UNIVERSITÉ DE NEUCHÂTEL Institut de statistique

Time Series Analysis Project

Master of Science in Statistics

Faculty of Science

Institute of Statistics

University of Neuchâtel

Forecasting Silver Pricing Data

By : **Kevin Nyamai**

May 2023

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ADF	Augmented Dickey Fuller
ARIMA	Autoregressive Integrated Moving Average
GBP	Great Britain Pound
GDP	Gross Domestic Product
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
MAPE	Mean Absolute Percentage
PACF	Partial Autocorrelation Function
RevIN	Reversible Instance Normalization
UK	United Kingdom
USD	United States Dollar

Introduction

Silver, as among the most expensive precious metals in the world, has historically been used a medium of exchange but currently mostly used as a portfolio diversifier and hedge against inflation (Fernando, 2022). The project is aimed at modelling data to be able to forecast silver prices, 6 months ahead, which in turn would aid in portfolio diversification. Data analysis and modelling will be done using R version 4.3.3 software

1 Data Description

1.1 Data Overview

Data is collected from London Buillon Market Association (LBMA) from 2nd January 1968 to 21st April 2022, from the London Metal Stock Exchange (Baskar, 2022).The data consists of daily prices of silver in Great British Pound (GBP), United States Dollar (USD), and EURO currencies.

On inspection of the data, all the silver prices apart from the British pound, have missing values as shown in the Table 1. To avoid obvious expectation of foreign exchange loss leading to inaccurate results and modelling (CFI, 2023), analysis was done in GBP, instead of initially proposed USD base owing to its status a world's reserve currency (Brusuelas, 2023).

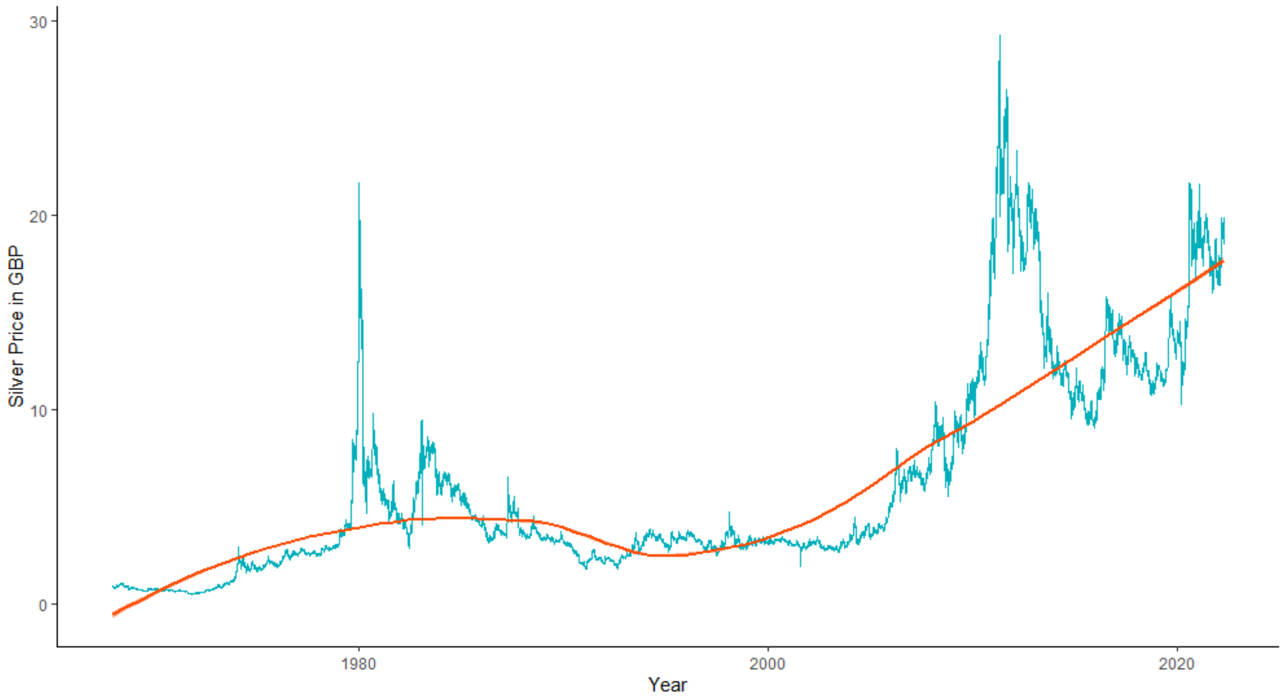
Table 1.1. Missing Data by Currency

Currency	**GBP**	**USD**	**EURO**
$Number of Missing Values$	None	19	7847

1.2 Data Exploration

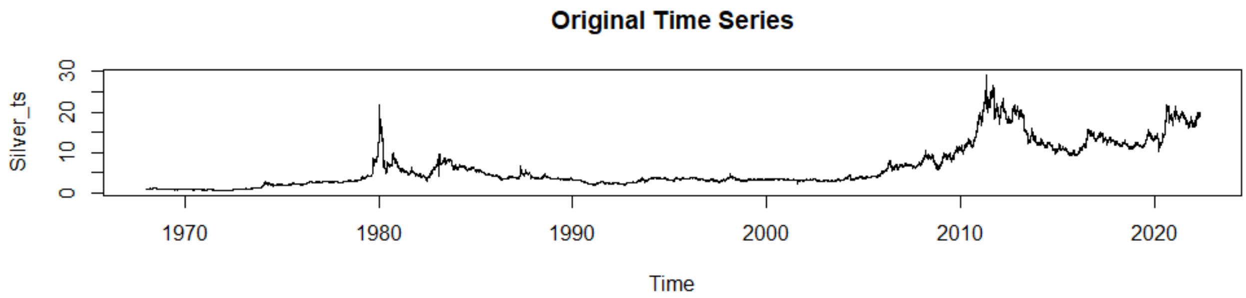
The data had 13,736 data points with missing dates, hence padding was done with imputation thereafter to fill in missing values for the newly added dates bring the data points to 19,834. A plot of the time series object was then created to observe the data into a year with the frequency 7*4*12 days per year as shown in Figure 1.1

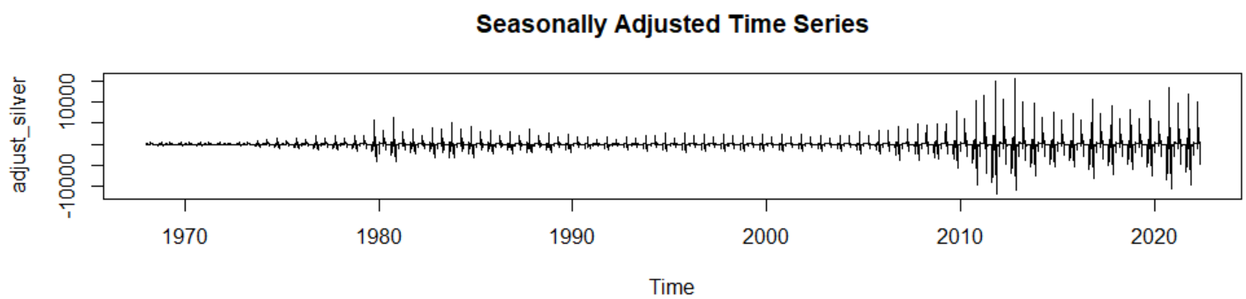
Figure 1.1. Time Series Plot



The plot above shows a non-linear increasing trend with distinct shifts in the 80's and 2010 that may not necessarily imply seasonality. To confirm this assertion, the time series was adjusted for seasonality and comparisons made. This involved decomposing the series that suggested an additive model and later modified to multiplicative as seen below in Figure 1.2.

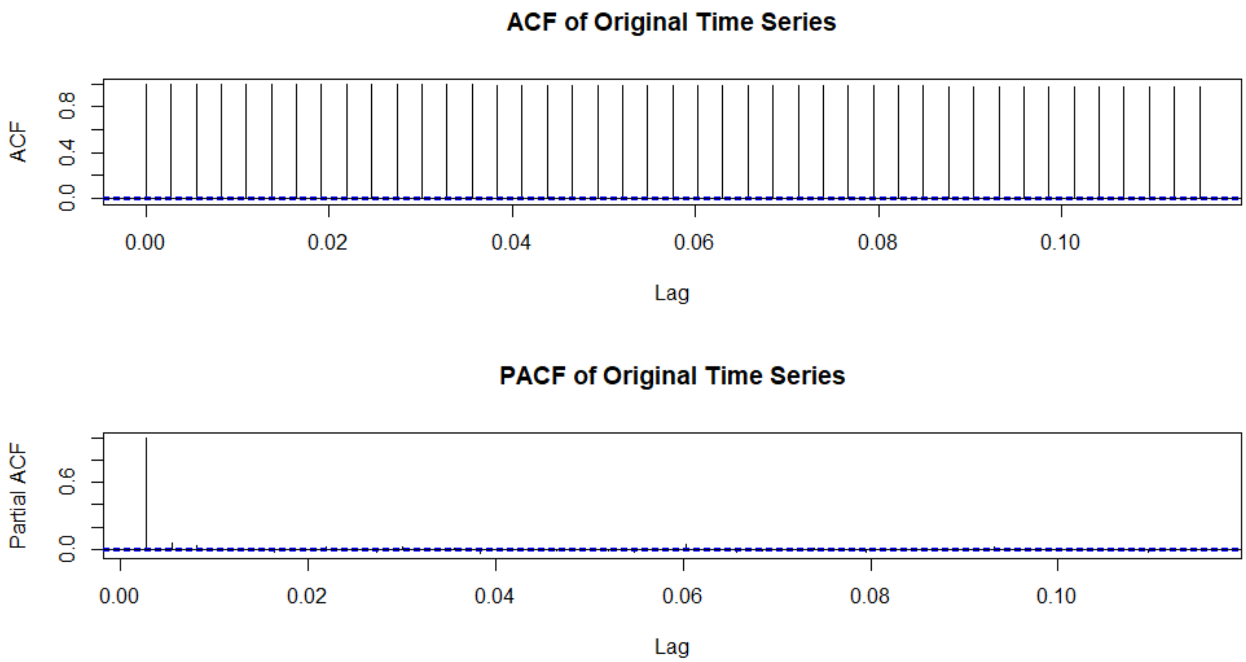
Figure 1.2. Plot Original Time Series vs Seasonally Adjusted Series





On plotting the additive seasonally adjusted series side by side with the original series from Figure 1.2 above,an increasing amplitude of the seasonal effects with the trend level in the original time series is evident from around 2010 going upwards, indicating multiplicative seasonality is likely more appropriate. This suggests that seasonal fluctuations change proportionally with the level of the series .Moreover, the autocorrelation function (ACF) plot, taking a lot of time to die off and partial autocorrelation function (PACF) plot, having only one significant lag as shown below, Figure 1.3, suggests a seasonal autoregressive model.

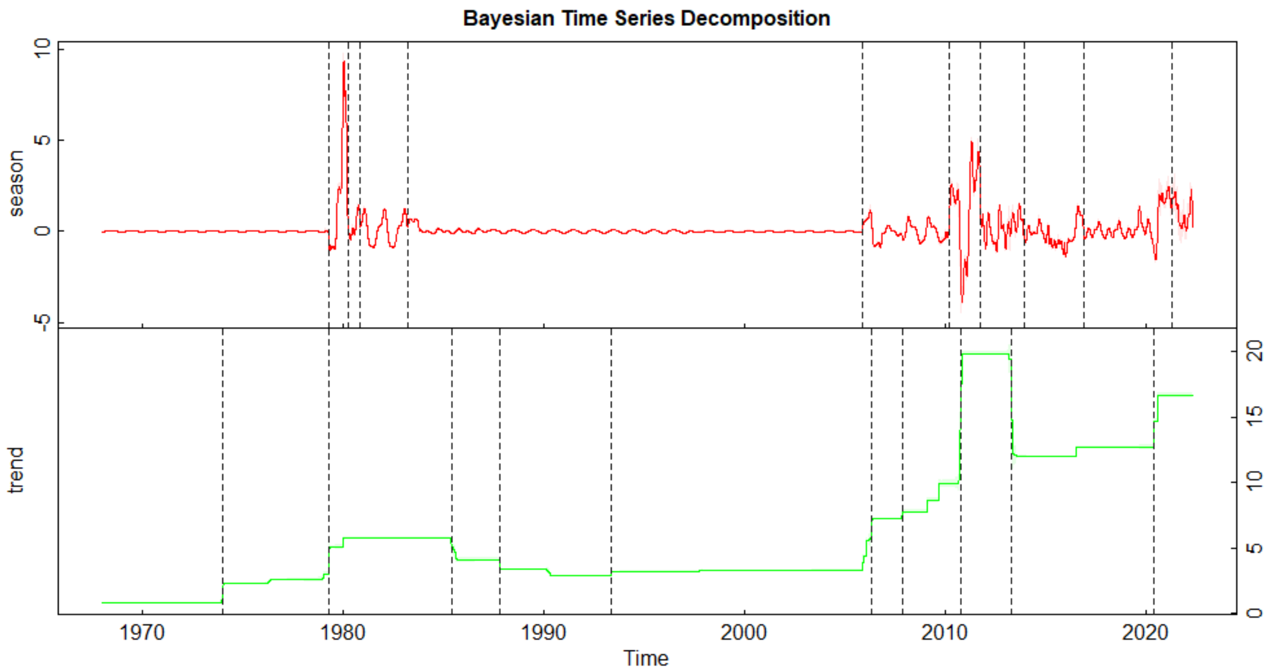
Figure 1.3. ACF and PACF Plots of Original Time Series



There are sharp shifts in the variance and/or mean at two different time points. This is inspected by a Bayesian ensemble algorithm for time series decomposition for changepoint (Kaiguang, 2022) as shown in Figure 1.4 below. The plot shows significant abrupt disruptions in the late 1970's to the mid 1980's as well as 2010 to 2011 relating to sharp silver prices around period of recessions(BullionByPost, 2022) i.e. :

- **Mid-1970s recession**: In late 1973 and early 1974, stagflation drove the United Kingdom's (UK) Gross Domestic Product (GDP) negative. This coincided with the silver price rising
- **Early 1980s recession**: The UK was in a recession for five quarters. The larger spike seen on the chart related to Silver Thursday but still experienced a price climb the year after.
- **The 'Great Recession'**: The financial crisis in the late-2000s pushed gold to a record price in Dollars. Silver also climbed to its peak Sterling price of £29.26 per ounce.

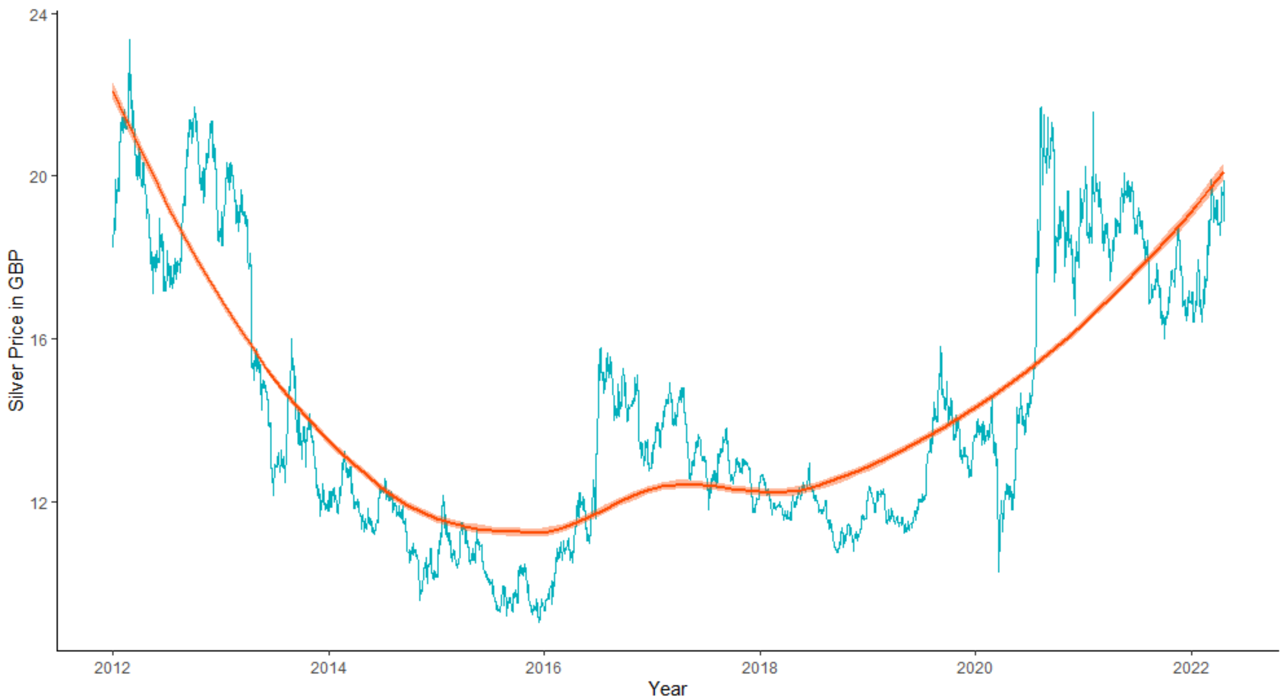
Figure 1.4. Bayesian Time Series Decomposition for Changepoint



Variance often changes over time in time series, i.e., time-series data suffer from a distribution shift problem. This change in temporal distribution is one of the main challenges that pose a risk to accurate time-series forecasting. To address this issue, (Kim et al., 2022), proposes a simple yet effective normalization method called reversible instance normalization (RevIN), a generally applicable normalization-and-denormalization method with learnable affine transformation. This method is symmetrically structured to remove and restore the statistical information of a time-series instance, leading to significant performance improvements in time-series forecasting.

However, naively discarding time points where data had disruptions by only considering the silver prices from 2012 onwards, because recession can in no way be considered a sure indicator of an impending price of for silver(BullionByPost, 2022).This at least provides assurance that of no distortion of our modelling owing to outliers. Subsetting the data from 1st January 2012, we get 3764 data points as shown by Figure 1.5 below.

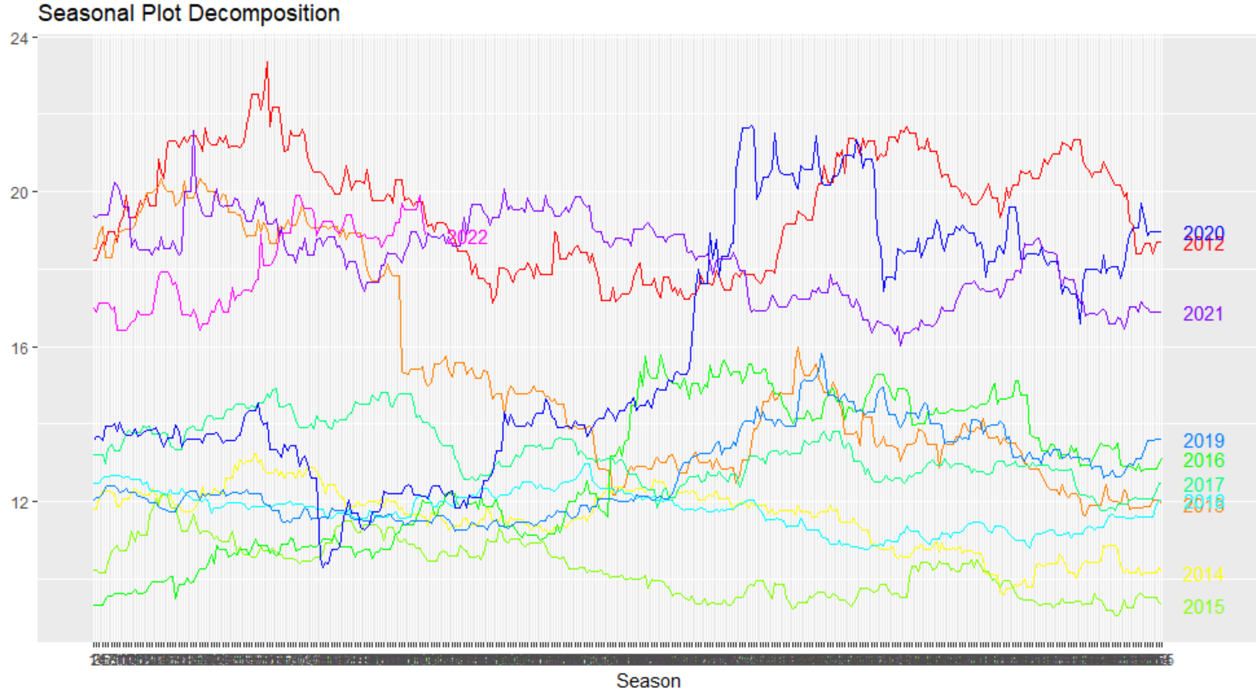
Figure 1.5. New Time Series Plot (Data Points from 1st Jan 2012- 21st April 2022)



The time series, has a non-linear trend, first decreasing till 2016 and increasing thereafter.

Checking for the seasonal component from Figure 1.6, there seems to be no apparent seasonal component across board, with similarities for groups i.e. group 1 (2012, 2020 and 2021), group 2 (2013,2016,2017,208,2018) and group 3 (2014 and 2015). The presence of there being a seasonal component is hence ruled out for modelling purposes.

Figure 1.6. Seasonal Decomposition Plot



2 Modelling/Analysis

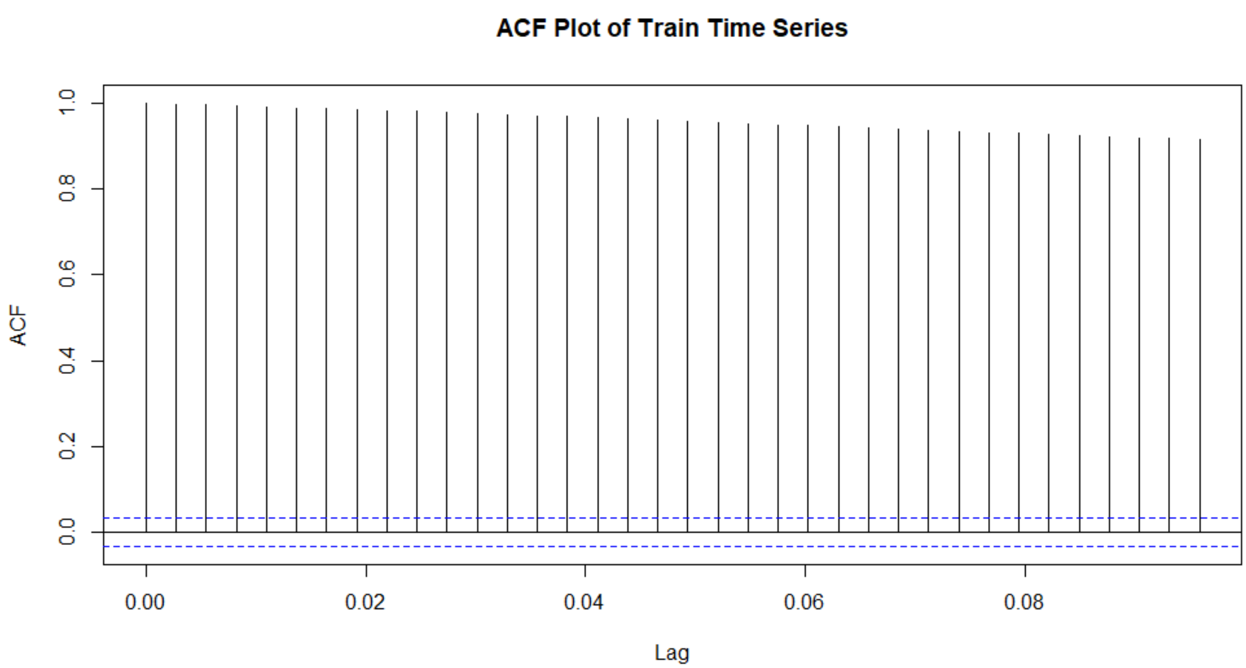
2.1 Data Training

The new data is split into train and test data sets, whereby the test will constitute of the last 6 months of our data (7*4*8=168 days), we will try to forecast authenticating our model based on prediction accuracy

2.2 Stationarity

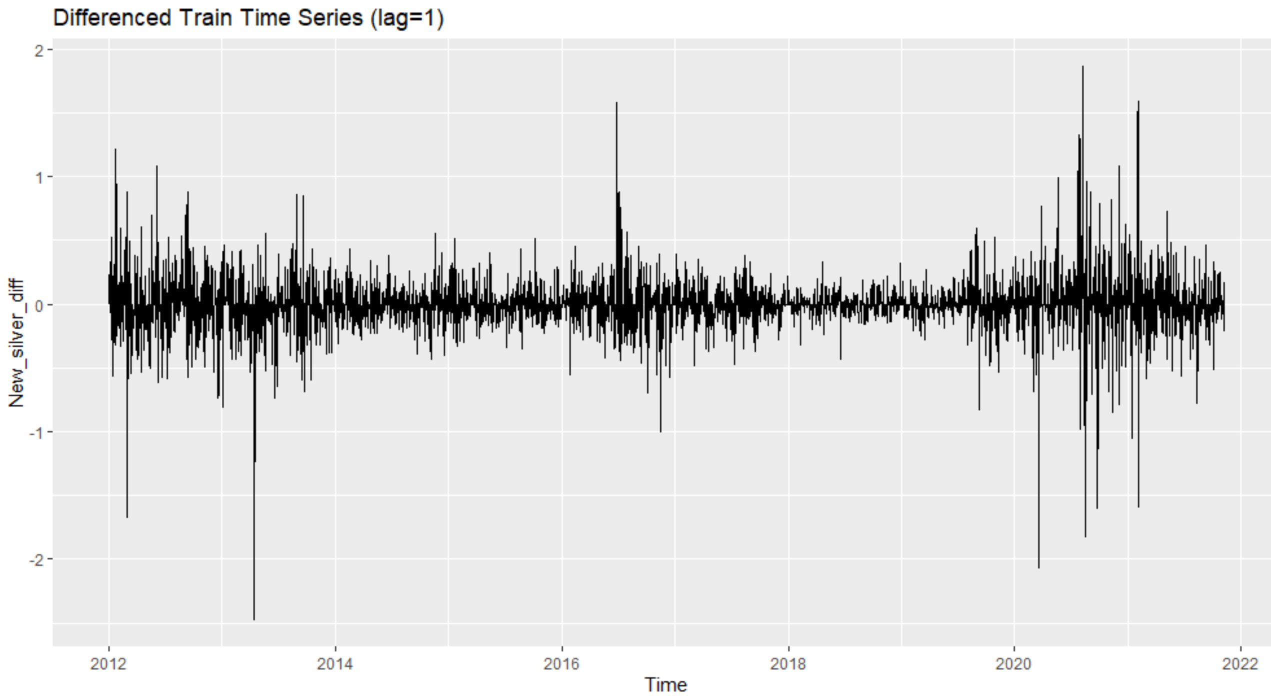
As seen from the ACF plot in Figure 2.1 the time series based on train data set takes a long time to die off and in hindsight of a non-linear trend, the series is differenced at lag 1, as seen in Figure 2.2.

Figure 2.1. ACF Plot of Train Time Series



The time series is centred with $mean = -0.0002223922$ (closer to zero) and appears stationary. To confirm this, a Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for trend stationarity is performed followed by Augmented Dickey–Fuller (ADF) test for stationarity.

Figure 2.2. Plot of Differenced Train Time Series at lag=1



For the KPSS test, the p-value, 0.1> =0.05, hence we fail to reject the null hypothesis that the time series trend is stationary.

```
KPSS Test for Trend Stationarity
data:  New_silver_diff
KPSS Trend = 0.033887, Truncation lag parameter = 9, p-value = 0.1
```

For the ADF test, the p-value, 0.01 is less than =0.05 level of significance, hence we reject the null hypothesis of non-stationarity and accept the alternative of stationarity of the series.

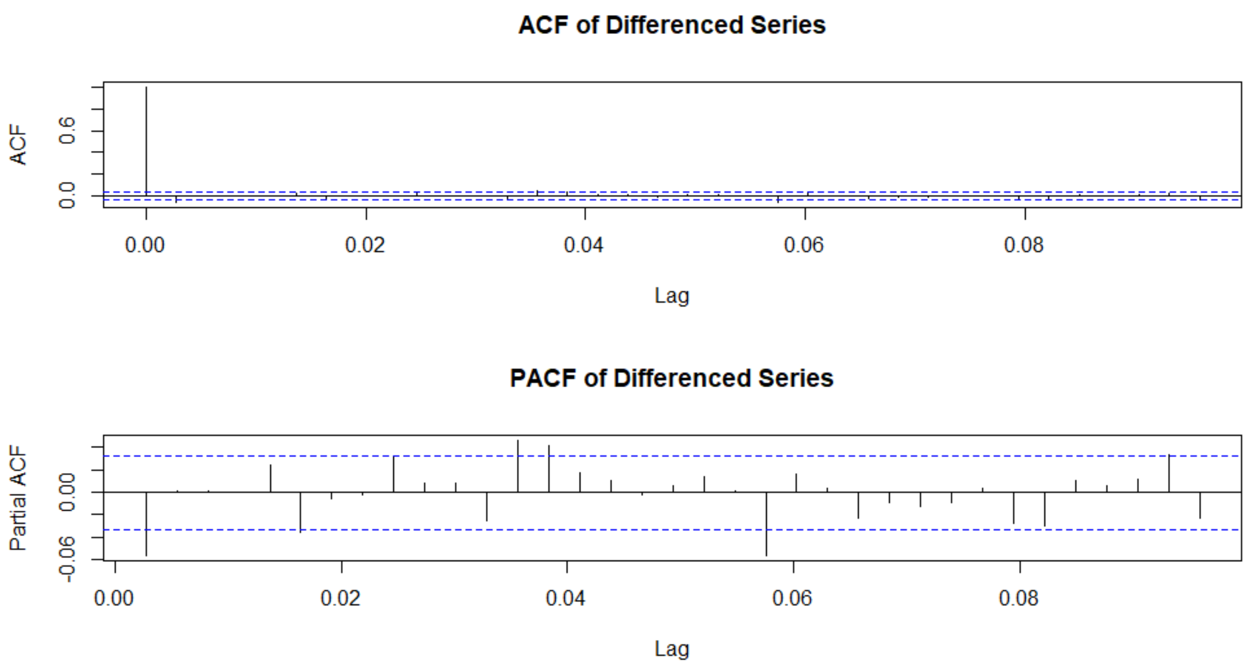
```
Augmented Dickey-Fuller Test

data:  New_silver_diff
 Dickey-Fuller = -13.674, Lag order = 15, p-value = 0.01
 alternative hypothesis: stationary
```

2.3 Modelling

After achieving stationarity, based on inspection of the ACF and PACF plots, in predicting best possible models. The ACF plot shows, one significant spike implying no moving average process, while PACF shows 4 clearly distinct spikes in an oscillating pattern as seen in Figure 2.3. With this hindsight, we predict best model based on the Akaike Information Criterion (AIC) value using *auto.arima* function in *forecast* package.

Figure 2.3. ACF and PACF Plots of the Differenced Series



An ARIMA (1,1,0) model was the best model as below:

$$X_t = -0.05619002X_{t-1} + Z_t,$$

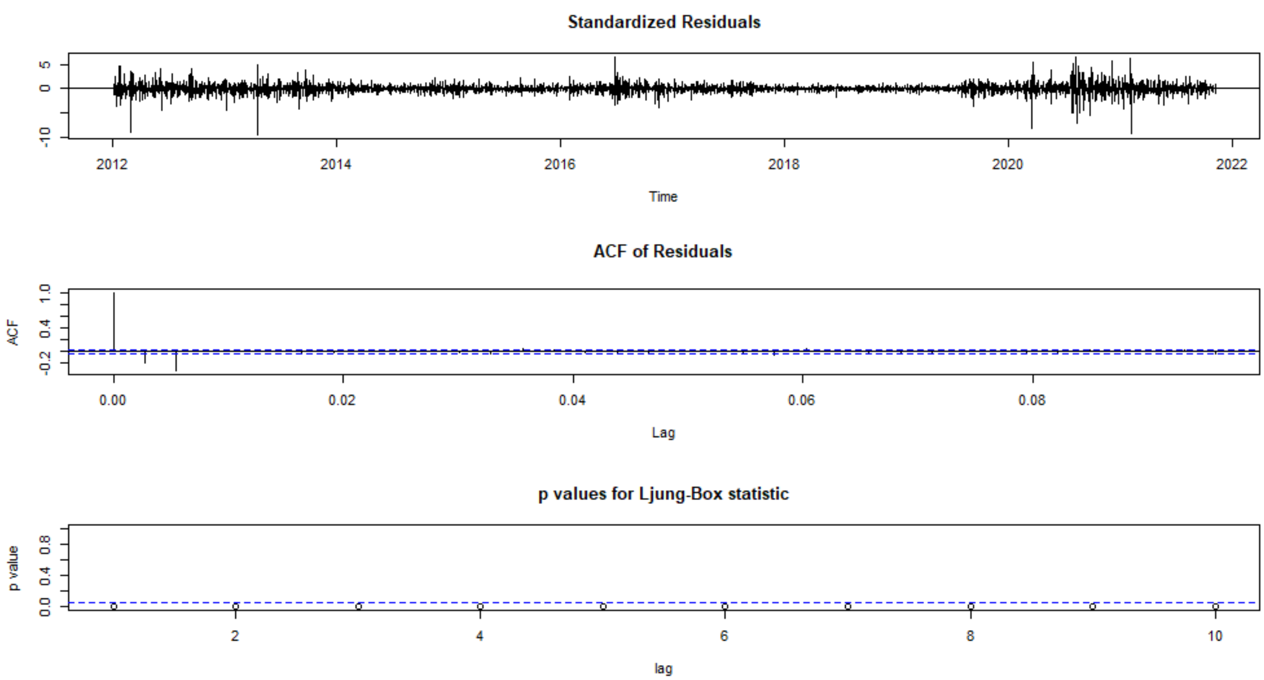
where $Z_t \sim \mathrm{WN}(0, \sigma^2)$.

Table 2.1. Best Model Metrics

Best Model	**AIC Value**	**Variance A (ar1-ar1)**
ARIMA(1,1,0)	449.6303	0.0002772914

2.4 Model Diagnostics

Figure 2.4. Model Diagnostics Plots



Based on the diagnostic plots above:

- The standardized residuals appear stationary.
- The ACF plot has one significant spike confirming our assertion in building the model.
- The p-values of the Ljung-Box statistic are above level significance implying zero or no autocorrelation.

Using the Box-Pierce, the residuals' stationarity can also be confirmed.The p-value of 0.9991 is greater than level of significance=0.05, hence we fail to reject the null hypothesis for the residuals having no autocorrelation.

```
Box-Ljung test

data:  obj1$residuals
X-squared = 1.0623e-05, df = 1, p-value = 0.9974
```

The Shapiro-wilk test confirms the normality of the residuals. The p-value is less than less than level of significance,0.05, hence we reject the null hypothesis of non-normality of the residuals.

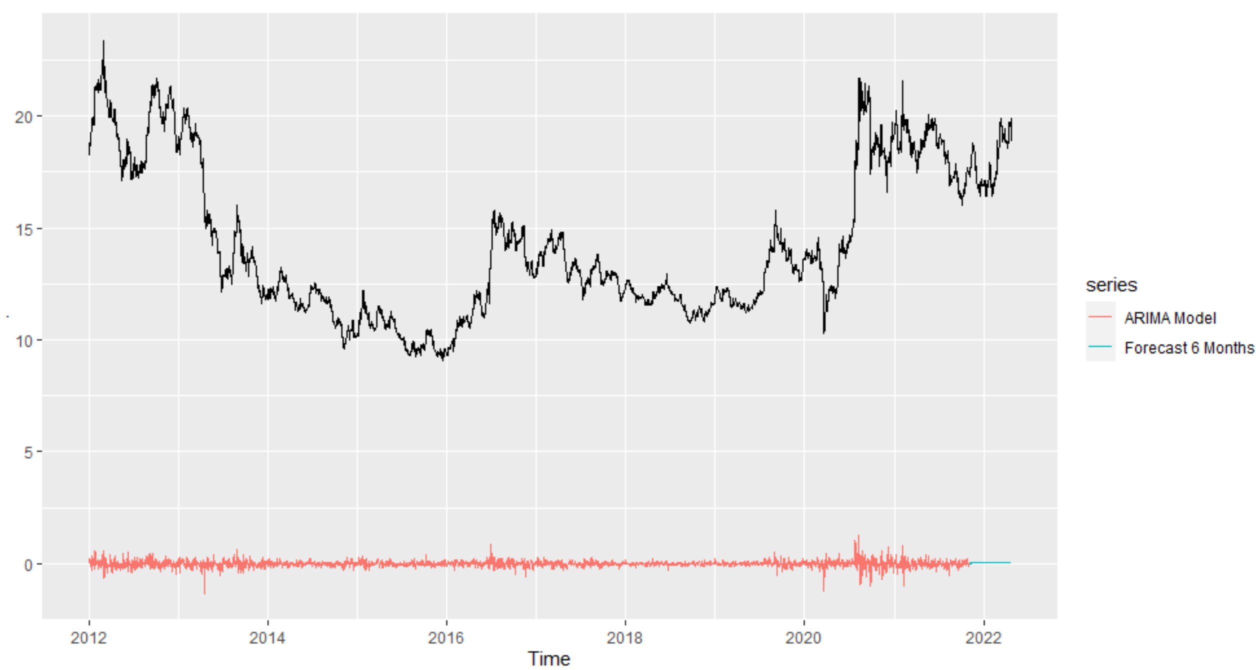
```
Shapiro-Wilk normality test

data:  obj1$residuals
W = 0.79984, p-value < 2.2e-16
```

3 Forecasting

Assured of the model, it is used to forecast for next 168 days (6 months) and make a comparison with the actual test data as seen in Figure 3.1.

Figure 3.1. Comparison of Forecasted and Actual Data



From Figure 3.1, the red line our ARIMA (1,1,0) model with train data is quite similar with our original data despite having a deviation denoted by the blue line. Consequently, there is need to check for the mean absolute percentage error regression loss (MAPE). The prediction accuracy of the model is 0.79% with a forecasting error of 5.05% which is relatively a good result owing to the data modifications done.

4 Conclusion

The final model appears to feasible despite the numerous modifications of the initial dataset. To improve on the model, there is need to implement reversible instance normalization (RevIN), to address the huge shifts at the two points in time i.e. mid 80s and 2008-2010 and perhaps, would improve the performance of the model in forecasting.

References

B. Baskar. Silver prices 1968-2021. https://www.kaggle.com/datasets/balabaskar/silver-prices-1968-2021, 2022. Accessed: 2023-04-24.

Joseph Brusuelas. Why the dollar remains the world's reserve currency, and will stay that way. https://realeconomy.rsmus.com/why-the-dollar-remains-the-worlds-reserve-currency/#:~:text=Other%20IMF%20analysis%20notes%20that,the%20global%20economy%20is%20secure, April 17 2023. Accessed: 2023-04-24.

BullionByPost. Silver price during recession. https://www.bullionbypost.co.uk/index/silver/silver-price-during-recession/, May 2022. Accessed: 2023-04-24.

CFI. Foreign exchange gain/loss. https://corporatefinanceinstitute.com/resources/accounting/foreign-exchange-gain-loss/, March 13 2023. Accessed: 2023-04-24.

Jason Fernando. Precious metals: Trade. https://www.investopedia.com/terms/p/preciousmetal.asp, June 12 2022. Accessed: 2023-04-24.

Kaiguang. *Bayesian Changepoint Detection Time-Series Decomposition*, July 5 2022. Accessed: 2023-04-24.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022.

LBMA. Lbma. https://www.lbma.org.uk/. Accessed: 2023-04-24.