Statistical Modeling Assignment Report

MS1415 Robusta metoder

19 May 2025

Abdalrahman Mohammed, abmm22@student.bth.se

Simon Lindqvist, siln22@student.bth.se

Contents

Introduction

This assignment focuses on modelling and forecasting the monthly useful water volume of the Caconde Reservoir in São Paulo, Brazil, using time series analysis. The dataset, spanning January 2015 to February 2024, records the proportion of the reservoir's operational capacity, scaled between 0 and 1.

The goal is to explore the data's temporal structure identifying trends, seasonality, and autocorrelation and to build and compare AR, ARMA, SARMA and SARIMA models. Model performance will be evaluated using both in-sample and out-of-sample metrics to determine the most suitable approach for accurate forecasting.

Used R libraries

The following r libraries were used:

- tidyverse: Used for plotting.
- forecast: Used for box-cox transformation, plotting, and the Arima function.
- tseries: Used for the augmented dickey fuller test.
- lmtest: Used for testing significance of coefficients.
- nortest: Used for the Anderson-Darling test.
- car: Used for QQ plots.
- FinTS: Used for the ARCH test

Data exploration and Data preparation

Initial data exploration

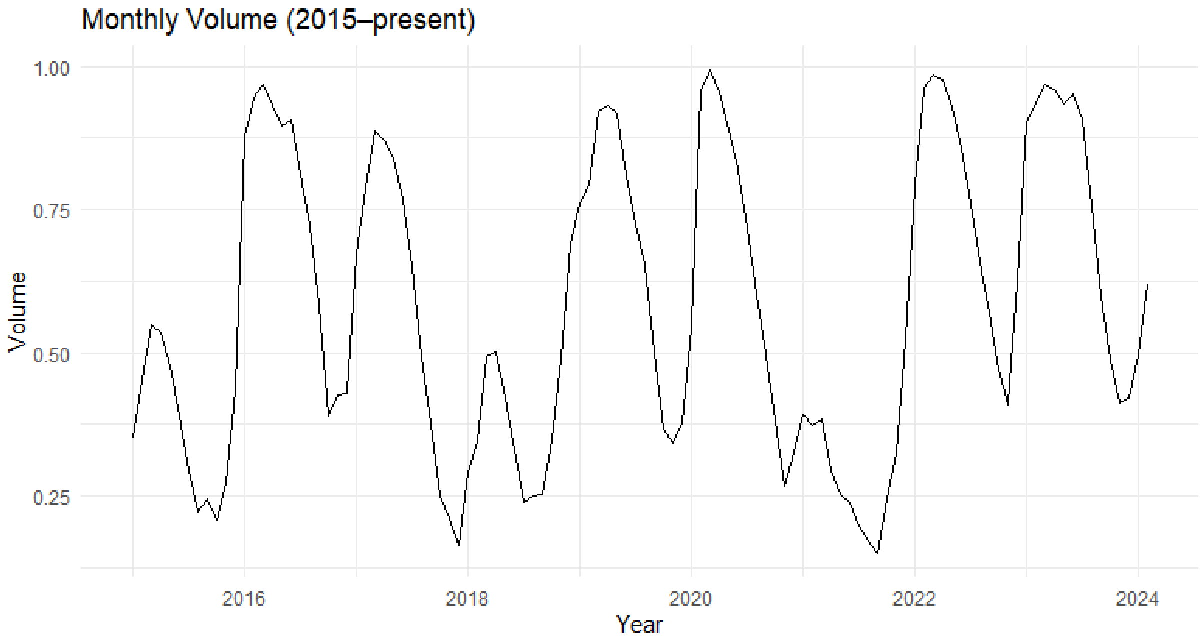

Figure 1 Plot of the CANCONDE data as a time series

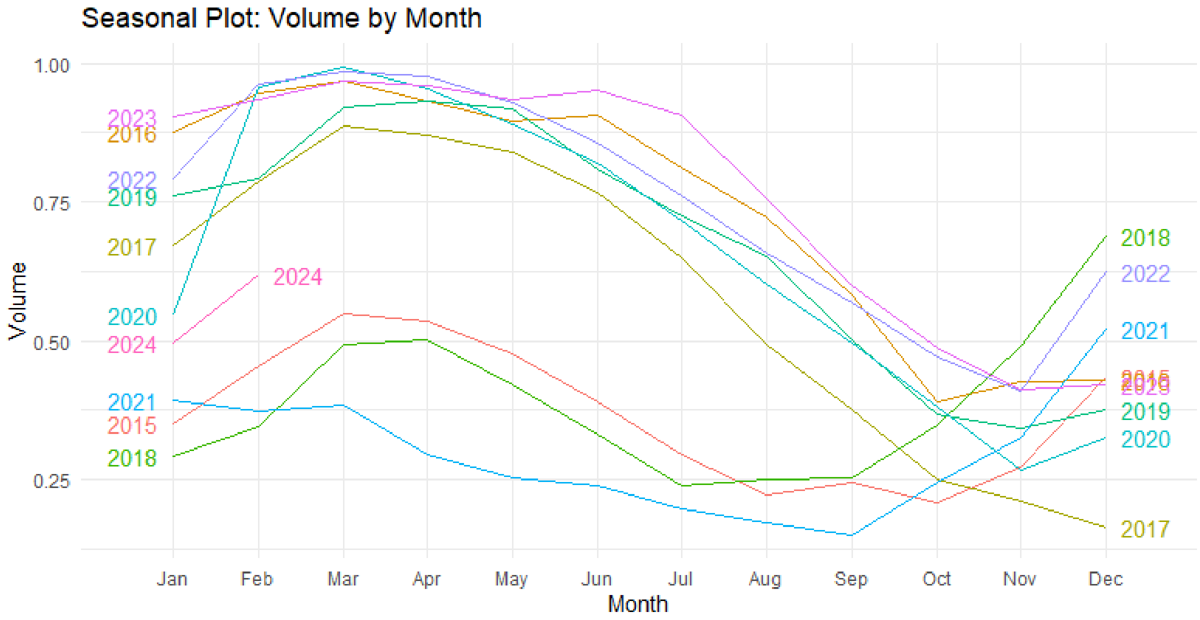

Figure 2 Seasonal plot of the data comparing years

Firstly, the time series was plotted as-is by loading the data as a time series with a frequency of 12 and January 2015 as the start data and plotting it. It appears that the mean averages out to around the same over time and the variance seems relatively stable, hinting at the possibility of the data being stationary. In the plot a pattern where some countries seemed to have vastly lower water levels on a semi-regular basis so an additional seasonal plot where each year is plotted monthly separately was made.

In the year-seasonal plot with volume per month it was observed that 2021, 2018, and 2015 had vastly lower peak water levels compared to other years. It was also observed that the current year, 2024, seemed to be in the middle of the normal years and the dry years, indicating that this year might be a somewhat dryer year.

Decomposition plot, ACF, and PACF

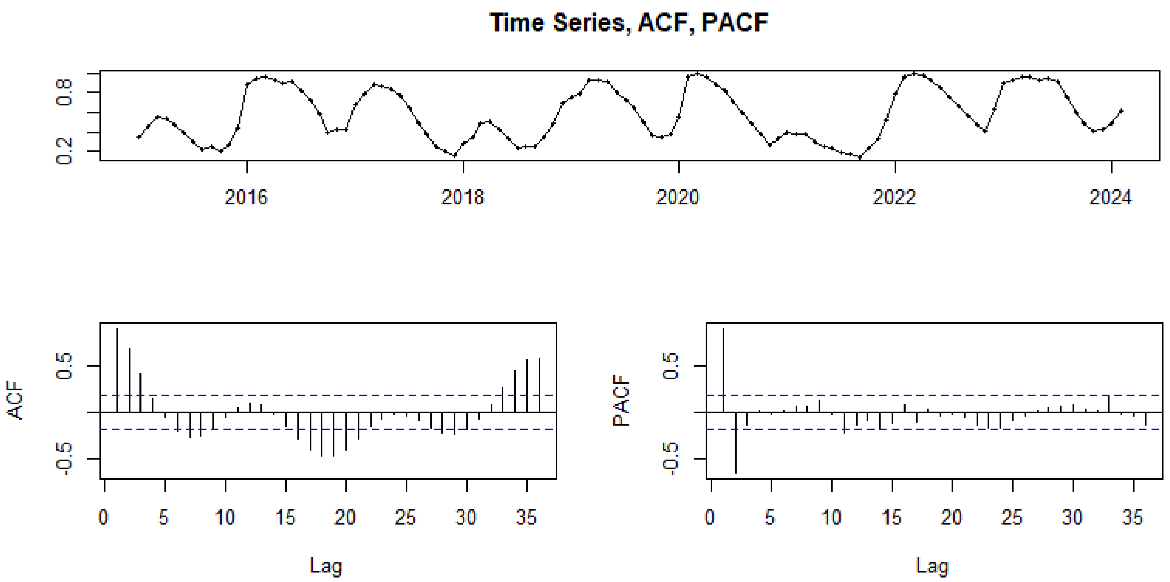

Figure 3 ACF and PACF plot of the data

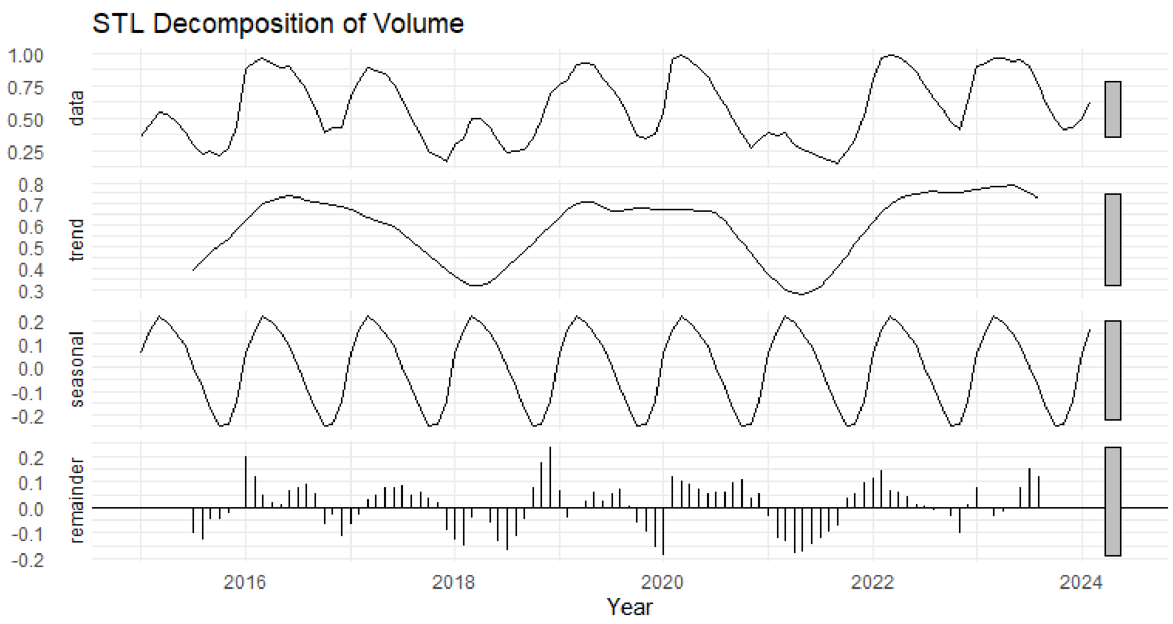

Figure 4 Decomposition plot of the data

An autocorrelation function plot and a partial autocorrelation function plot was made followed by a decomposition plot. The autocorrelation function showed a decent to zero in 4 lags of time, which isn't sublime, but acceptable for the sake of verification of stationarity. Later lags show at some kind of seasonality pattern. The partial autocorrelation shows no significant spikes above the confidence interval except for the first lag. This means that there is a strong autocorrelation between a value in the time series and the previous value, potentially hinting at an AR(1) component. The decomposition plot shows that the trend exhibits a stable mean over time due to fluctuations not permanently rising, which rather points towards some form of seasonality. The seasonality component shows a clear seasonal pattern with approximately one year between two high points. The remainder component mayhap shows remnants of some kind of pattern, but it looks more random than not. An augmented Dickey-Fuller test was performed on the time series, confirming that the data indeed is stationary, meaning no differentiation was performed.

Train and validation split

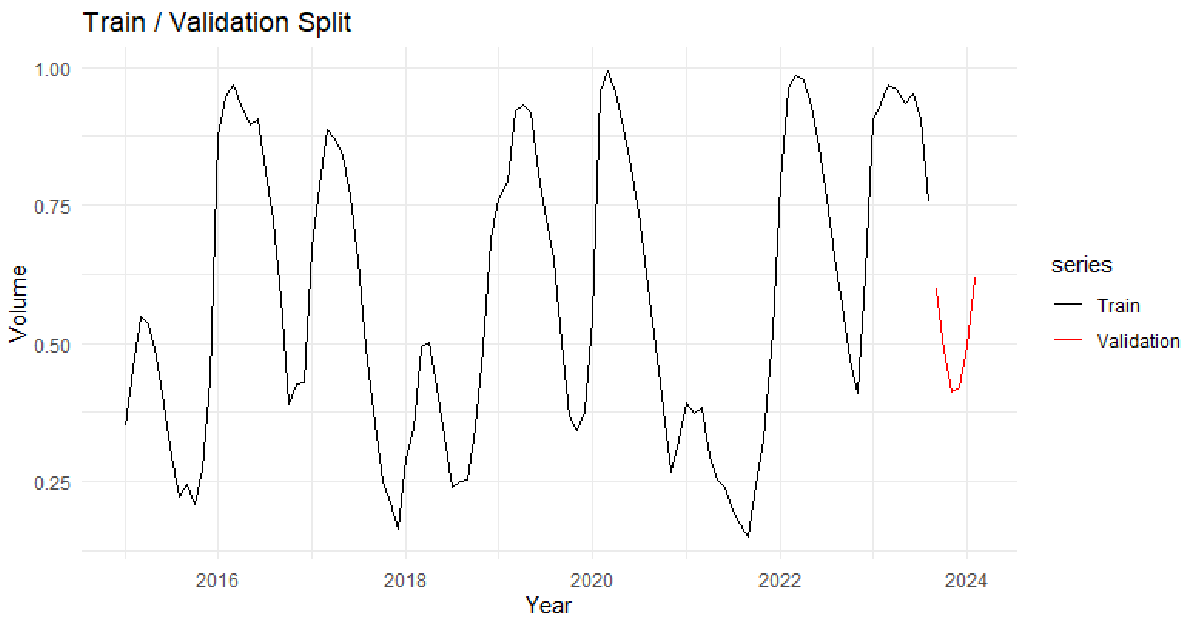

Figure 5 Plot showing the training and validation data split

Before proceeding to model selection and training the time series was divided up into a training and a validation dataset. The cutoff was set to:

$$cutoff = n - 6$$

As per assignment instructions. A plot for showcasing purposes was made to showcase the training and validation data in black and red respectively. This allowed for easier visual comparisons with the predicted values when selecting models.

Models

Model validation method

Before comparing and scoring models, each model must first be validated. For a model to be considered valid, certain criteria must be met. Firstly, all model coefficients should be statistically significant, meaning they contribute meaningfully to explaining the variation in the data. Secondly, the residuals of the fitted model—assuming all coefficients are significant—must behave as white noise. This means the residuals (i.e., the difference between observed and predicted values) should appear random, with no discernible patterns. Any structure or autocorrelation in the residuals would suggest that the model has failed to capture all relevant information in the data. Additionally, the residuals should exhibit constant variance (i.e., no heteroscedasticity) and follow a normal distribution.

The method/pipeline used when validating models in this report is as follows:

1. Run coefficient test to see if all coefficients are statistically significant to our model using the coeftest() function in R.
2. Visual inspection of raw residuals and standardized residuals using the qqPlot() function in R.
3. Determining whiteness (nonsignificant autocorrelation) by visual ACF inspection of the residuals and by performing the Box-Ljung test using the Box.test() function in R with a minimum of 10 for the lag parameter.
4. Determining if the data follows normal distribution by visual histogram inspection of the residuals as well as performing normality tests:
    a. Jarque-Bera test using the jarque.bera.test() function in R
    b. Shapiro-Wilks test using the shapiro.test() function in R
    c. Anderson-Darling test using the ad.test() function in R

    A model is said to pass normality if it passes a minimum of one of the normality tests.

5. Determining heteroscedasticity (constant variance) of the residuals using the ARCH test using the ArchTest() R function.
6. Acquiring AIC and BIC for later comparisons using the AIC() and BIC() functions in R.

AR

AR models of order 1, 2, 3, and 6 were fitted on the training data. 1 and 2 were chosen due to our previous look at the data showing a strong autocorrelation close in time. 3 and 6 were chosen due to the nature of the data being monthly, regarding water levels, and that the data exhibited seasonal patterns with a wavelength where quarterly or bi-yearly lags in time could provide the model information. None of these models were however validated due to failing all the normality tests in the test pipeline. It was thus decided that a Box-Cox transformation was to be applied to help reduce the skew hoping this would stabilize the residuals distribution enough for either of the normality tests to pass validation. AR(2) and AR(3) performed the best and exceedingly like each other. AR(3) did however perform slightly better with a better AIC score and slightly better RMSE, MAPE and MASE scores. Since it performed the best and passed validation AR(3) was chosen for the AR candidate of the model comparison.

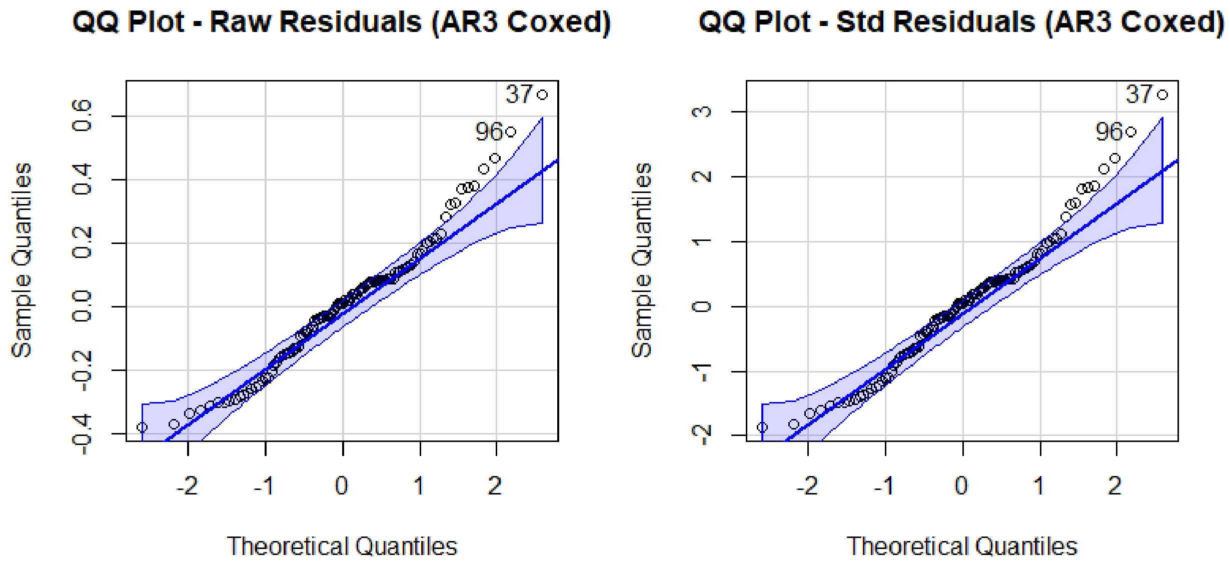

Figure 6 QQ plots for raw and std residuals of the best AR model

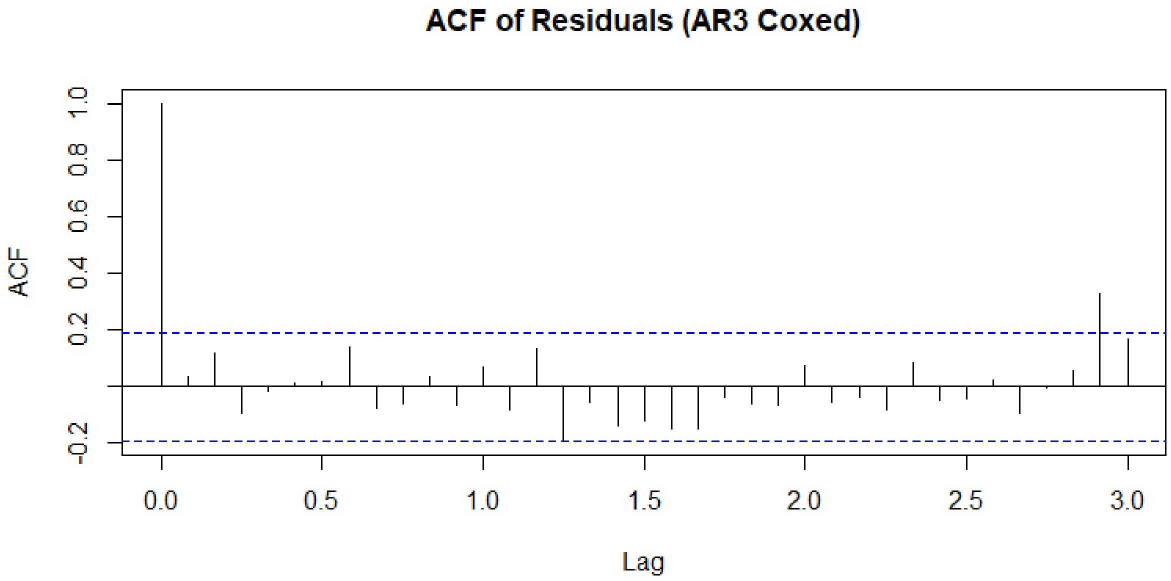

Figure 7 ACF plot of the residuals of the best AR model

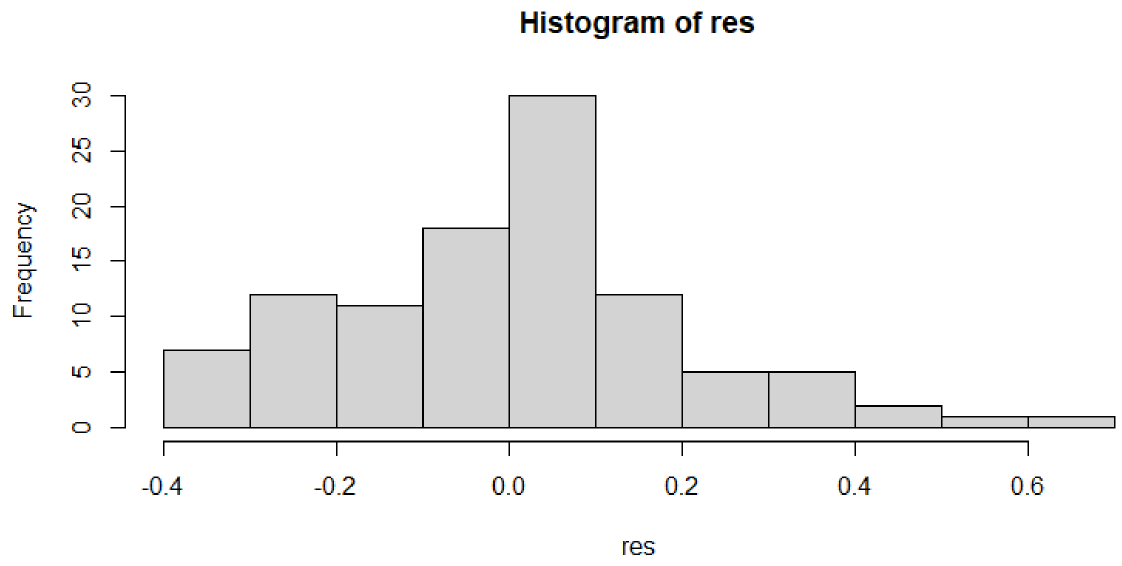

Figure 8 Histogram of residuals of the best AR model

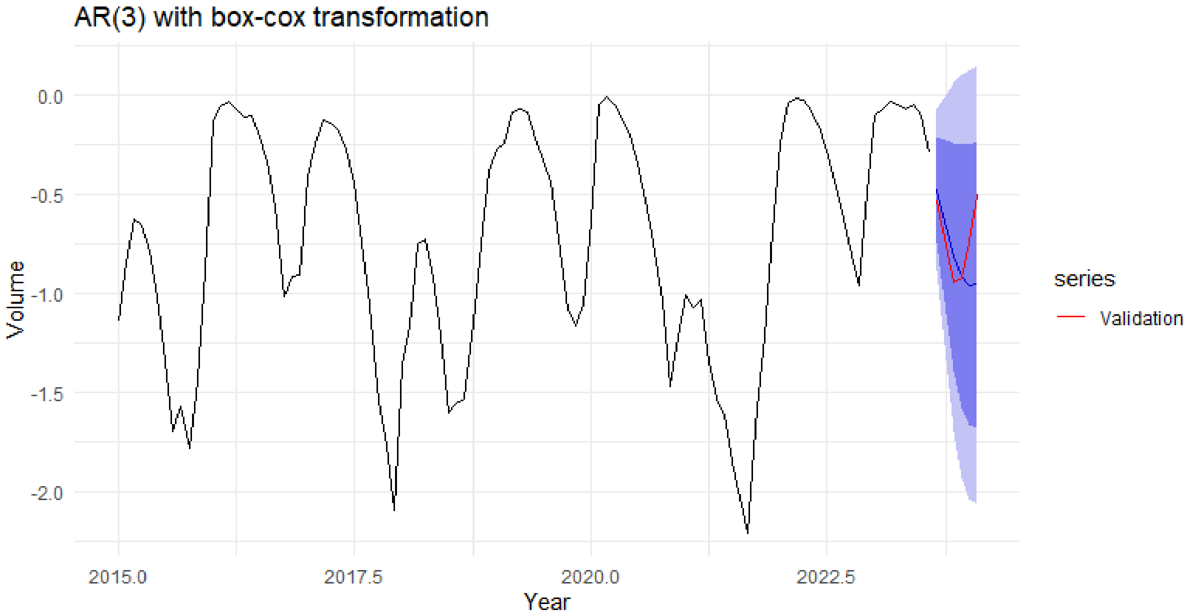

Figure 9 Fitted vs Validation plot for the best AR model

ARMA

For ARMA models no models were able to be validated due to failing all normality tests. This meant a Box-Cox transformation was once again required. After transformation three candidates for the ARMA model candidate were trained and validated. Out of these the ARMA model with 2 autoregressive parts and 3 moving average parts, and seasonal components compromised of both a cosine and sine part performed the best, yielding it as the ARMA model for future cross model comparisons.

The seasonal components were of a 12-period cycle and are defined as:

$$cos_t = \cos\left(\frac{2\pi t}{12}\right),$$

$$sin_t = \sin\left(\frac{2\pi t}{12}\right)$$

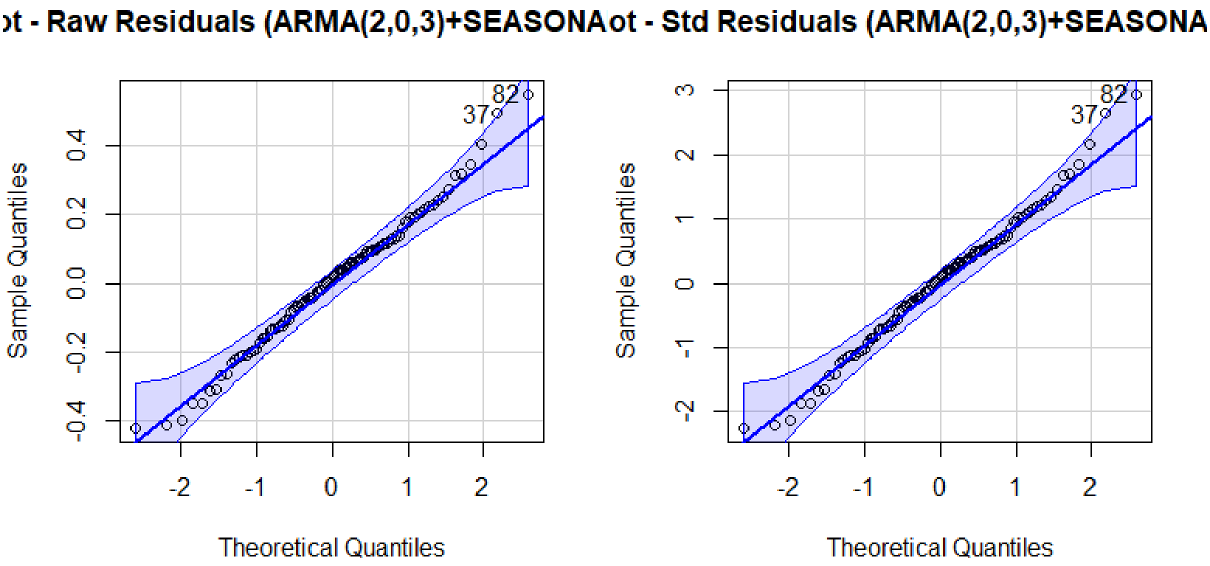

Figure 10 QQ plot for raw and std residuals for the best ARMA model

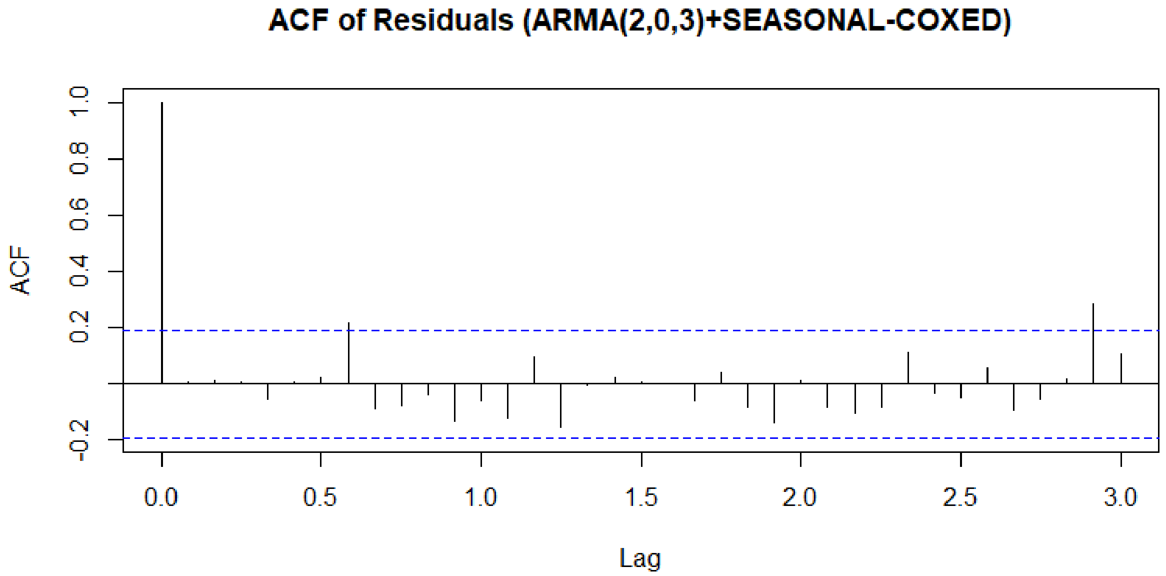

Figure 11 ACF plot for the residuals of the best ARMA model

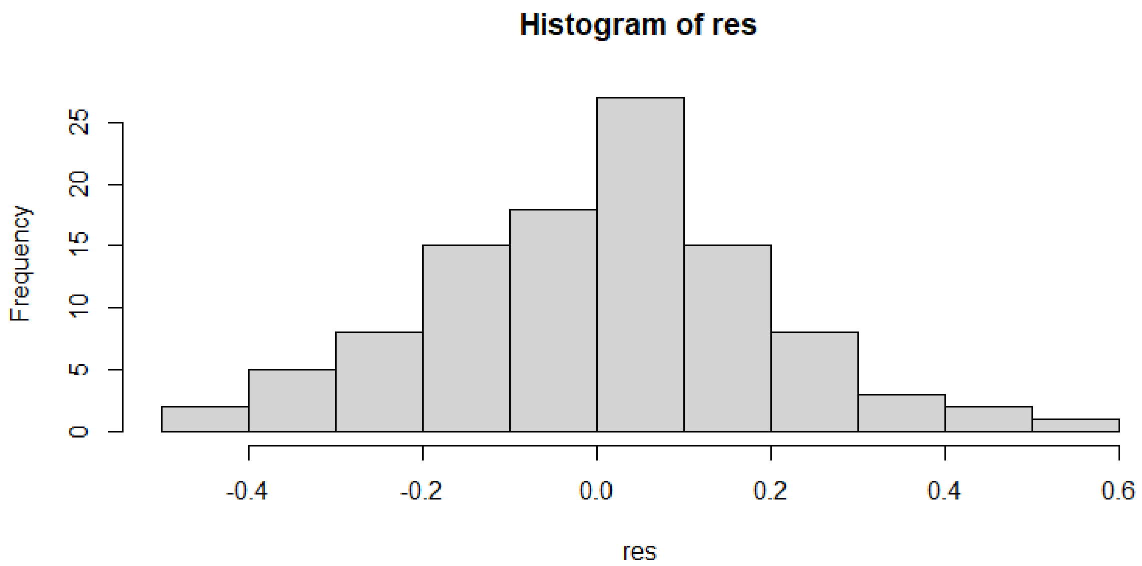

Figure 12 Histogram of the residuals for the best ARMA model

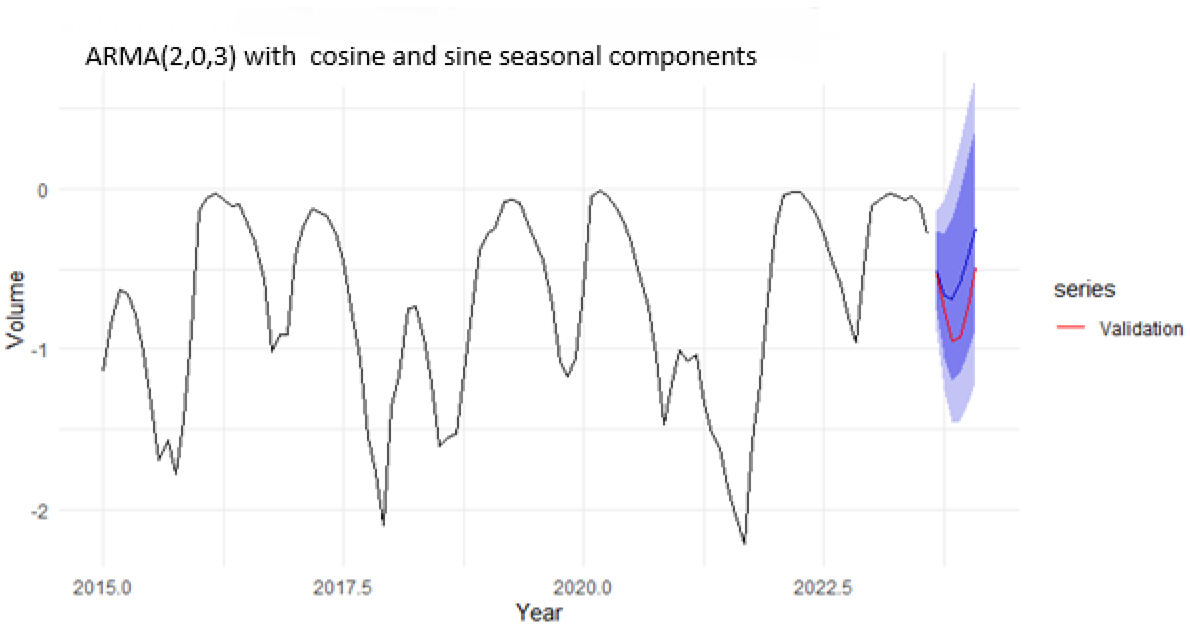

Figure 13 Fitted vs Validation plot for the best ARMA model

SARMA

For the SARMA model once again no models were validated unless we performed a Box-Cox transformation. Several runs of the R function auto.arima() were run with various parameters and testing all combinations up to order 6 of each relevant ARIMA parameter yielding a few candidates from where the optimal SARMA was chosen. The best SARMA had 2 autoregressive components and 2 moving average components, and 1 autoregressive and 1 moving average component for the seasonal order with a period of 12 where the first moving average coefficient was removed using the parameter “fixed” in the R function ARIMA.

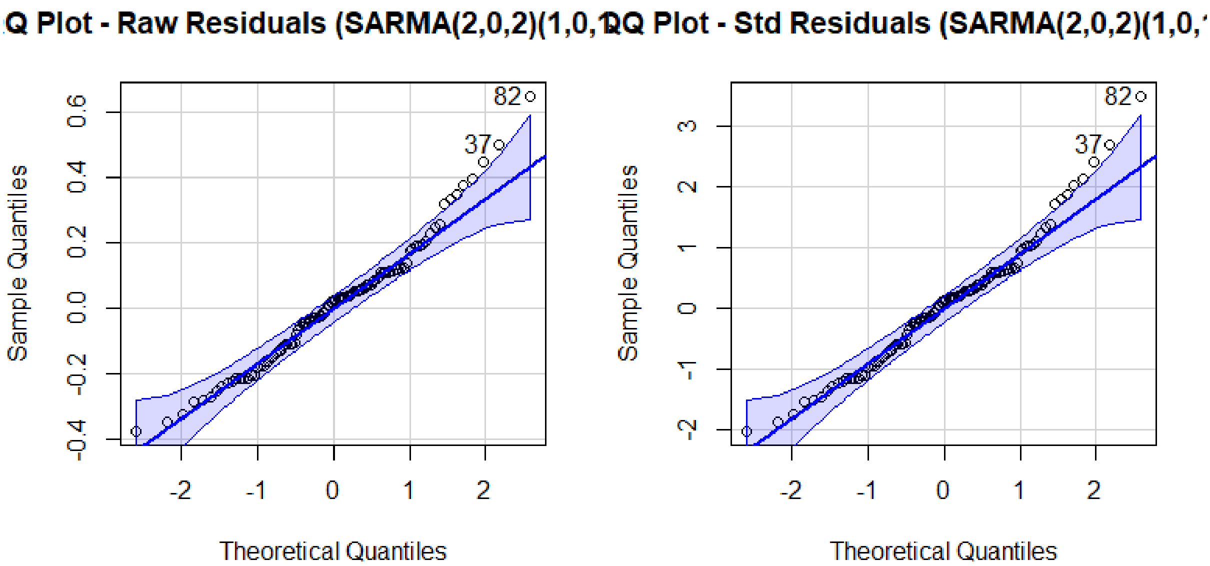

Figure 14 QQ plot for raw and std residuals of the best SARMA model

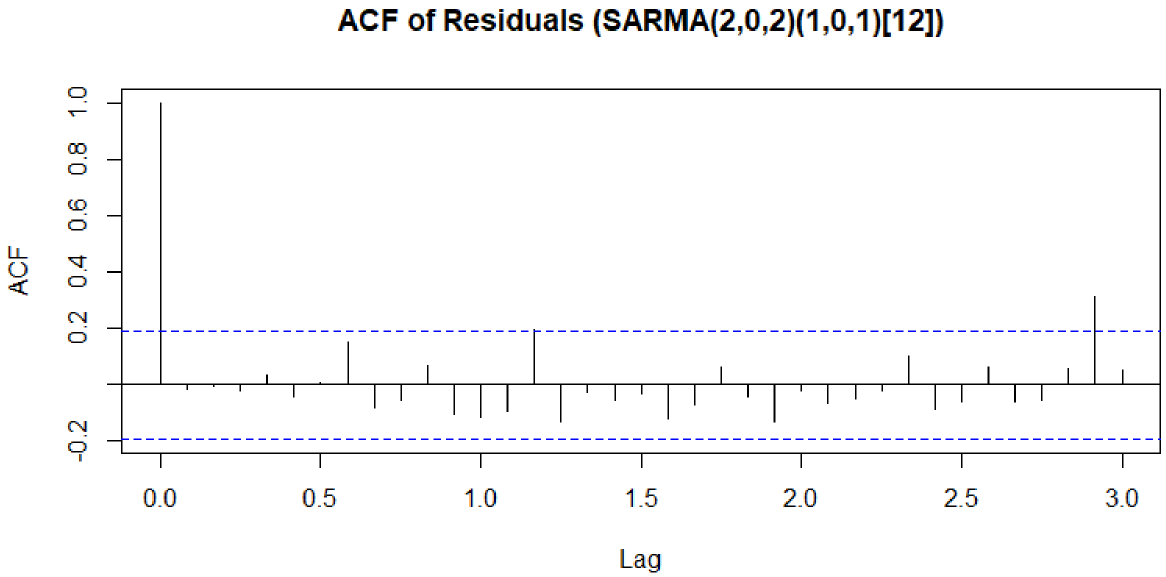

Figure 15 ACF plot of the residuals for the best SARMA model

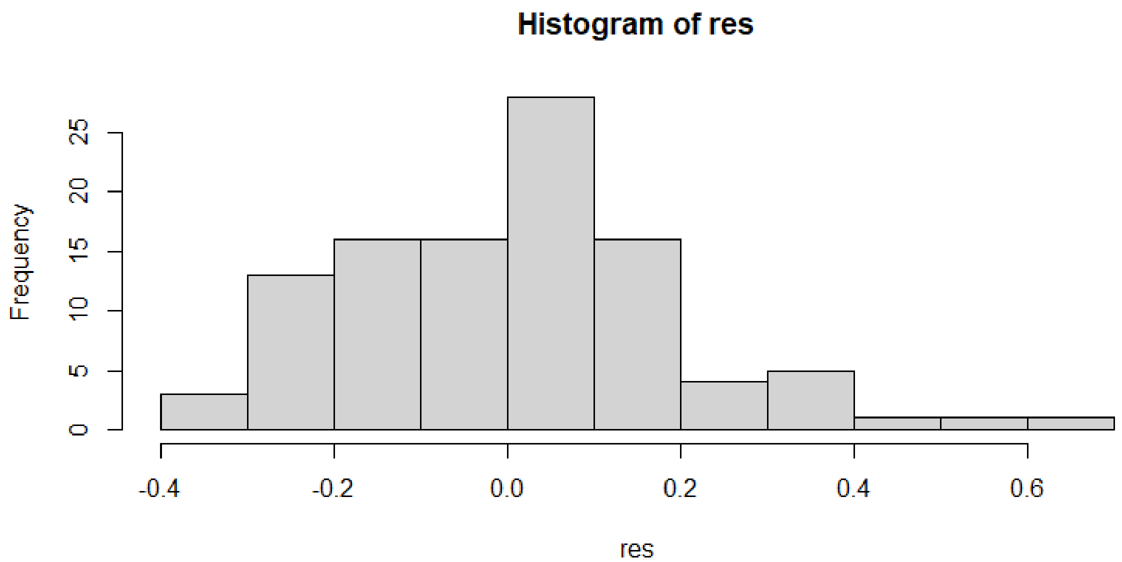

Figure 16 Histogram for the residuals of the best SARMA model

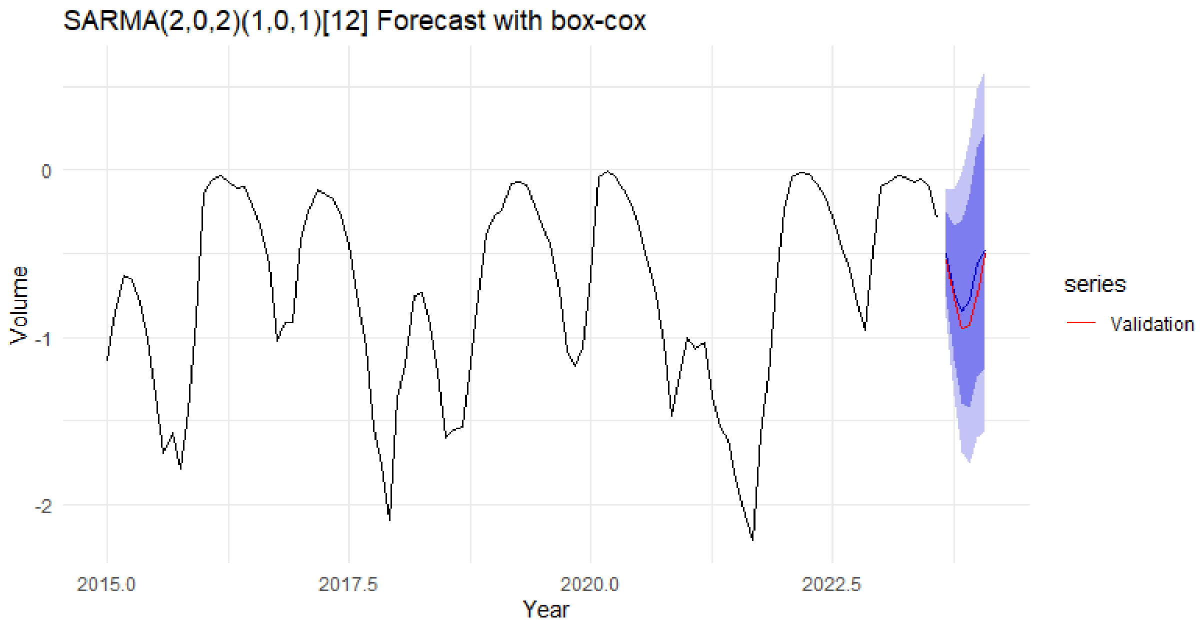

Figure 17 Fitted vs Validation plot for the best SARMA model

SARIMA

As per assignment instructions there were also the question on other potential models. Two interesting models were decided to be the ARIMA and SARIMA models. Even though the data was validated as stationary, mayhap a potential differentiation or seasonal differentiation could help improve performance, or would it lead to a worse fit due to overparameterization and applying differentiation when none was necessary. Box-Cox transformation was once again necessary and after several runs of the R function auto.arima() several interesting models were found. The most interesting one was the SARIMA(2,0,0)(2,2,1)[12] model which showed good error metrics but poor AIC and BIC. It was validated and decided to be used in the model comparison as a candidate for any other models.

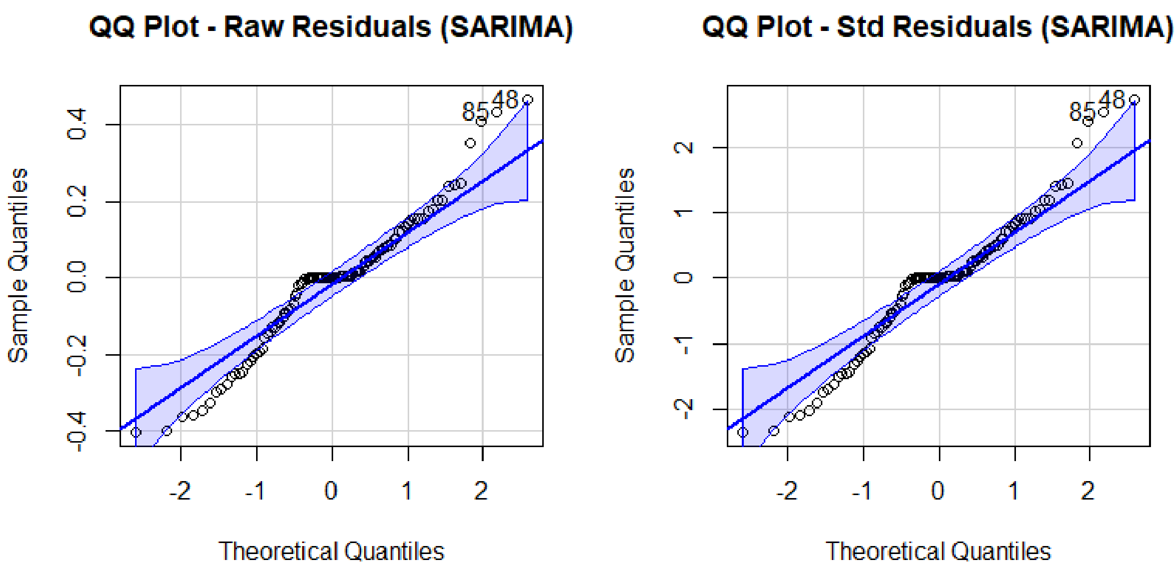

Figure 18 QQ plot of raw and std residuals of the SARIMA model

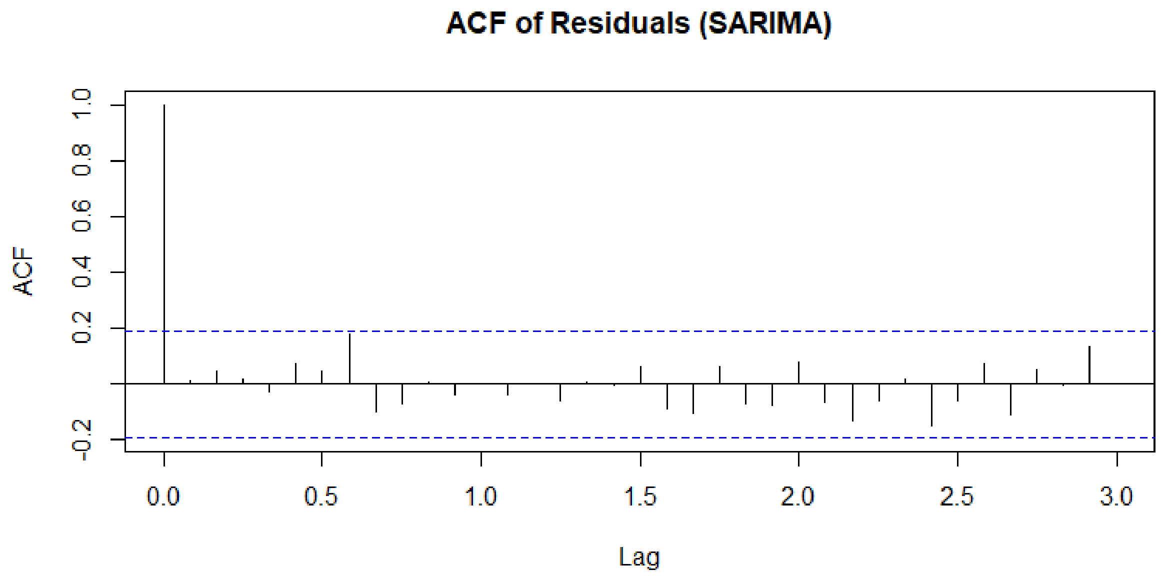

Figure 19 ACF of the residuals of the SARIMA model

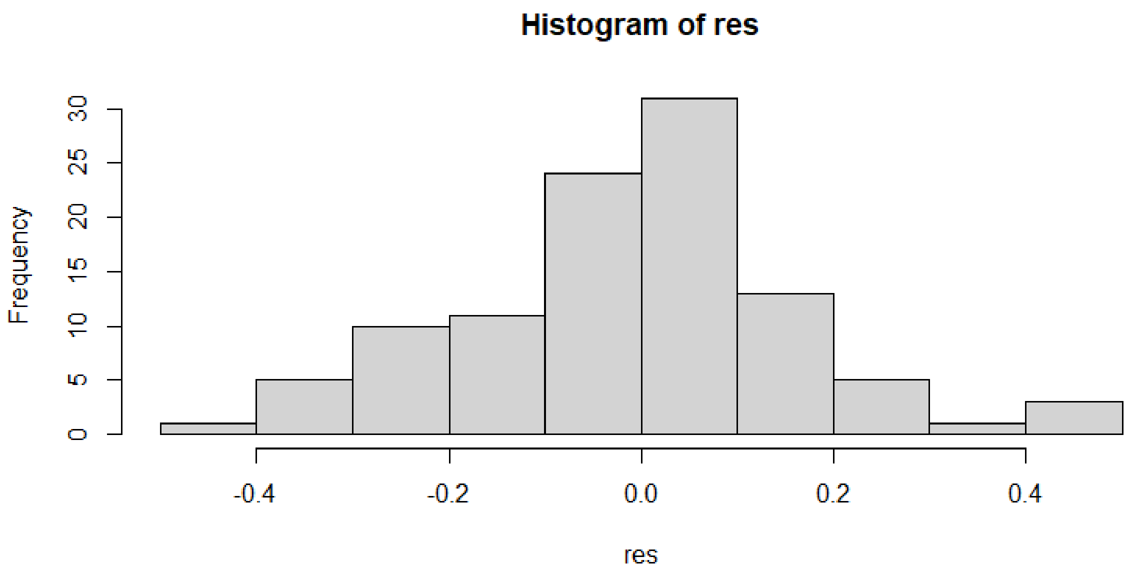

Figure 20 Histogram of the residuals of the SARIMA model

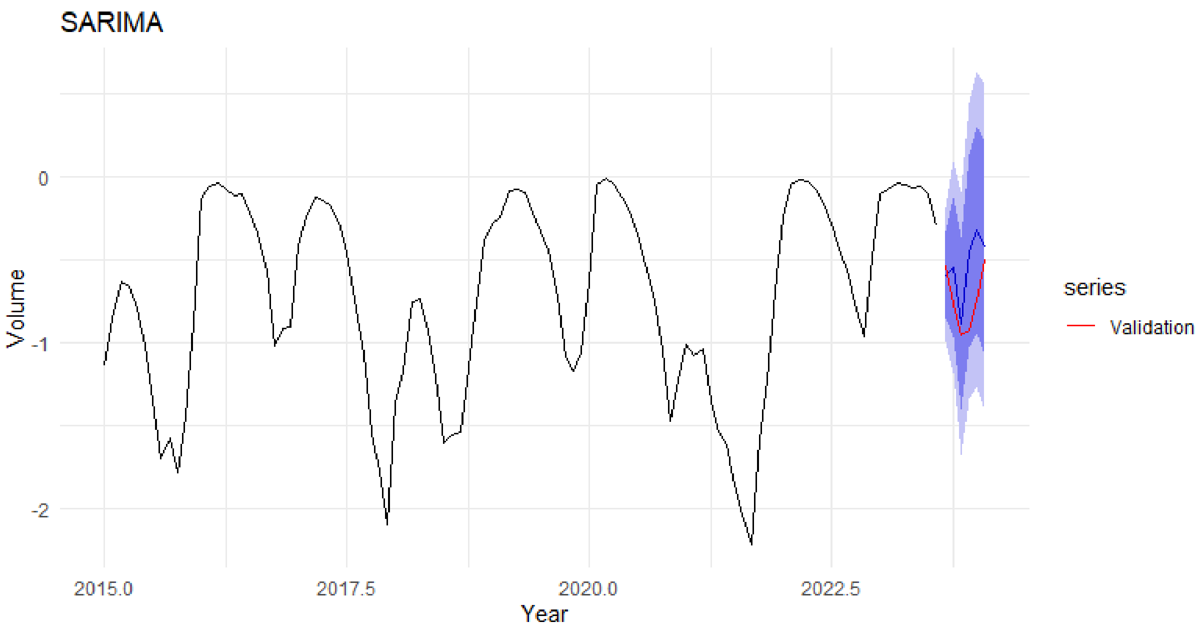

Figure 21 Fitted vs Validation plot for the SARIMA model

Results

Table 1 Table of AIC, BIC, RMSE, MAPE, and MASE scores for the best model of each type.

Metric	AR	ARMA	SARMA	SARIMA
AIC	-23.23	-35.39	-30.1	23.71
BIC	-10	-11.59	-11.59	38.01
RMSE	0.2036623	0.1851362	0.1849285	0.1709636
MAPE	91.20405	93.57528	94.3286	108.4546
MASE	0.2351551	0.2180856	0.2131581	0.1814181

Discussion

Model Fit – AIC and BIC

The ARMA model yielded the lowest AIC value (-35.39), indicating the most favourable balance between model complexity and goodness of fit. SARMA produced a comparable BIC value (-11.59), but its AIC (-30.1) is not as low as that of ARMA, suggesting that ARMA is better suited for the given data. In contrast, SARIMA recorded positive AIC (23.71) and BIC (38.01) values, which reflect a poorer fit, likely due to over-parameterization.

Forecast Accuracy – RMSE, MAPE, and MASE

Among all models, SARIMA achieved the lowest RMSE (0.17096) and MASE (0.18142), suggesting that it delivers the smallest average forecast errors and performs well relative to a naïve baseline. However, SARIMA also produced the highest MAPE (108.45%), indicating a high percentage error that may be attributed to the presence of small actual values in the series. The SARMA model, on the other hand, offers a more balanced performance across all accuracy metrics. It reports a low RMSE (0.18493), relatively low MAPE (94.33), and a favorable MASE (0.21316). Combined with its reasonable AIC and BIC scores, SARMA demonstrates both stable fit and consistent forecasting accuracy.

Conclusion

Based on the evaluation metrics, the ARMA model demonstrates the best overall fit to the data, supported by the lowest AIC and a competitive BIC. While the SARIMA model achieves the lowest RMSE and MASE, indicating strong raw error performance, its high AIC, BIC, and MAPE suggest potential overfitting and instability. The SARMA model offers the most balanced performance across all metrics, combining reasonable model fit with consistent forecast accuracy. Therefore, SARMA emerges as the most reliable choice when both model quality and predictive performance are considered.