Simon Fraser University

STAT 485

Group 19 Final Project Report

Time Analysis on SALR Data

Author

Alexander Li

Mirrien Li

June Nguyen

Professor

Gary Parker

December 9, 2023

1 Project Background

The real estate market in the Greater Vancouver area has experienced various fluctuations over the past decade. Understanding these changes is crucial for market stakeholders, including investors, realtors, and policy makers. There is a need for a comprehensive analysis of the **Sales to Active Listings Rate** (SALR) to understand its trends, seasonal patterns, and the probability and impact of extreme market events. We propose a practical approach by conducting a time series analysis of the SALR from 2011 to 2023. This will include trend analysis, seasonality study, forecasting future values, and particularly focusing on the identification and analysis of extreme events in the market.

2 About Dataset

The SALR values are collected and provided by The Real Estate Board of Greater Vancouver (REBGV) for public usages. The REBGV is a member-based professional association of 15,000 REALTORS® and their companies who live and work in communities from Whistler to Maple Ridge to Tsawwassen and everywhere in between.

Here we perform initial exploratory data analysis (EDA) to understand the basic characteristics of the data. This includes plotting the time series, checking for missing values, and basic descriptive statistics. Our primary focus will be on the feature `Year_month` and `Calculated_SAL`.

From the data plot [*Figure 1.*], we have some initial observations:

- The descriptive statistics show that the average SALR over this period is approximately 25.95%, with a standard deviation of 12.78%. The minimum SALR observed is 8.23%, and the maximum is 70.30%.
- In the plot, the SALR seems to fluctuate periodically across the years (unlike white noises). There seems to be some seasonality patterns (about every 5 years). There are 2 spikes 60 months apart, where we suspect that they are more likely due to non-recurring events (e.g., property purchase policy changes, or the pandemic) rather than some natural cycles.
- The series seems to be non-stationary. There are significant swings in SALR over time, indicating a high variability among the data points.

3 Trend, Seasonality, and Stationarity Analysis

Due to the nature of the SALR variable, it is unlikely that there is a deterministic trend where it can apply to the time series forever as the domain of the feature is ranged between 0 and 1. We hypothesize that there is a stochastic seasonal trend and that the time series may be suitable for seasonal adjustment and/or forecasting using models that can account for both trend and seasonality, such as Multiplicative Seasonal ARIMA (as introduced in Chapter 10). First, we will check if a slightly upward linear trend exists.

3.1 Testing for a Linear Trend

The summary output [*Table 1.*] is for a linear trend model fitted to the SALR data. The model suggests a significant relationship ($p = .000342 < .05$) between time and SALR. The coefficient for time (0.9794% per year) is positive, indicating an upward trend over the years. The multiple R-squared value is 0.08116, which implies that about only 8.1% of the variance in SALR can be explained by the time trend alone. This is relatively low, suggesting that other factors (possibly captured in the seasonal or other irregular components such as economic cycles, policy changes, etc.) may play a significant role in explaining the variability in SALR. The F-statistic (13.43) and its p-value ($0.0003423 < 0.0001$) suggest the model is statistically significant. Note that the linear model is assuming uncorrelated iid errors, whereas we potentially have correlated errors, but we are unsure at this point yet.

Generally, if the trend component is significant and persistent over time, even within a bounded series, detrending could be part of the process to achieve stationarity. However, based on the nature of the variable, to preserve the integrity of the data's natural boundaries, and to reflect a realistic understanding of the variable's behavior, we conservatively choose not to detrend the series by removing the linear trend.

3.2 Identify MA order: ACF

From the ACF plot (with unit Year on the x-axis) [*Figure 1.*], we find:

- **Significant Autocorrelation at Various Lags**: The plot shows several spikes (at months 1-15, 26-45, 60, 66-73) that exceed the significance bounds. We also observe that the sample autocorrelation does not exceed the bounds after 73 months (lag 6) and becomes smaller and smaller and eventually stabilizes after 120 months (lag 10).
- **Seasonal Effects**: The presence of peaks at approximately regular intervals suggests a seasonal pattern. However, these peaks do not seem to occur at each integer lag (i.e., an annual seasonality in this monthly-sampled data), but rather at roughly every 5-6 years. Because the dataset records SARL from January 2011 to October 2023, we are only observing two cycles, which is one of the major limitations we must acknowledge in this time series analysis.
- **Long Range Dependence**: The fact that the sample autocorrelation function remains significant over many lags (where they slowly taper off but don't cut off sharply) suggests that the series may have long-range dependence or a slowly decaying trend. It suggests that a pure MA process is likely not appropriate (but might need an MA component), and that differencing might be needed to achieve stationarity.
- **Non-Stationarity**: The slow decay of the autocorrelation function is typical of non-stationary data. For non-stationary series, the ACF typically fails to die out rapidly as the lags increase. There might need some seasonal adjustments on the series.
- An additional tool to test stationarity is the Augmented Dickey-Fuller test introduced in Ch 6.4. We may execute `adf.test(SALR_ts)` in the `tseries` package and find that the p-value is .1696, indicating that we do not reject the null hypothesis of non-stationarity.

3.3 Identify AR order: PACF

From the PACF plot (with unit Year on the x-axis) [*Figure 1.*], we find:

- **Many Non-Significant Lags**: Most lags are within the confidence bounds and are therefore considered non-significant, with a few exceptions at month 1, 11, 13, and 49. There are no values exceeding the noise bounds from month 50 and beyond.
- **No Clear Seasonal Pattern**: Unlike the ACF plot, the PACF plot does not show the periodic behavior. Besides the first month (lag .083), the big spike at lag 1 may be hinting that a pure AR(1) may be an appropriate model to start with fitting, or as a non-seasonal component of a potential SARIMA model.

3.4 Identify ARMA orders: EACF

The triangle region of zeros shown in the EACF [*Table 2.*] is not a clear cut for lower orders of p and q. We see many x's around 11, 12, and 13, as well as some x's at 23 and 25, indicating a possible 12-month seasonality (with a decaying seasonal effect) in the data generating process. The matrix suggests that ARMA(1,0), ARMA(1,1), or ARMA(2,1) might be a good point to start. The MA(1) model, as expected, does not seem appropriate, while we would nevertheless try to fit it to see if any insights exist.

4 Model Fitting

4.1 Fitting ARMA Models

Initially, we do not apply differencing and assume that the SALR time series is stationary. We try to fit ARMA models (non-seasonal at this point) to look for some insights on the undifferenced data. We will use the dynamics method but starting at MA(1), AR(1), and ARMA(1,1), ignoring an AR(0) model as the series is unlikely to be white noise. Then we check ARMA(2,1) and see if it improved from ARMA(1,1).

From the comparative ACF Plot [*Figure 2.*], we can see that the MA(1) is significantly less favorable than the other two as evidenced in the dissimilarity in the ACFs and the highest AIC score of 1076.17 among the models. The sample ACF plot seems to only resemble the ACF plots of the AR(1), ARMA(1,1), and ARMA(2,1) for the first 16 lags, failing to account for any of the periodic patterns beyond. This would be consistent with the earlier analysis that suggested the presence of seasonality or other complex patterns. ARMA(1,1) does not improve significantly from AR(1), while ARMA(2,1) seems to be worth further analysis. We will choose AR(1) and ARMA(2,1) and see what their residuals look like.

The plot of the residuals [*Figure 3.*] of the AR(1) model looks a bit more white-noisy with some extremes around 2011, 2016 and 2021, indicating that potential periodic trends may still persist. No apparent systematic pattern (e.g., a funnel shape) shown in the residuals vs fitted values plot, aligning with the assumption of constant variance of errors in the AR(1) model. The histogram shows approximately symmetrical distribution around zero, suggesting that the model is likely not biased. Potential outliers in the residuals are seen at both tails. The Q-Q plot shows normality in the central part but heavy deviations at the ends of the residuals' distribution, indicating potentially more outliers than a normal distribution would have, which aligns with what we observed in the histogram. These two plots demonstrate that the AR(1) model residuals do not perfectly follow a normal distribution.

The ACF plot shows that most autocorrelations are within the confidence bounds, suggesting that AR(1) has captured the main AR structure of the series. The few big spikes seem to occur around year 1, 2, 4, 5, and 6, indicating that the residuals might not be random and a more complex model (e.g., a SARIMA with a period of 12 months) might be necessary to capture the rest of the pattern. The AR(1) model appears to be a suitable choice given the sharp cutoff in the PACF plot of the original data, which indicates an AR(1) process is present. Only a few significant spikes with smaller magnitude compared to the sample PACF suggest that the need for additional AR terms might be low. We will test this in subsequent analysis.

We also compare the residuals between AR(1) and ARMA(2,1) and we see similar results except for a heavier right-skewness in the histogram and an improvement in the ACF and PACF of the ARMA(2,1) residuals, where we observe less spikes in both ACF and PACF in the negative direction. For the level of complexity ARMA(2,1) adds on top of the AR(1) model, we tentatively keep both models as a candidate for adding a seasonal component.

4.2 Fitting Seasonal ARMA Models

Given the results of the AR(1) and ARMA(2,1) residuals analysis, we find it appropriate to add the seasonal component to both models (i.e., a SAMIRA model with difference = 0) of period 12 (i.e., the yearly seasonality seen in the residuals) and 60 months (i.e., the 5-year cycle seen in the observations).

We immediately notice that a seasonal period of 60 months does not work here. The sufficiency of an annual seasonality is overriding the needs of trying to look for a 5-year periodic correlation in the time series. In addition, we only have 2 cycles in our dataset, limiting our capability of finding such seasonality. Therefore, from now on we will not be considering a period of 60 months.

The SARIMA$(1, 0, 0) \times (1, 0, 0)_{12}$ [*Figure 4.*] appears to require least number of parameters among the models yet performs fairly well among the models we considered. It has a non-seasonal AR coefficient of .9, a seasonal AR coefficient of .5, and low standard errors for both terms. Its response function tells us that it forgets 50% of a shock after 7 months, 80% 17 months, and 95% 31 months. This is similar to what we observed in the plot of the raw data points, where the shocks at 2016 and 2021 were impacting for about the next 15-25 months.

The SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$ [*Figure 5.*] seems to give a better fit. The model has a correlation of 0.45 and 0.38 for the non-seasonal AR terms, 0.75 for the non-seasonal MA term, and 0.56 correlation every 12 months. However, they receives a low level of significance, indicating less reliable estimates of the coefficients. The response function suggest that it forgets about 50% of a shock after 8 months, 80% after 15 months, and 95% after 26 months, which aligns better with the observations we see in the actual data plot.

The residual plots of the two SARIMA models has the same general look. The presence of outliers at 2016 and 2021 in both plots indicates that there might be extraordinary events not captured by either model, or that these could be natural, rare fluctuations in the data. The histogram of SARIMA(2,0,1)x(1,0,0)12 is more centralized around 0. Its Q-Q plot seems to stay closer to the line at both tails than in SARIMA(1,0,0)x(1,0,0)12 residuals. We reach the conclusion that the SARIMA(2,0,1)x(1,0,0)12 model residuals is better at following a normal distribution.

The ACF and PACF plot of the SARIMA(2,0,1)x(1,0,0)12 model shows a reduction in the amount of significant spikes by about 50% compared to those for the SARIMA(1,0,0)x(1,0,0)12 model. It indicates that the SARIMA(2,0,1)x(1,0,0)12 is capturing more information than the other.

In conclusion, the residual analysis for the SARIMA models seems to suggest that SARIMA(2,0,1)x(1,0,0)12 is a better fit for the data than the SARIMA(1,0,0)x(1,0,0)12 model. The ACF and PACF plot show the sufficiency in the former model to capture most of the patterns in the data, leaving some spikes observed in the plots. The response function [*Figure 7.*] of the winner model also mimics more closely to our expectations of how the time series would be responding to a shock, based on how it has responded to the extreme events at 2016 and 2021. Therefore, we would choose SARIMA(2,0,1)x(1,0,0)12 as our best model.

4.3 Exploring Differencing

In our initial analysis, we treated the SALR data as if it were stationary, allowing us to fit stationary ARMA models and incorporate a seasonal component to address potential seasonality. While this approach has provided valuable insights, we acknowledge that, as argued in the stationary analysis, the time series does exhibit non-stationary behavior. To further refine our understanding and model fitting, we investigate the presence of a stochastic trend by applying differencing to the SALR data.

We will begin with identifying the level of differencing and apply the appropriate differencing to remove linear trends to attempt achieving stationarity. This process will possibly lead to an improved SARIMA model that better captures the underlying dynamics of the data.

From the above plots [*Figure 8, 9, 10.*], we conclude that the first differencing seems more appropriate than the second differencing for several reasons:

- The second differencing shows signs of over-differencing, which is indicated by the increased number and magnitude of spikes in the ACF plot and the early significant negative spikes in the PACF plot. The over-differencing issue typically manifests as a series of alternating positive and negative spikes in the PACF plot, especially if the underlying process is not of a higher order.
- The first differencing maintains a reasonable level of variability without introducing the artificial patterns seen in the second differencing.

- The increase in the number of significant lags in the ACF and PACF for the second differenced series suggests that it may be capturing noise rather than the true signal.

For the first differencing, the ACF plot with 8 significant lags may indicate some remaining autocorrelation, but the overall reduction in autocorrelation is a positive sign. The presence of some significant lags in its PACF plot may suggest the need for additional AR or MA terms to capture the autocorrelation structure. Given these points, we proceed with the first differencing and consider adding AR or MA terms.

4.4 Fitting SARIMA Models

We first compare several ARIMA models: ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1), and ARIMA(2,1,1). From the ACF plots [*Figure 11.*], we observe that all of the models by themselves are not sufficiently capturing the long-term dependency in the data. However, the ARIMA(1,1,1) model seems to capture more of the complexity in the data more appropriately, as evidenced in the ACF plot with a more pronounced decay across several lags. Therefore, we will advance with ARIMA(1,1,1) for residual analysis.

We have similar observations as seen in the residual analysis [*Figure 12.*] for the AR(1) model, except that:

1. The histogram has achieved a closer approximation to the expected bell-shaped curve with the outliers still persist at both tails
2. The Q-Q plot has a reduction in the deviations from the theoretical quantiles near the right tail
3. Both ACF and PACF for ARIMA(1,1,1) are only marginally better than those of AR(1) residuals.

The similarities in the residuals analysis suggest that the added complexity has not captured substantially more information of the data structure.

Next, we incorporate seasonal components.

By incorporating both non-seasonal and seasonal AR components while differencing once, the SARIMA$(1,1,1)\times(1,1,1)_{12}$ model *Figure 13.*] seems to capture the short-term dynamics and the seasonality of the series effectively. Given the substantial improvement in AIC compared to the previously best model's AIC of 918.58 (from SARIMA$(2,0,1)\times(1,0,0)_{12}$, it seems that adding a first differencing in both seasonal and non-seasonal component collectively has significantly improved the fit of the model. In addition, this differenced model has lower standard errors, suggesting more confidence in the coefficient estimates. The lower variance of the residuals also indicates a better fit to the data in terms of explaining the variability. We look at its residuals:

The first couple residuals (up to 2012) closely cluster around zero, suggesting a sign of overfitting where the model captures the noise of the early data too well and do not generalize to the rest of the data or new data. Both the ACF and PACF plots indicate that there are no significant autocorrelations or partial autocorrelations in the residuals, suggesting that the SARIMA(1,1,1)x(1,1,1)[12] model has captured the underlying process well and that the residuals are essentially white noise.

4.5 Selecting the Best Model

In the comparative analysis of the ARIMA$(1,1,1)\times(1,1,1)_{12}$ and ARIMA$(2,0,1)\times(1,0,0)_{12}$ models, diagnostic checks including the ACF and PACF indicated that the residuals of both models are close to white noise, suggesting an adequate fit. However, given our practical understanding of the data, where significant spikes in 2016 and 2021 are more likely attributed to non-recurring external events rather than systematic seasonal patterns, a simpler model is favored. The ARIMA(2,0,1)x(1,0,0)[12], while statistically less optimal based on AIC, avoids overfitting by not overemphasizing seasonal differencing, which could inappropriately model these one-off events as seasonal phenomena. This aligns with our theoretical knowledge of the dataset and the specific nature of the observed anomalies. Moreover, the simpler model is preferred for its ease of interpretation, potential robustness in out-of-sample forecasting, and its pragmatic fit with the time series' characteristics, making it a prudent choice for our analytical objectives.

5 Simulation

We're simulating 100 series [*Figure 13.*] into the future and the simulated series from SARIMA$(2,0,1) \times (1,0,0)_{12}$ model capture similar patterns with our original time series. These simulations give us an idea of the possible distributions/paths of our observations over time. Most of the simulated series have observations reverted to the rate of 25%, with reasonable fluctuations around the mean. Some simulated series predict extremes that may happen in the next 5-year period.

6 Forecasting and Model Evaluation

6.1 Forecasting on Existing Time Period

One of the major goals of time series analysis is to predict the future data trend with high confidence and accuracy. In the previous section we have detected candidate models that fit the data relatively well, we will now test the model's performance on future data via a training/validation split. From the plot [*Figure 14.*], The ARIMA$(2,0,1) \times (1,0,0)_{12}$ does extremely well forecasting the data behavior and volatility in the testing period, although it does have a slightly slower mean reversion compared to AR(1) and the real data.

In the longer forecast [*Figure 15.*], we can see our model perform less well on the dataset. Particularly, although it does revert to the mean, the model no longer captures the volatility as accurately as the 24 month testing set. This may be a sign of overfitting, which we are unable to find a solution for within the scope of the course.

6.2 Forecasting on Future Time Period

We used our model to predict 5 years into the future. From the forecasting plot [*Figure 16.*] , our model suggests that market volatility would still remain high within the first two years, and slowly disappear assuming no major shock happens. It is impossible for our model to predict when or whether a shock would happen, hence it would be necessary for the model to be re-fitted whenever a shock happens for a more reliable forecast over a shock response.

7 Conclusion

Modeling efforts led to the selection of SARIMA$(2,0,1)\times(1,0,0)_{12}$ as the preferred choice due to its ability to capture short-term dynamics and seasonal patterns. This model, despite a relatively higher AIC compared to SARIMA$(2,0,1) \times (1,0,0)_{12}$, showcased a more reasonable fit by appropriately handling presumably non-recurring shocks in 2016 and 2021 without overfitting.

While proficient in short-term forecasting, the chosen model displayed reduced accuracy in longer forecasts, possibly indicative of overfitting. Predictions suggested a gradual decrease in market volatility, assuming no major shocks.

In conclusion, the SARIMA$(2,0,1)\times(1,0,0)_{12}$ model was favored for its alignment with observed anomalies and practicality. Continuous refinement is crucial for enhancing the model's reliability in predicting the dynamic trends of the Greater Vancouver real estate market.

7.1 Limitation and Recommendation

1. If experts believe that the shocks are regularly recurring events, we might need to re-consider the feasibility of the SARIMA$(1,1,1)x(1,1,1)_{12}$ model.
2. It is possible to obtain daily/weekly observations instead of monthly. In that case, we may derive different models and insights of how the market will behave in the future.

Appendix

```
Call:
   lm(formula = SALR_ts ~ time(SALR_ts))

   Residuals:
       Min        1Q    Median       3Q       Max
   -0.17793 -0.08132 -0.03085  0.06088  0.45529

   Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
   (Intercept)   -19.498040   5.392227  -3.616 0.000406 ***
   time(SALR_ts)   0.009794   0.002673   3.664 0.000342 ***
   ---
   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

   Residual standard error: 0.1229 on 152 degrees of freedom
   Multiple R-squared:  0.08116,Adjusted R-squared:  0.07511
   F-statistic: 13.43 on 1 and 152 DF,  p-value: 0.0003423
```

Table 1. Summary Output for `lm()`

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o	o	o	o	o	o	x
1	o	o	x	o	o	o	o	o	o	x	o	x	o	o	o	o	o	o	o	o	o	o	o	x	o	o
2	x	o	o	o	o	o	o	o	o	x	o	x	x	o	o	o	o	o	o	o	o	o	o	x	o	o
3	x	x	o	o	o	o	o	o	o	o	o	x	x	o	o	o	x	o	o	o	o	o	o	x	o	o
4	x	o	x	o	o	o	o	o	o	o	o	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o
5	x	o	x	x	o	o	o	o	o	o	o	x	o	x	o	o	o	o	o	o	o	o	o	o	o	o
6	x	x	x	x	o	o	o	o	o	o	o	x	o	o	o	o	x	o	o	o	o	o	o	o	o	o
7	o	x	x	o	o	o	o	o	o	o	o	x	x	o	o	o	x	o	o	o	o	o	o	o	o	o
8	o	o	o	x	o	o	o	o	o	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o
9	x	x	o	x	o	o	o	o	o	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o
10	x	o	o	x	o	o	o	o	o	o	o	x	x	o	o	o	o	o	o	o	o	o	o	x	o	o
11	o	x	o	x	o	x	o	x	o	o	o	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o
12	o	x	o	x	o	x	o	x	x	x	x	x	o	x	o	o	o	o	x	o	o	o	o	x	o	o
13	x	o	x	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
14	x	o	x	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
15	x	x	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
16	x	x	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
17	x	x	x	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
18	x	x	x	o	x	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
19	x	x	x	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
20	x	x	o	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
21	o	x	o	x	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
22	o	x	o	x	o	x	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
23	x	x	o	x	x	x	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
24	o	x	x	o	o	o	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
25	o	x	x	o	o	o	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o

Table 2. EACF Result

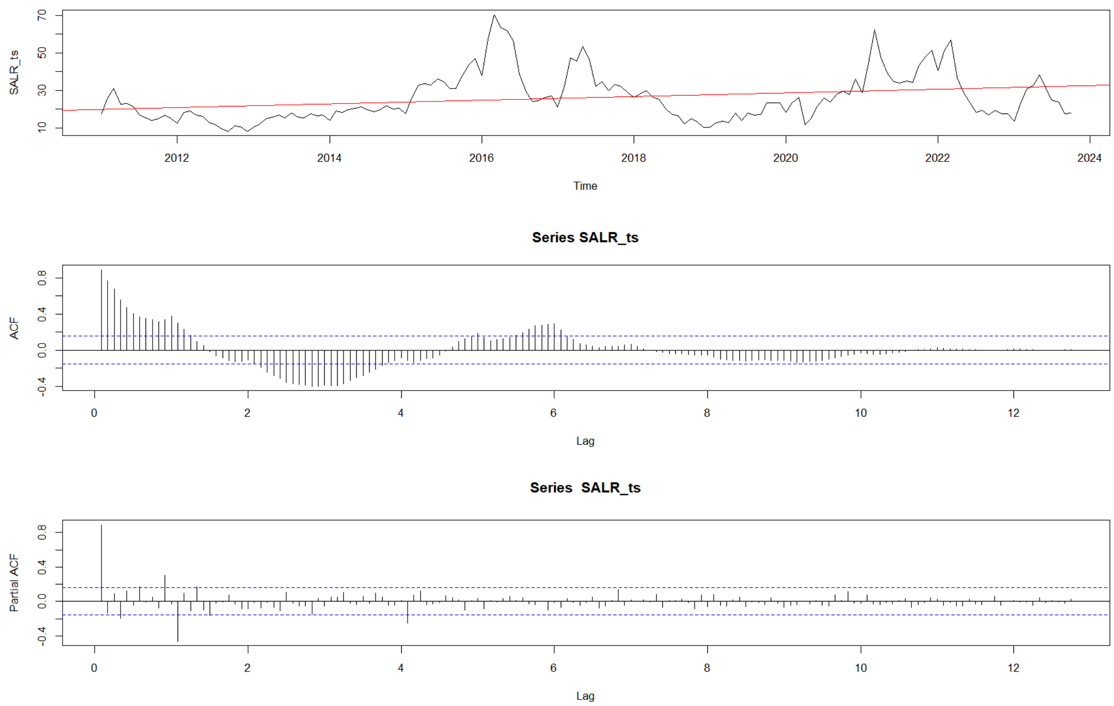

Figure 1. Data Information Plot

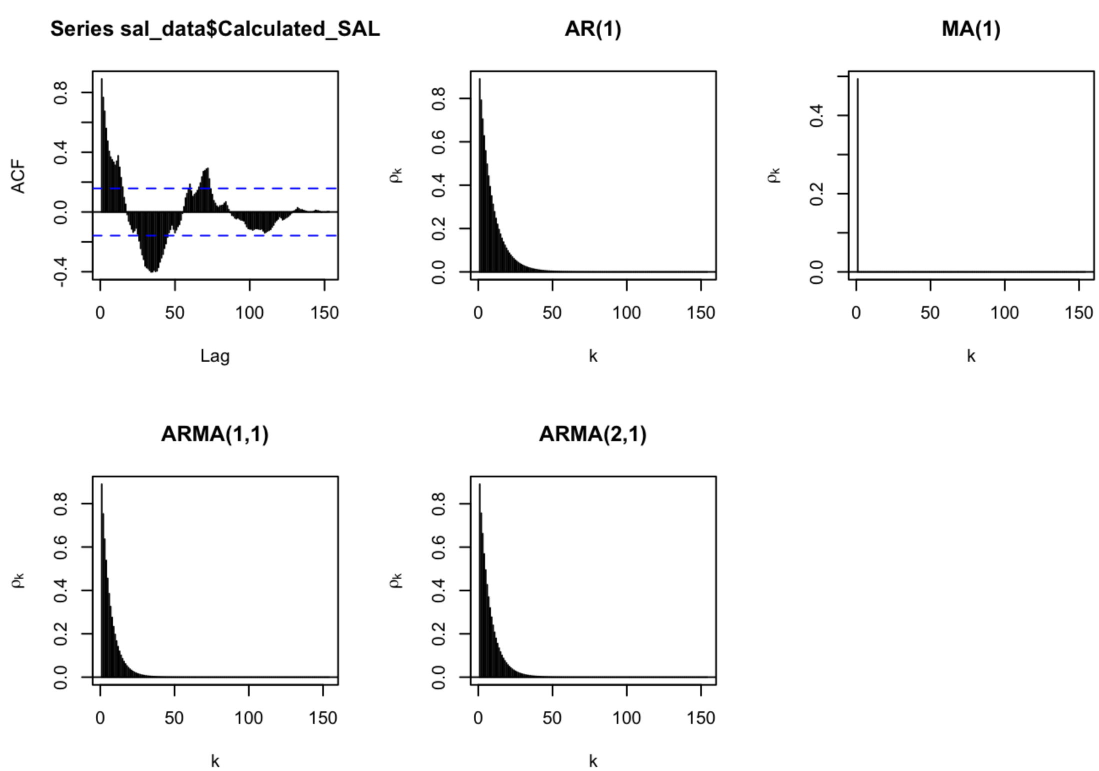

Figure 2. Comparative ACF Plot

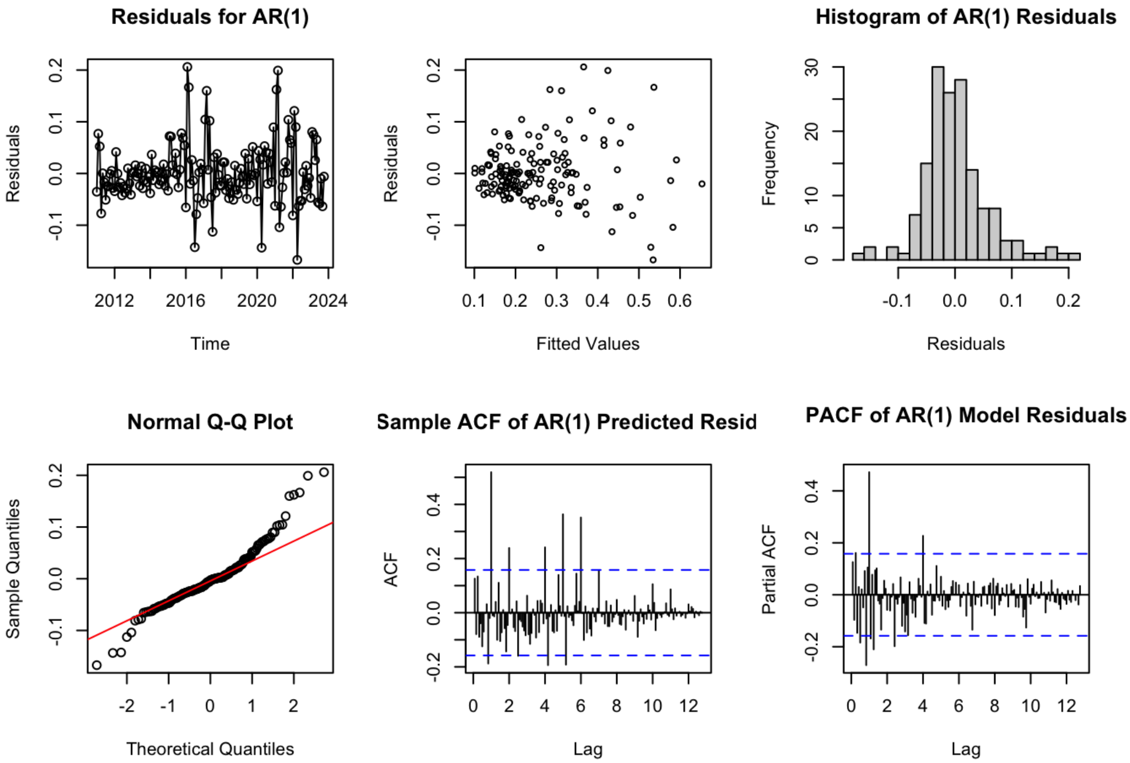

Figure 3. Residual Analysis of AR(1)

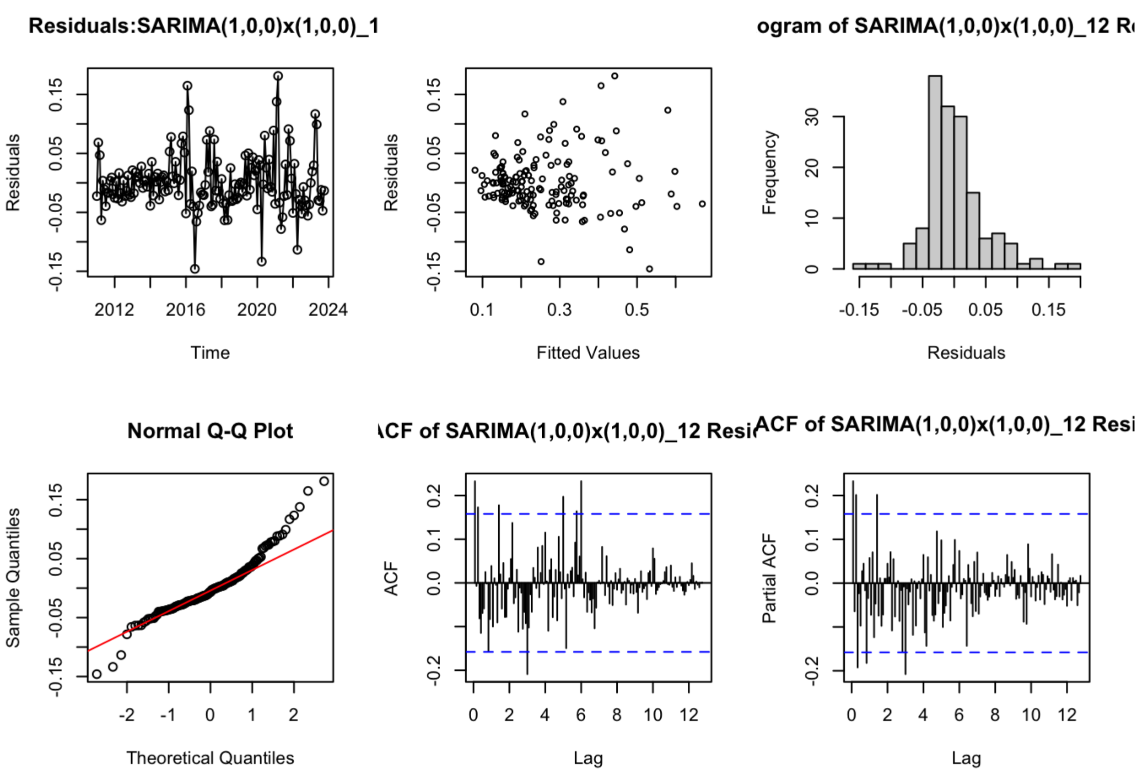

Figure 4. Residual Analysis of ARIMA$(1, 0, 0) \times (1, 0, 0)_12$

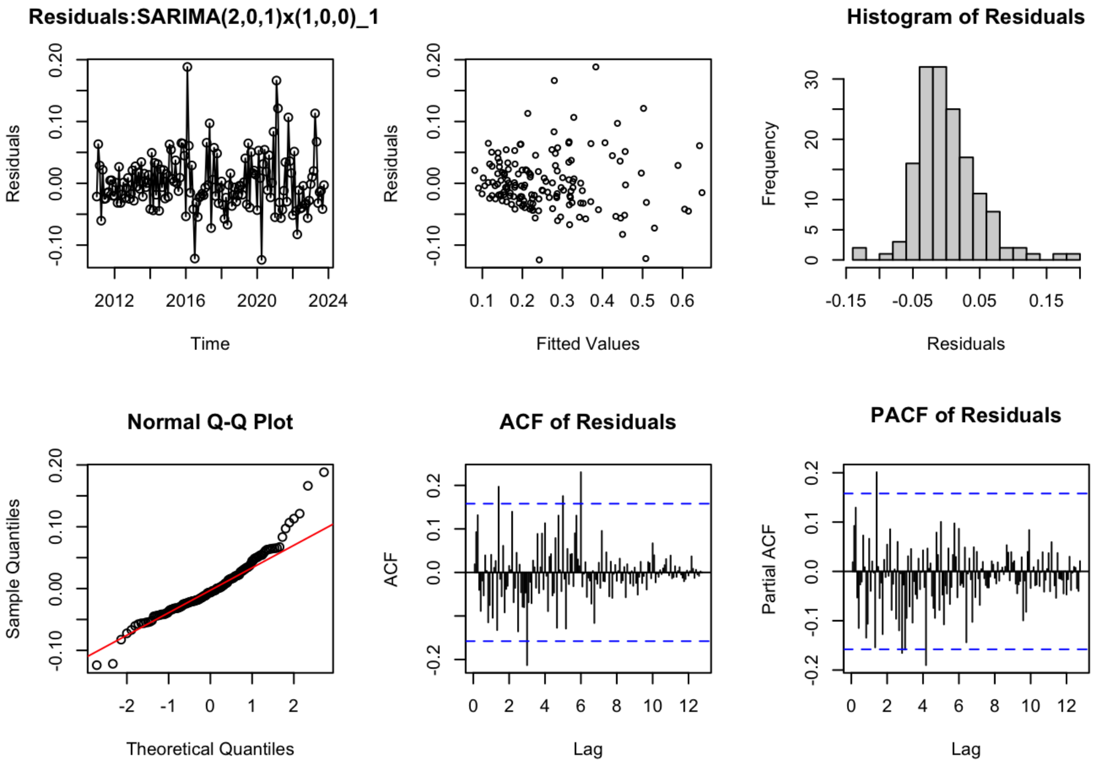

Figure 5. Residual Analysis of $\mathrm{ARIMA}(2,0,1) \times (1,0,0)_1 2$

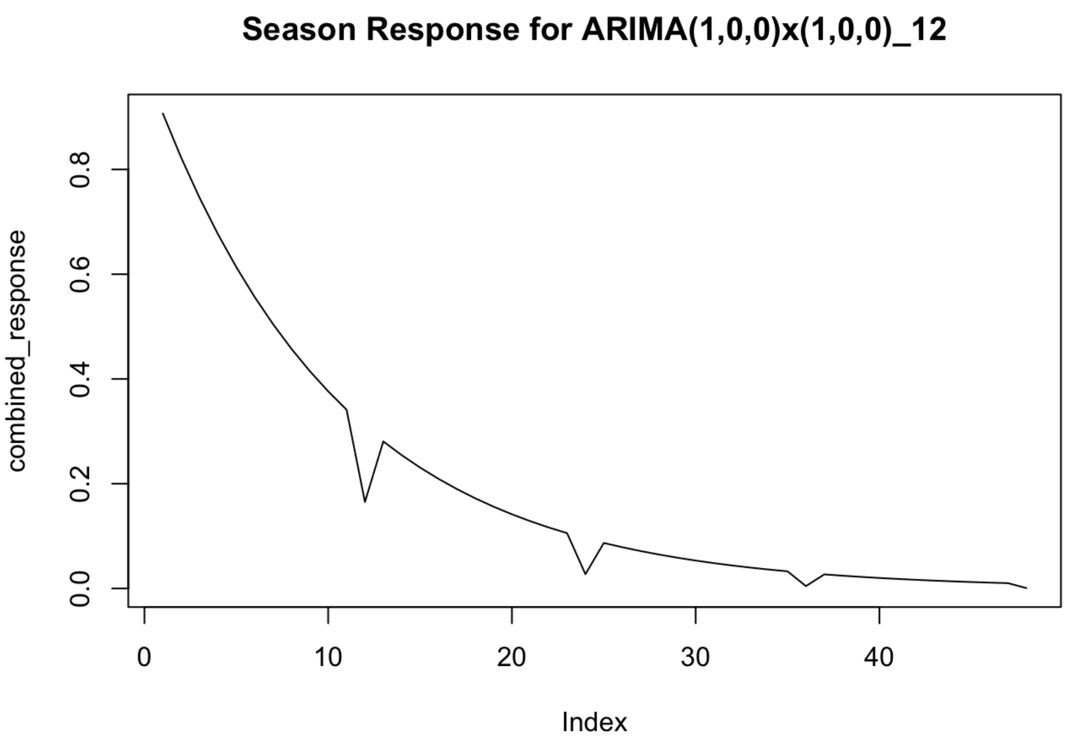

Figure 6. Response Function of $\mathrm{ARIMA}(1,0,0) \times (1,0,0)_1 2$

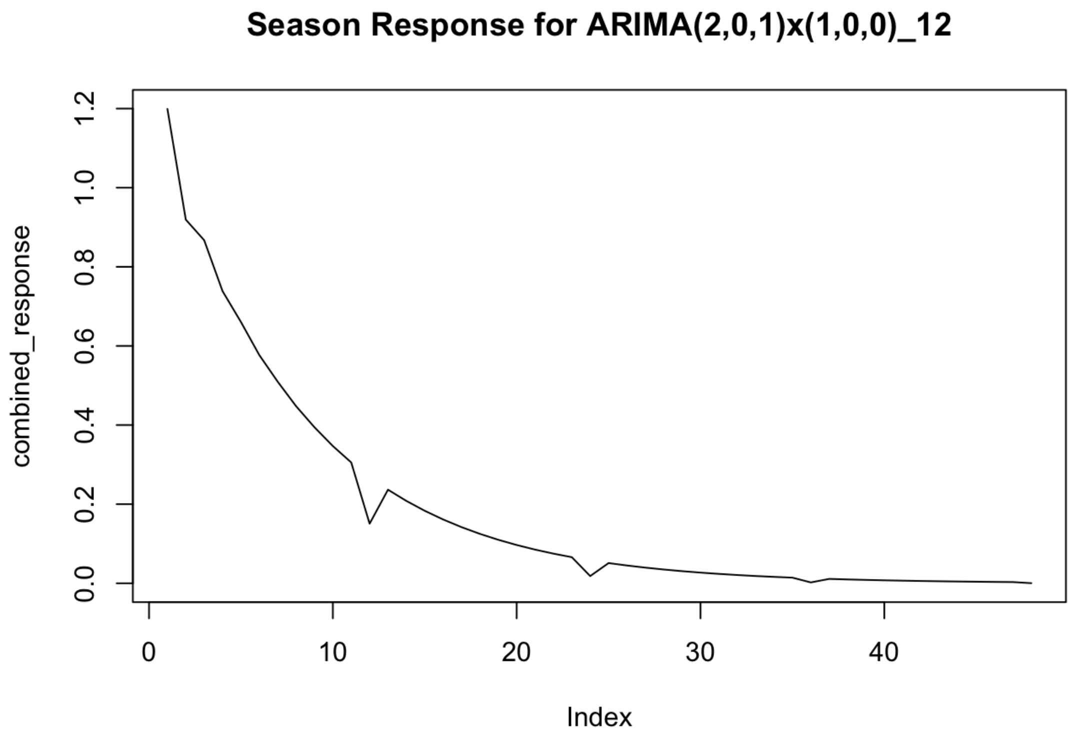

Figure 7. Response Function of $\text{ARIMA}(2,0,1) \times (1,0,0)_12$

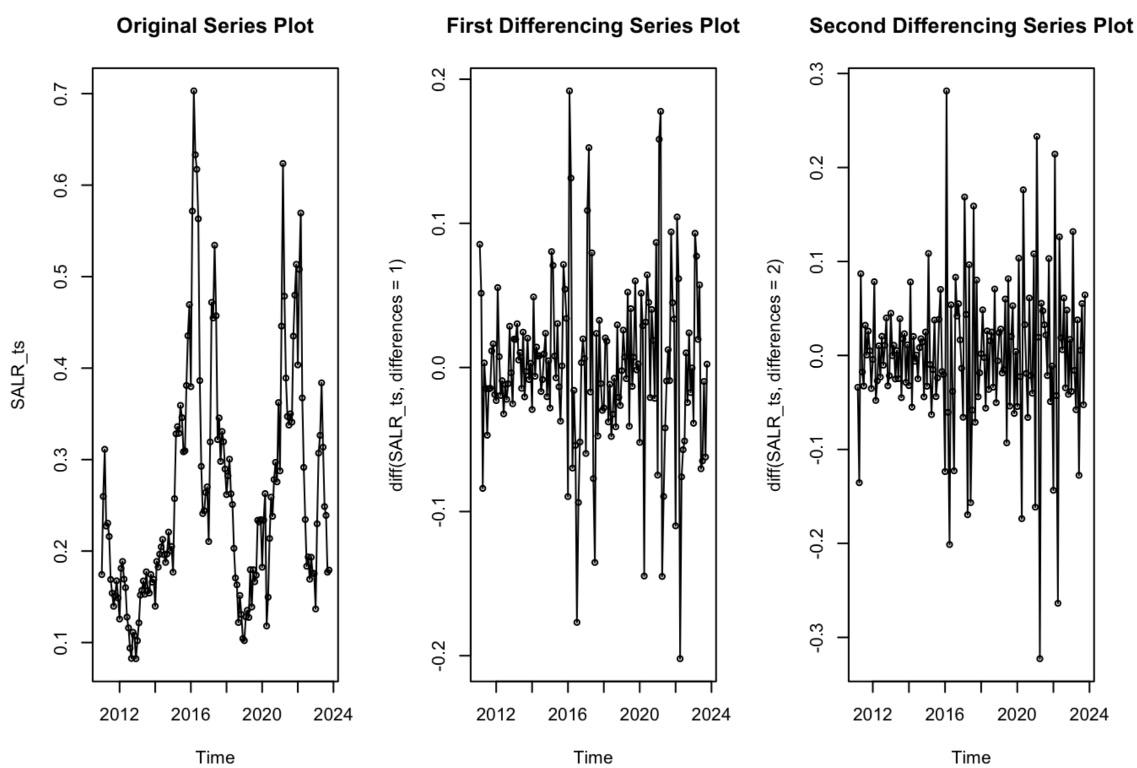

Figure 8. Differencing Dataset

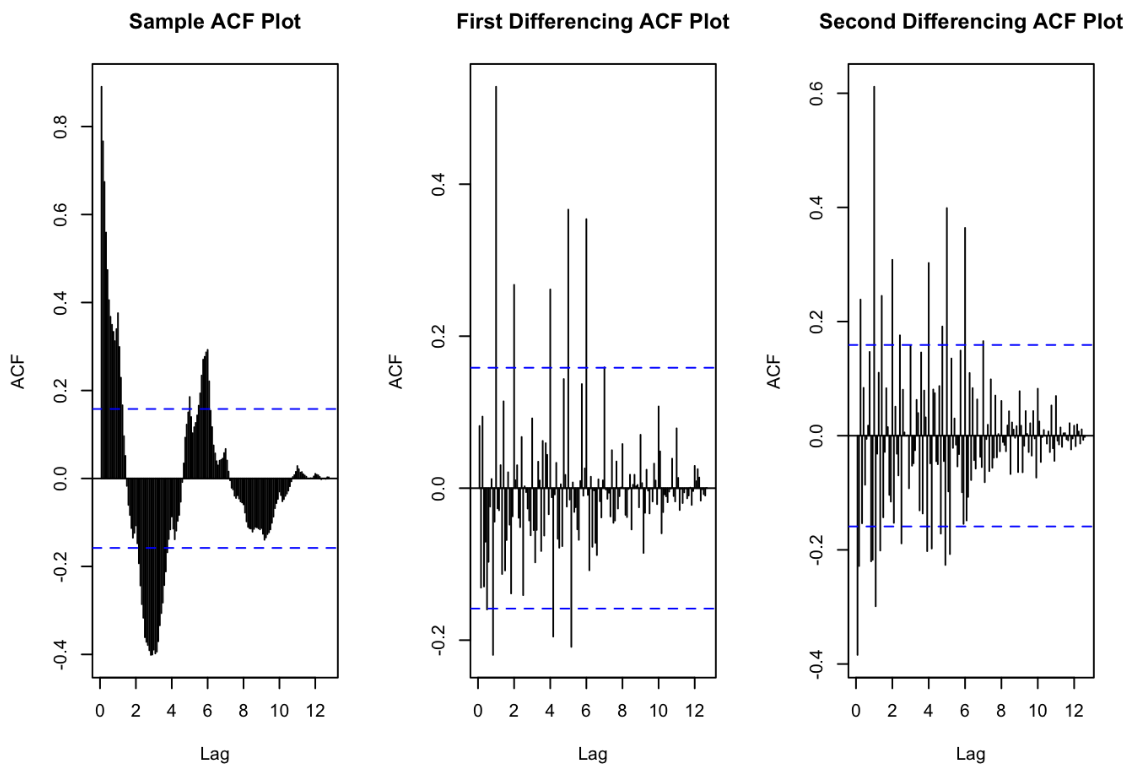

Figure 9. Differencing ACF

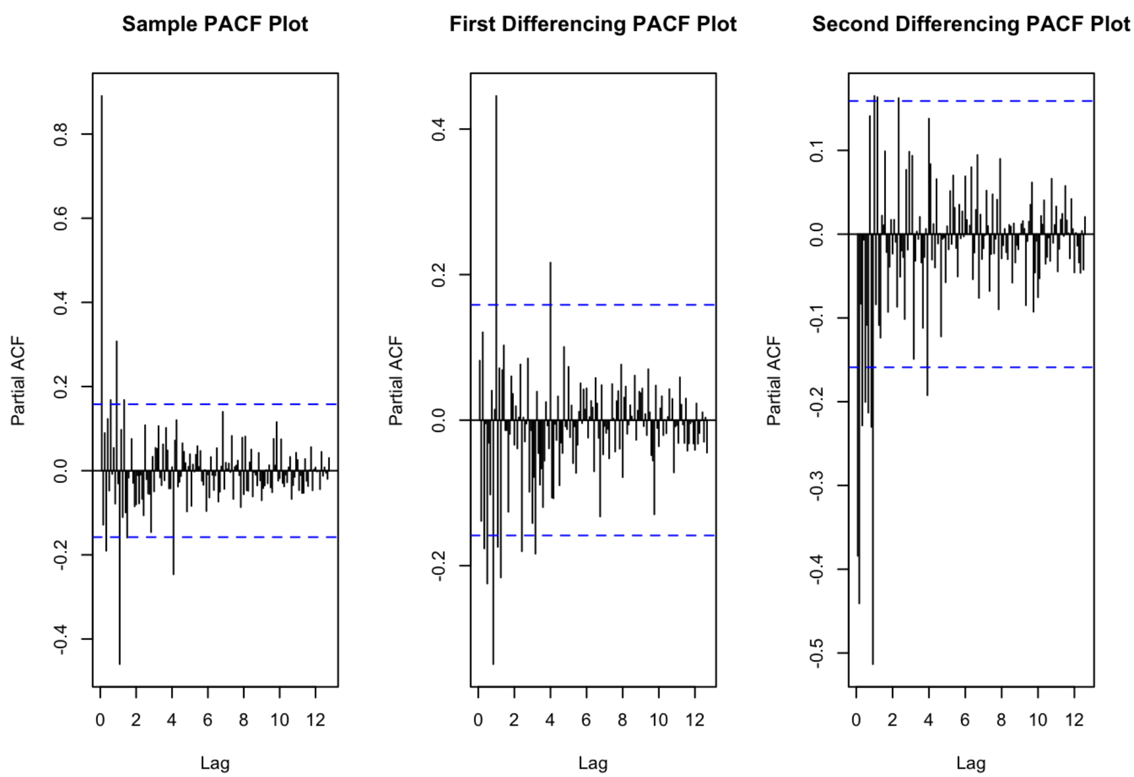

Figure 10. Differencing PACF

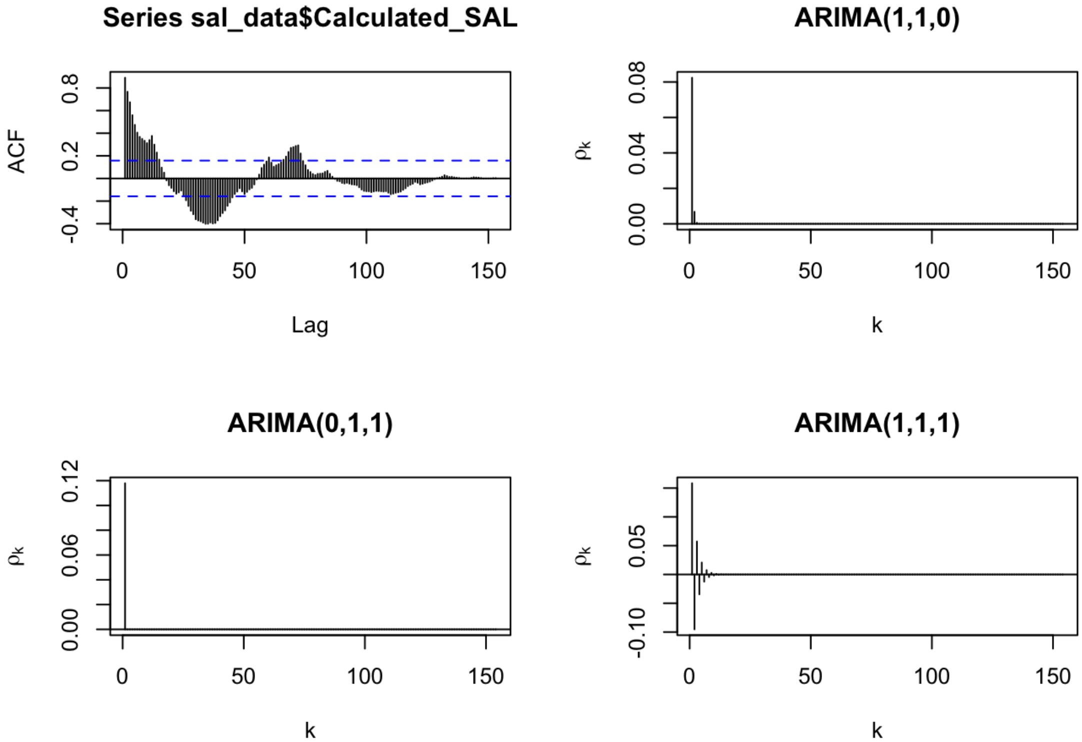

Figure 11. ACF Comparison of ARIMA Models

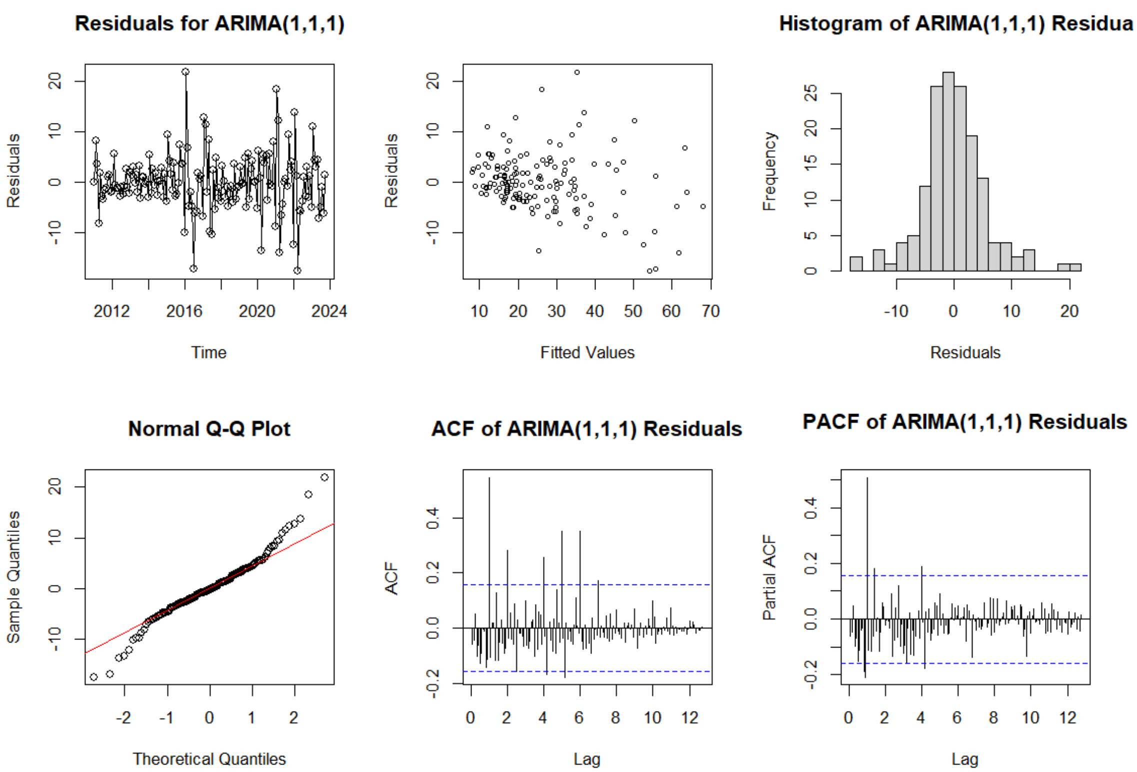

Figure 12. Residual Analysis For ARIMA(1,1,1)

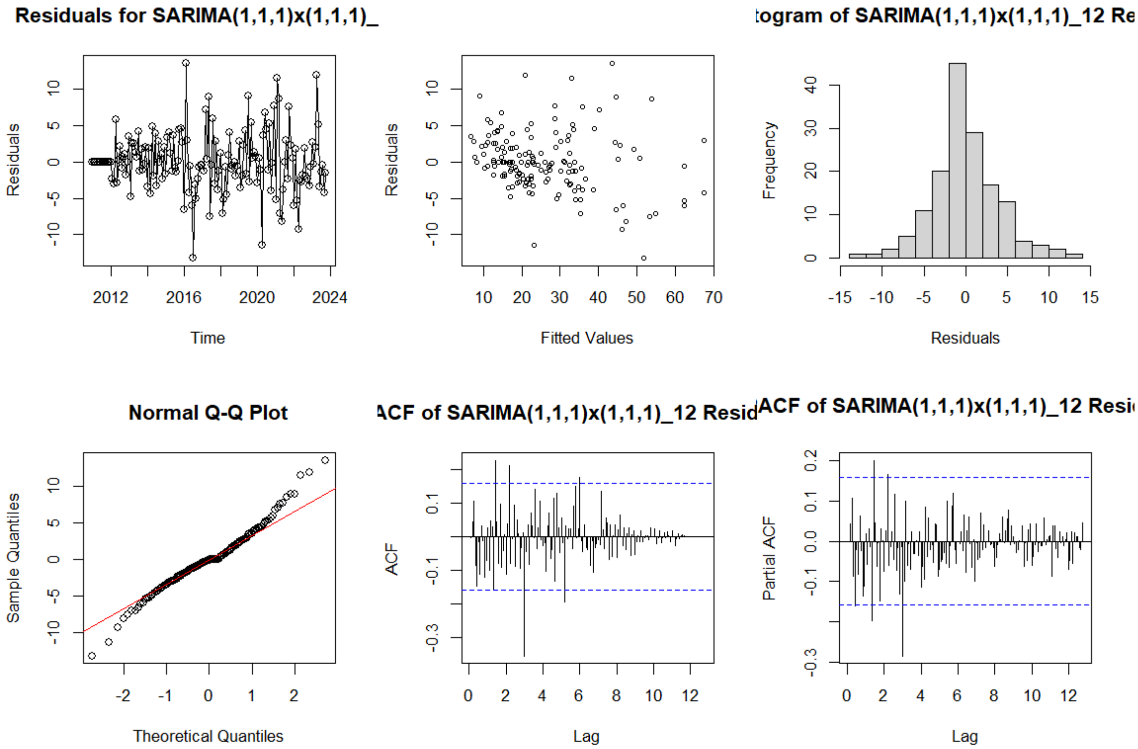

Figure 13. Residual Analysis For SARIMA(1,1,1)

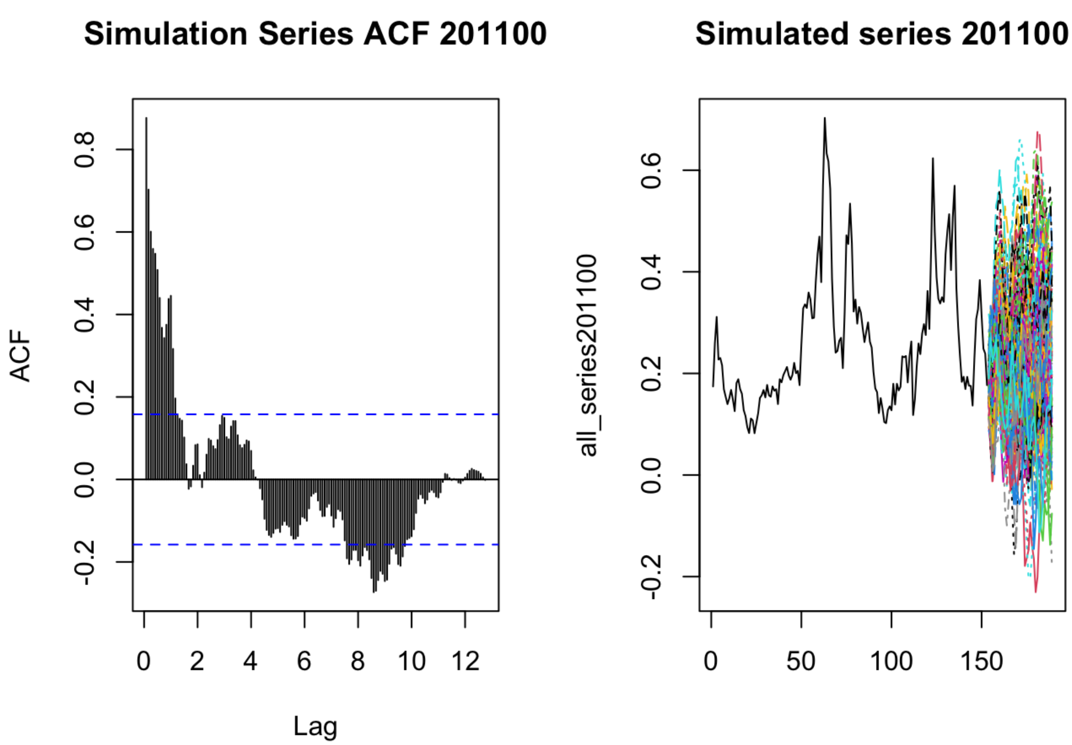

Figure 14. Simulations

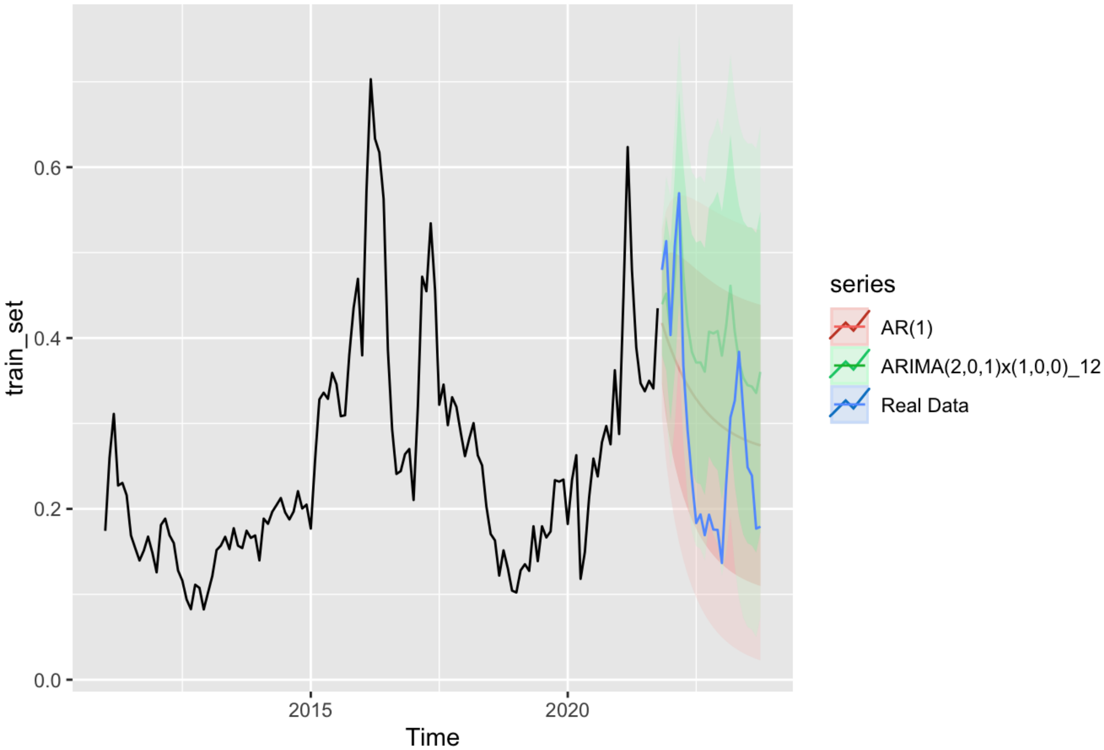

Figure 15. Forecasting 12 Months (Train/Valid)

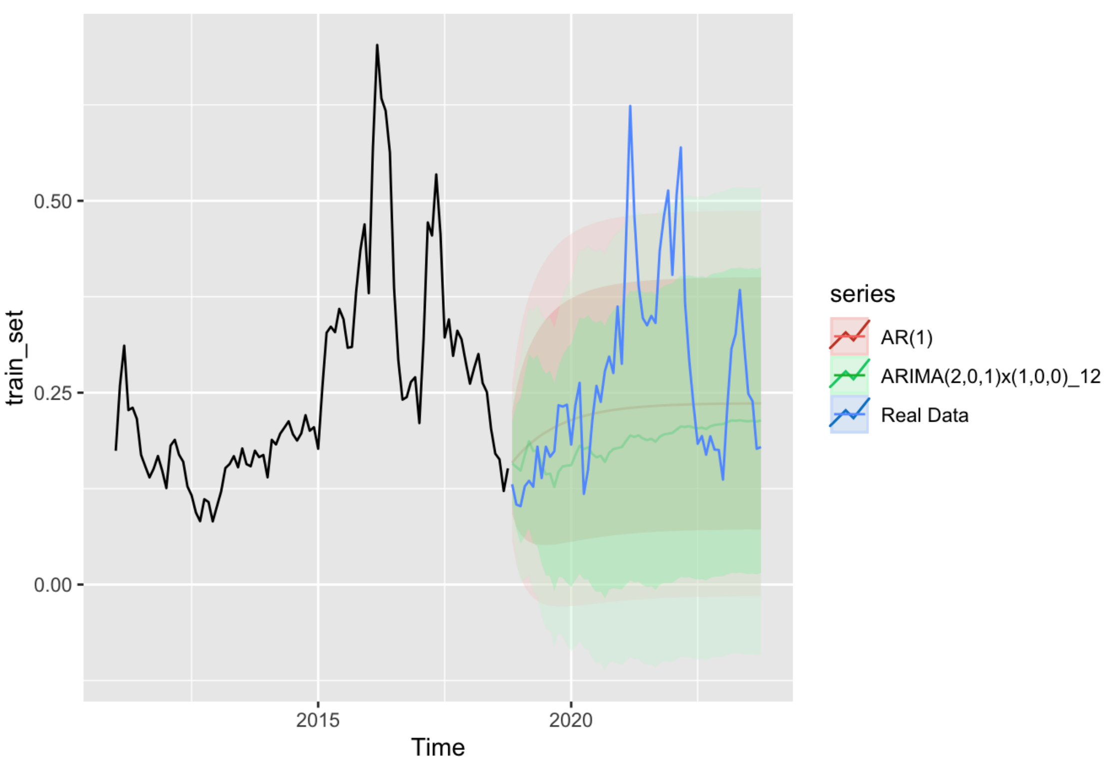

Figure 16. Forecasting 60 Months (Train/Valid)

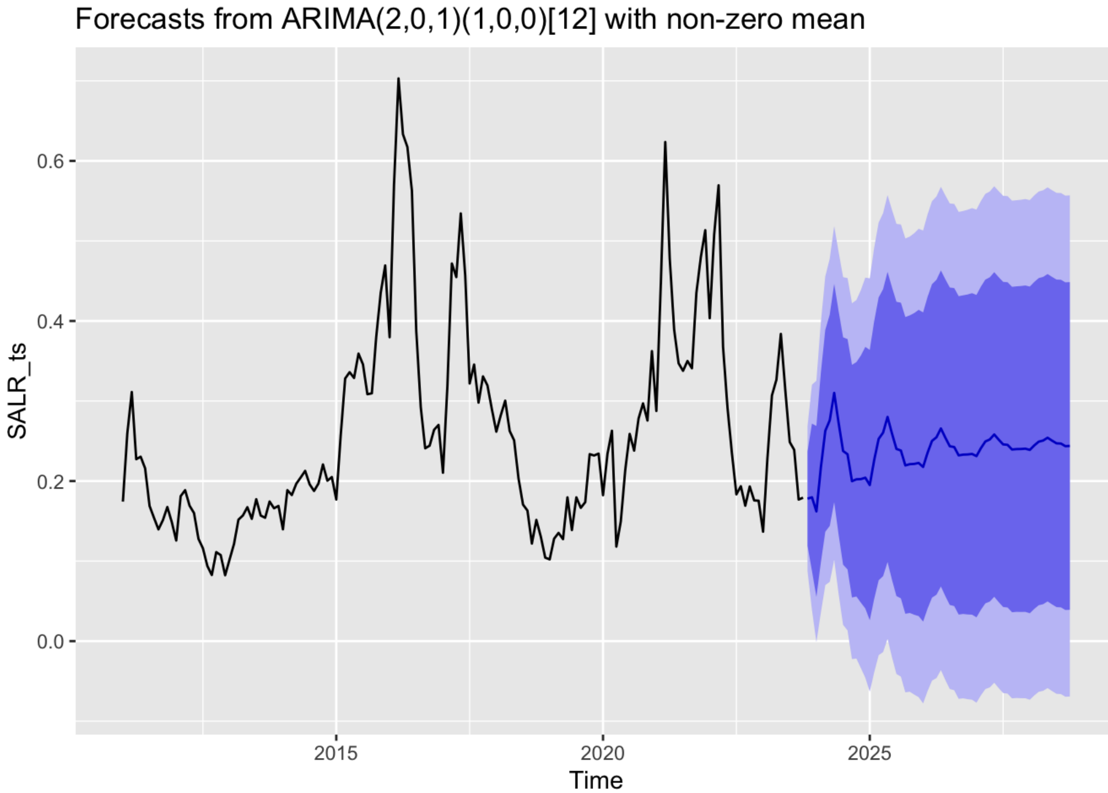

Figure 17. Forecasting Future 5 Years