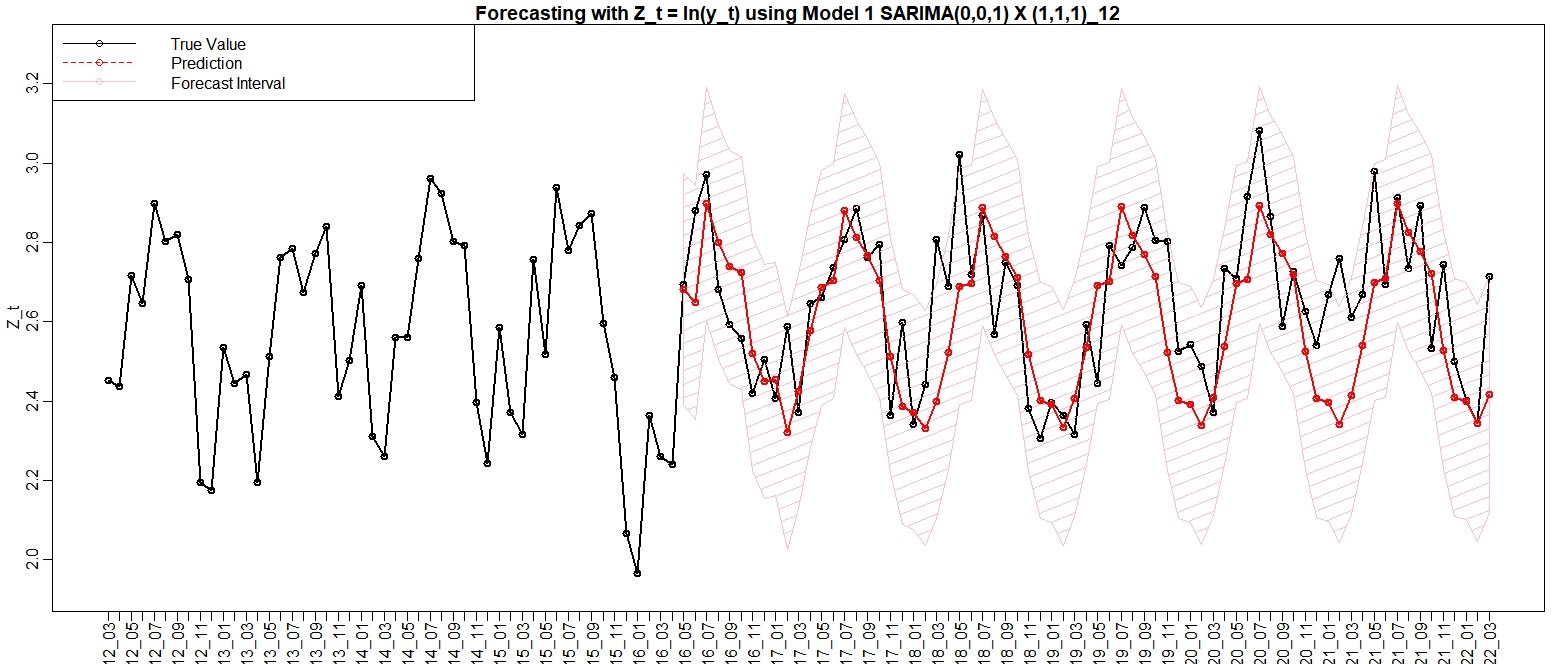


Fig. 3.7. Predictions for model 1 with with 95% forecast interval.

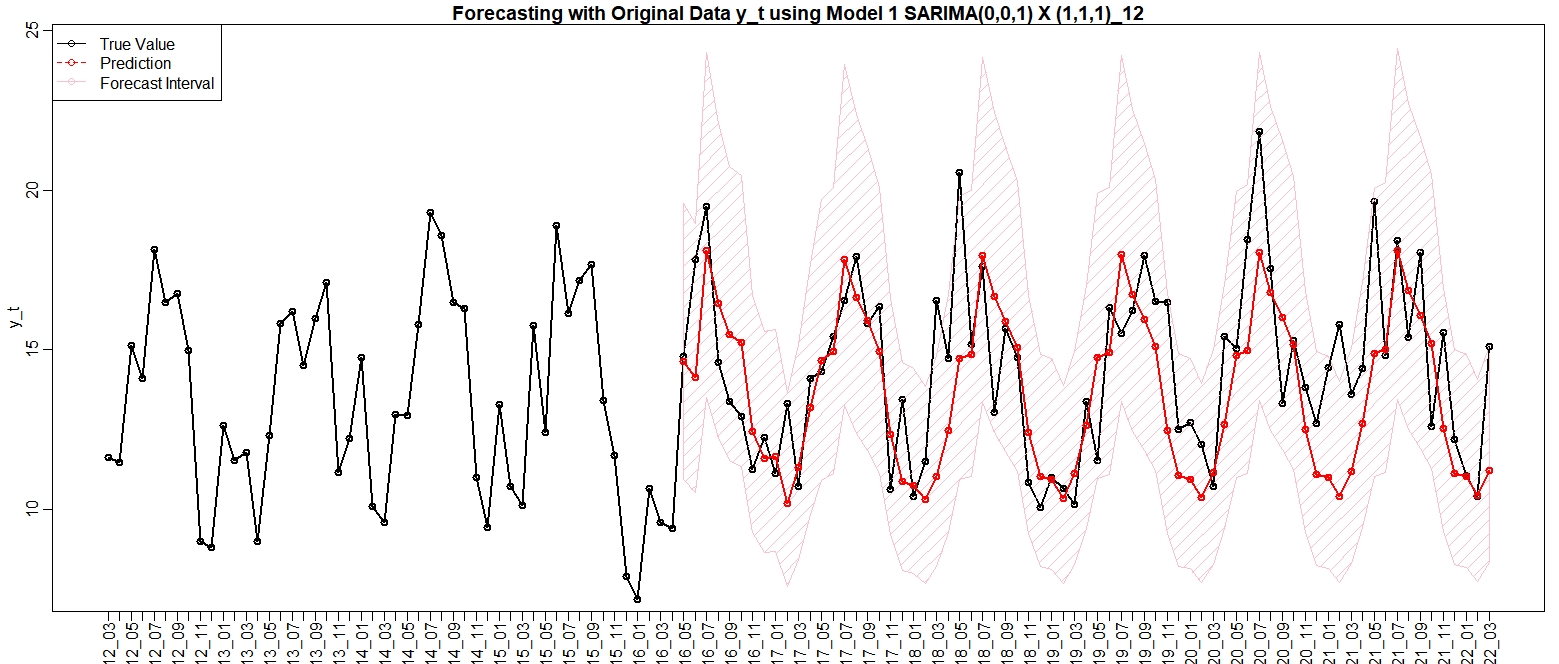


Fig. 3.8. Predictions for model 1 with with 95% forecast interval.

From the above plots, model 1 has shown reasonable forecasting values, having most of the predicted values lying in the forecast interval. Finally, in table 3.3, we use the mean-squared error to summarize the differences of predictions and true data.

|  |  |  |
| --- | --- | --- |
|  | Log values | Original values |
| Model 1 | 0.02465061 | 4.988274 |

Table 3.3. The mean squared error for model 1.

3.3.6 Conclusion

The MGSR model follows the seasonal ARIMA model . Particularly, it has the coefficients

where is the original MGSR data with unit .

4 Reference

[1] Yahoo! (2022, May 14). Csop Hang seng index daily (-2x) inverse product (7500.HK) stock historical prices &amp; data. Yahoo! Finance. Retrieved May 15, 2022, from https://finance.yahoo.com/quote/7500.HK/history

[2] Daily global solar radiation. Daily global solar radiation | DATA.GOV.HK. (n.d.). Retrieved May 15, 2022, from https://data.gov.hk/en-data/dataset/hk-hko-rss-daily-global-solar-radiation

[3] Thoughts on the Ljung-Box Test. Portrait of the author. (2014, January 24). Retrieved May 15, 2022, from https://robjhyndman.com/hyndsight/ljung-box-test/