DNSC 6319
SPRING 2024

TIME SERIES PROJECT REPORT

GROUP 3

Team Members

Abhijit Kudtarkar
Akshay Reddy
Manuel Chavez
Shubham Patil
Zhenqi Zheng

1. INTRODUCTION

The Time Series data we have is the US Auto sales data dating from January 1995 to December 2020. The data has 5 variables namely Month, Autosales, Inflation, Unemploy, PPI, and Gas Prices. In all our analysis the independent variable is Autosales and we used a training sample of 288 observations and a holdout sample of 24 months to test the data.

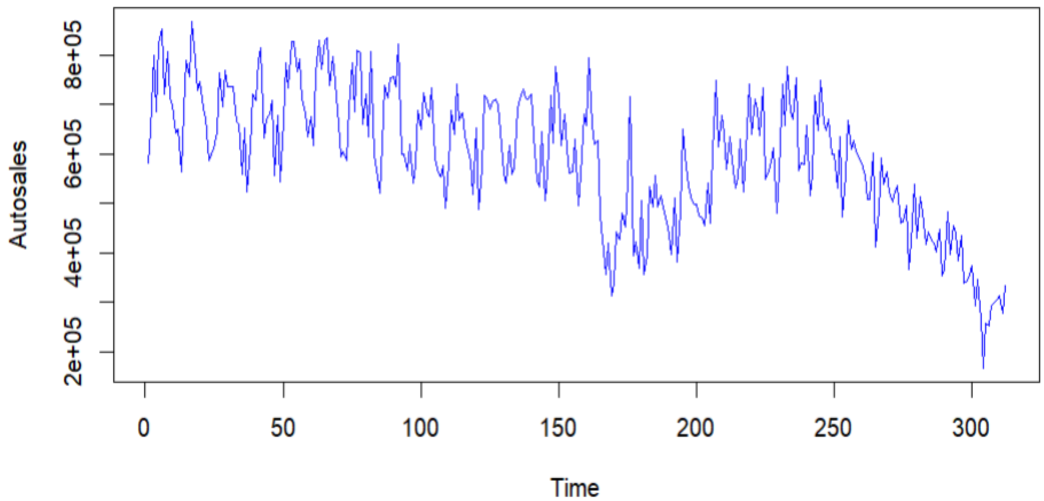


Figure 1.1. Time Series Plot Auto Sales

The time series plot in Figure 1.1, seems to have a switching trend, since we have changes in the trend at certain points. The series can be described as in three pieces, the first piece showing a downward trend, second piece an upward trend, and the third piece a downward trend.

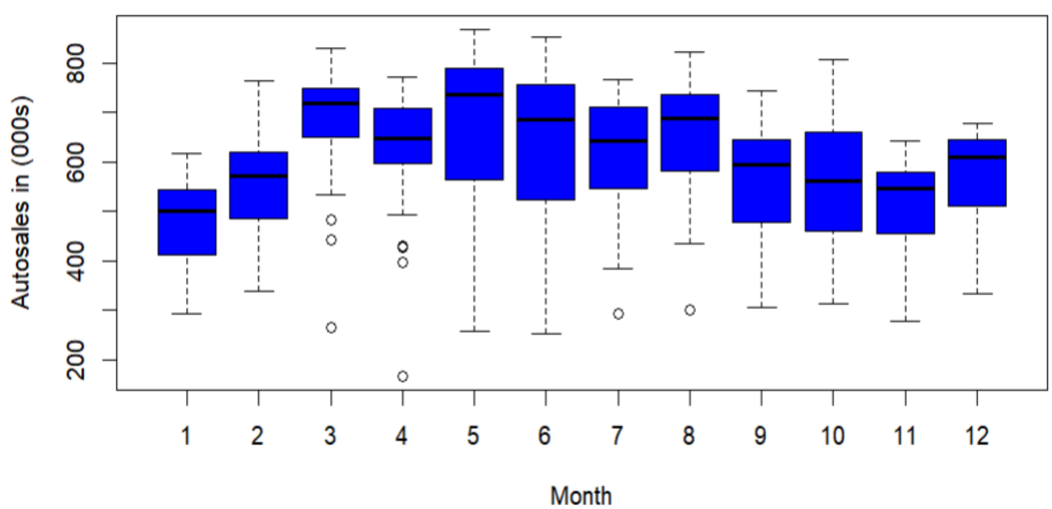


Figure 1.2. Box Plot Auto Sales

From the Box plot of the series in Fig 1.2, There is clear variation in median sales between months, which is indicative of seasonality. Some months show a higher median, suggesting periods of higher demand or sales. The circles represent outliers that are significantly higher or lower than the rest of the data for that month. These could be due to important events (holidays). The length of the boxes and whiskers shows the spread or variability in the sales data for each month. Some months have a wider range, indicating more variability in sales within that month.

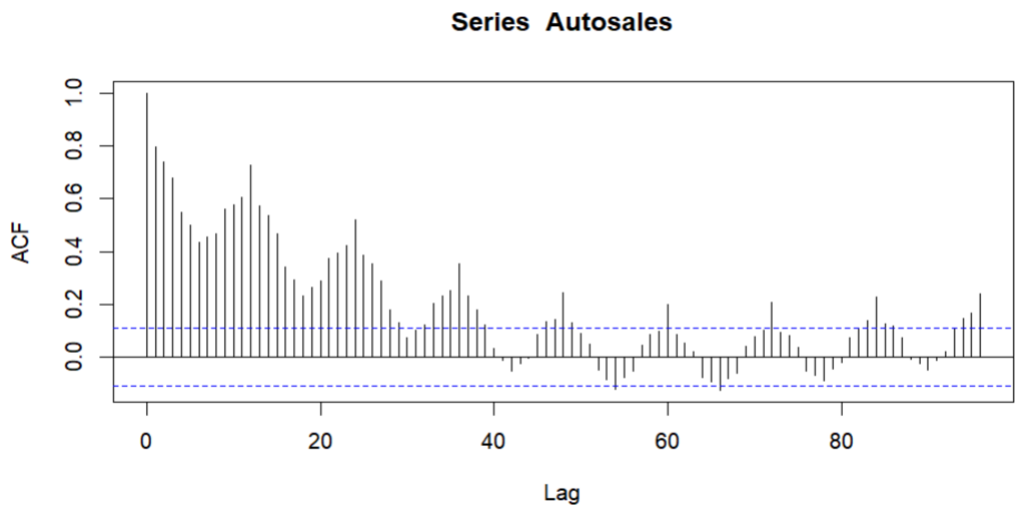
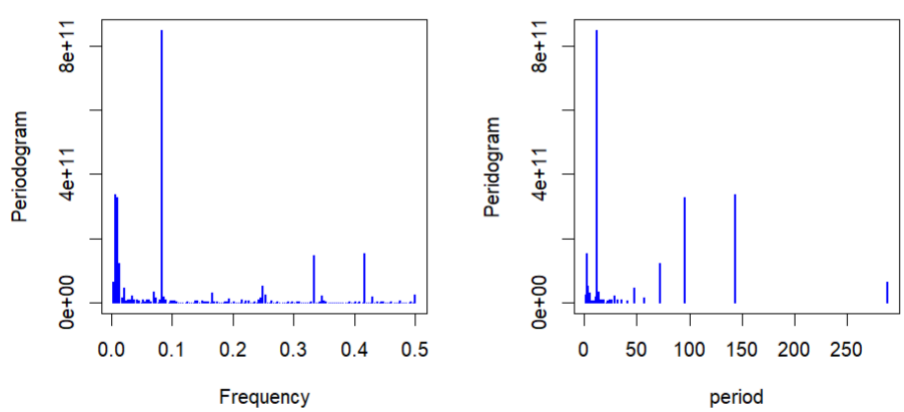


Figure 1.3. Sample Autocorrelation of Autosales

The ACF plot of series in Fig 1.3, The Autocorrelation values start high and gradually decay as the lags increase, which often indicates a trend in the data. In this context, it suggests that recent sales figures are more closely related to the current ones, which is typical in non-stationary data that contain a trend. The spikes in the ACF at regular intervals suggest a seasonal pattern. Spikes at lags of approximately 12, 24, 36 (and so on) suggest yearly seasonality, which also suggests a non-stationary behavior at these seasonal lags. The ACF values that extend beyond the blue dashed confidence bounds are statistically significant. The first lag is particularly high, which could be indicative of a strong correlation from one period to the next.



```
       Harmonic     period   frequency     amplitude
 [1,]        24  12.000000 0.083333333 848569755433
 [2,]         2 144.000000 0.006944444 335570459830
 [3,]         3  96.000000 0.010416667 328249653234
 [4,]       120   2.400000 0.416666667 153263748438
 [5,]        96   3.000000 0.333333333 147492623395
 [6,]         4  72.000000 0.013888889 121442449879
 [7,]         1 288.000000 0.003472222  65911571742
 [8,]        72   4.000000 0.250000000  51275837905
 [9,]         6  48.000000 0.020833333  46489221784
[10,]        20  14.400000 0.069444444  32864642453
[11,]        48   6.000000 0.166666667  30624468420
```

Figure 1.4

Here from the above output we can see, the first 2 dominant periods i.e periods (12 and 144) with high amplitude represent the frequencies at which significant variations or patterns occur in our auto sales data. The high amplitude value indicates that the fluctuations in auto sales data at this particular frequency (0.0834 cycles per month) are substantial. It suggests that the variations in auto sales during this 12-month period are significant and may have a considerable impact on overall sales performance.The periods range from as short as 2 months to as long as 288 months, indicating the presence of multiple influencing cycles within the autosales data. This periodogram is necessary in the context of auto sales analysis to uncover underlying patterns, trends, and seasonal effects, facilitating better understanding and decision-making within the automotive sector.

2. UNIVARIATE TIME SERIES MODELS

2.1 Deterministic Time Series Models

Seasonal Dummies and Trend

```
lm(formula = n_Autosales ~ time + as.factor(n_MONTH))

Residuals:
    Min      1Q  Median      3Q     Max 
-233706  -31446    5390   50371  185847 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          599702.05   16191.64  37.038  < 2e-16 ***
time                   -745.76      50.67 -14.719  < 2e-16 ***
as.factor(n_MONTH)2   77774.92   20618.87   3.772 0.000198 ***
as.factor(n_MONTH)3  203529.02   20619.05   9.871  < 2e-16 ***
as.factor(n_MONTH)4  147341.44   20619.36   7.146 8.05e-12 ***
as.factor(n_MONTH)5  219483.03   20619.80  10.644  < 2e-16 ***
as.factor(n_MONTH)6  186987.12   20620.36   9.068  < 2e-16 ***
as.factor(n_MONTH)7  146862.05   20621.04   7.122 9.32e-12 ***
as.factor(n_MONTH)8  184999.47   20621.85   8.971  < 2e-16 ***
as.factor(n_MONTH)9   99757.73   20622.79   4.837 2.20e-06 ***
as.factor(n_MONTH)10  82503.49   20623.85   4.000 8.13e-05 ***
as.factor(n_MONTH)11  38024.25   20625.03   1.844 0.066318 .  
as.factor(n_MONTH)12 104528.34   20626.33   5.068 7.40e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 71430 on 275 degrees of freedom
Multiple R-squared:  0.627,	Adjusted R-squared:  0.6107 
F-statistic: 38.52 on 12 and 275 DF,  p-value: < 2.2e-16
```

Figure 2.1.1 LM for Seasonal and Trend

By looking at the *Figure 2.1.1*, with seasonal dummies and the time component p-values, we can conclude that they are significant ,in addition,the R-squared , suggest that the model explains *62,70%* of the auto sales.

In the model we are using 24 year (288 observations) for the training and also a hold-out sample of 2 years (24 observations)

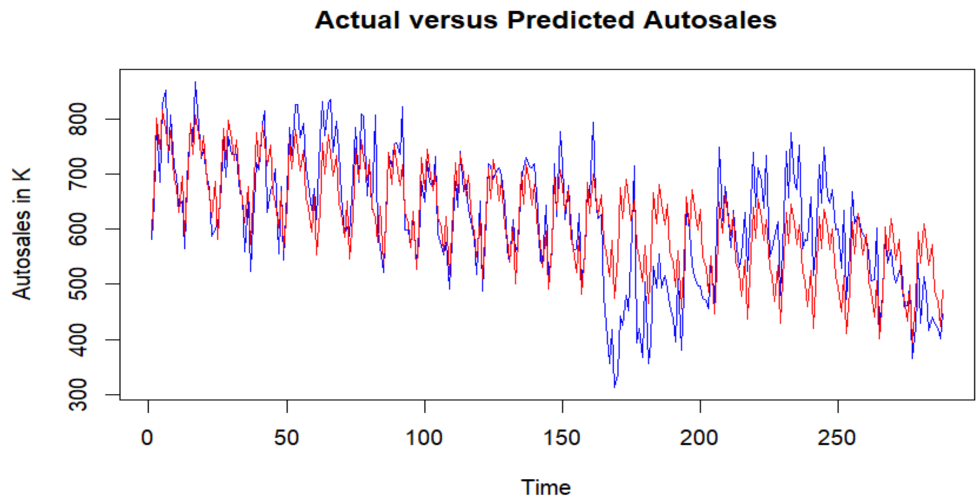


Figure 2.1.2. Actual vs Predicted Auto Sales

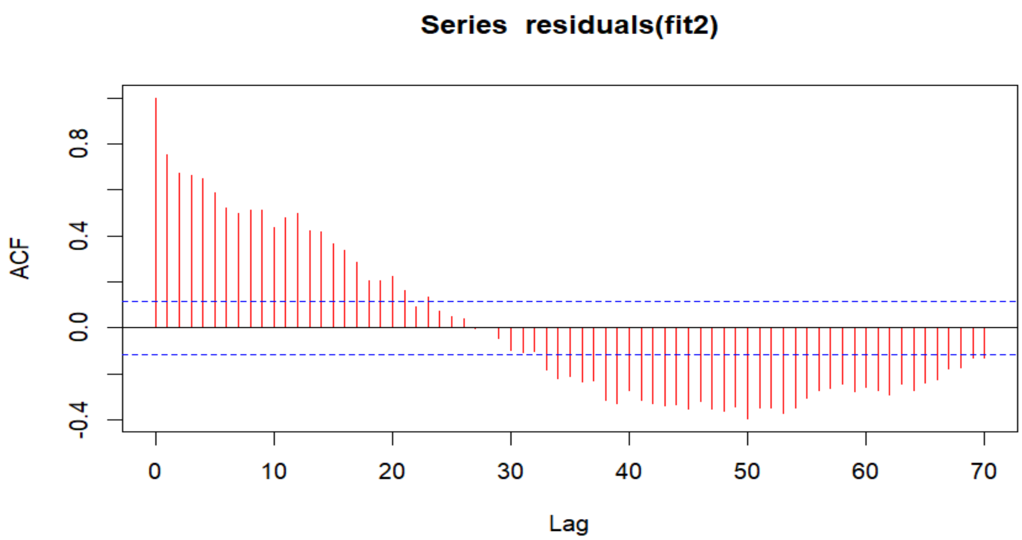


Figure 2.1.3. Actual vs Predicted Auto Sales

By looking at the *figure 2.1.2*, we can see that the model is not capturing the Auto sales well. Moreover, by looking at the residuals at *figure 2.1.3,* we can see that the residuals are not white noise, which suggest that still exists patterns or structures in the residuals that the model has not been able to capture.

MAPE (Training Sample)	MAPE (Test Sample)
9.55%	55.30%

Figure 2.1.4. MAPE for Seasonal Dummies and Trend

By looking at the MAPE's at the *Figure 2.1.4 ,* we can conclude that the model is performing well in the training set (by achieving a 9.55% as MAPE), but the model is not predicting properly in the test data (by achieving a 55.30% as MAPE). It can suggest that the model is overfitting to the train data.

Seasonal Dummies , Trend and Interaction

```
Call:
lm(formula = n_Autosales ~ time + as.factor(n_MONTH) + dummy1 *
    time + dummy2 * time)

Residuals:
    Min      1Q  Median      3Q     Max
-163347  -30274   -1574   27009  209554
```

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          612510.33   11855.01  51.667  < 2e-16 ***
time                   -750.82      73.42 -10.227  < 2e-16 ***
as.factor(n_MONTH)2   73088.15   13725.54   5.325 2.12e-07 ***
as.factor(n_MONTH)3  198425.55   13722.33  14.460  < 2e-16 ***
as.factor(n_MONTH)4  141821.30   13720.12  10.337  < 2e-16 ***
as.factor(n_MONTH)5  213546.20   13718.93  15.566  < 2e-16 ***
as.factor(n_MONTH)6  180633.61   13718.74  13.167  < 2e-16 ***
as.factor(n_MONTH)7  140091.85   13719.57  10.211  < 2e-16 ***
as.factor(n_MONTH)8  177812.60   13721.40  12.959  < 2e-16 ***
as.factor(n_MONTH)9   92154.17   13724.23   6.715 1.10e-10 ***
as.factor(n_MONTH)10  74483.25   13728.07   5.426 1.28e-07 ***
as.factor(n_MONTH)11  29587.32   13732.92   2.154   0.0321 *
as.factor(n_MONTH)12  95674.73   13738.78   6.964 2.50e-11 ***
dummy1              -930310.91   55996.00 -16.614  < 2e-16 ***
dummy2              1090998.64  136241.17   8.008 3.46e-14 ***
time:dummy1            4421.47     279.83  15.800  < 2e-16 ***
time:dummy2           -4101.73     517.74  -7.922 6.06e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 47490 on 271 degrees of freedom
Multiple R-squared:  0.8375,    Adjusted R-squared:  0.8279
F-statistic:  87.3 on 16 and 271 DF,  p-value: < 2.2e-16
```

Figure 2.1.5. LM for Seasonal Dummies ,Trend and Interactions

By looking at Figure *2.1.5*, with seasonal dummies, trend and interactions p-values, we can conclude that they are significant ,in addition,the R-squared , suggest that the model explains *87,3%* of the auto sales.

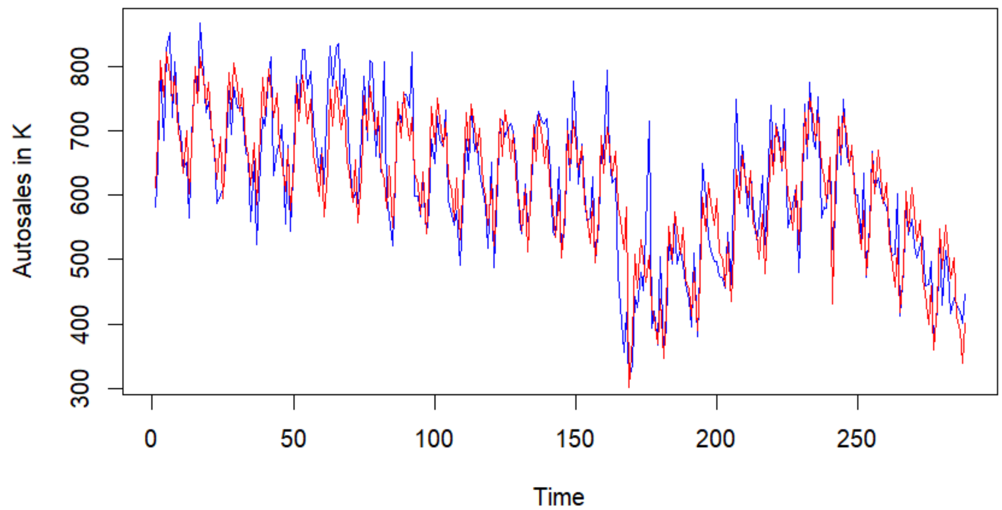


Figure 2.1.6. Actual vs Predicted Auto Sales

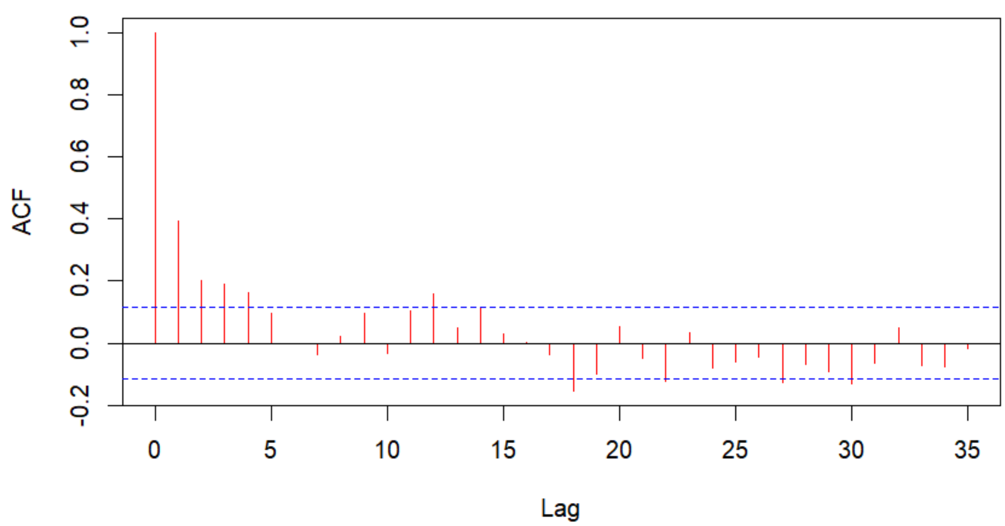


Figure 2.1.7. Actual vs Predicted Auto Sales

By looking at the *figure 2.1.6*, we can see that the model is capturing the Auto sales much better , since we use interactions with dummy variables associated with time, by using the piecewise trend. By looking at the residuals at *figure 2.1.7,* we can see that the residuals are not white noise, which suggest that still exists patterns or structures in the residuals that the model has not been able to capture.

Cyclical Model

```
Call:
lm(formula = n_Autosales ~ time1 + cos2 + sin2 + sin3 + cos3 + 
    sin4 + cos4 + sin24 + cos24 + sin96 + cos96 + cos120 + sin120)

Residuals:
    Min      1Q  Median      3Q     Max 
-145190  -32893    -762   31271  184262 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 740508.1     6676.4 110.914  < 2e-16 ***
time1         -859.9       41.4 -20.768  < 2e-16 ***
cos2        -25452.5     4191.0  -6.073 4.17e-09 ***
sin2        -46058.5     4600.4 -10.012  < 2e-16 ***
sin3          8292.8     4377.5   1.894  0.05922 .  
cos3        -46177.4     4191.0 -11.018  < 2e-16 ***
sin4          8484.3     4296.8   1.975  0.04932 *  
cos4         26972.6     4191.0   6.436 5.47e-10 ***
sin24        13634.4     4193.7   3.251  0.00129 ** 
cos24       -75356.6     4191.0 -17.980  < 2e-16 ***
sin96       -20789.1     4190.9  -4.961 1.24e-06 ***
cos96        24498.8     4191.0   5.846 1.43e-08 ***
cos120       19421.5     4191.0   4.634 5.55e-06 ***
sin120       26265.7     4190.8   6.267 1.42e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 50290 on 274 degrees of freedom
Multiple R-squared:  0.8157,	Adjusted R-squared:  0.807 
F-statistic: 93.31 on 13 and 274 DF,  p-value: < 2.2e-16
```

Figure 2.1.8 Cyclical Model for Autosales series

From Figure 2.1.8, we estimated the cyclical model for the Autosales series using the dominant periods. By looking at the R-squared value of the model of 0.8157, it explains about 82% of the variance in the series. Hence, we can conclude that the cyclical model captures a significant portion of the seasonal patterns inherent in the data. The highly significant p-value (less than 0.05), strongly supports the model's fit. The coefficients for the sinusoidal and cosinusoidal terms at various frequencies—particularly at annual cycles (sine and cosine terms for 24, 96, and 120 months)—are statistically significant, indicating their crucial role in modeling seasonal fluctuations in autosales. This model provides a robust framework for understanding and predicting cyclical variations in the autosales data, suggesting its utility for strategic planning and decision-making in related business sectors.

	Harmonic	period	frequency	amplitude
[1,]	24	12.000000	0.083333333	848569755433
[2,]	2	144.000000	0.006944444	335570459830
[3,]	3	96.000000	0.010416667	328249653234
[4,]	120	2.400000	0.416666667	153263748438
[5,]	96	3.000000	0.333333333	147492623395
[6,]	4	72.000000	0.013888889	121442449879
[7,]	1	288.000000	0.003472222	65911571742
[8,]	72	4.000000	0.250000000	51275837905
[9,]	6	48.000000	0.020833333	46489221784
[10,]	20	14.400000	0.069444444	32864642453
[11,]	48	6.000000	0.166666667	30624468420

Figure 2.1.9

Figure 2.1.9 presents a detailed harmonic analysis of a time series, where various cycles are identified by their periods, frequencies, and amplitudes. The periods range from as short as 4 months to as long as 288 months, indicating the presence of multiple influencing cycles within the autosales data. For example, the 24-month cycle with a very high amplitude suggests a significant biennial influence, potentially linked to broader economic or industry-specific factors.

Similarly, the substantial amplitudes at 96 and 288 months point to longer-term cycles that could reflect more substantial market or economic shifts. The analysis also captures shorter cycles, such as those spanning 4 and 6 months, likely reflecting quarterly financial cycles and semi-annual trends. This comprehensive breakdown helps in understanding the various cyclic patterns affecting Autosales, providing essential insights for strategic planning and forecasting in the industry.

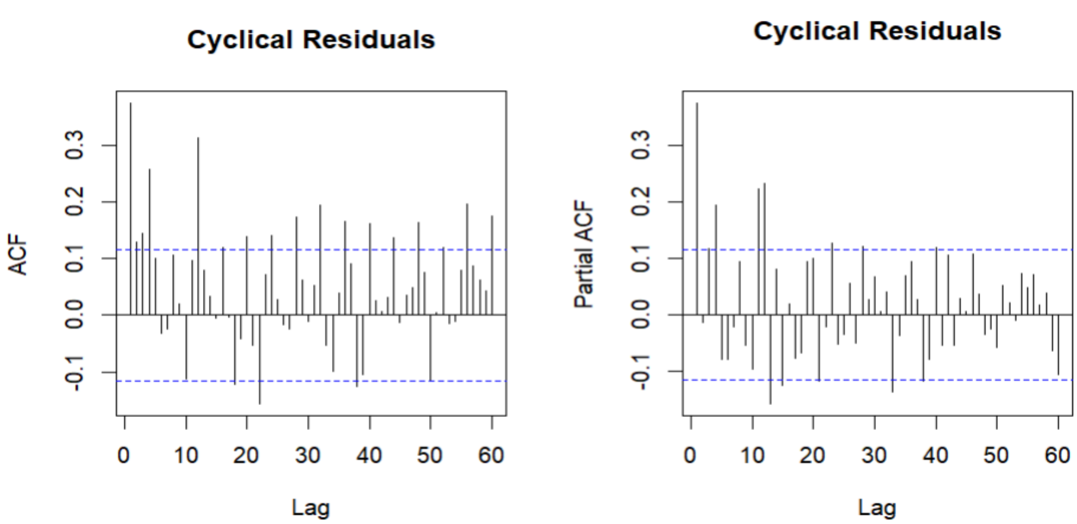


Figure 2.1.10

Figure 2.1.10 shows the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots for the residuals of a cyclical model. These plots are used to assess the adequacy of the model in capturing the cyclical dependencies within the time series data. From the ACF plot we can see that some correlation still exists and a possible trend. Hence, we can conclude that the residuals are not stationary / white noise. Therefore, concluding the cyclical model does not do a great job in capturing the series.

MAPE (Training Sample)	MAPE (Test Sample)
29.08%	34.12%

2.2.Comparison of candidate models

MODEL	MAPE Training	MAPE Test	Model Fit
Seasonal Dummies +Trend	9.55%	55.30%	62.7%
Piecewise LM	5.83%		83.75%
Cyclical	29.08%	34.12%	81.57%

3. Time Series Regression Models

In this section we are going to model the time series using a multiple regression model. Our dependent variable is Autosales and the predictor variables are Inflation, Unemployment, PPI, and Gas prices.

To better understand the variables we are going to use some visualizations like the overlay time series plot to see how they covary over time compared to the dependent variable.

3.1 Discussion of independent variables.

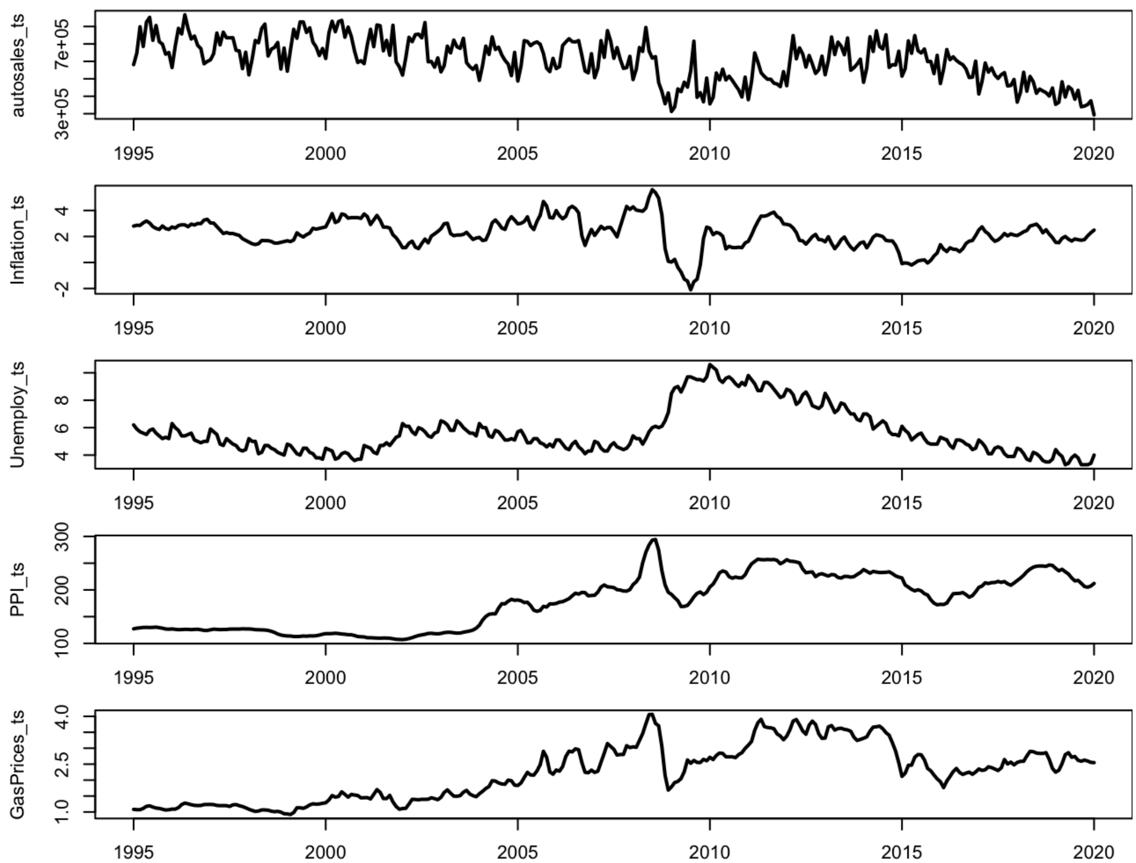


Fig 3.1

Auto Sales: Auto sales follows a seasonal pattern with a downward trend with a sudden dip in the year 2008 during the global financial crisis, which gradually recovered by 2015 followed by another downward trend.

Inflation: Inflation has no seasonal pattern and appears volatile. Inflation suddenly increased in the year 2008 during the global financial crisis but suddenly decreased in the year 2009.

Unemployment: Unemployment follows a seasonal pattern, increasing and decreasing overtime

Producer Price Index for Iron and Steel: Producer price index for iron and steel has an increasing trend with some fluctuations.

Gas Price: Gas Price has an increasing trend with some fluctuations.

3.2 Correlation Matrix

```
> cor(newdata)
           Autosales   Inflation   Unemploy         PPI   GasPrices
Autosales  1.0000000  0.28551283 -0.3326053 -0.50440364 -0.29808729
Inflation  0.2855128  1.00000000 -0.3189847 -0.04095798  0.05473201
Unemploy  -0.3326053 -0.31898472  1.0000000  0.35156671  0.36330844
PPI       -0.5044036 -0.04095798  0.3515667  1.00000000  0.91703510
GasPrices -0.2980873  0.05473201  0.3633084  0.91703510  1.00000000
```

Fig. 3.2.1

Here are some potential insights from correlation matrix (fig 3.2.1)

- **Auto Sales Vs Inflation:** There appears to be weak positive correlation between auto sales and inflations. The correlation value is 0.2855 that means for as inflation increases, auto sales also increases but slightly.
- **Auto Sales vs Unemployment:** There appears to be a weak negative correlation between Auto Sales and unemployment. The correlation value is -0.3326 that means as unemployment increases, auto sales decreases but slightly.
- **Auto Sales vs producer price index for iron and steel:** There appears to be a moderate negative correlation between auto sales and producer price index for iron and steel. The correlation value is -0.5044 that means as the producer price index for iron and steel increases, auto sales decreases.
- **Auto Sales vs Gas Prices:** There appears to be a weak negative correlation between auto sales and Gas Prices. The correlation value is -0.2980 that means as the gas prices increases, auto sales decreases but slightly.

3.3 Multiple Linear Regression: PPI for Iron and Steel, Unemployment, Gas Prices and Inflation

A Multiple linear regression model has been built with Auto Sales as the dependent variable and here, the independent variables include PPI for Iron and Steel, Unemployment, Gas Prices, and Inflation.There are 312 observations, of which 288 observations are used to train the mode, and the remaining 24 observations are used as hold out sample.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 932211.9    31423.4  29.666  < 2e-16 ***
n_PPI        -2922.9      286.6 -10.199  < 2e-16 ***
n_Unemploy  -16480.8     3727.7  -4.421 1.40e-05 ***
n_GasPrices 124956.4    16966.8   7.365 1.95e-12 ***
n_Inflation   8643.2     4862.8   1.777   0.0766 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 87800 on 283 degrees of freedom
Multiple R-squared:  0.4199,    Adjusted R-squared:  0.4117
F-statistic: 51.21 on 4 and 283 DF,  p-value: < 2.2e-16
```

Fig 3.3.1

Based on fig 3.3.1, it is evident that the p value is <2.2e-16 which is less than 0.05 therefore is statistically significant. On the other hand, all independent variables except inflation are significant as their P value is less than 0.05. The R squared value is 0.4199 that means the model is a weak predictor. The model covers only 41.99% of the variability of dependent variables.

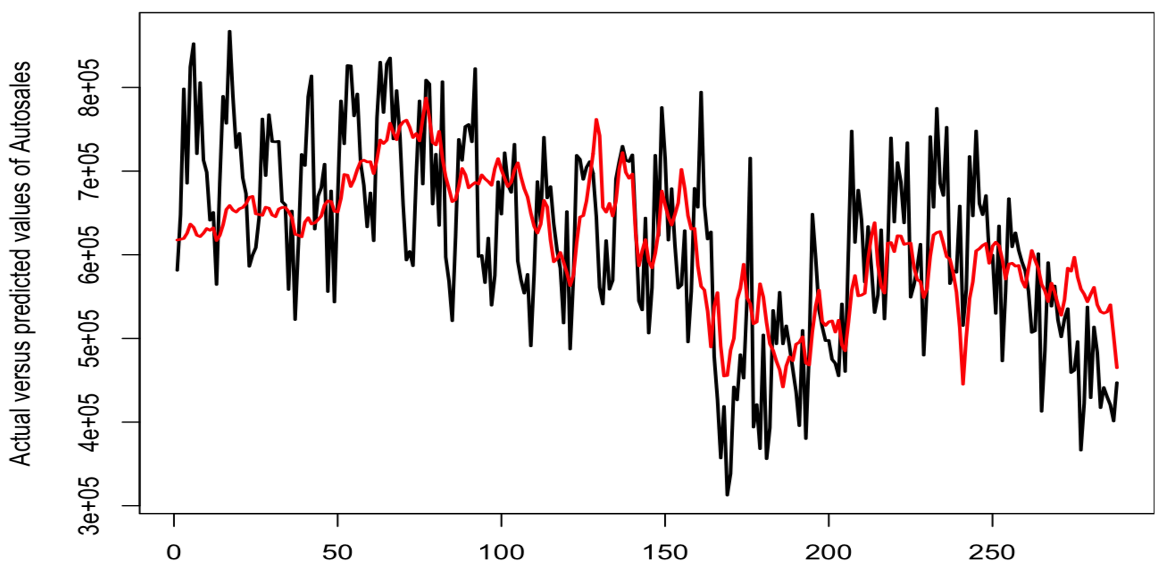


Fig 3.3.2

In Fig 3.3.2, it is clear that, model is not doing a perfect job in predicting the Auto Sale.Although it captures the general trend somewhat accurately, it struggles to predict the seasonal ups and downs accurately.

3.3.1 Mean Absolute Percentage Error (MAPE)

For Training Data: The mean absolute percentage error (MAPE) for this model is 12.13% which implies that, on average, the predictions are 12.13% off from the actual values.

For Hold Out Sample: The mean absolute percentage error (MAPE) for this model is 56.78% which implies that, on average, the predictions are 56.78% off from the actual values.

MAPE (Training Sample)	MAPE (Test Sample)
12.13%	56.78%

By looking at the MAPE's in the above table , we can conclude that the model is performing well in the training set (by achieving a 12.13% as MAPE), but the model is not predicting properly in the test data (by achieving a 56..78% as MAPE). It can suggest that the model is overfitting to the train data.

Fig 3.3.3

From Fig 3.3.3, the autocorrelation function (ACF) of residuals from this model falls outside the standard error bounds, indicating that the residuals are not a white noise series. Additionally, from the ACF graph, it is evident that the residuals are not stationary, as the ACF gradually decreases when observing ACF values at lag 12,24 and 36. Furthermore, there are seasonal components present in the residuals.

The variance inflation factor for PPI, unemployment, gas prices and inflation are calculated to be 7.7467, 1.4740, 8.3086 and 1.1988 respectively . Since these values are below 10, there is no indication of multicollinearity in this model.

4.1 - Residuals of Deterministic Time Series Models

The Fig 4.1.1 below shows the ACF and PACF of the residuals from the Deterministic time series model with Trend and Seasonality.

From the ACF plot, we could say that the series is stationary, and the Autocorrelation decays after lag 1. Similarly, from the PACF plot the Autocorrelation seems to cut off after lag1. This suggests that an AR(1) process might be a good fit for this series. So we are going to fit an AR(1) process using the switching trend and seasonal dummies as the regressors.

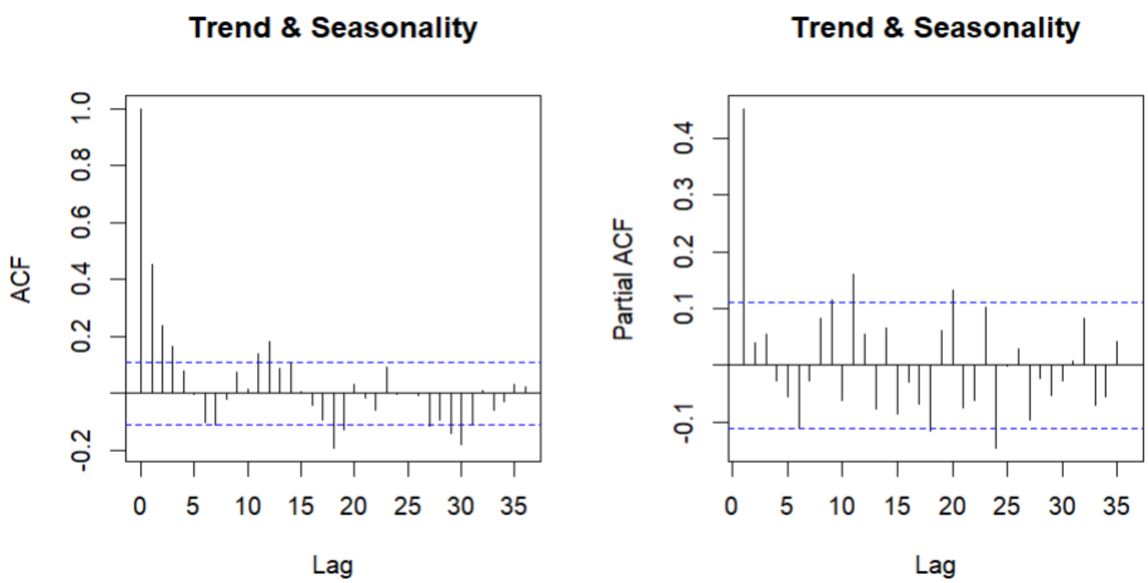


Fig 4.1.1

```
Series: n_Autosales
Regression with ARIMA(1,0,0) errors

Coefficients:
         ar1  intercept        time
      0.4062  608889.93   -708.4537   73469.05   198621.93   141940.35   213630.37
s.e.  0.0564   13616.64    103.0306   10330.56    12220.51    12904.01    13169.24

      180689.18  140119.84  177813.37   92115.36   74371.11   29341.62   95149.31
s.e.   13268.01   13294.57   13271.67   13178.34   12926.66   12269.17   10444.81
         dummy1     dummy2       int1        int2
     -863083.70   981200.6  4075.5821  -3724.1548
s.e.   82762.84   196355.0   411.7291    749.0085

sigma^2 = 1.906e+09:  log likelihood = -3476.45
AIC=6990.9   AICc=6993.74   BIC=7060.5

Training set error measures:
                   ME     RMSE      MAE        MPE     MAPE      MASE        ACF1
Training set 44.12685 42266.88 32548.89 -0.5006706 5.548546 0.4725272 -0.02576027
```

Fig 4.1.2

Fig 4.1.2, is the result after applying an AR(1) process. We can conclude that all the coefficients are statistically significant based on the t-statistic. The MAPE of 5.54% denotes that the model predictions are 5.54% off from the actual observation in the training sample.

From the Fig 4.1.3 below, we can see that the predictions from this model do a decent job of predicting the actual observations. But, we can still see that this is not best fit since there is a visible error in the predictions at some intervals.

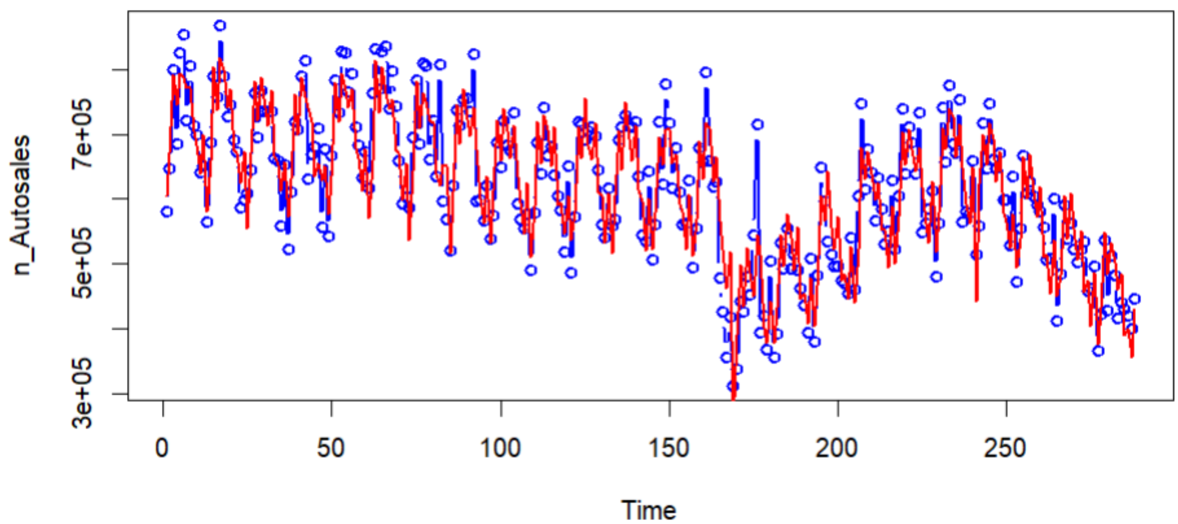


Fig 4.1.3

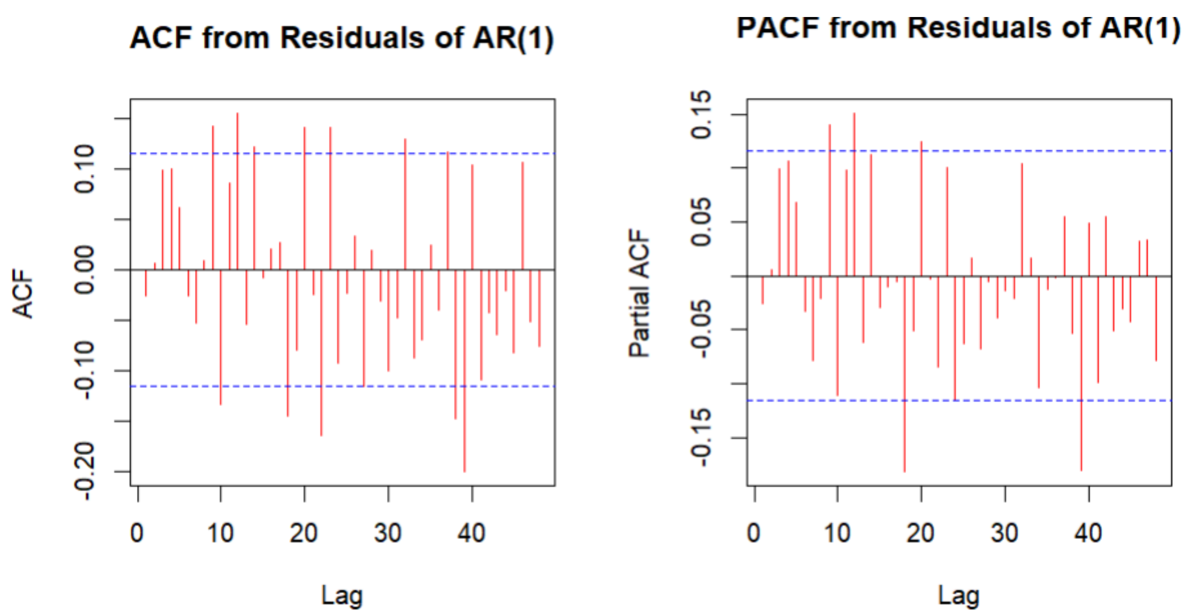


Fig - 4.1.4

The ACF and PACF plots from Fig 4.1.4, show some correlation between them. The residuals do not exhibit the properties of a white noise series. This could be because of the wrong assumption that the series was stationary and fitting an AR(1) process to it.

4.2 - Residuals of Time Series Regression Models

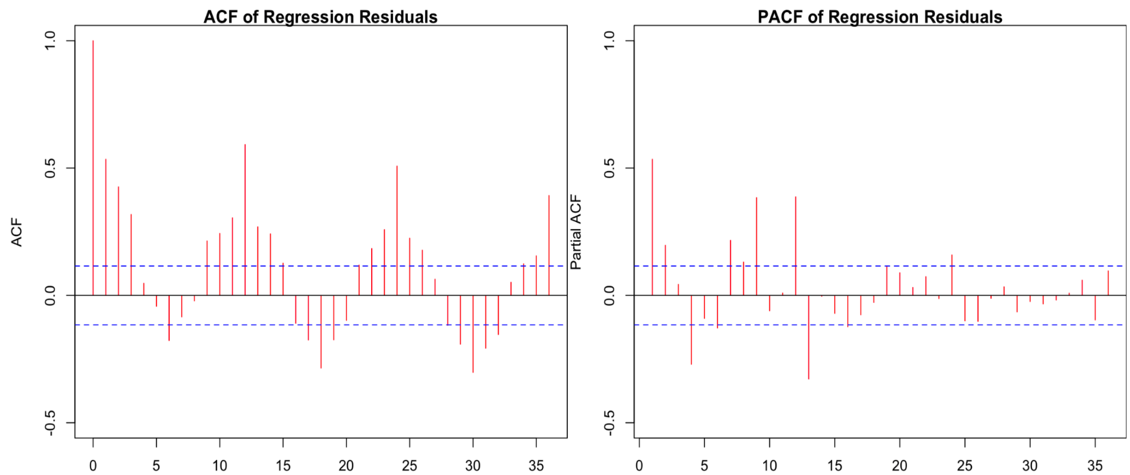


Fig 4.2.1

The analysis of the residuals from the regression model in Section 3.3 revealed some important findings. First, the autocorrelation function (ACF) of residuals falls outside the standard error bounds, indicating that the residuals are not a white noise series. Upon observing seasonal lags at 12, 24 and 36, a gradual decrease in th ACF of the regression residuals are noted, which is mirrored in the non-seasonal lags. This pattern signifies that the residuals exhibit non-stationarity. To make it non-stationary, both non-seasonal and seasonal differencing techniques will be applied to the residuals to make it non-stationary.

In the case of partial autocorrelation function (PACF) of residuals, the PACF is quite evident at lag 12 after which it gets chopped off.

4.2.1 Non-Seasonal and Seasonal Differencing of Regression Residuals:

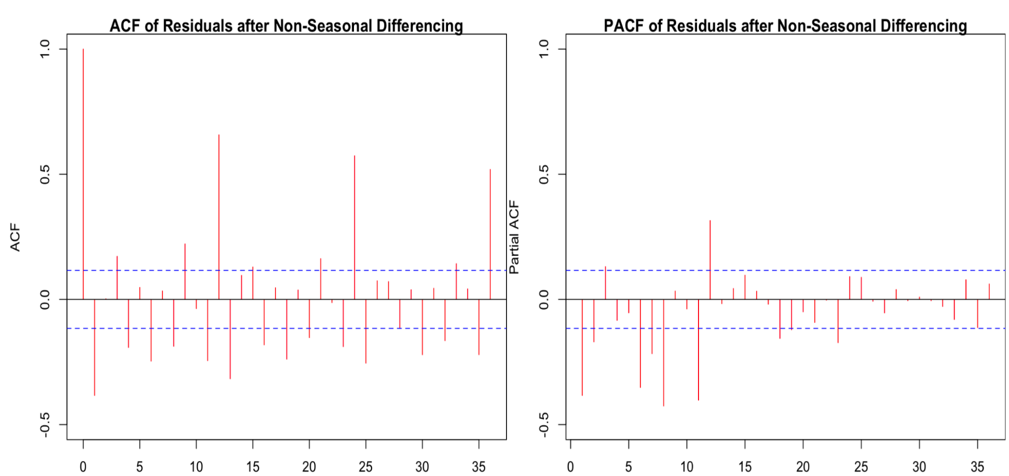


Fig 4.2.1.1

After applying non-seasonal differencing, it is observed from the autocorrelation function (ACF) graph that the lags at 12,24 and 36 are decaying gradually. Consequently, the regression residual is still non-stationary. In order to make it non-stationary, non-seasonal differencing is performed in the subsequent step of the analysis.

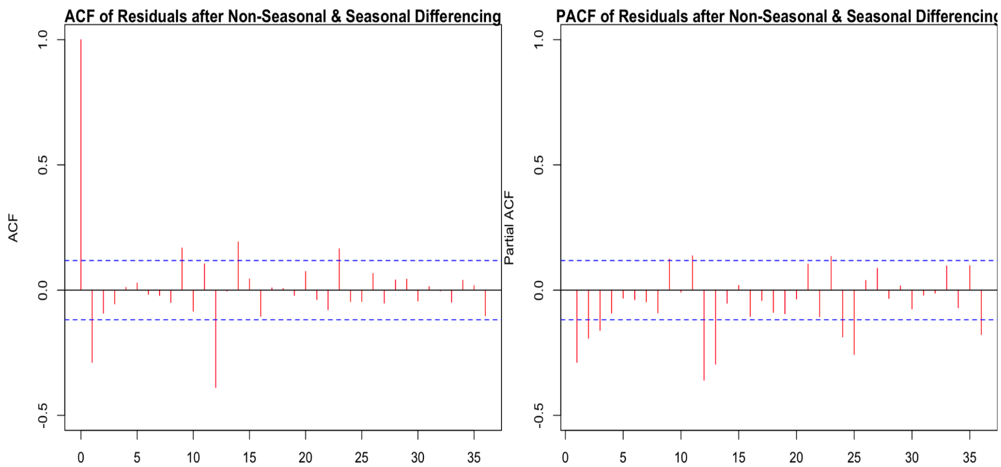


Fig 4.2.1.2

After the non-seasonal and seasonal differencing of the regression residuals, below are the observations from the ACF and PACF graphs:

- **Non-Seasonal Lags:** The ACF for non-seasonal lags that is lags other than 12, 24 and 36 gets chopped off after lag 1. Meanwhile the PACF graph shows exponential decline in the PACF value for non-seasonal lags. The non-seasonal lags after lag 1 are slightly greater than the absolute values of 2 standard errors: hence, they are considered as not significant.
- **Seasonal Lags:** The ACF for seasonal lags that is lags at 12, 24 and 36 gets chopped off at lag 12. Meanwhile the PACF graph shows exponential decline in the PACF value for seasonal lags.

Based on the preceding analysis, it is evident that the regression residuals follow an integrated moving average of order 1 process for both seasonal and non-seasonal lags. Thus modeling of regression residuals using MA(1) process for both seasonal and non-seasonal components is employed in section 4.2.2.

4.2.2 Seasonal and Non Seasonal MA(1) Differencing in Multiple Linear Regression

Fitting an MA(1) process for the seasonal and non-seasonal components using Inflation, Unemploy, PPI, and Gas prices as regressors.

```
Series: n_Autosales
Regression with ARIMA(0,1,1)(0,1,1)[12] errors

Coefficients:
          ma1     sma1  n_Inflation  n_Unemploy      n_PPI  n_GasPrices
      -0.7168  -0.7649   -11031.139  -23651.826  -373.5603     89376.93
s.e.   0.0450   0.0447     5043.253    8224.546   287.4926     17354.49

sigma^2 = 1.894e+09:  log likelihood = -3330.1
AIC=6674.19   AICc=6674.61   BIC=6699.51

Training set error measures:
                  ME     RMSE      MAE        MPE     MAPE      MASE       ACF1
Training set -1293.3 42060.17 31487.28 -0.4512548 5.310825 0.4571153 0.04935694
> phi=regdif$coef
> se_phi=rep(0,6)
> se_phi[1]=(regdif$var.coef[1,1])^0.5
> se_phi[2]=(regdif$var.coef[2,2])^0.5
> se_phi[3]=(regdif$var.coef[3,3])^0.5
> se_phi[4]=(regdif$var.coef[4,4])^0.5
> se_phi[5]=(regdif$var.coef[5,5])^0.5
> se_phi[6]=(regdif$var.coef[6,6])^0.5
>
> t_stat=abs(phi)/se_phi
> t_stat
      ma1        sma1 n_Inflation  n_Unemploy       n_PPI n_GasPrices
 15.931487   17.094297    2.187306    2.875761    1.299374    5.150076
```

Fig 4.2.1.1

From Fig 4.2.1.1, all the coefficients of the model except PPI for Iron and Steel are greater than 1.96, indicating they are statistically significant and their coefficients are different than zero. Conversely, for Inflation and PPI for Iron and Steel, their t-statistics are less than 1.96, indicating they are statistically insignificant, and we reject the null hypothesis that their coefficients are zero.

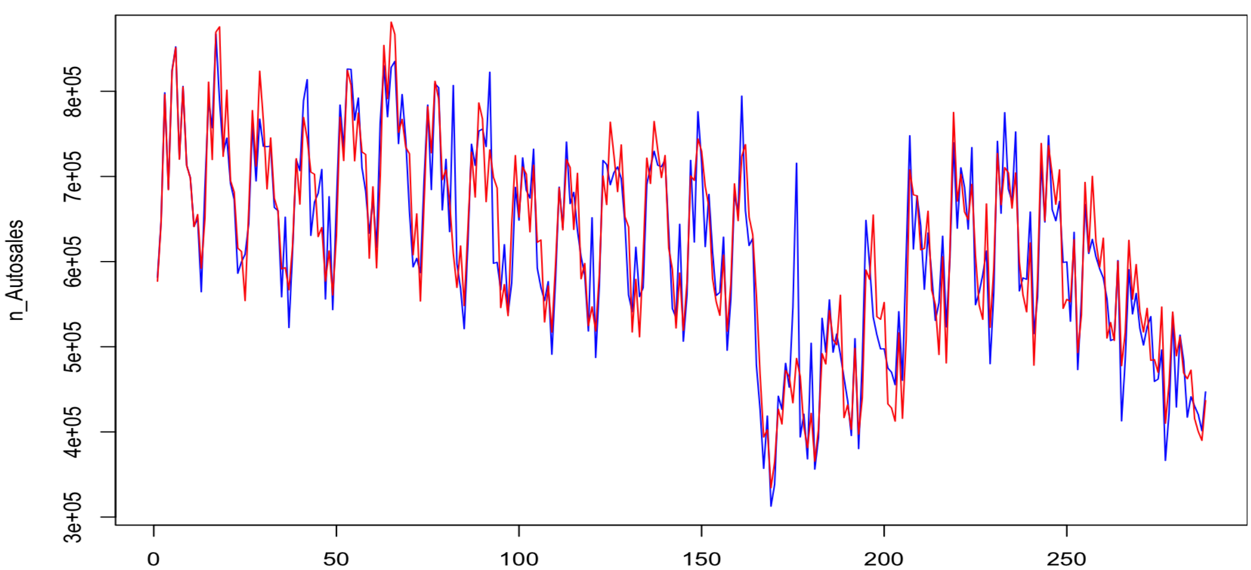


Fig 4.2.1.2

The comparison between the actual and predicted values shows that the model is attempting to account for all the variability in autosales. The same can be observed in the value of R- squared which is 0.8709. This indicates that 87.09% of the variation of autosales is captured by the model.

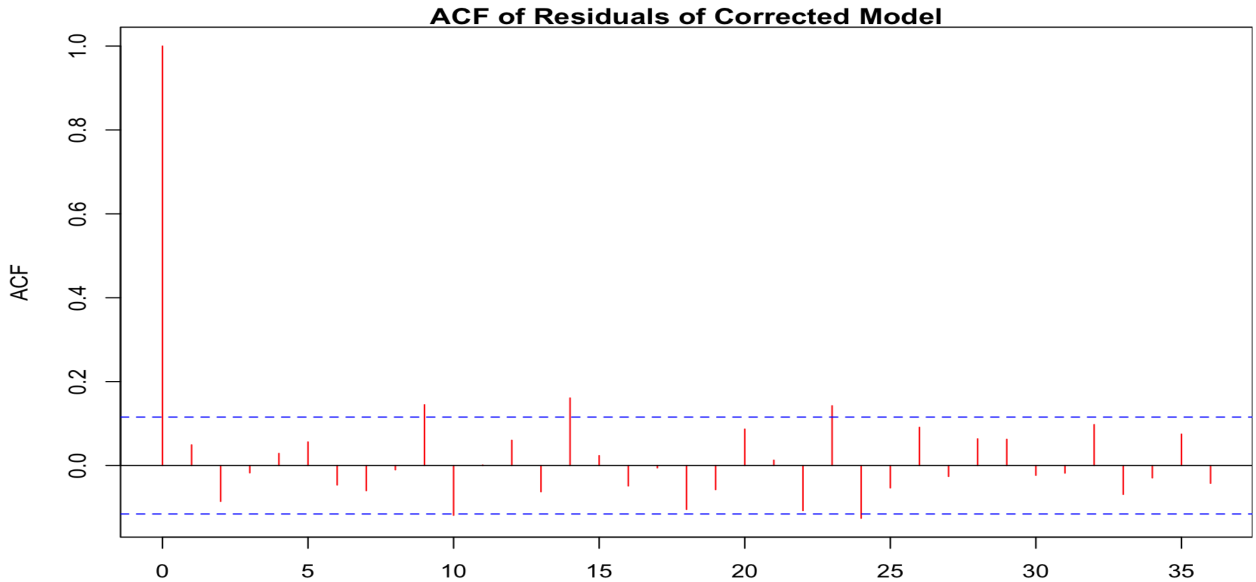


Fig 4.2.3

From Fig 4.2.1.3, the ACF values of the residual are slightly greater than the absolute values of 2 standard errors: hence, they are considered as not significant. Thus we conclude the residuals are white noise series.

4.2.2.1 Mean Absolute Percentage Error (MAPE)

For Training Data: The mean absolute percentage error (MAPE) for this model is 5.31% which implies that, on average, the predictions are 12.13% off from the actual values.

For Hold Out Sample: The mean absolute percentage error (MAPE) for this model is 6.86% which implies that, on average, the predictions are 6.86% off from the actual values.

4.3 ARIMA models (for the variable of interest)

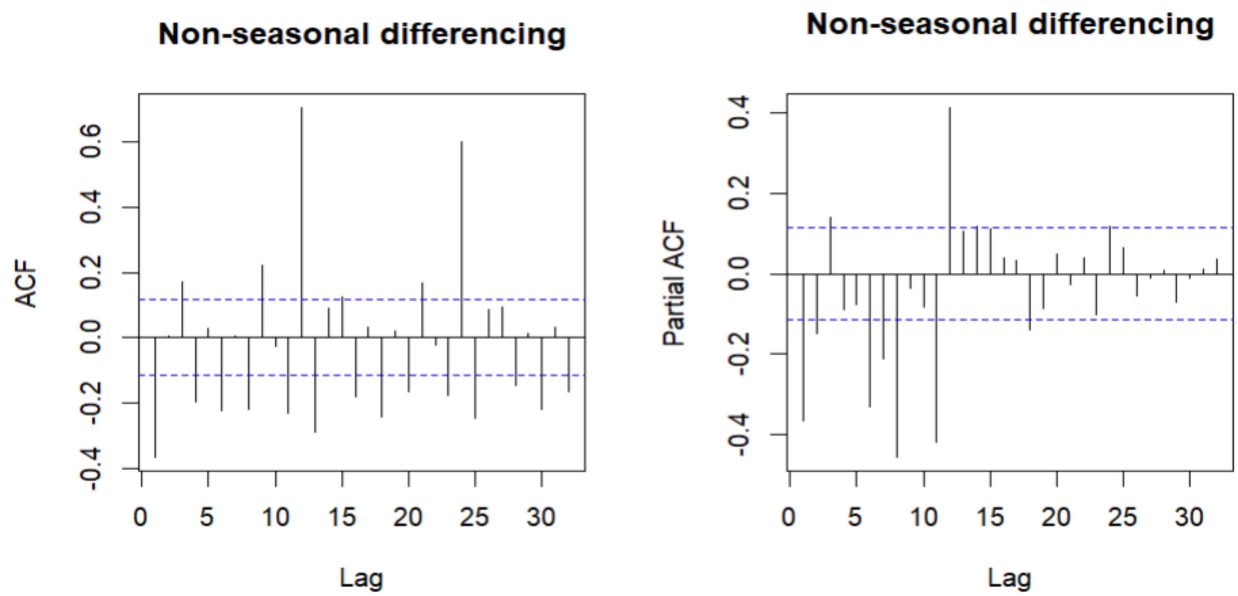


Fig4.3.1

The two plots in Fig4.3.1, shown are autocorrelation function (ACF) and partial autocorrelation function (PACF) plots for the autosales series after non-seasonal differencing, suggesting the presence of correlation at seasonal lags. The ACF plot on the left displays the correlation of the series at every 12th lag.

From Fig4.3.2, the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, both considering non-seasonal and seasonal differencing of a time series dataset.The plots shown are autocorrelation function (ACF) and partial autocorrelation function (PACF) plots for a dataset after applying both non-seasonal and seasonal differencing, suggesting they are being analyzed for a potential Moving Average (MA) process.

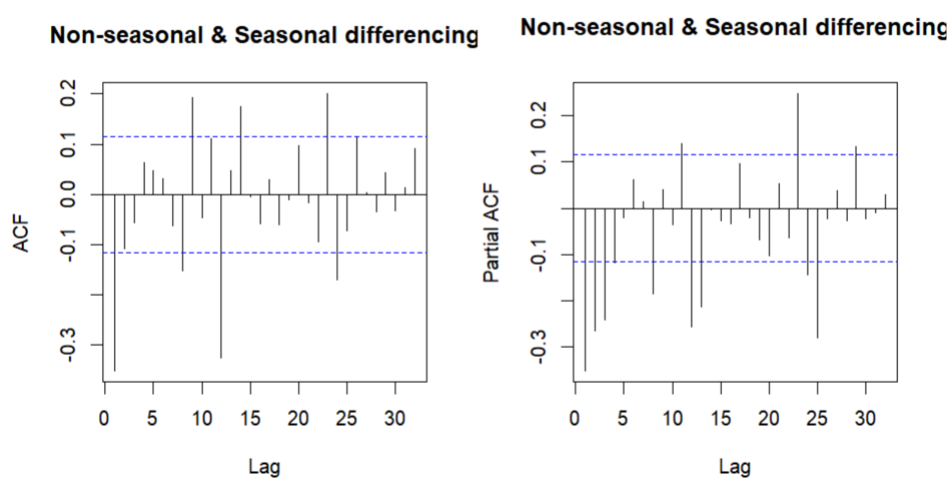


Fig4.3.2

The ACF plot on the left indicates that the autocorrelations chopped off and mostly remain within the confidence bounds (dashed blue lines), which is a characteristic often observed in MA processes. The PACF plot on the right shows a similar pattern with values decaying and remaining within the confidence bounds, indicating that any higher-order correlations have been accounted for, and the model might benefit from including specific MA terms at identified lags. This behavior in the ACF and PACF suggests that the series, after differencing, might be modeled effectively with a seasonal MA model with a period of 12.

```
> summary(sarima)
Series: n_Autosales
ARIMA(0,1,1)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
          ma1     sma1
      -0.5246  -0.7788
s.e.   0.0543   0.0510

sigma^2 = 0.005822:  log likelihood = 312.83
AIC=-619.66   AICc=-619.57   BIC=-608.81

Training set error measures:
                    ME     RMSE      MAE        MPE     MAPE      MASE        ACF1
Training set -710.6035 43378.17 32139.62 -0.4490702 5.445599 0.4665855 -0.02559113
```

Fig4.3.3

By looking at the SARIMA summary, This model is used for analyzing time series data that has both non-seasonal and seasonal differences. The coefficients shown include ma1 = -0.5246 for the non-seasonal moving average component and sma1 = -0.7788 for the seasonal moving average component, the coefficients of ma1 and sma1 are significantly different than 0 (by using the absolute values of the t_stat, higher than 1.96). indicating significant estimates. The output shows a small sigma squared (sigma^2 = 0.005822), suggesting low error variance, and robust log-likelihood (312.83), AIC (-619.66), and BIC (-608.81) values, which indicate a good model fit.

Additionally, the training set error metrics like ME, RMSE, MAE, MPE, MAPE, and MASE suggest the model's accuracy and precision in forecasting, while the ACF1 value close to zero (-0.02559113) confirms minimal autocorrelation in the residuals, supporting the model's adequacy in capturing the dynamics of the series.

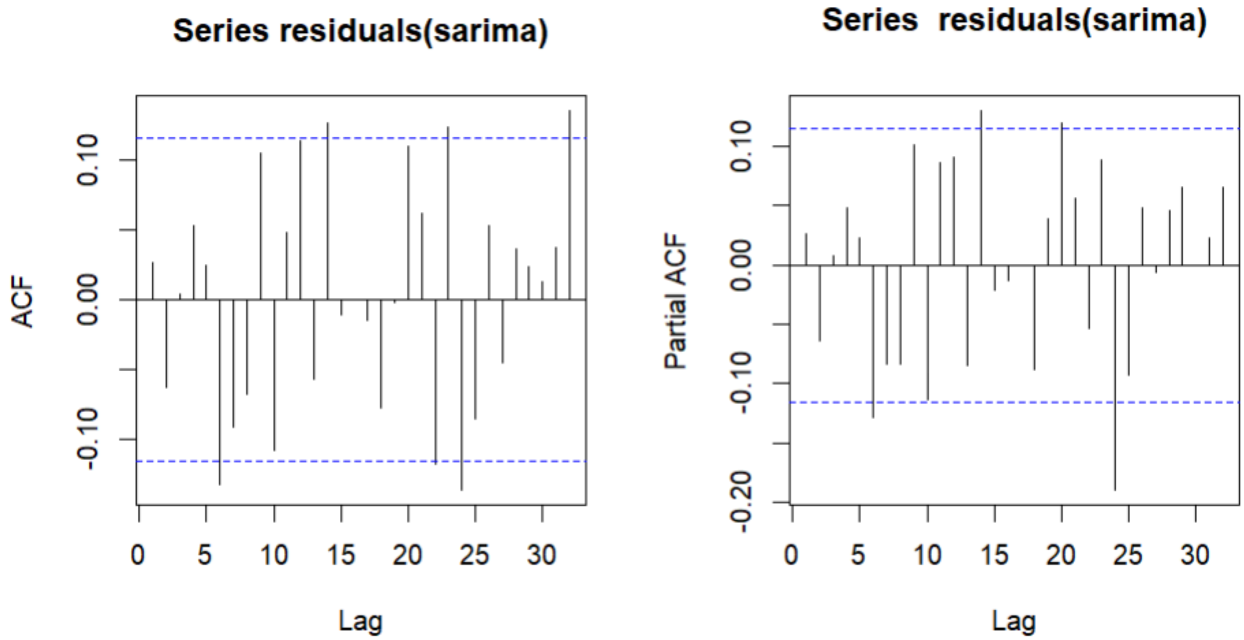


Fig4.3.4

In the figure 4.3.4, the two plots displayed are autocorrelation function (ACF) and partial autocorrelation function (PACF) plots for the residuals of a SARIMA model fit. Both plots are essential for diagnosing the effectiveness of the model in capturing the underlying patterns in the time series data. The ACF plot on the left shows that the autocorrelations of the residuals are mostly within the confidence bounds (dashed blue lines), suggesting that there is little to no autocorrelation at various lags. This indicates that the model residuals behave like white noise, which is an ideal outcome, implying that the model has successfully captured the main patterns in the data. Similarly, the PACF plot on the right confirms this finding, as the partial autocorrelations are also largely within the confidence bounds across all lags. This further supports the conclusion that the model is adequately specified.

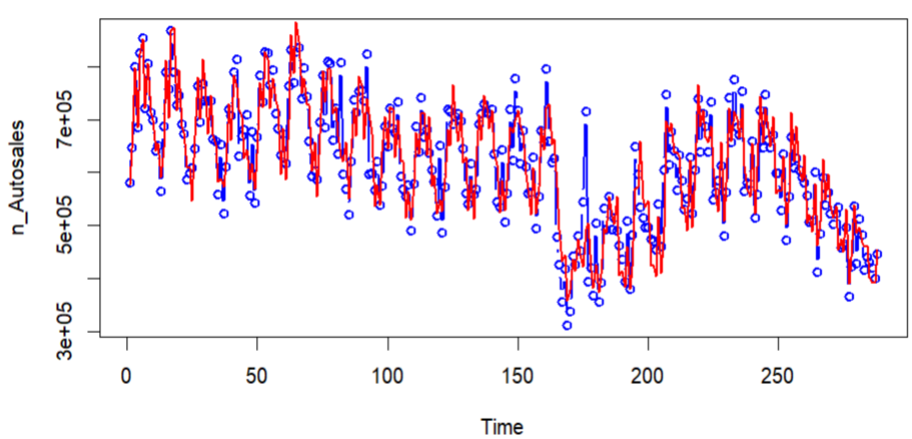


Fig4.3.5

By looking at the plot 4.3.5, it represents the time series data of "n_Autosales" along with its corresponding fitted values from a SARIMA model, depicted over time. The blue circles represent the actual sales data points, while the red line indicates the model's fitted values, showing how well the model conforms to the observed data across the given time span. The fluctuations and peaks in sales data are closely tracked by the red line, suggesting that the model has a reasonable fit. The time index on the x-axis shows the data points spanning over approximately 250 time periods, illustrating the model's ability to capture both the trend and any seasonal patterns present in the data. This visual comparison helps in evaluating the predictive accuracy and effectiveness of the SARIMA model in forecasting future sales based on historical data.

For Training Data: The mean absolute percentage error (MAPE) for this model is 5.46% which implies that, on average, the predictions are 5.46% off from the actual values.

For Hold Out Sample: The mean absolute percentage error (MAPE) for this model is 7.75% which implies that, on average, the predictions are 7.75% off from the actual values.

MAPE (Training Sample)	MAPE (Test Sample)
5.46%	7.75%

CONCLUSION:

After the use of Deterministic Time Series Models and the Stochastic Time Series models to model the US Auto Sales series. We can compare the results and conclude that the Stochastic Time Series Model fits the best for our data. The hold out sample accuracy of the Seasonal Auto regressive integrated Moving Average model i.e. ARIMA (0,1,1) (0,1,1) [12] produces the least error which is 7.75% which is the best fit from our entire analysis. We can also justify that by looking at the residuals from this model which behave like white noise series.

Overall, the SARIMA model captures the trend, seasonal and non-seasonal variations in our Auto Sales data.