**Gold Price Time Series Prediction**

**Ishwarya Ramkumar**

**Nischita Ramachandrappa Mulikeri**

**Rahul Suhas Galgali**

**Team D4**

**Department of Management Information Systems, San Diego State University**

**MIS 748: Time Series for Business Forecasting**

**Dr. Xialu Liu**

**May 07, 2024**

**Contents**

**Introduction …………………………………..…………………………………………………3**

**Discovery and Data preparation …………………………………………….…………………5**

**Model Building …………………………………..……………………………………………..13**

**Forecasting and prediction ………………..……………………………………...………..…..18**

**Conclusion ……………………………………………………………………………………..20**

**References .……………………………………………………………………………………..20**

**Introduction**

Gold is one of the most valuable minerals in the world. Gold has become more popular as well as a very useful commodity in terms of investment. Gold has been used as a national reserve for many years, and that makes it very crucial in the economics of any country. Most of the investors are running to gold as a safe area from uncertainty and political chaos. A gold reserve is an amount of gold held by the central bank of any country for the purpose of the guarantee to be used to pay or trade in the world market and hence increase the country economically. Amongst all minerals in the world, gold is the most popular selection for investment. The price of gold is affected by different factors, thus making the movement of price unstable. These factors include inflation rate, demand and supply, and political issues among others. Inflation is one of the signs of economic growth. When it increases it obviously pushes the gold price higher, while when having a low supply of any commodity, the price of that commodity increases. Moreover, when countries fear the value of the dollar will fall since the dollar is the world's marketing currency, the gold price will eventually increase since demand for gold will remain.

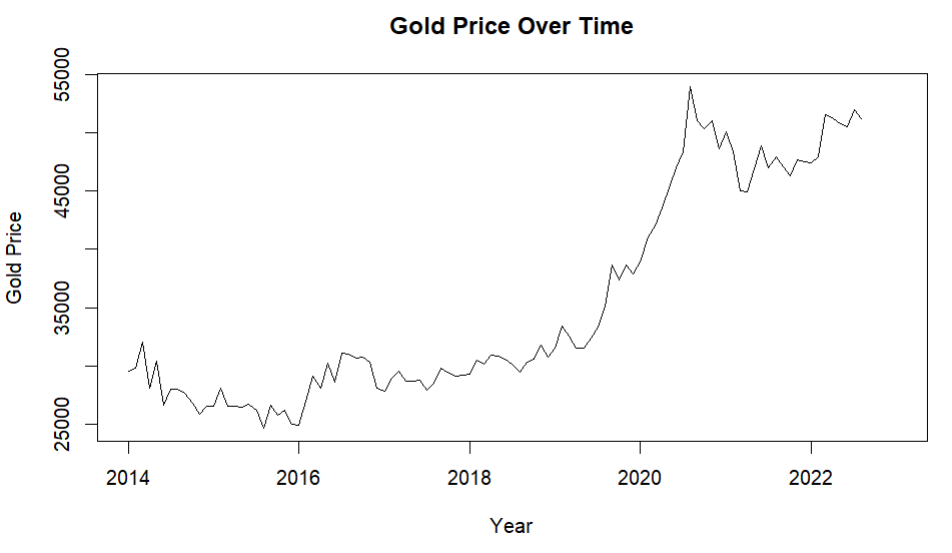
People investing in gold have mainly two primary objectives, one being it is a hedge against inflation as over a period of time, the return on gold investment is in line with the rate of inflation, next to mix your investment basket and hence diversify the risk and will help you reduce the overall volatility of your portfolio.

Determining the price movement of gold helps the investors to focus on their investments, and the government to make correct decisions about the economy since Gold price is a key element of the world economy.

This project gives an inside view of the application of ARIMA time series model to forecast the future Gold price in the Indian market based on past data from 2014 to 2022 to mitigate the risk in purchases of gold. Hence, to give guidelines for the investor when to buy or sell the yellow metal. This financial instrument has gained a lot of momentum in the recent past as the Indian economy is curbed with factors like changing political scenario, global clues & high inflation etc, so researchers, investors and speculators are in search of different financial instruments to minimize their risk by portfolio diversification.

Therefore, the price of gold moves up and down and is very uncontrollable. Figure1 shows the movement of the gold price for the period 2014 to 2022. Nevertheless, the gold price can be forecasted ahead, and that makes it possible to make the future decision. The movement of the gold price is time-series order and hence constantly changing with time, therefore performing forecasting with such data has been challenging for a while until recent years.

***Figure 1.*** The gold price over a period of time from 2014 to 2022.



**Data Discovery**

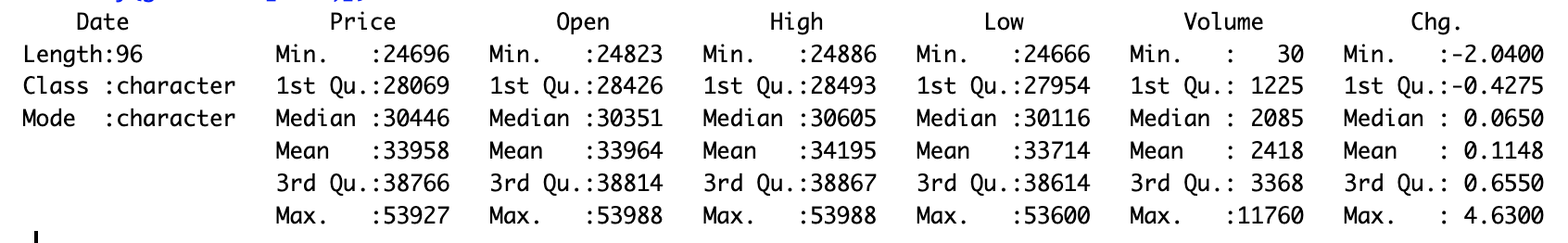
The daily gold price data set has been sourced (raw source) from “<https://in.investing.com/commodities/gold-mini>” and contains the daily gold price in India from 2014-01-01 to 2022-08-05. This dataset contains 2227 observations and includes opening price and closing price of gold. Each observation under consideration represents the closing price, which is considered as the final price. After analyzing the data from Figure1, we can say that the data does not seem to be stationary.

Table1 shows the description of data and Table2 shows the summary statistics of the data.

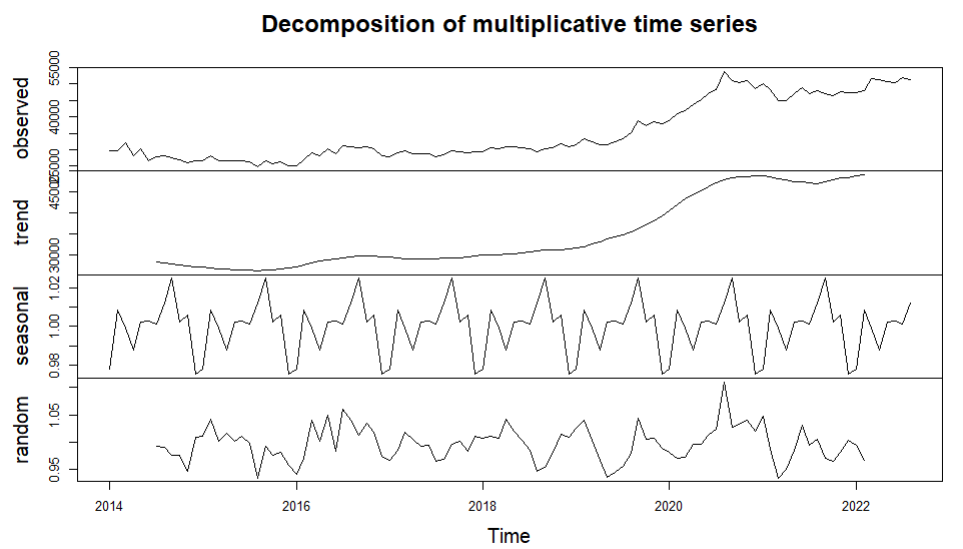
***Table 1.*** Description of the dataset.

|  |  |  |
| --- | --- | --- |
| **Column** | **Description** | **Data Type** |
| Date | The date on which the gold price was recorded on. | Date |
| Price | It is a close price which can be considered as the final price. | Integer |
| Open | Price at the time of market opening on that day. | Integer |
| High | Highest price during the whole day. | Integer |
| Low | Lowest price for the whole day. | Integer |
| Volume | Traded Volume | Integer |
| Chg% | Percentage change from previous price. | Decimal |

***Table 2.*** Summary Statistics of the data under study.



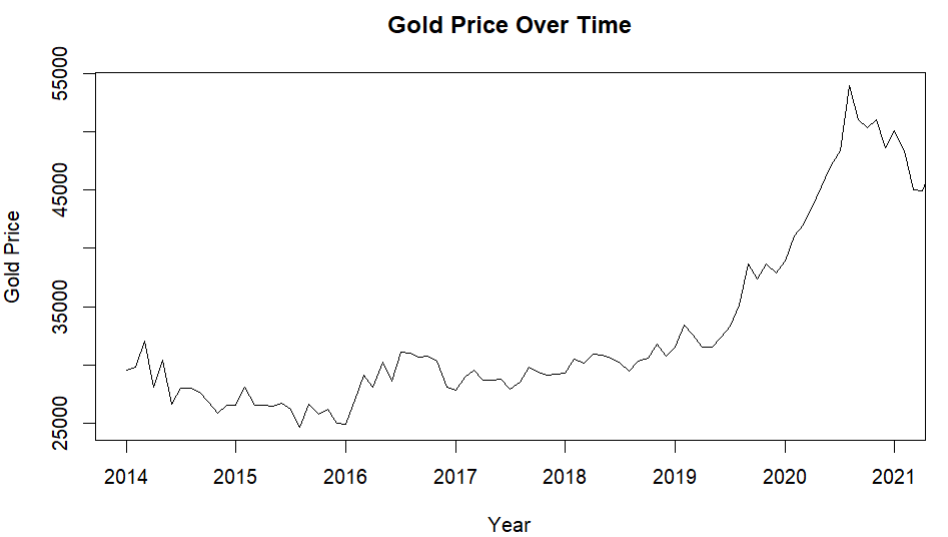
***Figure 2*.**



**Data Preparation.**

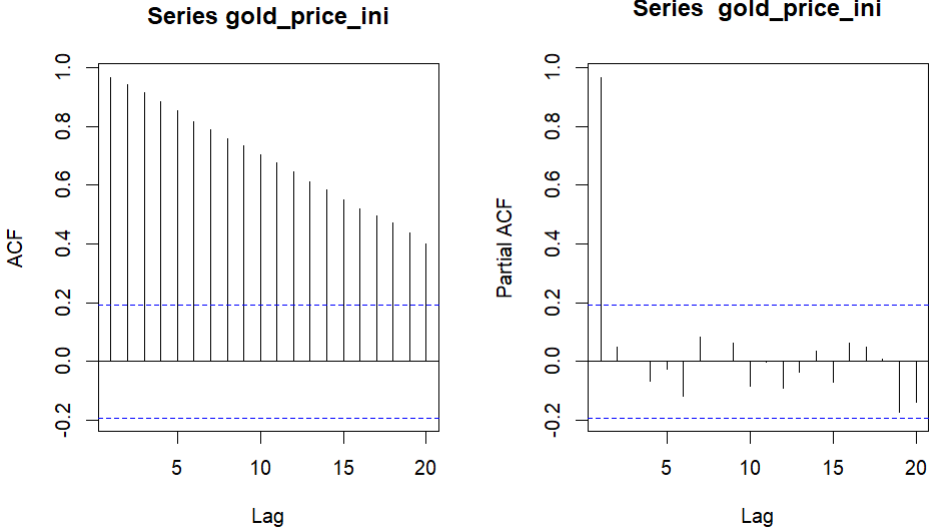
Our dataset contains daily gold prices data from 2014 to 2021 that we used for training. We then converted the daily price data into monthly data for better analysis and to reduce noise. To begin the analysis of data, the close price is considered as the final price of gold and is the most important variable in our analysis. Time series of the data is plotted as shown in Figure 3. The preliminary plot of the data shows an upward trend with a non constant mean and variance. Hence the data does not seem to be stationary at the first sight. Hence in this phase of data preparation, the data is stabilized using logarithmic and regular and seasonal difference transformations.

***Figure 3*.**  Price of gold in INR(Indian Rupee) from 2014 to 2021.

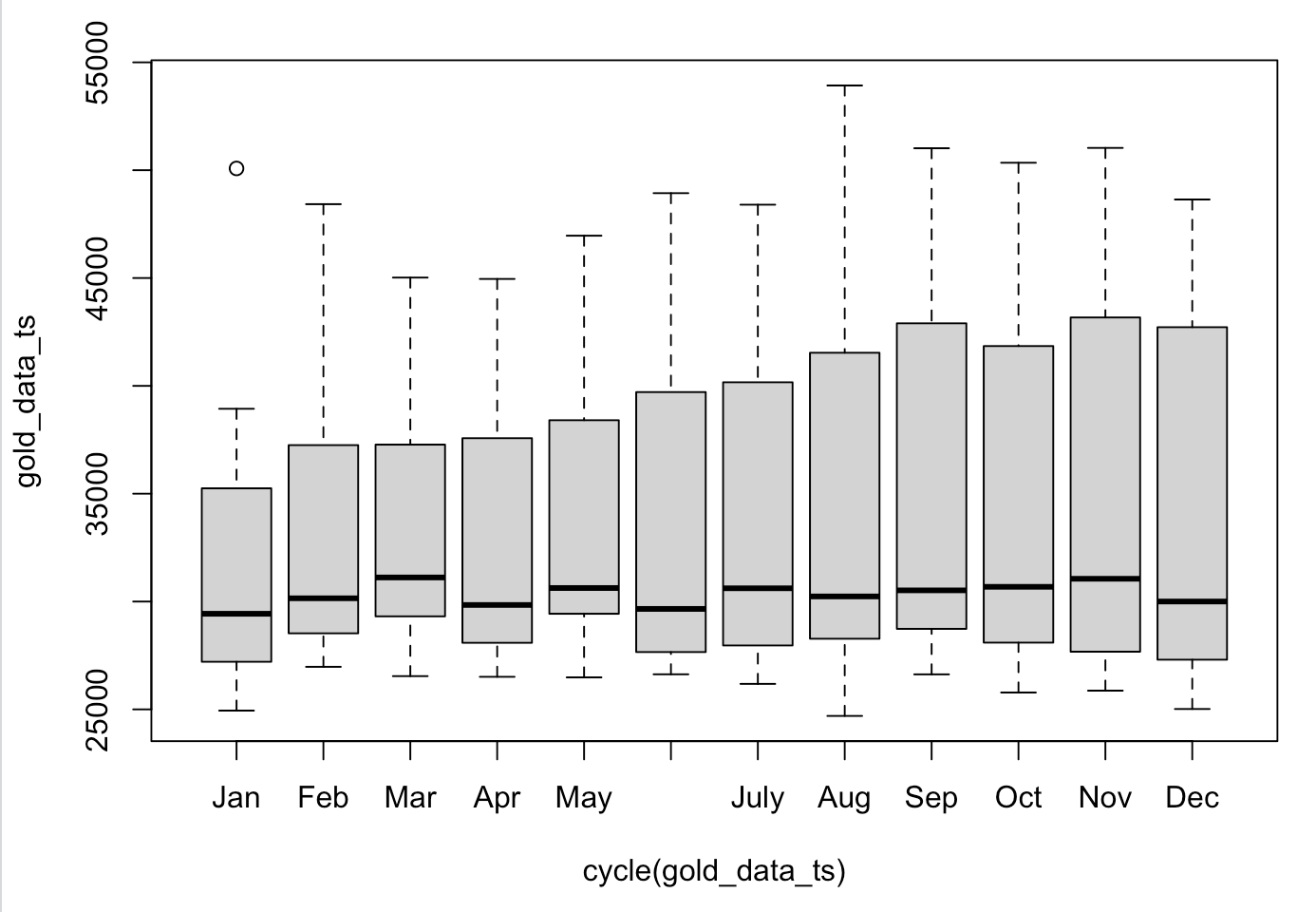


After performing the transformation, the ACF and PACF plots (Figure 4) indicated signs of seasonality. These were consistent with the box-plots as shown in Figure 5.

***Figure 4.*** ACF and PACF plots of initial data.



***Figure 5.***  Box-plot of data



The box-plot shows a seasonality in the price data. Seasonality is a regular pattern in a time series that repeats after S time periods. In this case of monthly data, the data has seasonality of s=12. From our observation, the price of gold increases every year during the second half of the year and the attributing factors may be the celebration of festivals in India and the wedding season. Celebrations as well as weddings in India are insufficient without gold; as a matter of fact, there are a few festivals like Akshay Tritya and also Dhanteras when it is considered auspicious to buy gold. After COVID-19 pandemic, when the lockdown was lifted, there seems to be a steep rise in the price of gold. The month of August has recorded the highest price of this yellow metal. It is during this time that Indians purchase gold for their personal usage or as an investment to diversify their portfolio.

From the above ACF and PACF, the ACF gradually decays to zero and PACF cuts off at lag 1 but eventually decays to zero.

To confirm the stationarity status of the data, Augmented Dickey Fuller test is performed using the function adf.test().

Augmented Dickey-Fuller Test

data: gold\_price\_ini

Dickey-Fuller = -2.3659, Lag order = 4, p-value = 0.425

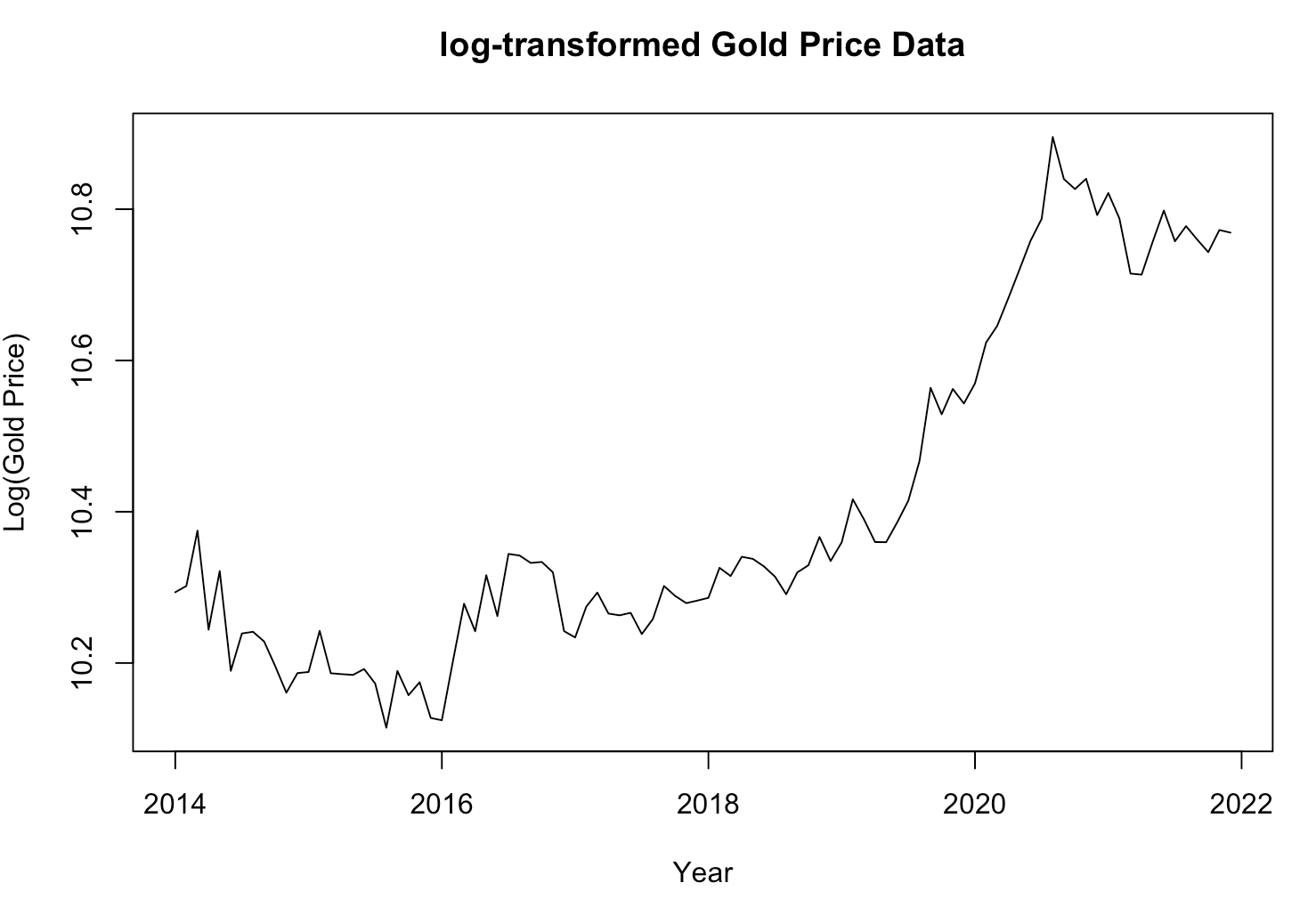
alternative hypothesis: stationary

From the adf test, p-value is 0.425 which is greater than 0.05. Hence we accept the null hypothesis, which further confirms that the data is non stationary.

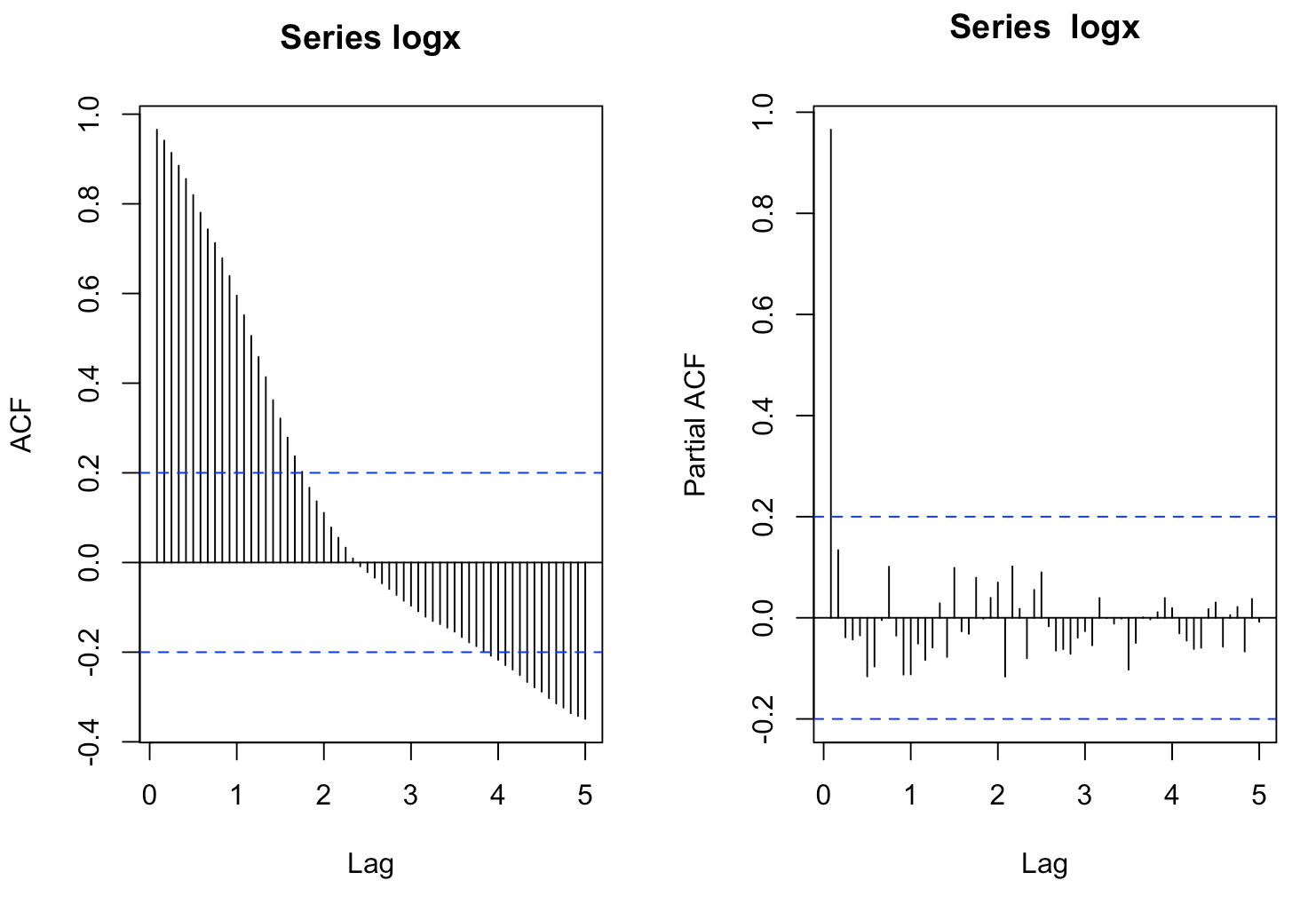
**Log transformation**:

The log transformed data (Figure 6) still shows an upward trend with a non constant mean and slightly stabilized variance. The ACF and PACF plots along with the ADF tests indicate the same results that the data is not stationary.

***Figure 6.***



***Figure 6.1***

****

Augmented Dickey-Fuller Test

data: logx

Dickey-Fuller = -2.6641, Lag order = 4, p-value = 0.3013

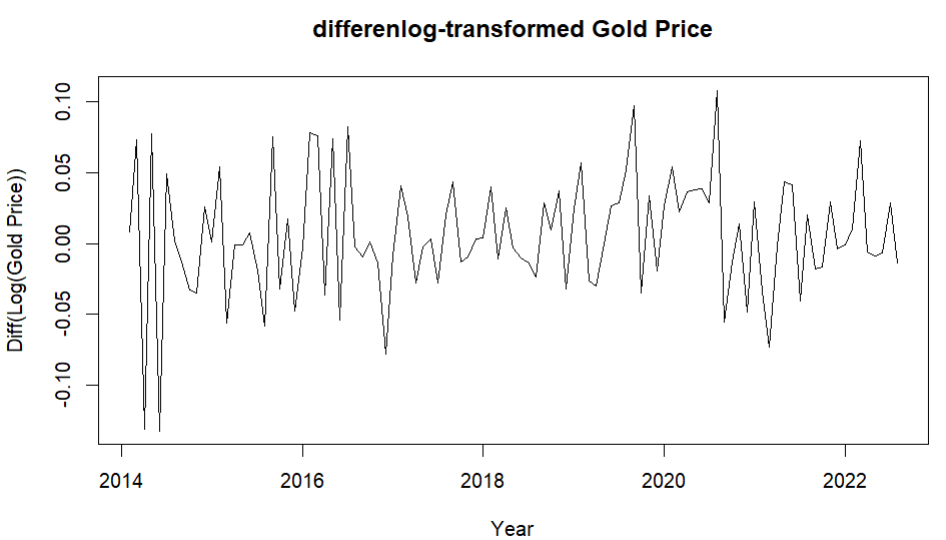
alternative hypothesis: stationary

According to adf test for log transformed data, p-value is 0.313 which is greater than 0.05, hence we accept the null hypothesis and conclude that the data is not stationary.

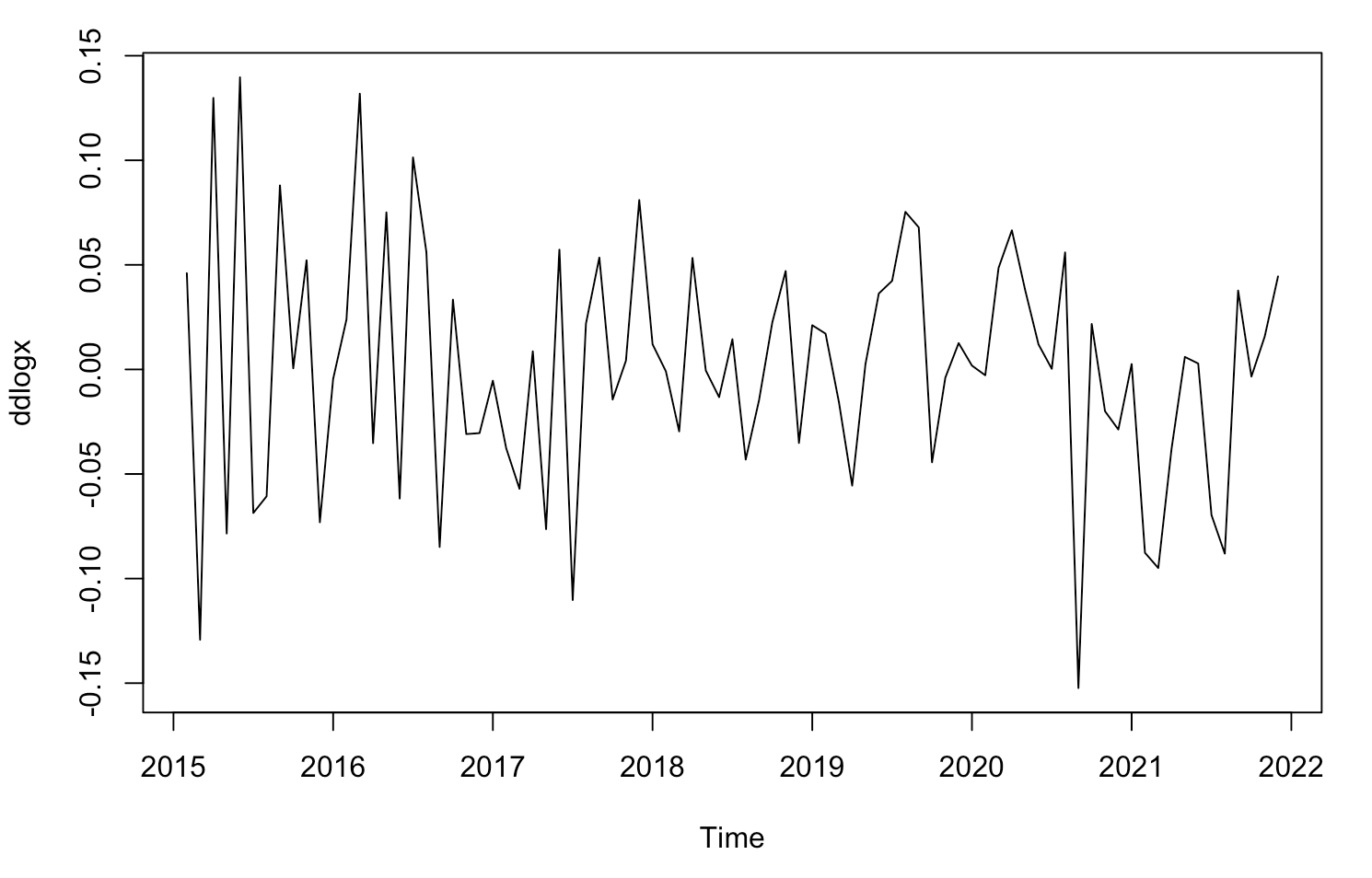
**Difference transformation:**

The data is stabilized using regular difference and seasonal difference transformation on the log data as shown in Figure 7 and 8. Figure 9 shows the seasonality component of the time series.

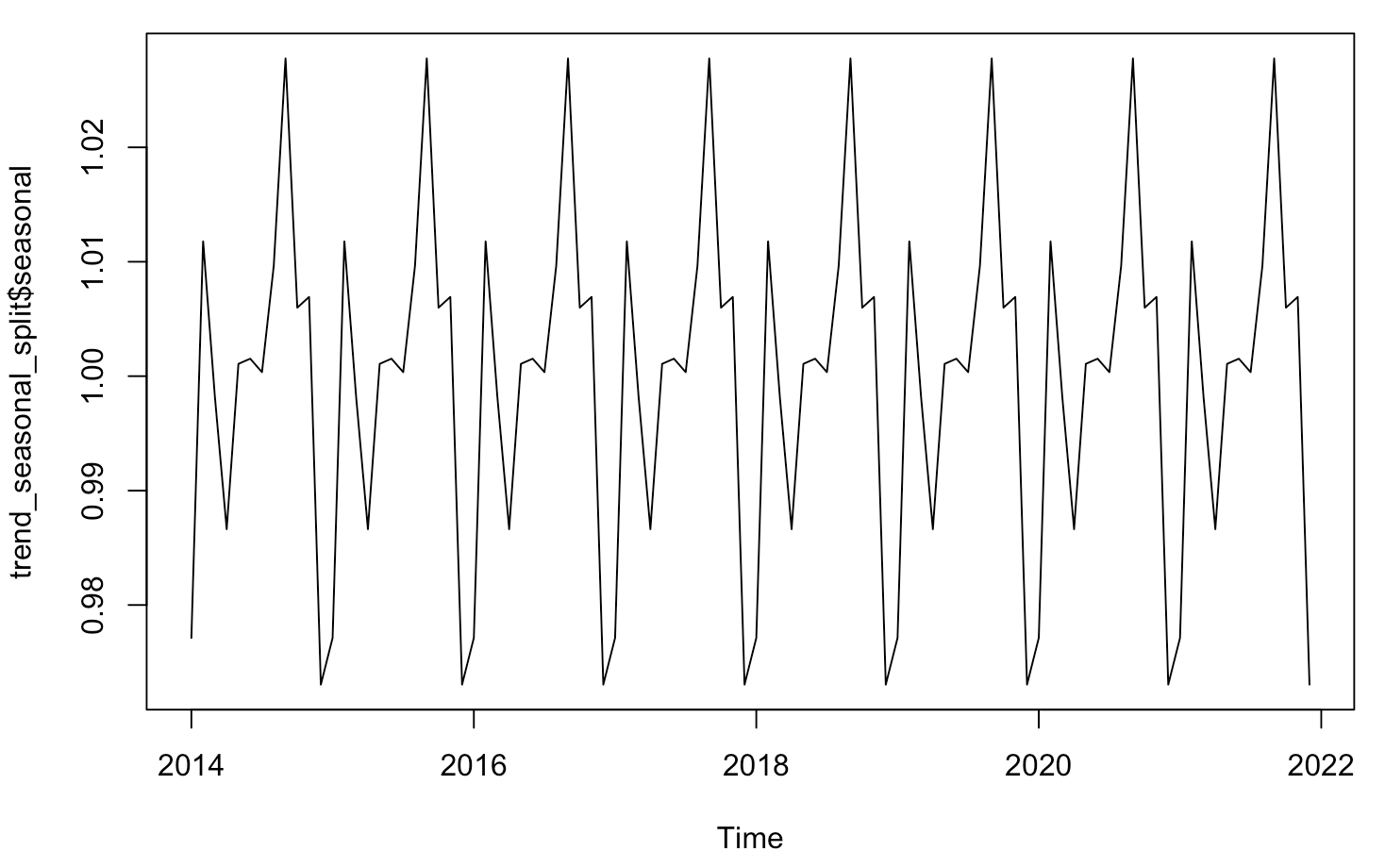
***Figure 7.*** Regular difference transformation.



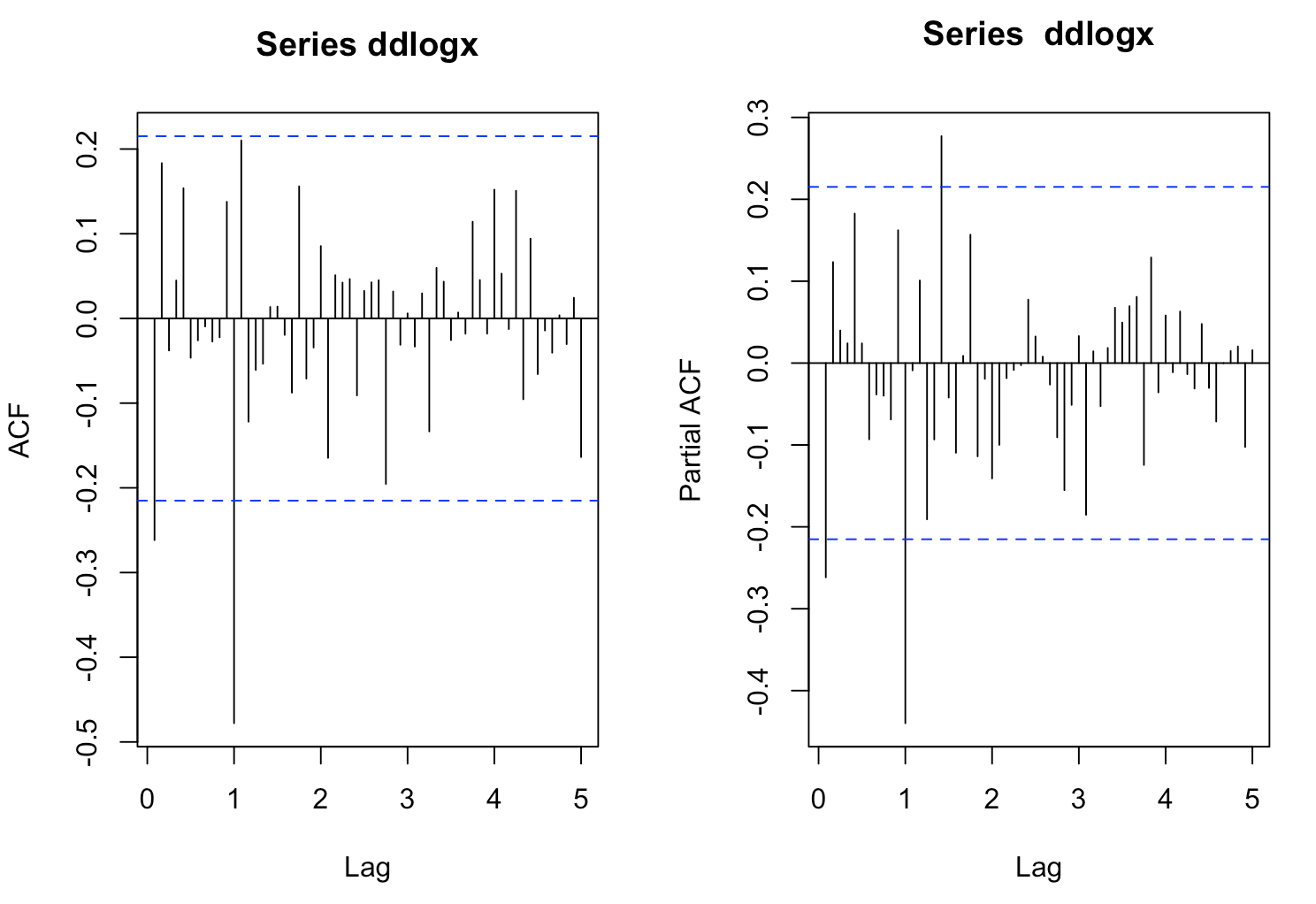
***Figure 8.*** Seasonal difference transformation.



***Figure 9.*** Seasonality component in the time series data.



***Figure 10*.** ACF & PACF plots.



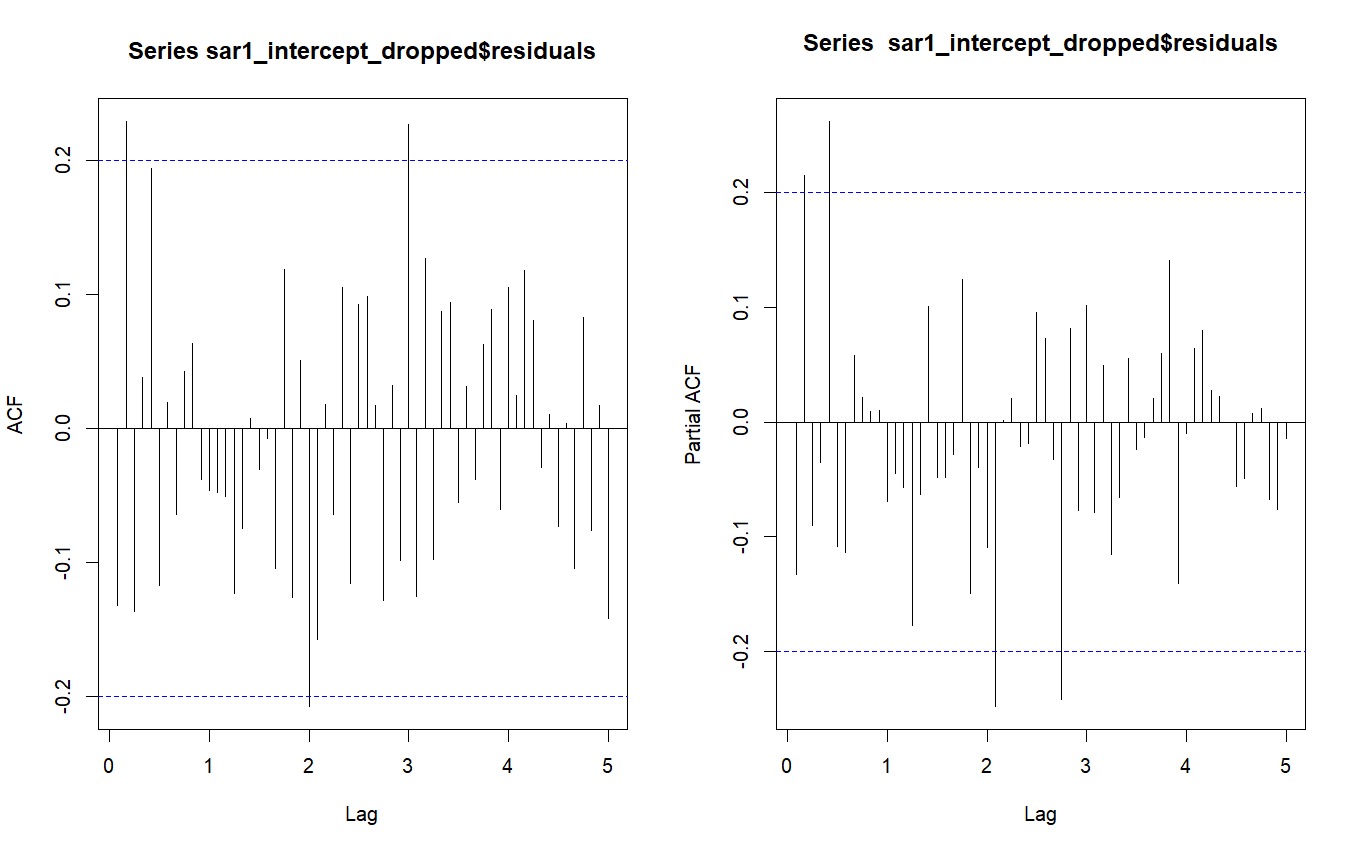
During the seasonality transformation, from the ACF and PACF plots we identified the Seasonal MA(1) and Seasonal AR(1) model.

|  |  |
| --- | --- |
| **Model** | **AIC** |
| Seasonal MA(1) | -267.88 |
| Seasonal AR(1) | -269.74 |

For further analysis and model building process, we tested these models and chose seasonal AR(1) model since it has the lowest AIC value.

**Model Building**

From ACF plot of seasonal AR(1) residuals, we choose MA(2) model with the seasonal AR(1) component.

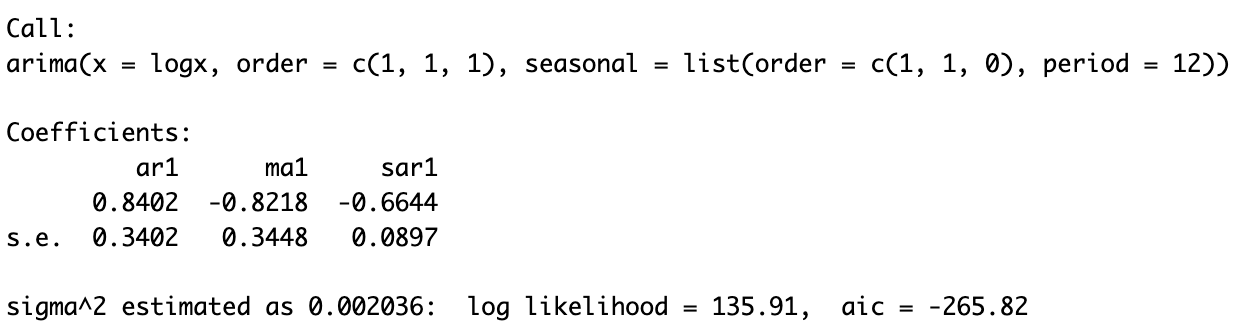


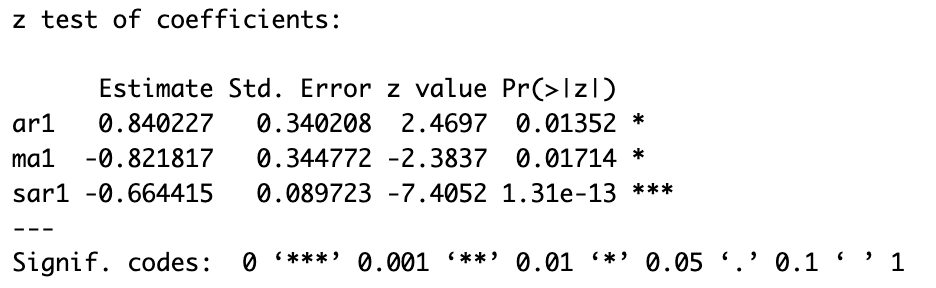
From the EACF plot shown below we choose the ARMA(1,1) model with the seasonal AR(1) component.



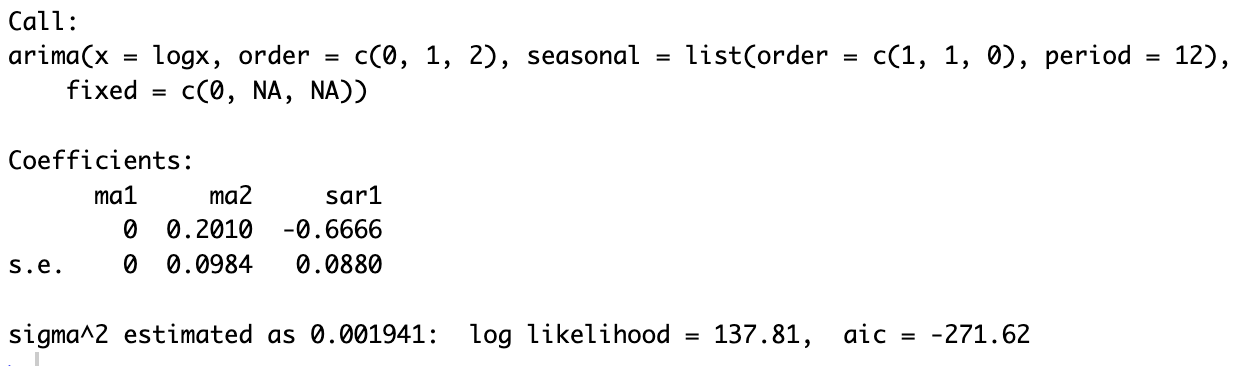
**Multiplicative Seasonal Model**

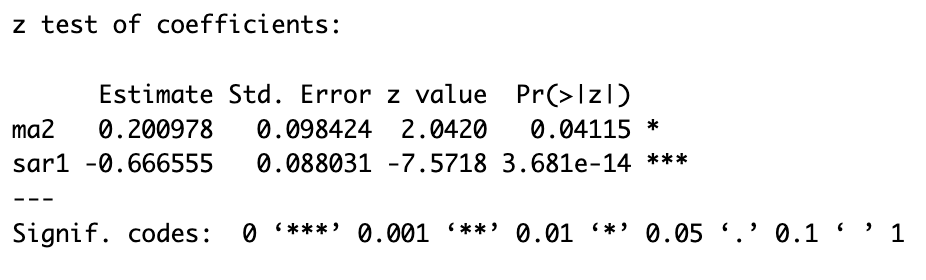
**ARIMA(1,1,1) X (1,1,0)S**





**ARIMA(0,1,2) X (1,1,0)S**

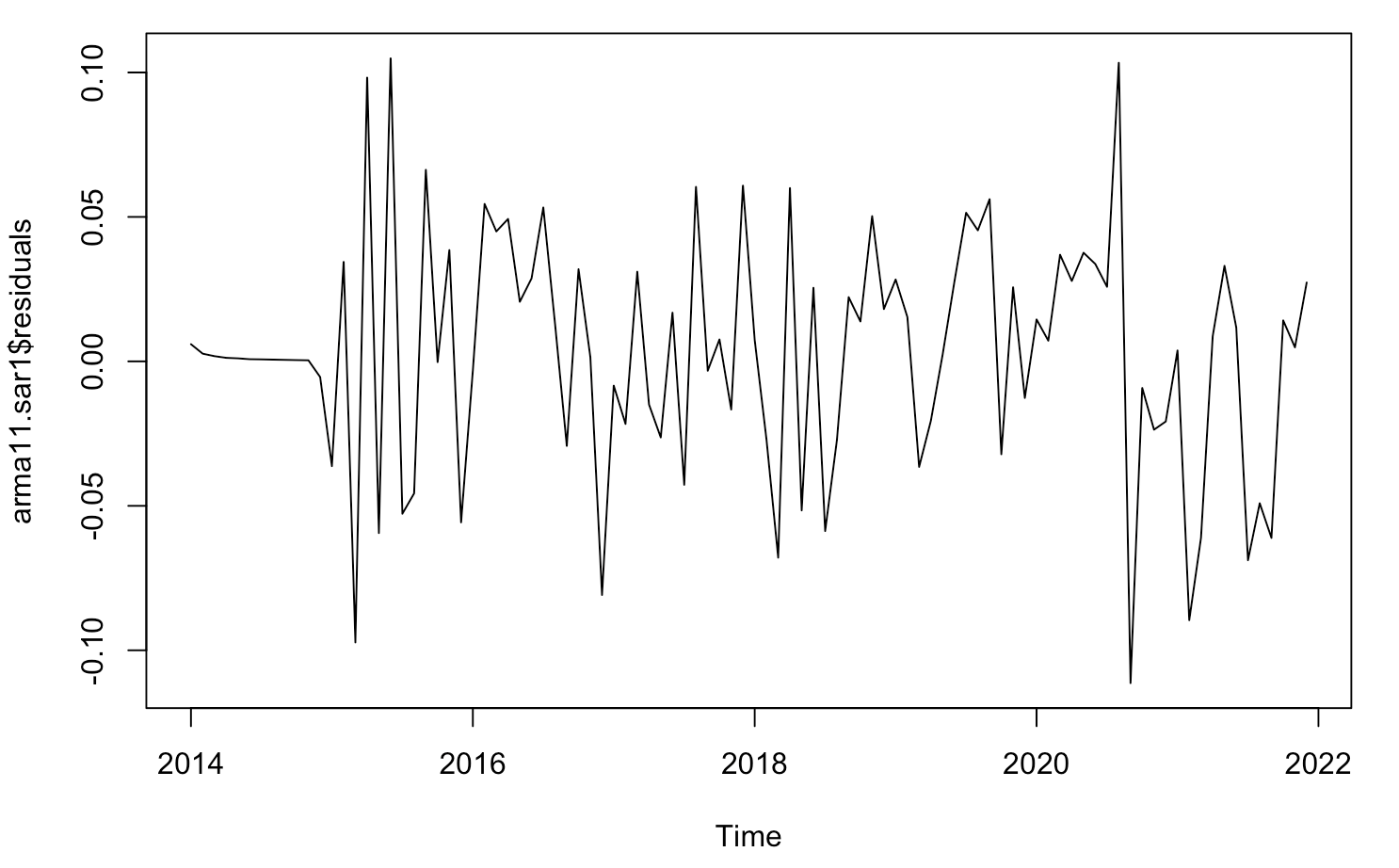


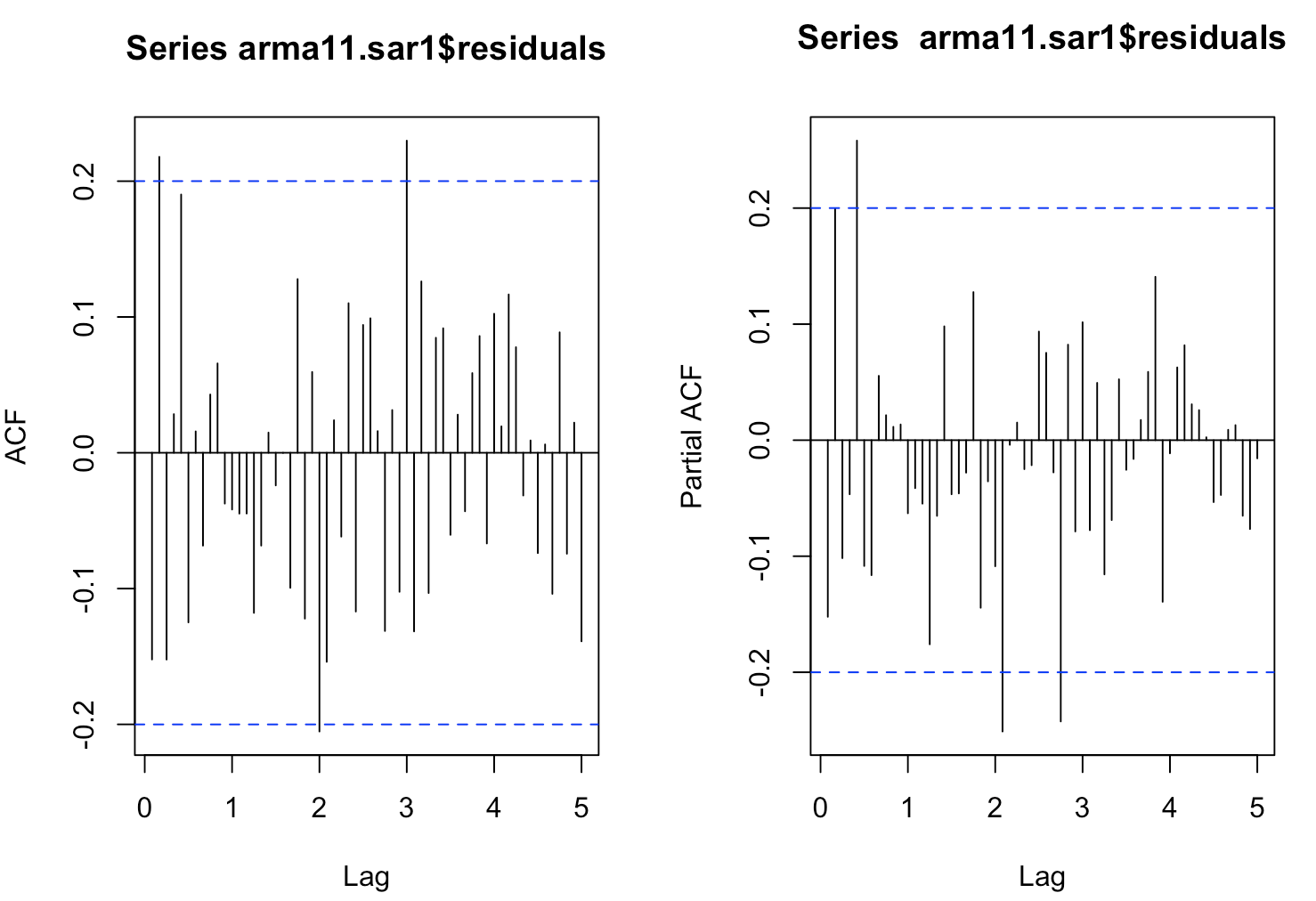


**Residual Analysis:**

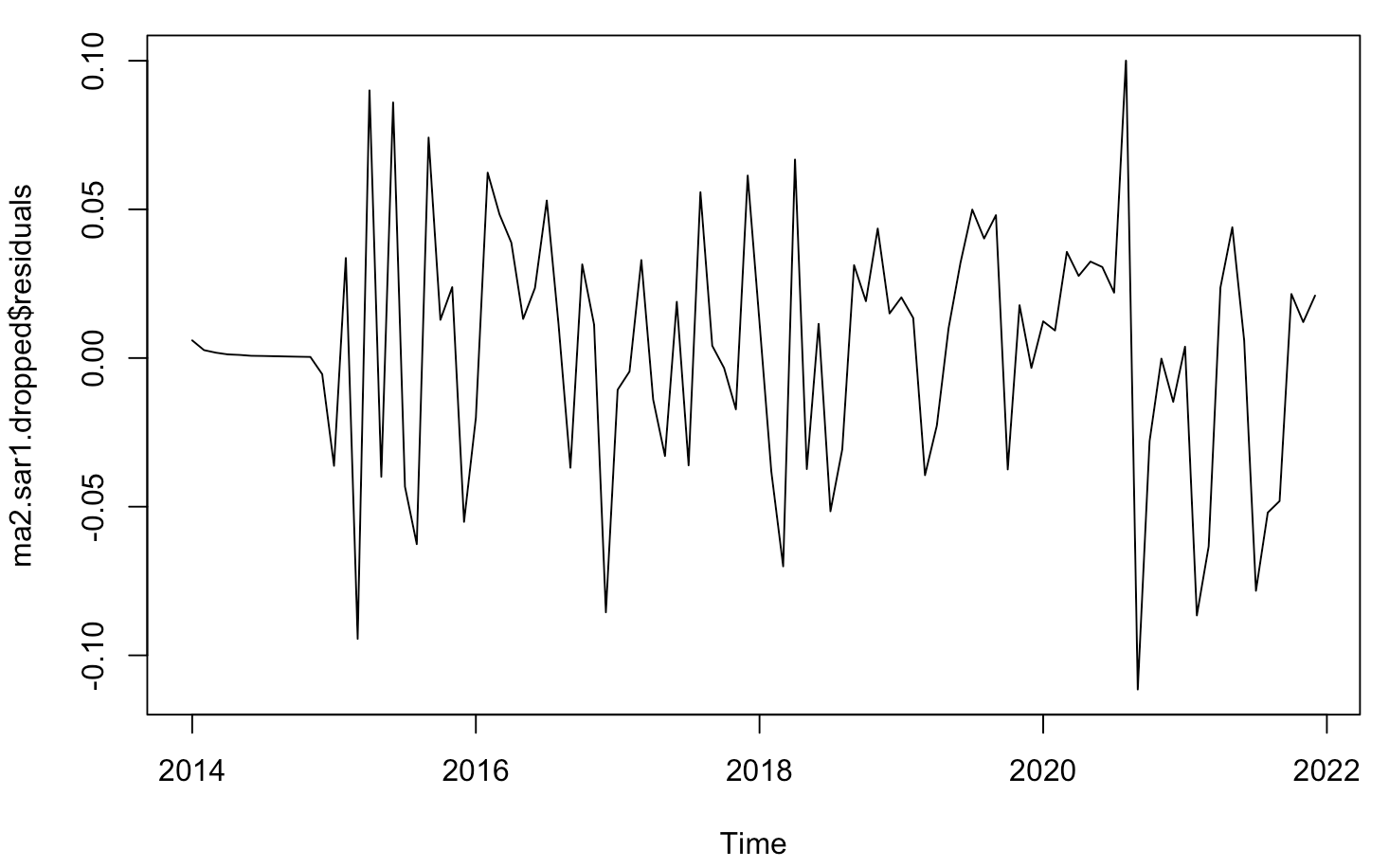
We performed residual analysis to ensure that our model is not underspecified, which means that there is no more valuable information left in the residuals. Ideally the residual series should behave like white noise, confirming that residuals have no valuable information.  
To assess the residuals, we plot the time series graph of the residual series and check the ACF & PACF plot to ensure that the plots are clean & insignificant. We also conduct Augmented Dickey Fuller test along with Box-Ljung test to verify the residuals behave like white noise.

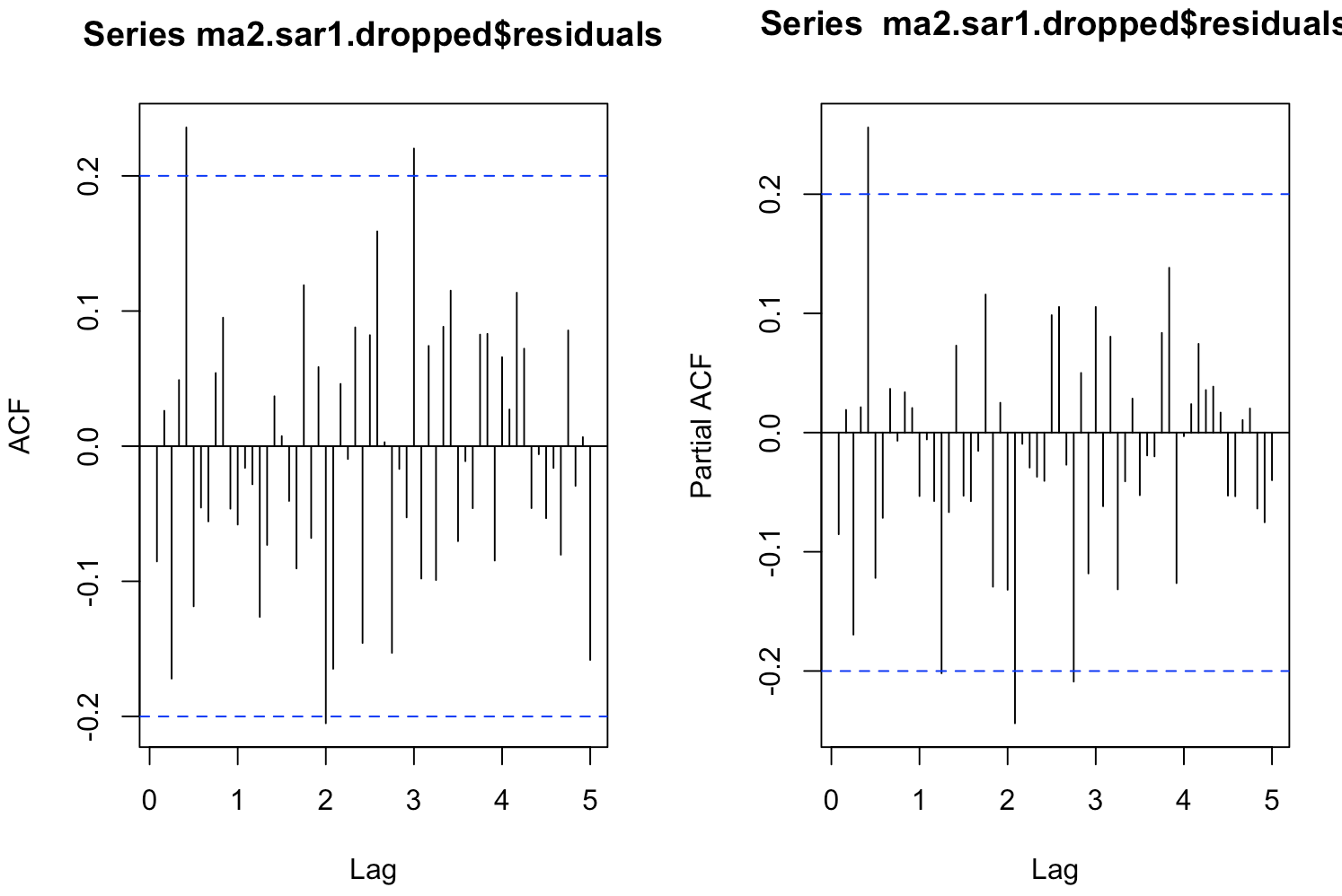
ARIMA(1,1,1) X (1,1,0)S - ACF & PACF





ARIMA(0,1,2) X (1,1,0)S - ACF & PACF





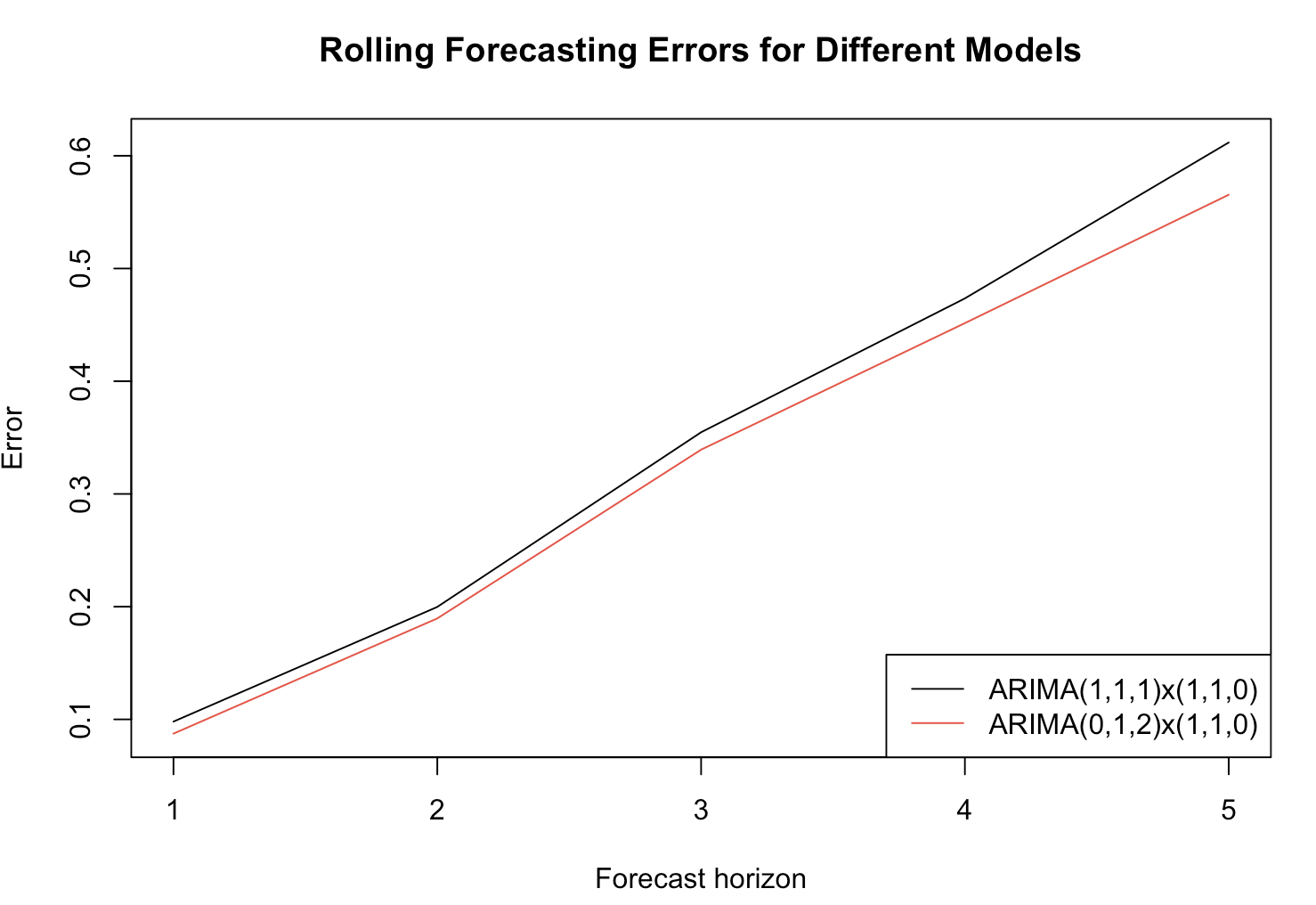
***Table 3.*** Residual plots and estimation results.

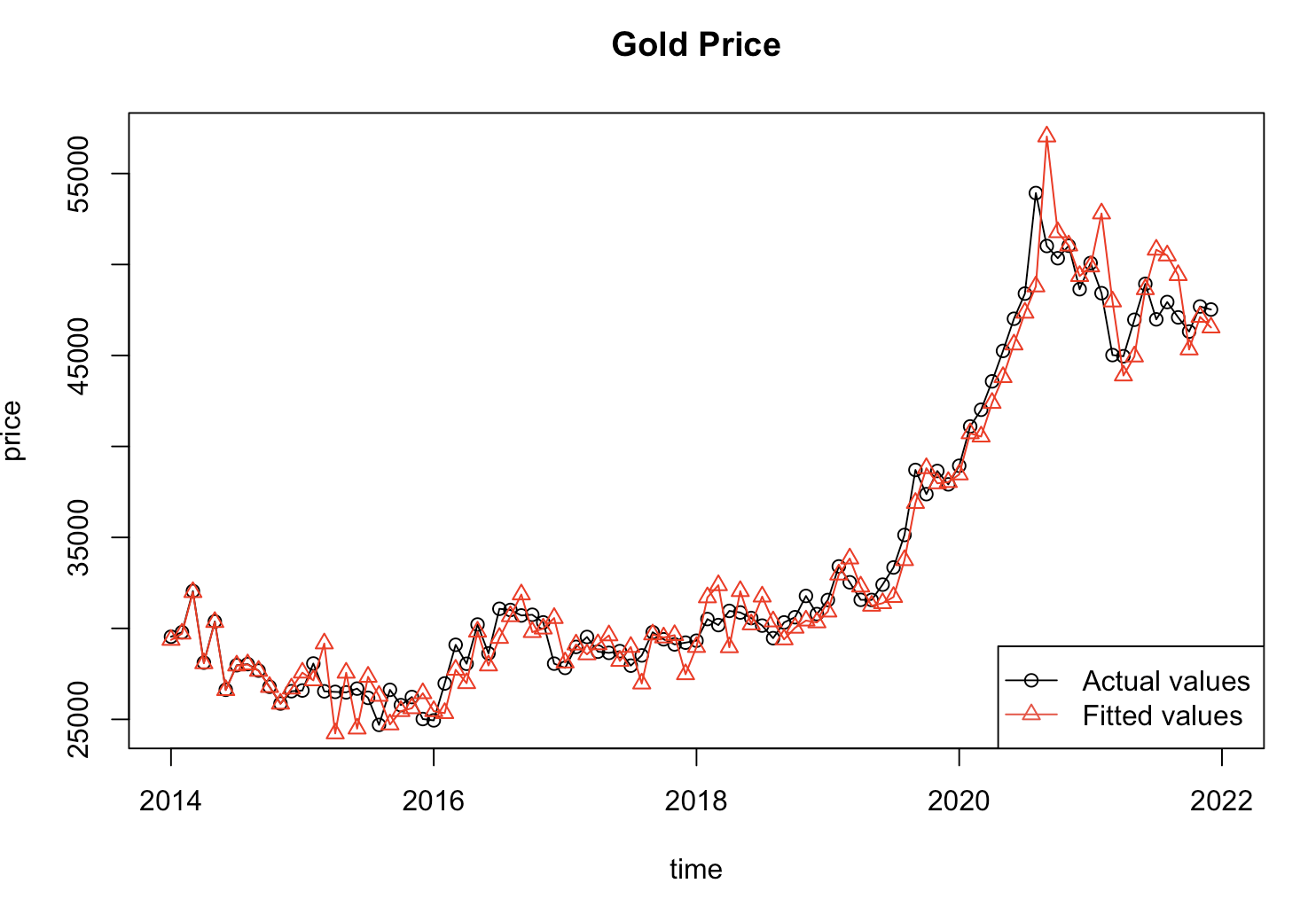
|  |  |  |
| --- | --- | --- |
|  | **ARIMA(1,1,1) X (1,1,0)S** | **ARIMA(0,1,2) X (1,1,0)S** |
| **AIC** | -265.82 | -271.62 |
| **ACF** | Clean | Clean |
| **PACF** | Clean | Clean |
| **Box-Ljung Test** | 0.1738 | 0.3193 |

SAR(1) had the least AIC value hence we considered proceeding with ARIMA(1,1,1) X (1,1,0)S and ARIMA(0,1,2) X (1,1,0)S.  We compare the related metrics in Table 3 and conclude that ARIMA(0,1,2) X (1,1,0)Shas the lowest AIC value and the ACF and PACF plots of residuals corresponding to this model are clean. The p-value in the Box-Ljung test is 0.3193, which indicates that the residual is white noise. All the terms in the model are significant. Hence we concluded that ARIMA(0,1,2) X (1,1,0)S  is the most suitable model for further analysis.

**Forecasting**

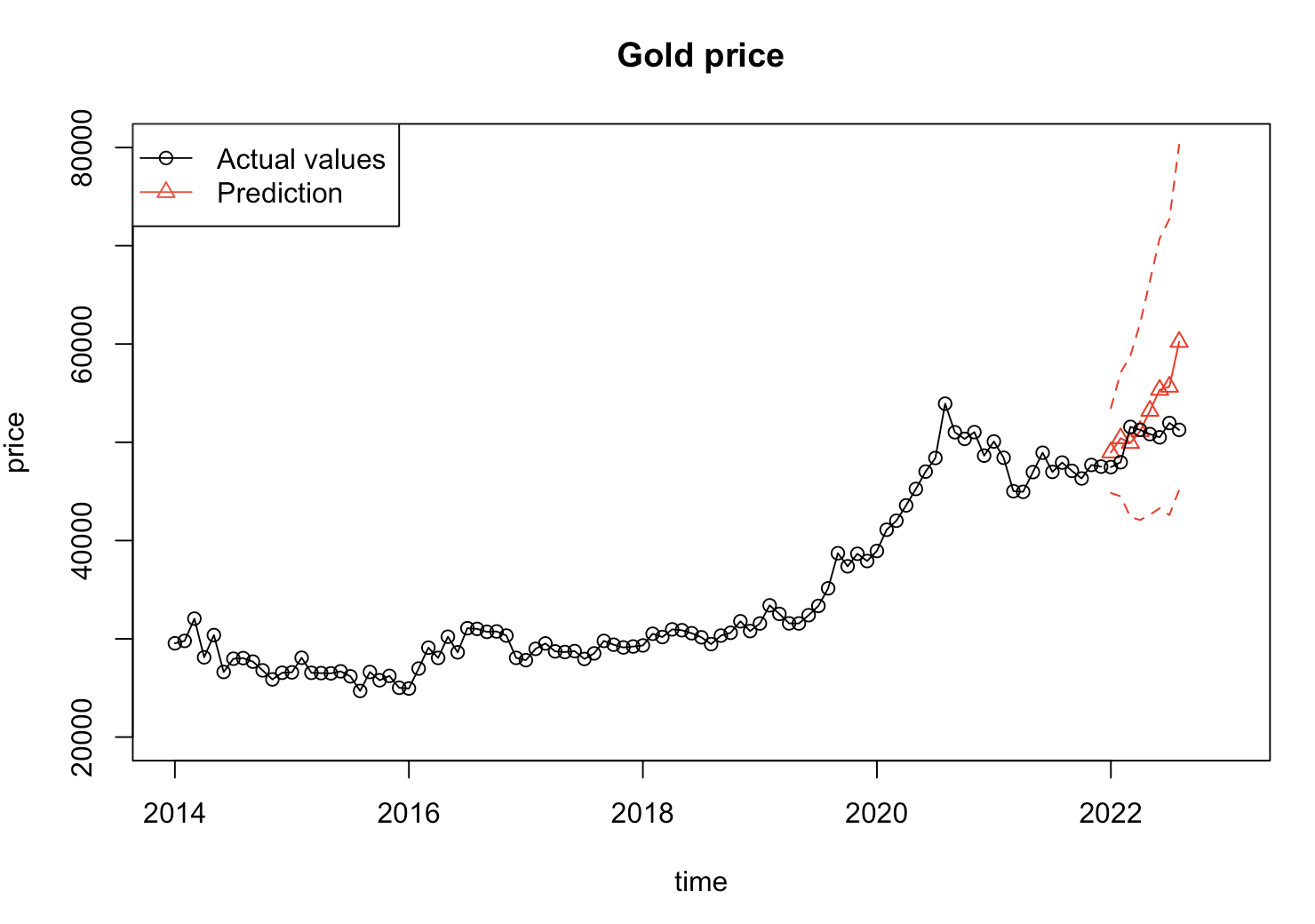
Forecasting is the process of predicting future values based on historical data. Rolling forecast errors provide insights into the accuracy and reliability of forecasting models. We performed a 5-step ahead rolling forecast for both models. The rolling forecast starts 50 observations before the end of this time series data. We identified the preferred model which has the lowest rolling forecast error for prediction and we chose the model “ARIMA(0, 1, 2) x (1, 1, 0)s”.





**Prediction**

Using the preferred model we predict the price of gold for 8 months in 2022. We observe that with ARIMA(0,1,2) X (1,1,0)Sthe predicted values are almost close to actual values and thus we can conclude that the selected model is the most suitable one.



**Conclusion**

The main aim of this project was to compare two types of time series forecasting models to understand which one would work better for future predictions. The project can be taken further by implementing deep learning to improve the accuracy of the prediction. We can also employ this model on a website which might be used to predict the future price of gold, helping people decide whether it is a good decision to invest or not.

**References**

1. Yang, X. (2019, January 1). *The prediction of gold price using Arima model*. Atlantis Press. https://www.atlantis-press.com/proceedings/ssphe-18/55911762
2. Makala1, D., & Li1, Z. (2021, February 1). *IOPscience*. Journal of Physics: Conference Series. https://iopscience.iop.org/article/10.1088/1742-6596/1767/1/012022/meta