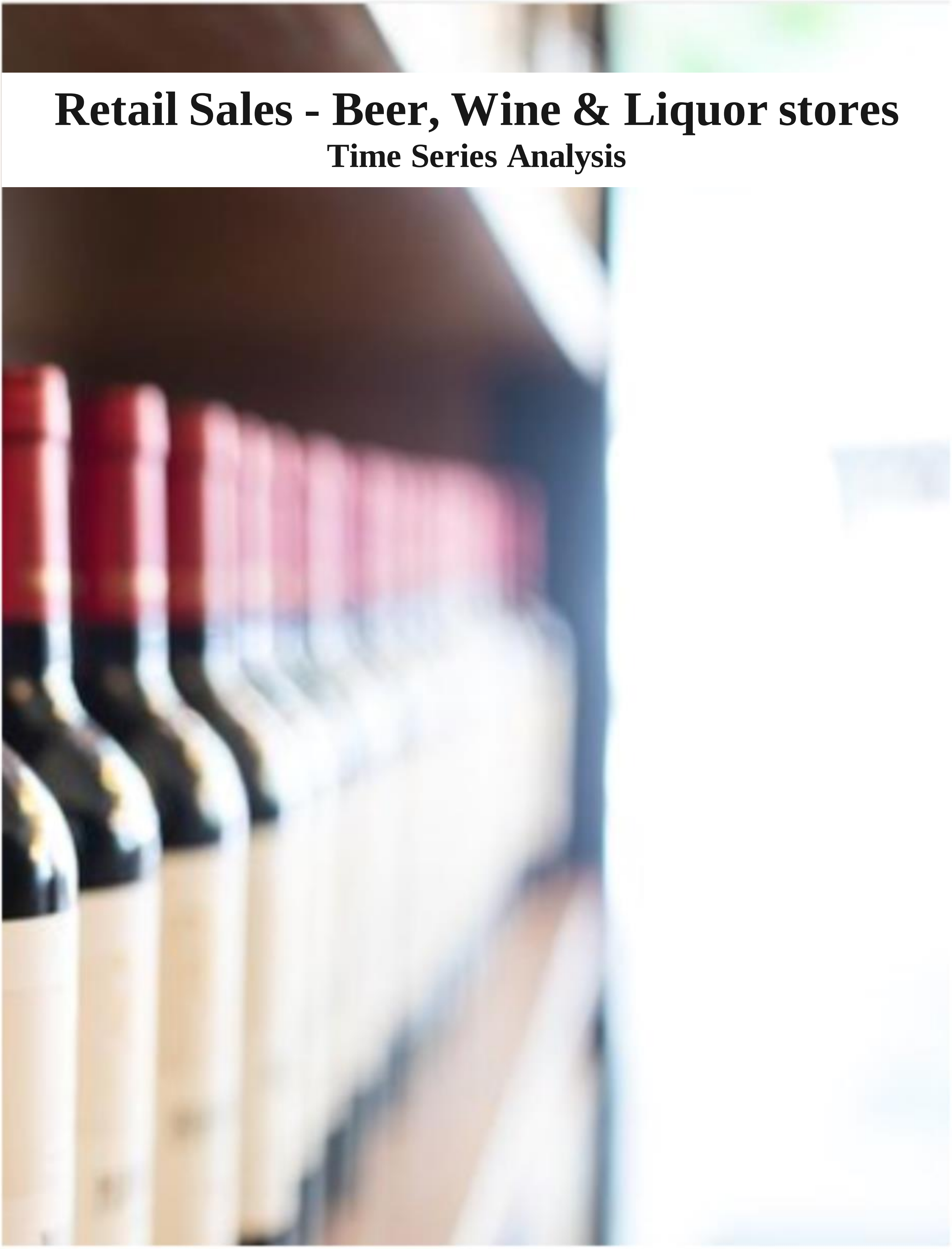
Retail Sales - Beer, Wine & Liquor stores

Time Series Analysis



Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

TABLE OF CONTENTS

EXECUTIVE SUMMARY

For this project, data of 'Retail Sales – Beer, Wines and Liquor Stores' is collected from Federal Reserve Economic Data and U.S Census Bureau. Retails sales is a monthly data measured in millions of dollars and is used to predict monthly retail sales across all stores in United States for next 2 fiscal years. Sales exclude sales taxes collected directly from customer and paid directly to a local, state, or federal tax agency. From visualizations we observed that retail sales data is having upward trend and multiplicative seasonality. From the box plot visualization, we have identified the retails sales is highest in December months compared to other months and lowest in February. The data is highly correlated and autocorrelation coefficients are significant in all lags.

Various models have been constructed on the retail sales data to forecast future values. Model based forecasting methods like Regression models such as linear and quadratic models with trend and seasonality, Auto Regressive models, Auto Regressive Integrated Moving Average models were utilized in this project. In addition to that data driven forecasting methods like smoothing methods like trailing MA, advanced exponential smoothing methods like Holt-winter's model are utilized. Model evaluation was based on accuracy measures like MAPE,RMSE,ACF1 etc. Typically models with low MAPE or RMSE values are considered as best forecasting models. Out of all the model built the best forecasting model that can be used to forecast future periods is two level forecasting model (Holt-Winter's Automatic Model with optimal parameters + AR(12) model for residuals) closely followed by ARIMA(3,1,2)(0,1,2).

INTRODUCTION

Demand & Sales forecasting is very important processes in business in which historic sales data or production data or any other data is used to develop an estimate of an expected forecast of demand. Critical business assumptions like turnover, profit margins, cash flow, capital expenditure, risk assessment and mitigation plans, capacity planning, etc. are dependent on forecasting.

Forecasting will reveal seasonal trends which helps in spotting the seasonal fluctuations, helps in panning the supply chain ,rationalize the cash flow and helps in preparing for the future. Demand and sales forecasting help drive smart business decisions.

During COVID-19 due to stay-at-home orders in United states the sales of beverages has increased a lot. Although away from home beverage sales has a large decline the online sales have sky-rocketed. So, it is important for manufacturer and retail owners to forecast demand and sales for future periods so manufactures can produce or supply according to the consumer demand.

The scope of this project is to forecast retail sales by using various time series models to analyze the historical monthly retail sales generated in united states. This forecasting results will greatly aid the manufacturers to plan their supplies, inventory & produce according to the demand and retailer in buying the stock according to demand.

EIGHT STEPS IN FORECASTING PROCESS

1) DEFINE GOAL

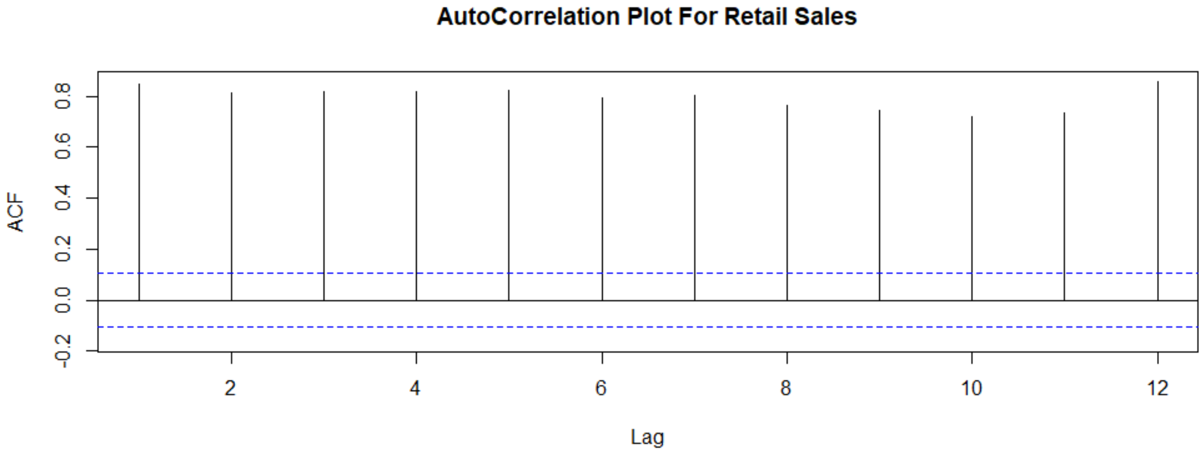
The goal of this project is to create numeric forecasts of the Retail Sales of Beer, Wine and Liquor stores in United states for the coming two years. The objective is to create a predictive model which will incorporates all the components like trend and seasonality of the historical data and effectively forecast the desired periods into future. Typically, the best forecasting model is selected based on the accuracy measures and several other metrics that explain the model. Since the data is generated monthly the forecasting models should be reevaluated for quarterly or at least semi-annually as the forecasting models utilize new data periods data when forecasting into future. Various forecasting methods like Smoothing methods, Holt-Winter's, ARIMA, Regression Models, Ensemble models are used to create various forecasting model to predict future values. The forecasts will be used to analyze the demand for beer, wines and liquor for future periods which can provides an estimate of the amount of goods and services that its end users will consume in the foreseeable future. Forecasting models for the project are developed in R language.

2) GET DATA

The data is collected from Federal Reserve Economic Data. The temporal frequency of the data is monthly .The data contains retails sales of beer, wine and liquor stores around USA from 1992 to April 2021.The data is monthly data with 352 data points available.

3) EXPLORE & VISUALIZE TIME SERIES

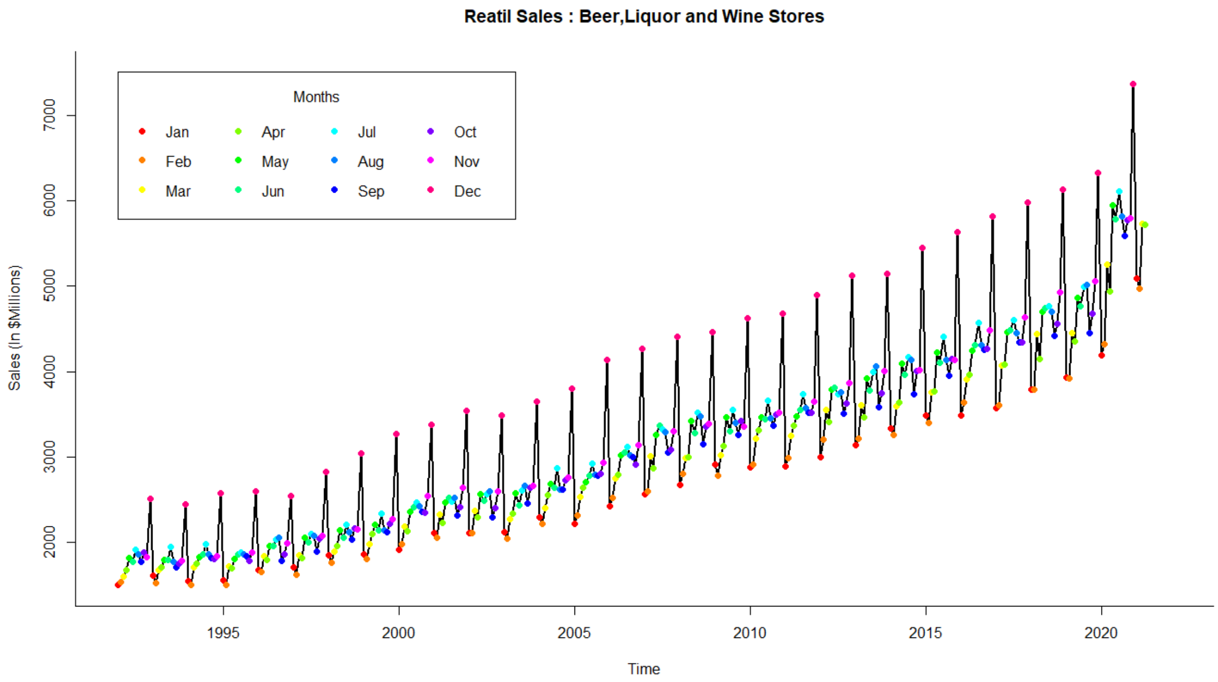
Data Visualization in single most important step in any project while dealing with data. Visualization will uncover the truth behind the data and will provide opportunity to analyze the patterns in the data.



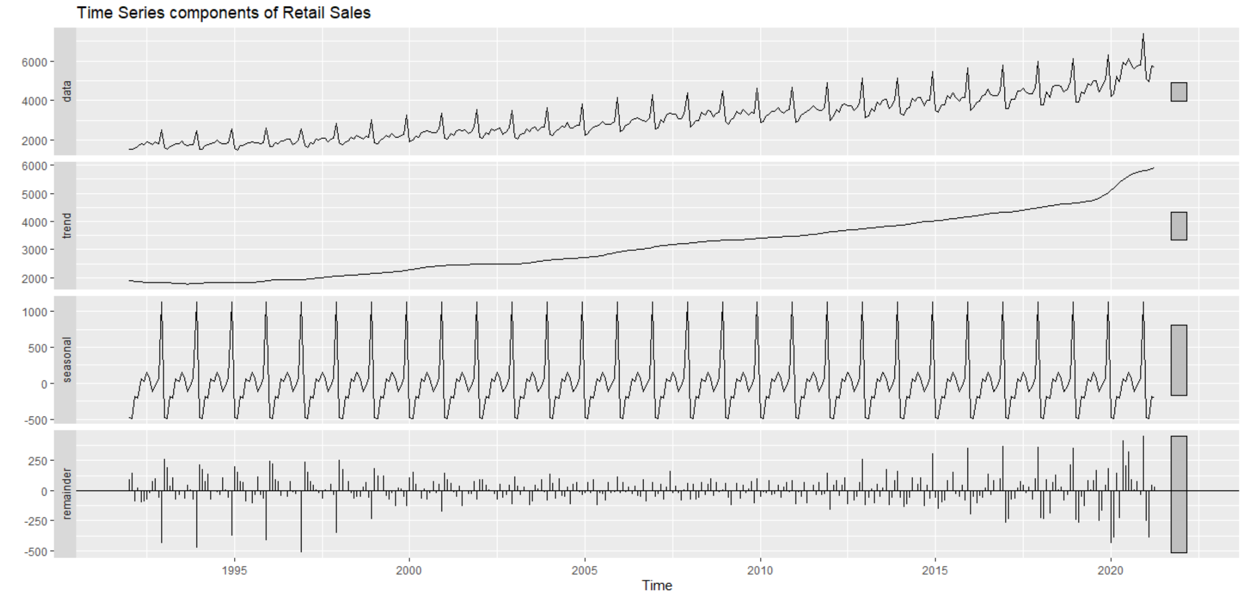
For time series data auto correlation plot is very crucial in analyzing the components of the time series data. As we can observed from the above plot all lags are significant.

Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

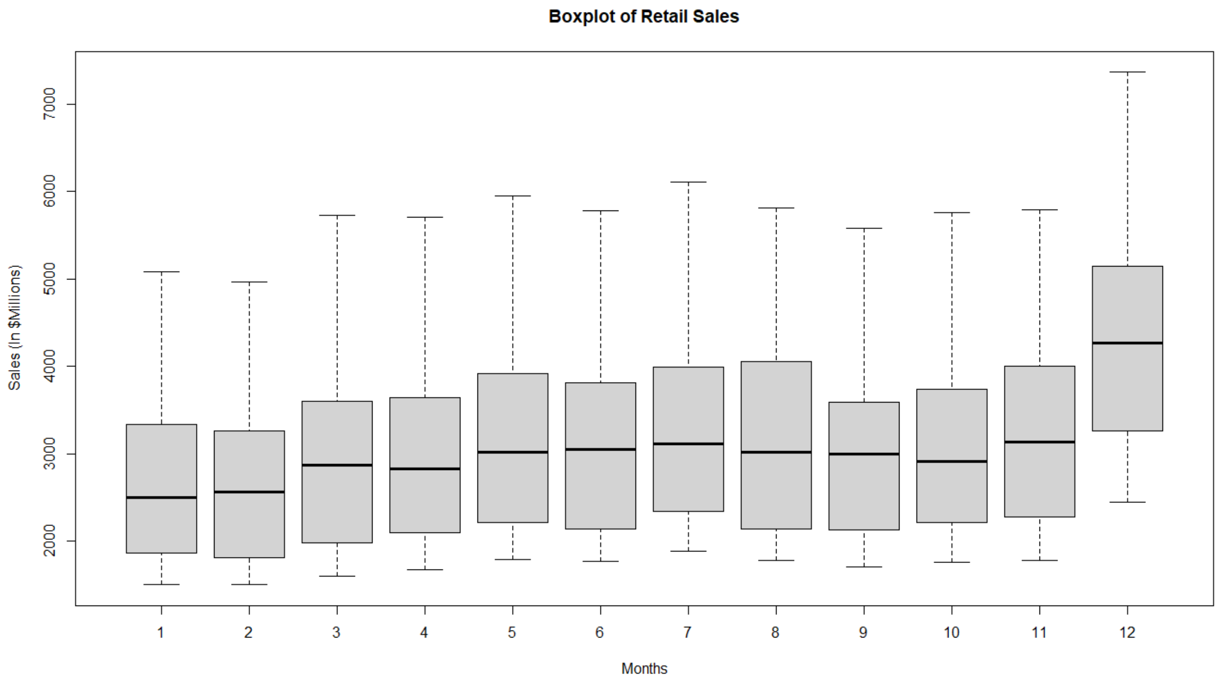
A significant coefficient at lag 1 indicates that there is trend in the data and significant lag at lag 12 indicates that there is seasonality in the data. Further patterns in the data can be uncovers by various visualizations.



As we can observe from the above plot the retails sales are growing linearly upward from 1992 with an increase in variance every year which indicates that we have multiplicate seasonality. So, we can interpret that retail sales data has upward trend with multiplicative seasonality.



Stl plot will uncover all the components of the data like trend, seasonality and remainder.



From the above box plot we can interpret that sales of beer, wine & Liquor stores are stable for all the months except for December. As we can observe, Median sales amount of December month's sales is more than 75th percentile of every other month from January to November.

4) Preprocessing

Preprocessing plays import role in any end-to-end data process. Preprocessing helps in detecting zero values, null values, outliers etc. It also helps in identifying if there are any gaps in the time periods. In retail sales data from Federal Reserve Economic Data we have clean data with no outlier or Null values, hence no preprocessing steps required.

5) Partition Series

Partitioning is important step in time series forecasting. Before applying any forecasting method, data should be parting into training (70-80%) and validation (20-30%) partitions. Main reason to partition the time series is that to avoid overfitting which will result in high train accuracies and poor test accuracies. The partition should not be a random as in cross section data because we require data set without any missing time periods and in orderly manner. For Retails sales data the data is partitioned into 282 rows for training and latest 70 rows into validation.

6) APPLY FORECASTING METHODS

Before we can build forecasting models, it is important to determine if the time series data is predictable or just a random walk. Implementing a forecasting models on random walk time series data would result in loss of time, money and effort.

Predictability :

This means time series data is predictable and an effort can be made to build a forecasting model to predict future values. Its historical data and patterns can be used to apply for important and crucial predictions with various forecasting models. Can be represented as below equation for an AR(1) model,

$$\boldsymbol{Yt} = \boldsymbol{\alpha} + \beta 1 * \boldsymbol{Yt} - \mathbf{1} + \boldsymbol{\varepsilon t}$$

Where $\beta 1 \neq 1$

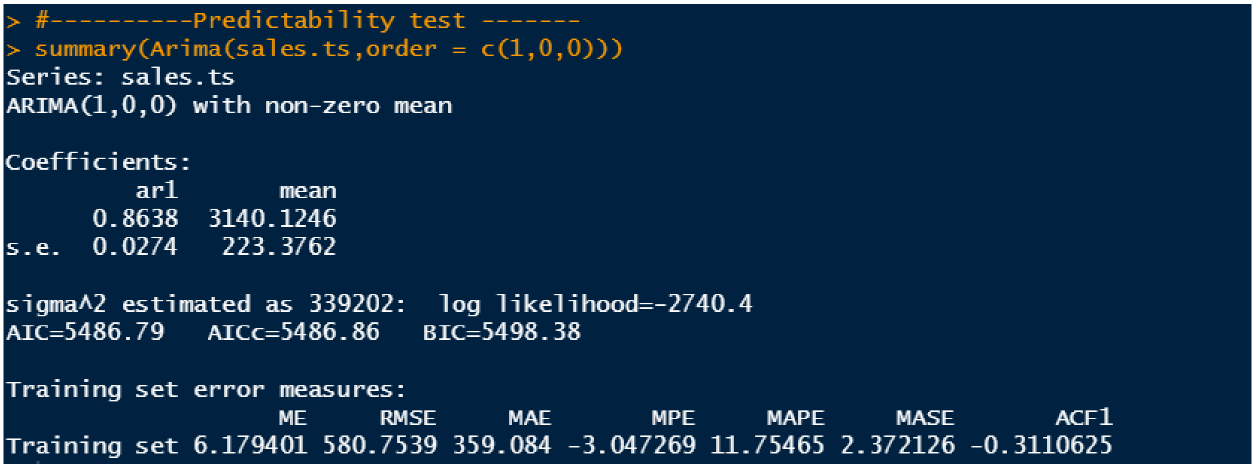
Random Walk:

Time series data which changes from one time period to next time period are random where current observation is equal to the previous observation with a random step up or step down. Can be represent as below equation where β1 =1,

$$\boldsymbol{Yt} = \boldsymbol{\alpha} + \boldsymbol{Yt} - \mathbf{1} + \boldsymbol{\varepsilon t}$$

Predictability Test:

Below is the model summary for AR(1) model.



```
> #----------Predictability test -------
> summary(Arima(sales.ts,order = c(1,0,0)))
Series: sales.ts
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1       mean
      0.8638  3140.1246
s.e.  0.0274   223.3762

sigma^2 estimated as 339202:  log likelihood=-2740.4
AIC=5486.79   AICc=5486.86   BIC=5498.38

Training set error measures:
                   ME     RMSE     MAE       MPE     MAPE     MASE       ACF1
Training set 6.179401 580.7539 359.084 -3.047269 11.75465 2.372126 -0.3110625
```

From the above model summary, we can infer that ARIMA model of order (1,0,0) has a mean or intercept of 3140.1246 and co-efficient(ar1) as 0.8638. The standard error for coefficient and intercept is 0.0274 and 223.3762 respectively. From this data we can build an equation for Yt as below,

Yt = 3140.1246 + 0.8638 * Yt-1

Where yt-1 is preceding value in the series

Hypothesis Test :

- Stating Null (H_0) and Alternate(H_a) Hypothesis

  $H_0 : \beta_1 = 1$, $H_a : \beta_1 \neq 1$

- Specifying level of significance

  Here we are considering a level of significance of as 0.05

- Calculating the test statistic (Z score)

  Now from the model summary the $\beta_1 = 0.8638$ and standard error is 0.0274.

  Z score = (0.8638 – 1)/(0.0274) = -4.9708

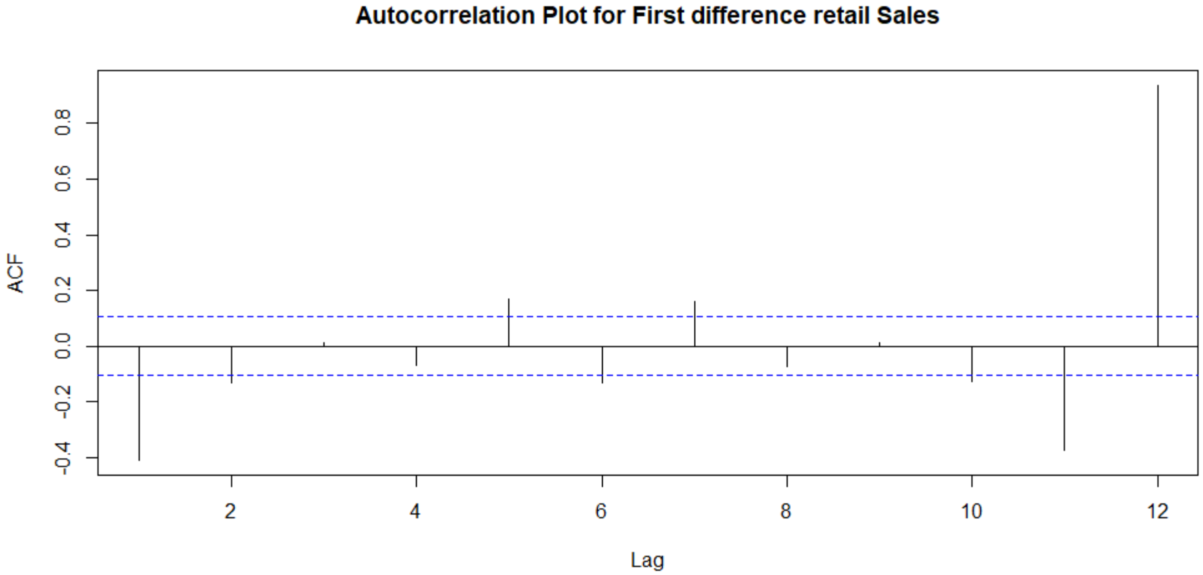
- Computing P-value

  Now from the Z-score table the p-value is 3.33381e-07.

- Now since p-value is less than 0.05 we can reject H_0 so, $\beta_1 \neq 1$ is true.

Another method to find out if the time series data is predictable or not is by using the differenced series (of lag1) i.e., using $Y_2\text{-}Y_1, Y_3\text{-}Y_2, \ldots, Y_t\text{-}Y_{t-1}$ .The idea is that if the original time series is a random walk then the differenced series will also behave like a random walk.

We can find the dependencies or relationships for the differenced series using Acf() which will provide autocorrelation coefficients at lags 1,2,3, …. so on and if all of the autocorrelation coefficients are between the horizontal thresholds then the time series is a random walk.



As we can interpret from the plot that most of the auto-correlation coefficients are above the horizontal threshold which indicates there is correlation between differenced data series at various lags which indicates that the retail sales time series data is predictable.

i) TIME SERIES REGRESSION MODELS:

The basic concept of regression models is that we forecast the values 'y' assuming it has a linear relationship with 'x'. The forecast variable y is sometimes called as dependent variable and x as independent variable or predictors. Below equation represents basic linear regression model equation,

$$\boldsymbol{Yt = \beta 0 + \beta 1\, t + \varepsilon}$$

where t represents periods i.e. 1,2,3, ….
Y_t represents the output variable for time series measurement
β_0 is intercept and β_1 represents coefficient of the equation
ε represents random component or error

Below statements will hold true for every regression model:

The sign of each coefficient indicates the direction of the relationship between the independent variable and the dependent variable.

- A positive sign indicates that as the independent variable increases, the dependent variable also increases.
- A negative sign indicates that as the independent variable increases, the dependent variable decreases.
- Coefficient value represents the change in dependent (Y_t) for every unit change in respective independent variable when all other values held constant.
- β_0 or intercept is the value of dependent variable or outcome of equation(Y_t) when all independent variables are zero.

➢ Regression model with linear trend and seasonality and Trailing MA for residuals:

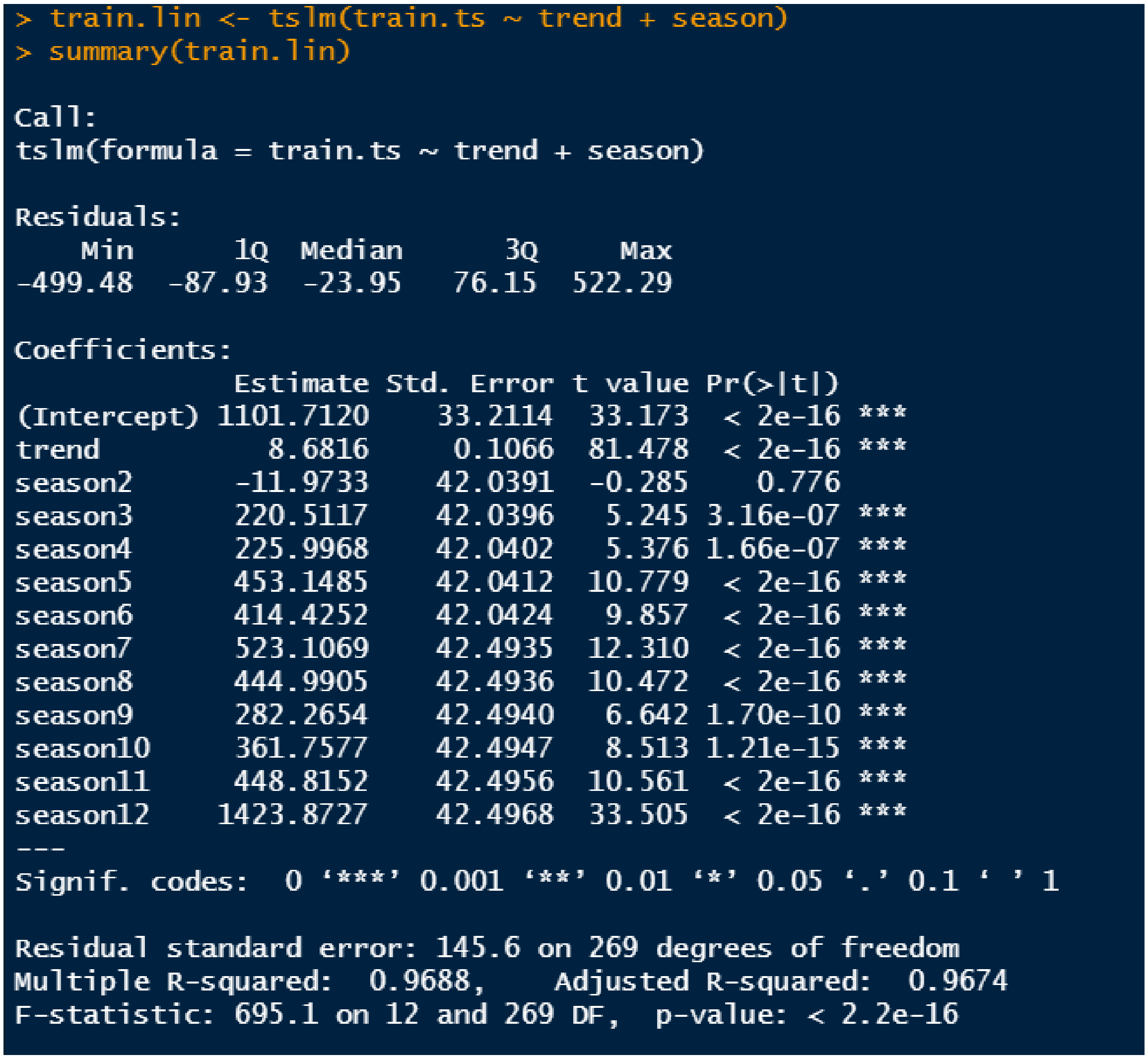
The model will fit a global trend and seasonality that applies to the entire training time series data and will use model to forecast in validation period. The seasonality is indicated as dummy variables. For monthly data we have 11 dummy variables namely D2,D3,D4….D12.Output is measured using below equation

$$\boldsymbol{Yt = \beta 0 + \beta 1\, t + \beta 2\, D2 + \beta 3\, D3 + \beta 4\, D4 \ldots\ldots + \beta 12 D12 + \varepsilon}$$

where Y_t represents the output variable at time t and t represents periods i.e. 1,2,3, ….
β_0 is intercept and β_1, β_2, β_3... B_{12} represents coefficients of the equation
Dummy variables take values as per season as mentioned in below table

	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
January	0	0	0	0	0	0	0	0	0	0	0
February	1	0	0	0	0	0	0	0	0	0	0
March	0	1	0	0	0	0	0	0	0	0	0
April	0	0	1	0	0	0	0	0	0	0	0
May	0	0	0	1	0	0	0	0	0	0	0
June	0	0	0	0	1	0	0	0	0	0	0
July	0	0	0	0	0	1	0	0	0	0	0
August	0	0	0	0	0	0	1	0	0	0	0
September	0	0	0	0	0	0	0	1	0	0	0
October	0	0	0	0	0	0	0	0	1	0	0
November	0	0	0	0	0	0	0	0	0	1	0
December	0	0	0	0	0	0	0	0	0	0	1

Below is the model summary for linear regression model with trend and seasonality for training partition.



```
> train.lin <- tslm(train.ts ~ trend + season)
> summary(train.lin)

Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-499.48  -87.93  -23.95   76.15  522.29 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1101.7120    33.2114  33.173  < 2e-16 ***
trend          8.6816     0.1066  81.478  < 2e-16 ***
season2      -11.9733    42.0391  -0.285    0.776    
season3      220.5117    42.0396   5.245 3.16e-07 ***
season4      225.9968    42.0402   5.376 1.66e-07 ***
season5      453.1485    42.0412  10.779  < 2e-16 ***
season6      414.4252    42.0424   9.857  < 2e-16 ***
season7      523.1069    42.4935  12.310  < 2e-16 ***
season8      444.9905    42.4936  10.472  < 2e-16 ***
season9      282.2654    42.4940   6.642 1.70e-10 ***
season10     361.7577    42.4947   8.513 1.21e-15 ***
season11     448.8152    42.4956  10.561  < 2e-16 ***
season12    1423.8727    42.4968  33.505  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 145.6 on 269 degrees of freedom
Multiple R-squared:  0.9688,	Adjusted R-squared:  0.9674 
F-statistic: 695.1 on 12 and 269 DF,  p-value: < 2.2e-16
```

From the model summary we can interpret that intercept is 1101.7120 and is significant as p-value is very low. Coefficient for trend component is 8.6816 and is statistically significant. Y_t increases with increase in any of the independent variables except season 2 as all other coefficients are positive.

Using the intercept and coefficients we can build the equation to predict sales (Y_t) for future time periods.

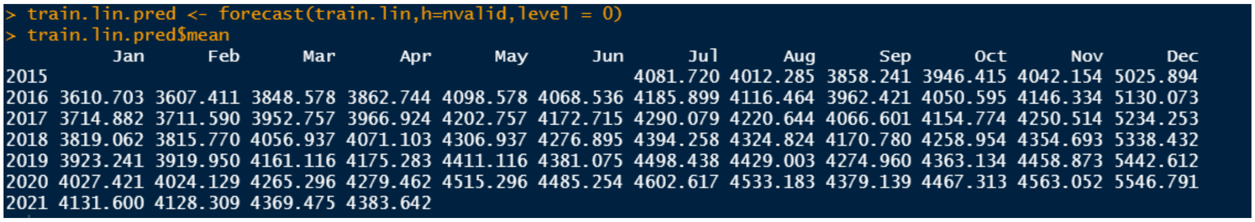
$$\begin{aligned} Yt = {} & 1101.7120 + 8.681*t + -11.9733*D2 + 220.5117*D3 + 225.9968*D4 + 453.1485*D5 \\ & + 414.4252*D6 + 523.1069*D7 + 444.9905*D8 + 282.2654*D9 + 361.7577*D10 + \\ & 448.8152*D11 + 1423.8727*D12 \end{aligned}$$

Substituting respective values for dummy variables depending on the season and time period (t) in the above equation will provide us sales for desired time period.

The summary table has R^2 and F-statistic which measure the overall explainability of the independent variables (t) over dependent variable (Y_t). R^2 or Coefficient of determination values lies between 0 and 1 and it is a measure of that indicates how much variance in Y_t can be explained by t. Having higher R^2 means that independent variable can explain most of the variance in dependent variable. In this model we have R^2 as 0.9688 which indicates that in historical data set t can explain 96.88% of variance in Y_t . Adjusted R-squared (0.9674 or 96.74%) also used similar to R-squared but adjusted R^2 prefers fewer independent variable by penalizing the excess independent variables.

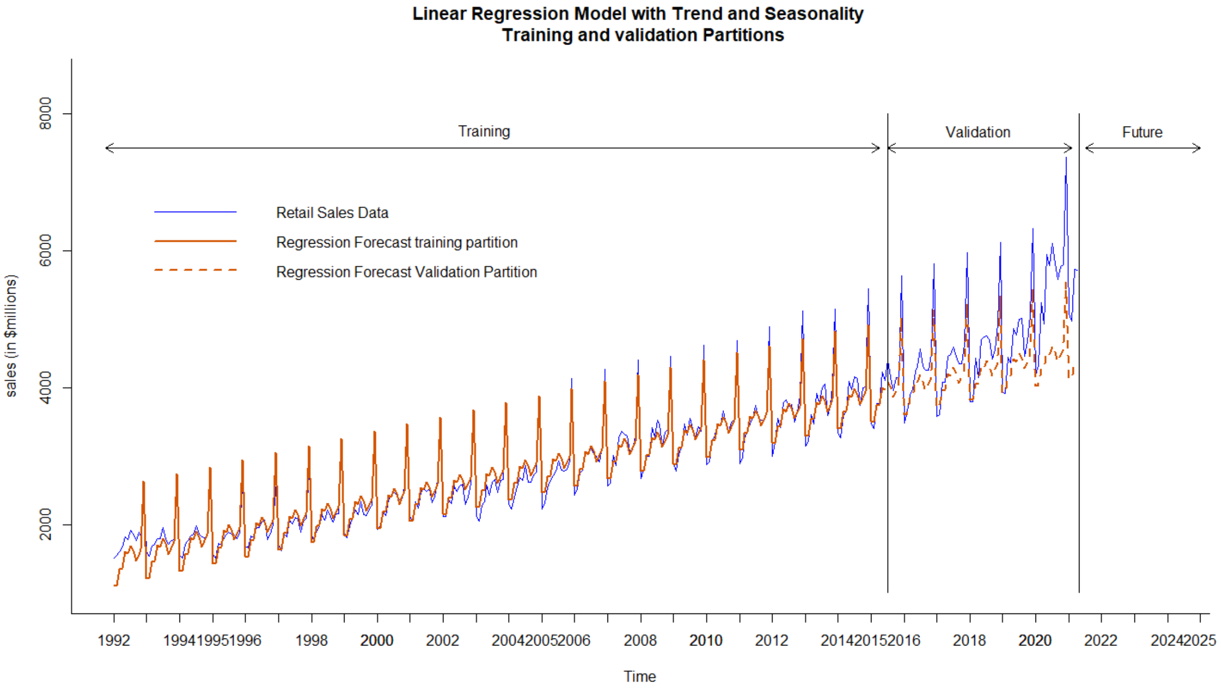
F-statistic indicates if model is fit by chance or not. A low F-statistic indicates that the independent variables do not explain dependent variable well. For the above model we have a F-statistic value of 695.1 indicates that overall is good fit and is significant as p-value is less than 0.05.

Below is the point forecasted values for validation period using linear trend and seasonality model.

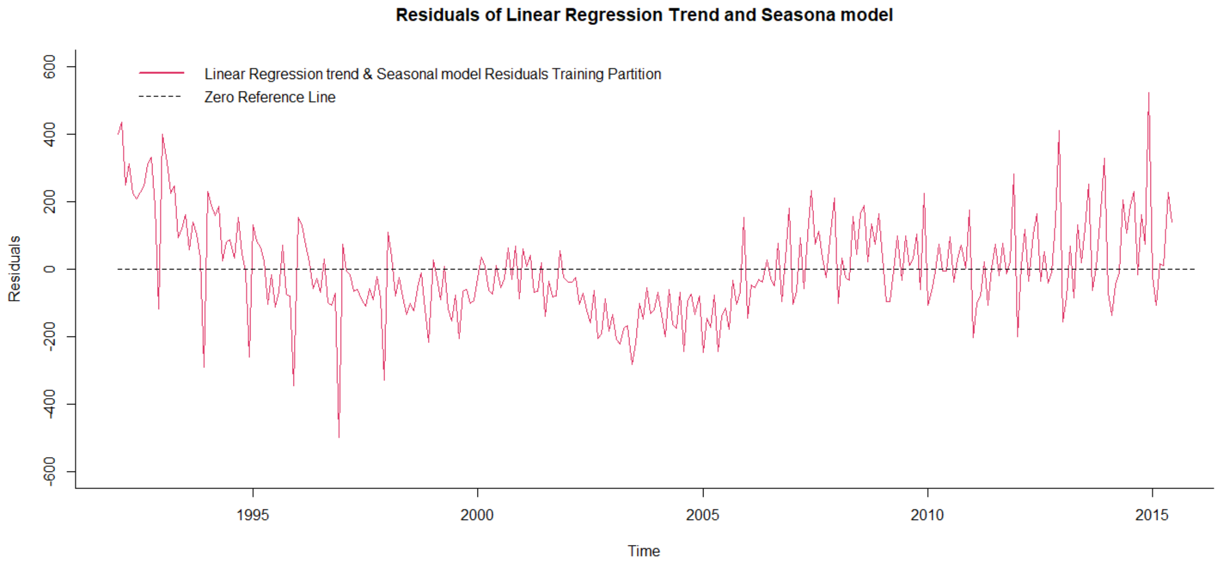


```
> train.lin.pred <- forecast(train.lin,h=nvalid,level = 0)
> train.lin.pred$mean
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015                                                       4081.720 4012.285 3858.241 3946.415 4042.154 5025.894
2016 3610.703 3607.411 3848.578 3862.744 4098.578 4068.536 4185.899 4116.464 3962.421 4050.595 4146.334 5130.073
2017 3714.882 3711.590 3952.757 3966.924 4202.757 4172.715 4290.079 4220.644 4066.601 4154.774 4250.514 5234.253
2018 3819.062 3815.770 4056.937 4071.103 4306.937 4276.895 4394.258 4324.824 4170.780 4258.954 4354.693 5338.432
2019 3923.241 3919.950 4161.116 4175.283 4411.116 4381.075 4498.438 4429.003 4274.960 4363.134 4458.873 5442.612
2020 4027.421 4024.129 4265.296 4279.462 4515.296 4485.254 4602.617 4533.183 4379.139 4467.313 4563.052 5546.791
2021 4131.600 4128.309 4369.475 4383.642
```

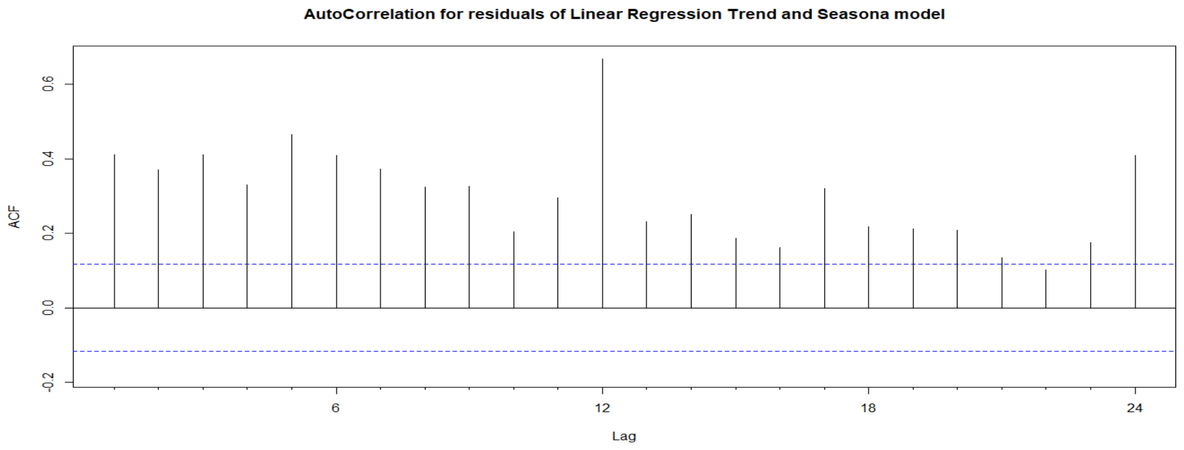
Visualizing retail sales using regression model:



Residuals of Linear Regression Model:



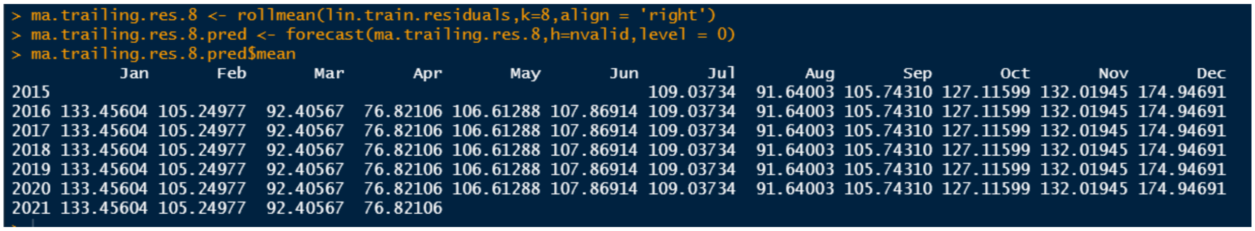
As observed from the below autocorrelation plot for residuals, there are a lot of significant relations not incorporated into the model. But to incorporate these relations using an AR(p) model for residuals we need a value for ‘p’ which is large like 12,24 …because seasonal lags like 12,24…have larger significant coefficients. If a model is built for residuals using AR(12) or AR(24) , the model will give 12 or 24 new independent variables in the auto regressive equation. All these independent variables will make the two-level forecast model more complex ,uninterpretable and is not parsimonious . For example, AR(15) for residuals will incorporate most of the dependencies in residuals but will produce an equation with 15 independent variables.



So instead of AR(p) model for residuals we are opting for a data driven method like trailing moving average to forecast the residuals and a two-level forecast model is created with linear regression model with trend and seasonality and trailing moving average.

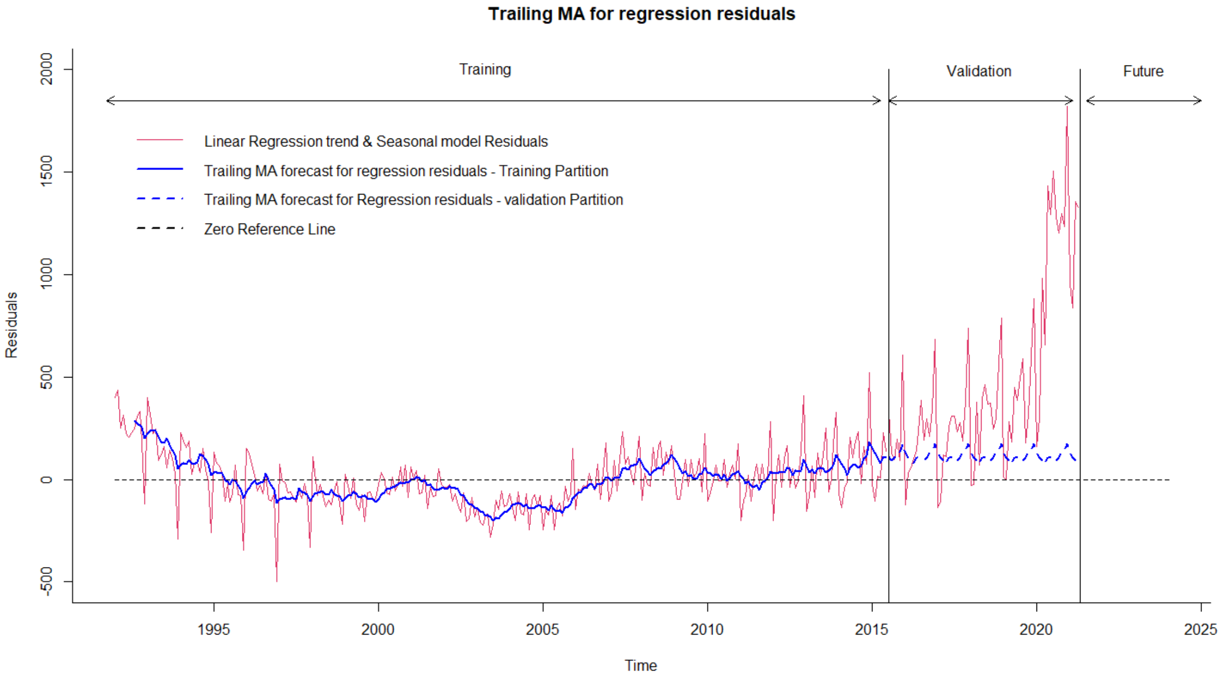
Trailing MA for regression residuals:

Below table represents point forecasted values of residuals for validation period using trailing moving average model with K=8.



```
> ma.trailing.res.8 <- rollmean(lin.train.residuals,k=8,align = 'right')
> ma.trailing.res.8.pred <- forecast(ma.trailing.res.8,h=nvalid,level = 0)
> ma.trailing.res.8.pred$mean
           Jan       Feb       Mar       Apr       May       Jun       Jul       Aug       Sep       Oct       Nov       Dec
2015                                                             109.03734  91.64003 105.74310 127.11599 132.01945 174.94691
2016 133.45604 105.24977  92.40567  76.82106 106.61288 107.86914 109.03734  91.64003 105.74310 127.11599 132.01945 174.94691
2017 133.45604 105.24977  92.40567  76.82106 106.61288 107.86914 109.03734  91.64003 105.74310 127.11599 132.01945 174.94691
2018 133.45604 105.24977  92.40567  76.82106 106.61288 107.86914 109.03734  91.64003 105.74310 127.11599 132.01945 174.94691
2019 133.45604 105.24977  92.40567  76.82106 106.61288 107.86914 109.03734  91.64003 105.74310 127.11599 132.01945 174.94691
2020 133.45604 105.24977  92.40567  76.82106 106.61288 107.86914 109.03734  91.64003 105.74310 127.11599 132.01945 174.94691
2021 133.45604 105.24977  92.40567  76.82106
```

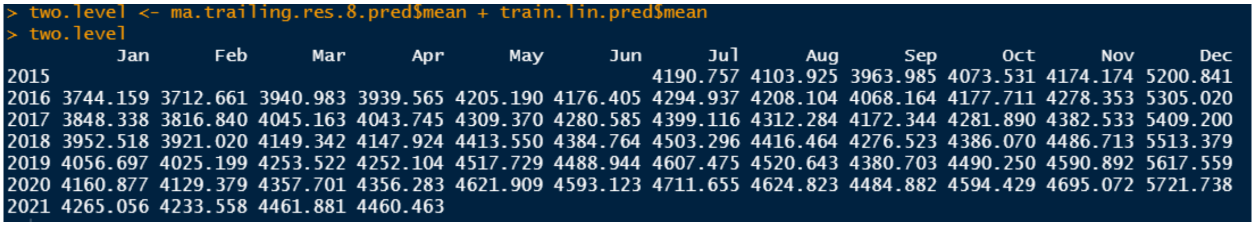
Visualizing Trailing MA residuals for Training and Validation partition:



Two-level Forecast with Trailing MA for Residuals:

Now once residuals are forecasted into validation, both residual forecast for validation and regression model forecast for validation are combined to from a two-level forecast.

Below table shows forecasted values for validation period by two-level forecast.



```
> two.level <- ma.trailing.res.8.pred$mean + train.lin.pred$mean
> two.level
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015                                                       4190.757 4103.925 3963.985 4073.531 4174.174 5200.841
2016 3744.159 3712.661 3940.983 3939.565 4205.190 4176.405 4294.937 4208.104 4068.164 4177.711 4278.353 5305.020
2017 3848.338 3816.840 4045.163 4043.745 4309.370 4280.585 4399.116 4312.284 4172.344 4281.890 4382.533 5409.200
2018 3952.518 3921.020 4149.342 4147.924 4413.550 4384.764 4503.296 4416.464 4276.523 4386.070 4486.713 5513.379
2019 4056.697 4025.199 4253.522 4252.104 4517.729 4488.944 4607.475 4520.643 4380.703 4490.250 4590.892 5617.559
2020 4160.877 4129.379 4357.701 4356.283 4621.909 4593.123 4711.655 4624.823 4484.882 4594.429 4695.072 5721.738
2021 4265.056 4233.558 4461.881 4460.463
```

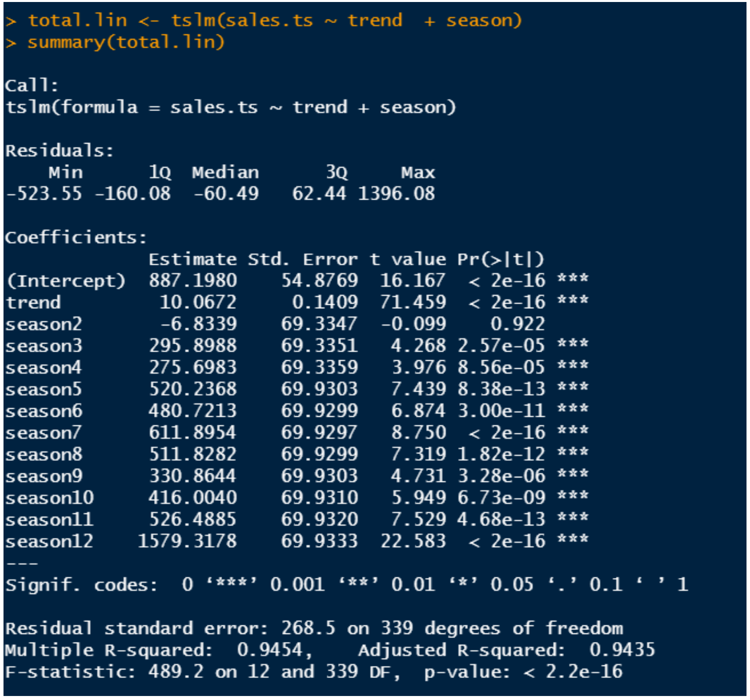
Below table shows accuracy of two-level forecast and regression model for trend and seasonality.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's
Two Level Forecast (Regression trend + seasonal and Trailing MA for residuals)	352.072	567.155	392.572	6.335	7.433	0.721	0.71
Regression Model	465.829	646.778	477.828	8.776	9.111	0.706	0.821

As we can observe two-level forecast has produced a good MAPE compared to the regression model for trend and seasonality in the validation period.

Two-level forecast for entire data With Trailing MA for Residuals:

Now training and validation are combined to make forecast for future periods.
Below is the summary of regression model for entire data.



```
> total.lin <- tslm(sales.ts ~ trend  + season)
> summary(total.lin)

Call:
tslm(formula = sales.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-523.55 -160.08  -60.49   62.44 1396.08 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  887.1980    54.8769  16.167  < 2e-16 ***
trend         10.0672     0.1409  71.459  < 2e-16 ***
season2       -6.8339    69.3347  -0.099    0.922    
season3      295.8988    69.3351   4.268 2.57e-05 ***
season4      275.6983    69.3359   3.976 8.56e-05 ***
season5      520.2368    69.9303   7.439 8.38e-13 ***
season6      480.7213    69.9299   6.874 3.00e-11 ***
season7      611.8954    69.9297   8.750  < 2e-16 ***
season8      511.8282    69.9299   7.319 1.82e-12 ***
season9      330.8644    69.9303   4.731 3.28e-06 ***
season10     416.0040    69.9310   5.949 6.73e-09 ***
season11     526.4885    69.9320   7.529 4.68e-13 ***
season12    1579.3178    69.9333  22.583  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 268.5 on 339 degrees of freedom
Multiple R-squared:  0.9454,	Adjusted R-squared:  0.9435 
F-statistic: 489.2 on 12 and 339 DF,  p-value: < 2.2e-16
```

From the model summary we can interpret that intercept is 887.1980 and is significant as p-value is very low. Coefficient for trend component is 10.0672 and is statistically significant. Y_t increases with increase in any of the independent variables except season 2 as all other coefficients are positive. Out of all the seasons ,season2 is not statistically significant.

Using the intercept and coefficients we can build the equation to predict sales (Y_t) for future time periods.

$$Yt = 887.1980 + 10.0672*t - 6.8339 *D2 + 295.58988 *D3 + 275.6983 *D4 + 520.2368*D5 + 480.7213*D6 + 611.8954 *D7 + 511.8282*D8 + 330.8644*D9 + 416.0040*D10 + 526.4885*D11 + 1579.3178*D12$$

Substituting respective values for dummy variables depending on the season and time period (t) in the above equation will provide us sales for desired time period.

. In this model we have R^2 as 0.9454 which indicates that in historical data set t can explain 94.54% of variance in Y_t . Adjusted R-squared (0.9435 or 94.35%) also used similar to R-squared but adjusted R^2 prefers fewer independent variable by penalizing the excess independent variables.

F-statistic indicates if model is fit by chance or not. A low F-statistic indicates that the independent variables do not explain dependent variable well. For the above model we have a F-statistic value of 489.2 indicates that overall is good fit and is significant as p-value is less than 0.05.

Now after developing regression model with trend and seasonality ,the residuals of the model are used for developing a trailing moving average model with K=8.

Below table shows values for future 24 periods for regression model forecast, residuals forecast and combined forecast

Time	Regression Forecast Trailing	Trailing MA residual Forecast Total	Combined Forecast
May-21	4961.174	974.7559	5935.93
June-21	4931.726	994.7081	5926.434
July-21	5072.967	1003.911	6076.878
August-21	4982.967	983.9872	5966.954
September-21	4812.07	1025.2816	5837.352
October-21	4907.277	1066.0957	5973.373
November-21	5027.829	1072.4856	6100.315
December-21	6090.726	1143.8751	7234.601
January-22	4521.475	1071.884	5593.359
February-22	4524.708	1020.6356	5545.344
March-22	4837.508	1013.064	5850.572
April-22	4827.375	995.6913	5823.066
May-22	5081.981	1032.3931	6114.374
June-22	5052.533	1043.4715	6096.004
July-22	5193.774	1045.1668	6238.941
August-22	5103.774	1018.8913	6122.665
September-22	4932.877	1054.8119	5987.689
October-22	5028.084	1091.0795	6119.164
November-22	5148.636	1093.6229	6242.259
December-22	6211.533	1161.7581	7373.291
January-23	4642.282	1087.0138	5729.296
February-23	4645.515	1033.436	5678.951
March-23	4958.315	1023.8937	5982.209
April-23	4948.182	1004.8536	5953.036

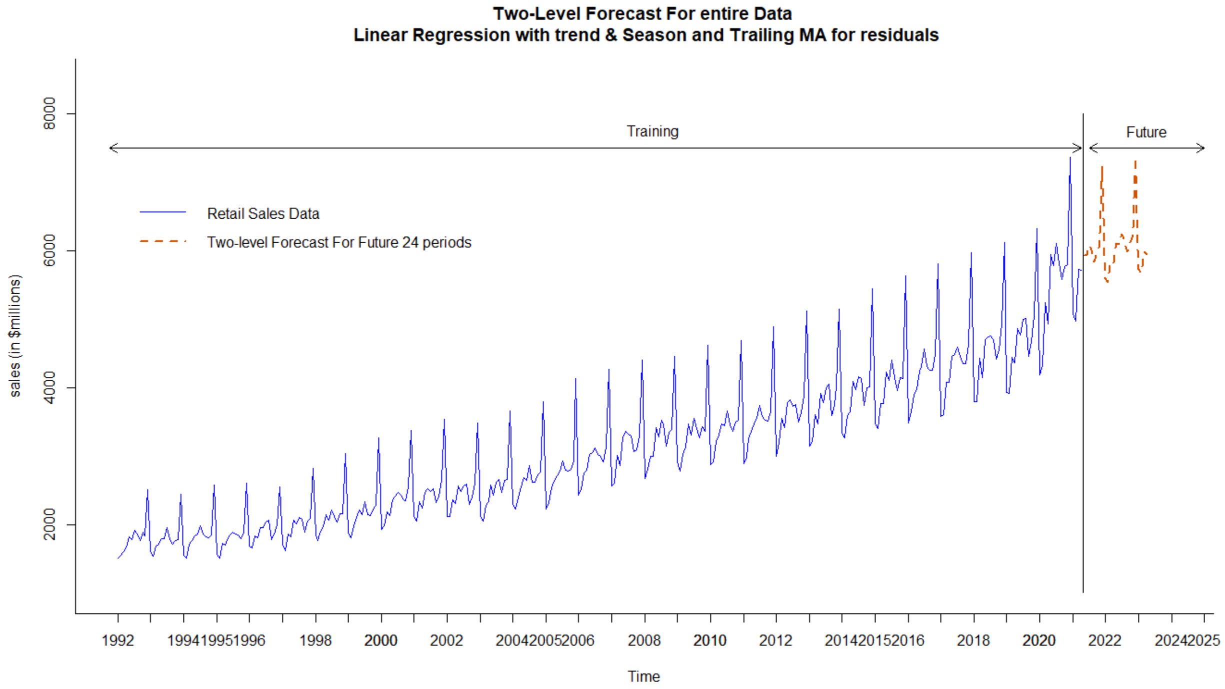
Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

Below table shows accuracy for linear regression model with trend and seasonality and two-level forecast for entire data.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Two level forecast entire data (Linear Regression + Trailing MA)	5.211	136.681	95.872	-0.024	3.174	0.145	0.287
Linear Regression Model with trend and seasonality on entire data	0	263.526	179.829	0.25	6.242	0.711	0.547

As we can observe from the above table two-level forecast performs better than the Linear regression model with trend and seasonality as it has low MAPE and RMSE values. So, we can conclude that two-level forecast can be used for forecasting into future periods.

Visualizing Two-level forecast model for future 24 periods:



- **Quadratic model with linear trend and seasonality:**

The model will fit a quadratic trend and additive seasonality that applies to the entire data and will use model for forecast future periods. The seasonality is indicated as dummy variables. For monthly data we have 11 dummy variables namely D2,D3,D4….D12.Output is measured using below equation

$$Yt = \beta 0 + \beta 1\, t + \beta 2\, t^2 + \beta 3\, D2 + \beta 4\, D3 \ldots \beta 12\, D12 + \varepsilon$$

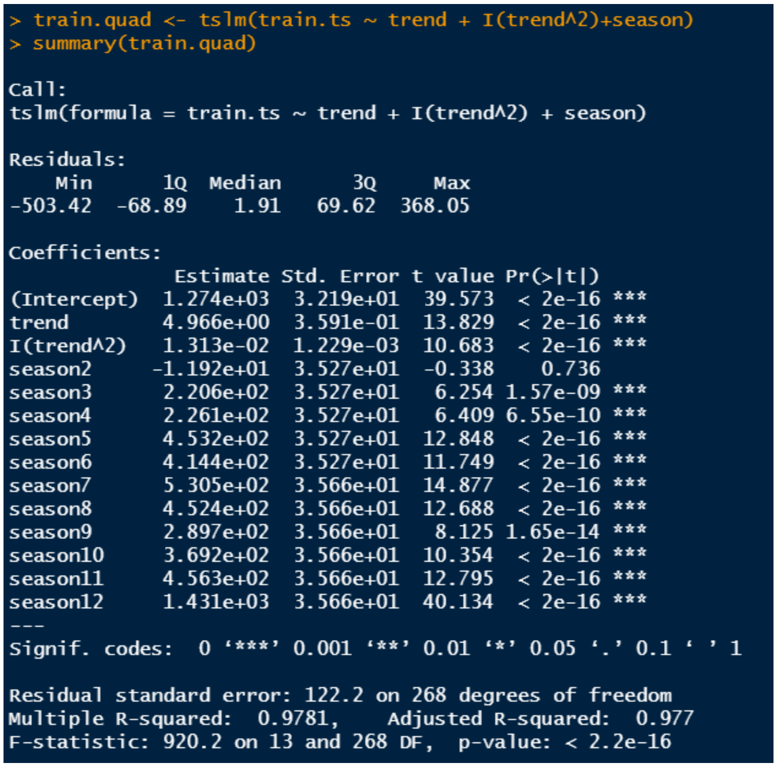
where t represents periods i.e. 1,2,3, ….

Y_t represents the output variable for time series measurement

β_0 is intercept and $\beta_1, \beta_2, \beta_3,$ ….β_{12} represents coefficients of the equation

Dummy variables take values as per season as mentioned in below table

	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
January	0	0	0	0	0	0	0	0	0	0	0
February	1	0	0	0	0	0	0	0	0	0	0
March	0	1	0	0	0	0	0	0	0	0	0
April	0	0	1	0	0	0	0	0	0	0	0
May	0	0	0	1	0	0	0	0	0	0	0
June	0	0	0	0	1	0	0	0	0	0	0
July	0	0	0	0	0	1	0	0	0	0	0
August	0	0	0	0	0	0	1	0	0	0	0
September	0	0	0	0	0	0	0	1	0	0	0
October	0	0	0	0	0	0	0	0	1	0	0
November	0	0	0	0	0	0	0	0	0	1	0
December	0	0	0	0	0	0	0	0	0	0	1



```
> train.quad <- tslm(train.ts ~ trend + I(trend^2)+season)
> summary(train.quad)

Call:
tslm(formula = train.ts ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-503.42  -68.89    1.91   69.62  368.05 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.274e+03  3.219e+01  39.573  < 2e-16 ***
trend        4.966e+00  3.591e-01  13.829  < 2e-16 ***
I(trend^2)   1.313e-02  1.229e-03  10.683  < 2e-16 ***
season2     -1.192e+01  3.527e+01  -0.338    0.736    
season3      2.206e+02  3.527e+01   6.254 1.57e-09 ***
season4      2.261e+02  3.527e+01   6.409 6.55e-10 ***
season5      4.532e+02  3.527e+01  12.848  < 2e-16 ***
season6      4.144e+02  3.527e+01  11.749  < 2e-16 ***
season7      5.305e+02  3.566e+01  14.877  < 2e-16 ***
season8      4.524e+02  3.566e+01  12.688  < 2e-16 ***
season9      2.897e+02  3.566e+01   8.125 1.65e-14 ***
season10     3.692e+02  3.566e+01  10.354  < 2e-16 ***
season11     4.563e+02  3.566e+01  12.795  < 2e-16 ***
season12     1.431e+03  3.566e+01  40.134  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 122.2 on 268 degrees of freedom
Multiple R-squared:  0.9781,	Adjusted R-squared:  0.977 
F-statistic: 920.2 on 13 and 268 DF,  p-value: < 2.2e-16
```

From the model summary we can interpret that 1273.910 is the intercept of the model and is significant as p-value is very low. 0.0131 and -11.920 are coefficients of trend(t) and trend^2 (t2) respectively and both are statically significant. Among all the seasons season2 is not statistically significant as p-value is greater than 0.05.

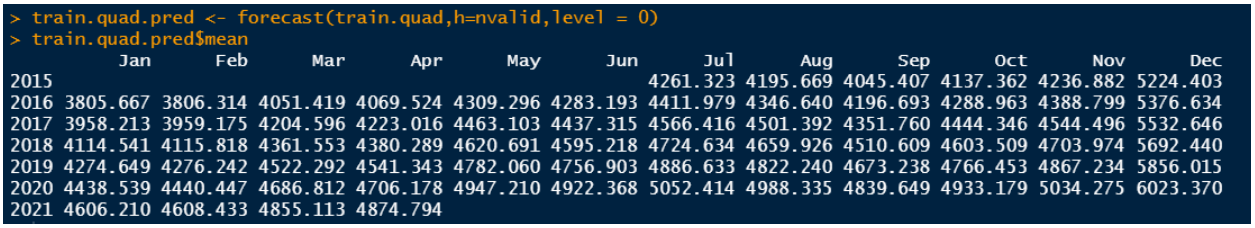
Using the intercept and coefficients we can build the equation to predict sales (Y_t) for future time periods.

$$Yt = 1273.910 + 4.9661*t + 0.0131*t^2 - 11.920*D2 + 220.590*D3 + 226075*D4 + 453.201*D5 + 414.425*D6 + 530.511*D7 + 452.447*D8 + 289.748*D9 + 269.241*D10 + 456.272*D11 + 1431.277*D12$$

Substituting respective values for dummy variables depending on the season and time period (t) in the above equation will provide us sales for desired time period.

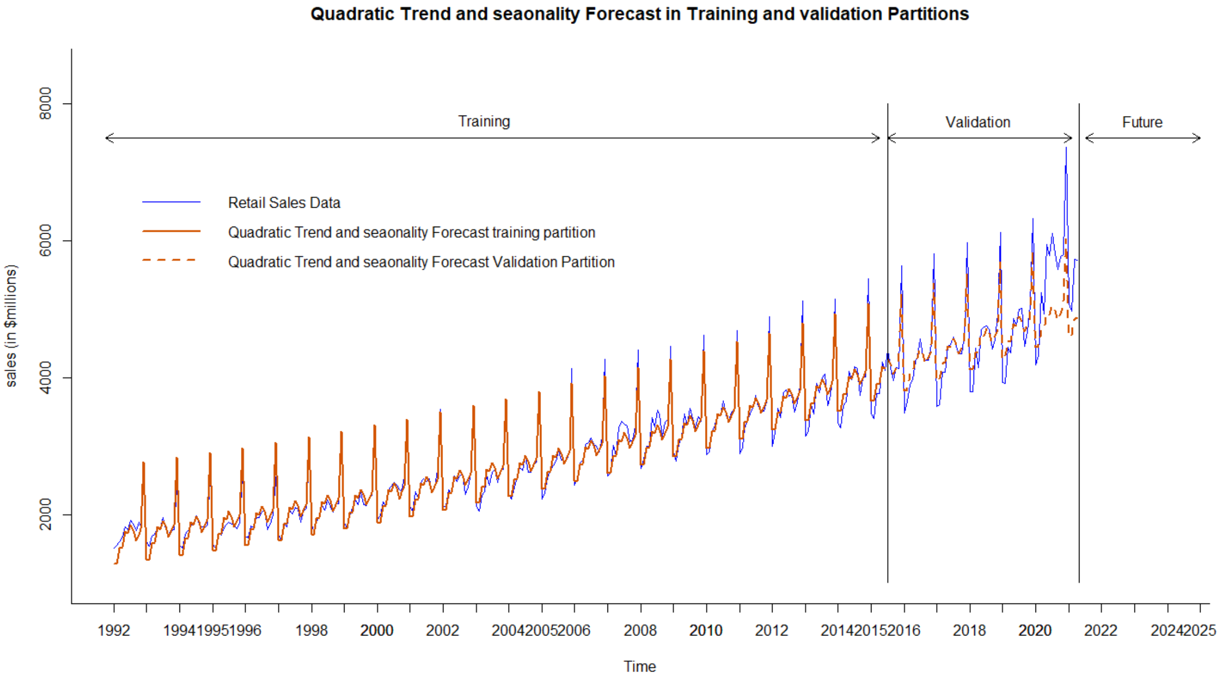
R^2 value is 0.9781 which indicates that independent variables can explain 97.81% of variance in Y_t ,which is a very good for the model. Adjusted R-squared is also high around 97.7%. F-statistic is 920.2 and is statistically significant which indicates that model is fit and independent variables can explain the dependent variable (Y_t) well.

Below is the point forecasted values for validation period using Quadratic trend and seasonality model.



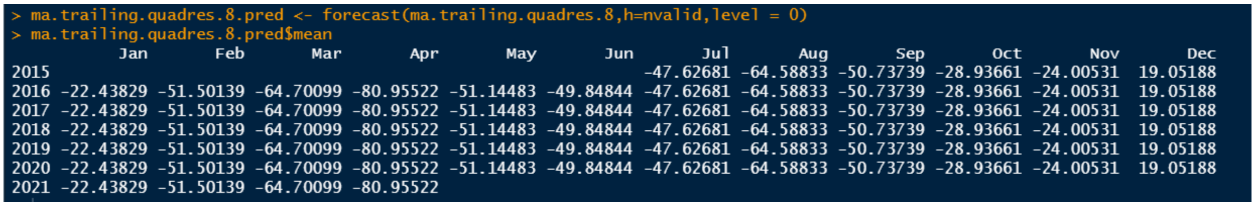
```
> train.quad.pred <- forecast(train.quad,h=nvalid,level = 0)
> train.quad.pred$mean
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015                                                    4261.323 4195.669 4045.407 4137.362 4236.882 5224.403
2016 3805.667 3806.314 4051.419 4069.524 4309.296 4283.193 4411.979 4346.640 4196.693 4288.963 4388.799 5376.634
2017 3958.213 3959.175 4204.596 4223.016 4463.103 4437.315 4566.416 4501.392 4351.760 4444.346 4544.496 5532.646
2018 4114.541 4115.818 4361.553 4380.289 4620.691 4595.218 4724.634 4659.926 4510.609 4603.509 4703.974 5692.440
2019 4274.649 4276.242 4522.292 4541.343 4782.060 4756.903 4886.633 4822.240 4673.238 4766.453 4867.234 5856.015
2020 4438.539 4440.447 4686.812 4706.178 4947.210 4922.368 5052.414 4988.335 4839.649 4933.179 5034.275 6023.370
2021 4606.210 4608.433 4855.113 4874.794
```

Visualizing retail sales using Quadratic model:



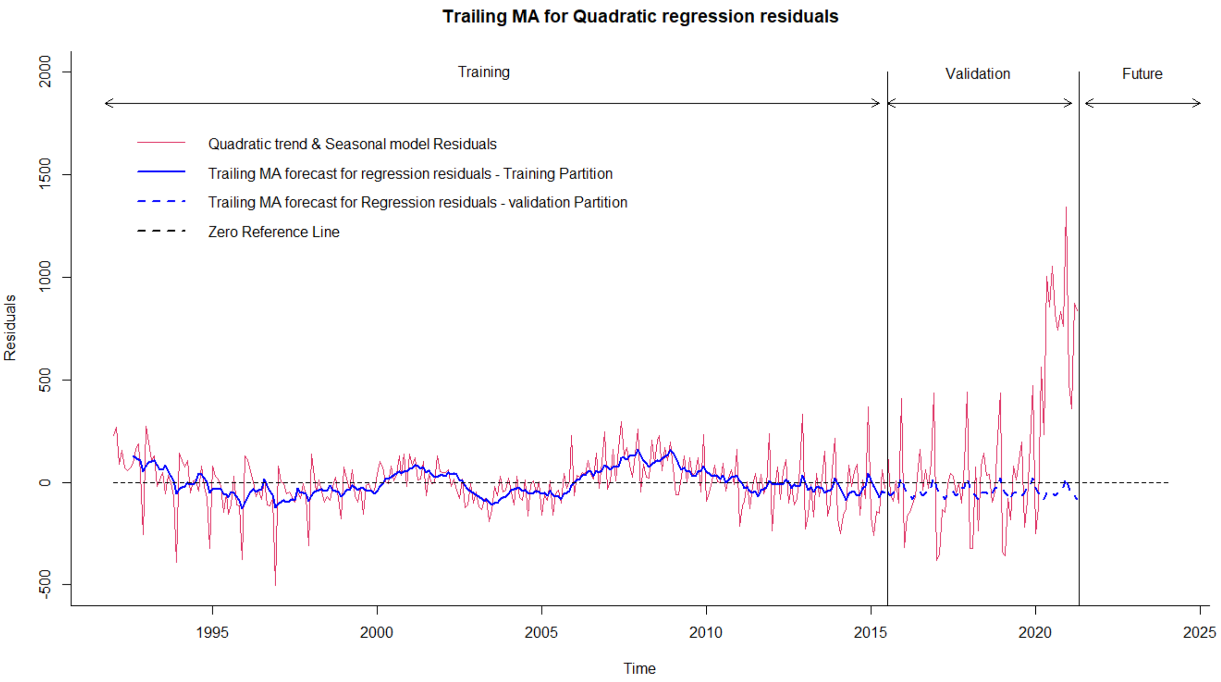
Trailing MA for Quadratic Model Residuals:

Below table represents point forecasted values of residuals for validation period using trailing moving average model with K=8.



```
> ma.trailing.quadres.8.pred <- forecast(ma.trailing.quadres.8,h=nvalid,level = 0)
> ma.trailing.quadres.8.pred$mean
           Jan       Feb       Mar       Apr       May       Jun       Jul       Aug       Sep       Oct       Nov       Dec
2015                                                              -47.62681 -64.58833 -50.73739 -28.93661 -24.00531  19.05188
2016 -22.43829 -51.50139 -64.70099 -80.95522 -51.14483 -49.84844 -47.62681 -64.58833 -50.73739 -28.93661 -24.00531  19.05188
2017 -22.43829 -51.50139 -64.70099 -80.95522 -51.14483 -49.84844 -47.62681 -64.58833 -50.73739 -28.93661 -24.00531  19.05188
2018 -22.43829 -51.50139 -64.70099 -80.95522 -51.14483 -49.84844 -47.62681 -64.58833 -50.73739 -28.93661 -24.00531  19.05188
2019 -22.43829 -51.50139 -64.70099 -80.95522 -51.14483 -49.84844 -47.62681 -64.58833 -50.73739 -28.93661 -24.00531  19.05188
2020 -22.43829 -51.50139 -64.70099 -80.95522 -51.14483 -49.84844 -47.62681 -64.58833 -50.73739 -28.93661 -24.00531  19.05188
2021 -22.43829 -51.50139 -64.70099 -80.95522
```

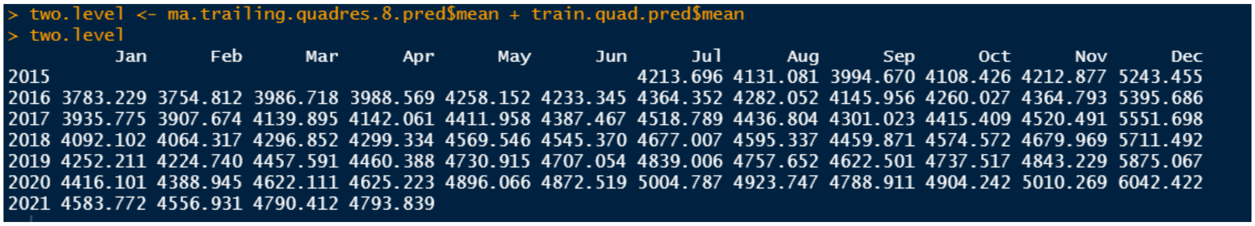
Visualizing Trailing MA residuals for Training and Validation partition:



Two-level Forecast with Trailing MA for residuals:

Now once residuals are forecasted into validation, both residual forecast for validation and quadratic model forecast for validation are combined to from a two-level forecast.

Below table shows forecasted values for validation period by two-level forecast.



```
> two.level <- ma.trailing.quadres.8.pred$mean + train.quad.pred$mean
> two.level
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015                                                       4213.696 4131.081 3994.670 4108.426 4212.877 5243.455
2016 3783.229 3754.812 3986.718 3988.569 4258.152 4233.345 4364.352 4282.052 4145.956 4260.027 4364.793 5395.686
2017 3935.775 3907.674 4139.895 4142.061 4411.958 4387.467 4518.789 4436.804 4301.023 4415.409 4520.491 5551.698
2018 4092.102 4064.317 4296.852 4299.334 4569.546 4545.370 4677.007 4595.337 4459.871 4574.572 4679.969 5711.492
2019 4252.211 4224.740 4457.591 4460.388 4730.915 4707.054 4839.006 4757.652 4622.501 4737.517 4843.229 5875.067
2020 4416.101 4388.945 4622.111 4625.223 4896.066 4872.519 5004.787 4923.747 4788.911 4904.242 5010.269 6042.422
2021 4583.772 4556.931 4790.412 4793.839
```

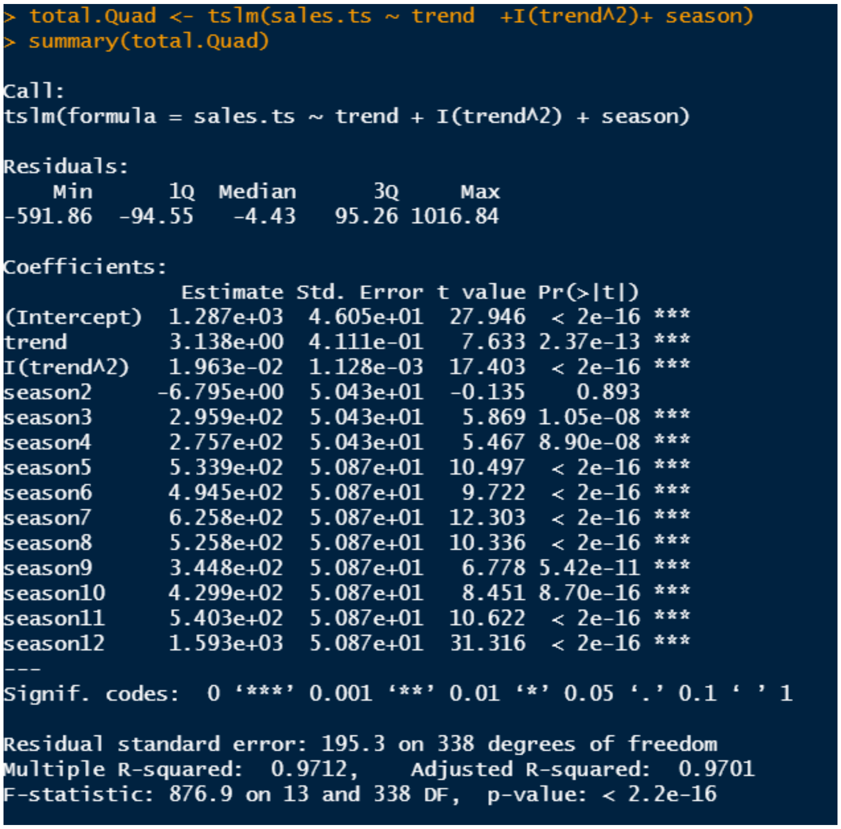
Below table shows accuracy of two-level forecast and regression model for trend and seasonality.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Quadratic trend and seasonal model	140.607	409.992	282.275	1.92	5.509	0.619	0.516
Two Level Forecast Quadratic + Trailing MA for residuals	183.516	422.357	285.545	2.879	5.502	0.637	0.527

As we can observe two-level forecast has produced a good MAPE compared to the quadratic trend and seasonality in the validation period.

Two-level forecast for entire data with Trailing MA for Residuals:

Below is the summary of quadratic model for entire data.



```
> total.Quad <- tslm(sales.ts ~ trend  +I(trend^2)+ season)
> summary(total.Quad)

Call:
tslm(formula = sales.ts ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-591.86  -94.55   -4.43   95.26 1016.84 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.287e+03  4.605e+01  27.946  < 2e-16 ***
trend        3.138e+00  4.111e-01   7.633 2.37e-13 ***
I(trend^2)   1.963e-02  1.128e-03  17.403  < 2e-16 ***
season2     -6.795e+00  5.043e+01  -0.135    0.893    
season3      2.959e+02  5.043e+01   5.869 1.05e-08 ***
season4      2.757e+02  5.043e+01   5.467 8.90e-08 ***
season5      5.339e+02  5.087e+01  10.497  < 2e-16 ***
season6      4.945e+02  5.087e+01   9.722  < 2e-16 ***
season7      6.258e+02  5.087e+01  12.303  < 2e-16 ***
season8      5.258e+02  5.087e+01  10.336  < 2e-16 ***
season9      3.448e+02  5.087e+01   6.778 5.42e-11 ***
season10     4.299e+02  5.087e+01   8.451 8.70e-16 ***
season11     5.403e+02  5.087e+01  10.622  < 2e-16 ***
season12     1.593e+03  5.087e+01  31.316  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 195.3 on 338 degrees of freedom
Multiple R-squared:  0.9712,	Adjusted R-squared:  0.9701 
F-statistic: 876.9 on 13 and 338 DF,  p-value: < 2.2e-16
```

From the model summary we can interpret that 1286.893 is the intercept of the model and is significant as p-value is very low. 3.138 and 0.0196 are coefficients of trend(t) and trend^2 (t2) respectively and both are statically significant. Among all the seasons season2 is not statistically significant as p-value is greater than 0.05.

Using the intercept and coefficients we can build the equation to predict sales (Y_t) for future time periods.

Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

Yt = 1286.893+ 3.138*t + 0.0196 *t^{a} -6.794 *D2 -6.794 *D3 + 295.938 *D4 + 275.698*D5 + 533.938*D6 + 494.540 *D7 + 525.765*D8 + 344.801*D9 + 429.901*D10 + 540.307*D11 + 1593.019*D12

Substituting respective values for dummy variables depending on the season and time period (t) in the above equation will provide us sales for desired time period.

R^2 value is 0.9712 which indicates that independent variables can explain 97.12% of variance in Y_t ,which is a very good for the model. Adjusted R-squared is also high around 97.01%. F-statistic is 876.9 and is statistically significant which indicates that model is fit and independent variables can explain the dependent variable (Y_t) well.

Now after developing Quadratic model with trend and seasonality ,the residuals of the model are used for developing a trailing moving average model with K=8.

Below table shows values for future 24 periods for quadratic model forecast, residuals forecast and combined forecast

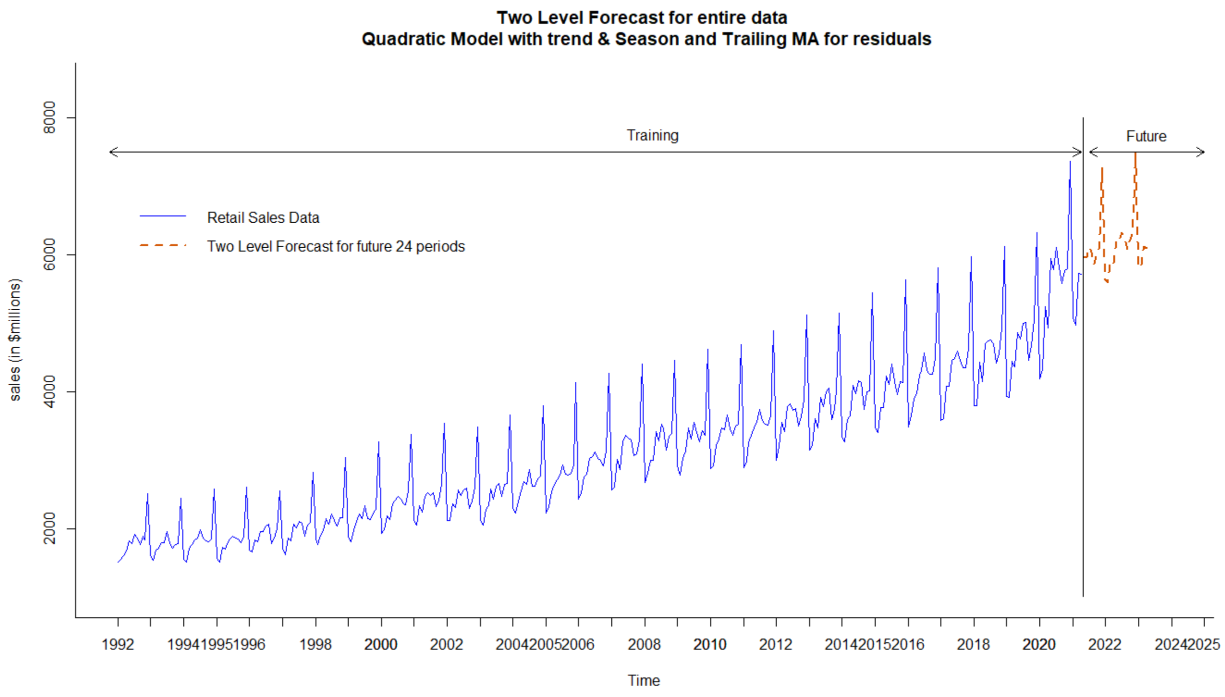
Time	Quadratic Trend and seasonality Forecast	Trailing MA residual Forecast	Total Forecast
May-21	5374.57	592.6425	5967.213
June-21	5352.189	605.5098	5957.699
July-21	5500.497	607.6195	6108.116
August-21	5417.563	580.8943	5998.458
September-21	5253.733	615.0776	5868.811
October-21	5356.007	650.9636	6006.971
November-21	5483.625	652.7801	6136.405
December-21	6553.588	723.8307	7277.419
January-22	4977.86	653.1999	5631.06
February-22	4988.396	602.9324	5591.328
March-22	5308.498	596.4311	5904.929
April-22	5305.667	579.7417	5885.408
May-22	5581.355	615.0424	6196.397
June-22	5559.444	623.43	6182.874
July-22	5708.223	621.9558	6330.179
August-22	5625.761	592.3635	6218.124
September-22	5462.402	624.2531	6086.655
October-22	5565.147	658.3041	6223.451
November-22	5693.236	658.6526	6351.889
December-22	6763.67	728.5288	7492.199
January-23	5188.413	656.9584	5845.372
February-23	5199.42	605.9392	5805.359
March-23	5519.993	598.8366	6118.83
April-23	5517.633	581.6661	6099.299

Below table shows accuracy for Quadratic model with trend and seasonality and two-level forecast for entire data.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Quadratic trend and seasonal model Entire data	0	191.381	132.707	-0.018	4.244	0.491	0.372
Two Level Forecast Quadratic + Trailing MA for residuals	5.134	134.422	94.33	0.145	3.184	0.119	0.287

As we can observe from the above table two-level forecast performs better than the quadratic model with trend and seasonality as it has low MAPE and RMSE values. So, we can conclude that two-level forecast can be used for forecasting into future periods.

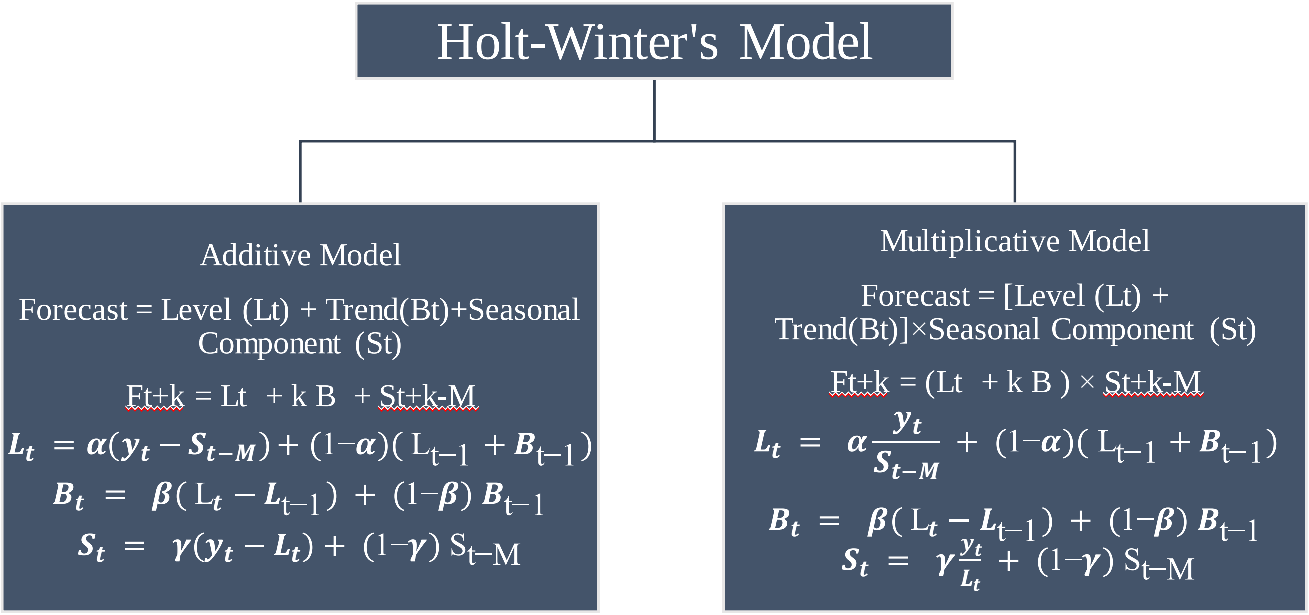
Visualizing Two-level forecast model for future 24 periods:



ii) HOLT-WINTER'S SEASONAL MODEL:

Holt-Winter's model is a data driven forecasting method and an exponential smoothing model. Forecasts produced using these exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. We can interpret this as most recent data is given higher weightage than the older data in the time series.

Holt-Winter's model can accommodate all the time series components while building a model and three smoothing equations namely level (l_t) ,trend (b_t) and season (s_t) with corresponding smoothing parameter α, β, γ.All values of the smoothing parameters lie between 0 and 1.Seasons are represented by M. There are two types of variation in Holt-Winter's model ,Additive and multiplicative.



Holt-Winter's Model

Additive Model

Forecast = Level (Lt) + Trend(Bt)+Seasonal Component (St)

Ft+k = Lt + k B + St+k-M

$$L_t = \alpha(y_t - S_{t-M}) + (1-\alpha)(L_{t-1} + B_{t-1})$$

$$B_t = \beta(L_t - L_{t-1}) + (1-\beta) B_{t-1}$$

$$S_t = \gamma(y_t - L_t) + (1-\gamma) S_{t-M}$$

Multiplicative Model

Forecast = [Level (Lt) + Trend(Bt)]×Seasonal Component (St)

Ft+k = (Lt + k B) × St+k-M

$$L_t = \alpha \frac{y_t}{S_{t-M}} + (1-\alpha)(L_{t-1} + B_{t-1})$$

$$B_t = \beta(L_t - L_{t-1}) + (1-\beta) B_{t-1}$$

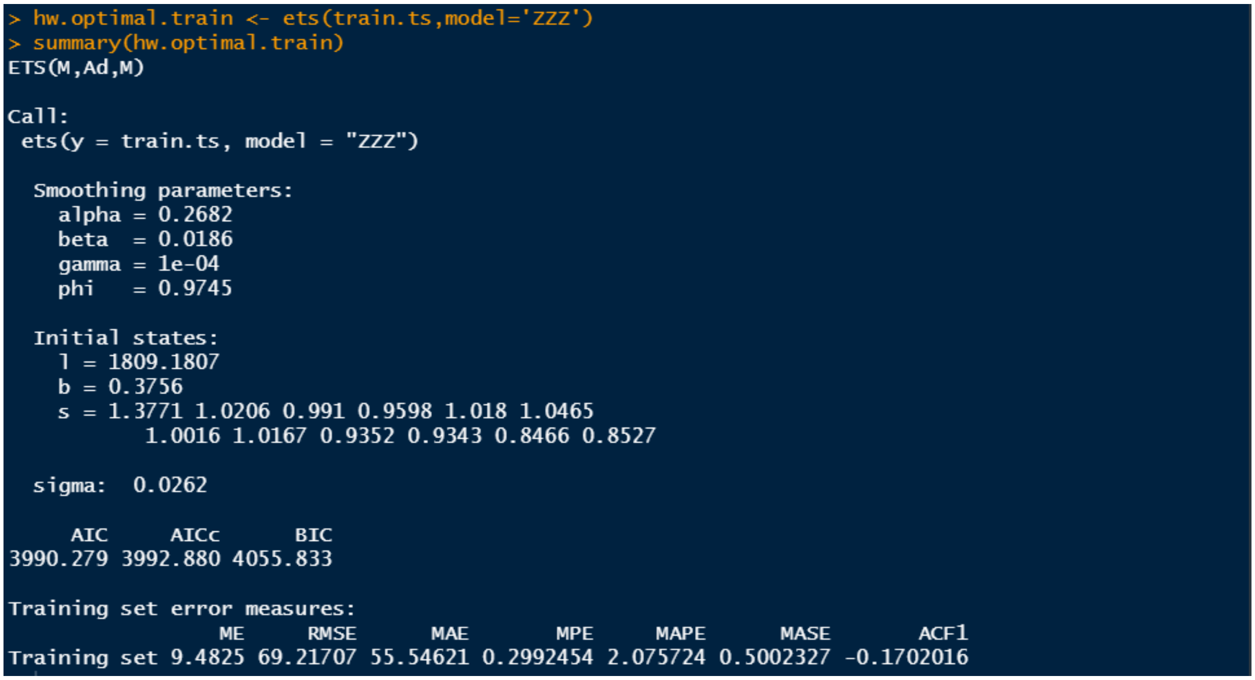
$$S_t = \gamma \frac{y_t}{L_t} + (1-\gamma) S_{t-M}$$

We can define a HoltWinter's model with automated selection of error, trend and seasonality options and automated selection of smoothing parameters by using a function ets('ZZZ') in R.27 variety combinations of models can be build using Holt-winter's model.

- First *Z* can be equal to *A* (additive) or *M* (multiplicative) or N (No) error
- Second *Z* can be equal to *A* (additive) or *M* (multiplicative) or N (No) trend
- Third Z can be equal to *A* (additive) or *M* (multiplicative) or N (No) seasonality

In Addition to these, sometimes Holts model adds a damping factor to trend which dampens the continuously increasing or decreasing trend. Damping parameter is represented by $\emptyset$. Damping is possible for both additive and multiplicative can be represented as ets(Z,A_d,Z) or ets(Z,M_d,Z) respectively.

- **Holt-Winter's Automatic Model with Optimal Parameters – Training Partition:**



```
> hw.optimal.train <- ets(train.ts,model='ZZZ')
> summary(hw.optimal.train)
ETS(M,Ad,M)

Call:
 ets(y = train.ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.2682
    beta  = 0.0186
    gamma = 1e-04
    phi   = 0.9745

  Initial states:
    l = 1809.1807
    b = 0.3756
    s = 1.3771 1.0206 0.991 0.9598 1.018 1.0465
           1.0016 1.0167 0.9352 0.9343 0.8466 0.8527

  sigma:  0.0262

     AIC     AICc      BIC
3990.279 3992.880 4055.833

Training set error measures:
                 ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 9.4825 69.21707 55.54621 0.2992454 2.075724 0.5002327 -0.1702016
```

Above summary shows the model options and smoothing parameters provided by ets('ZZZ') for training partition. The model options are **ets(M, Ad, M)** i.e. Multiplicative error/level, Additive trend with damping factor and Multiplicative seasonality and optimal smoothing parameters as below

α = 0.2682 , smoothing constant for exponential smoothing

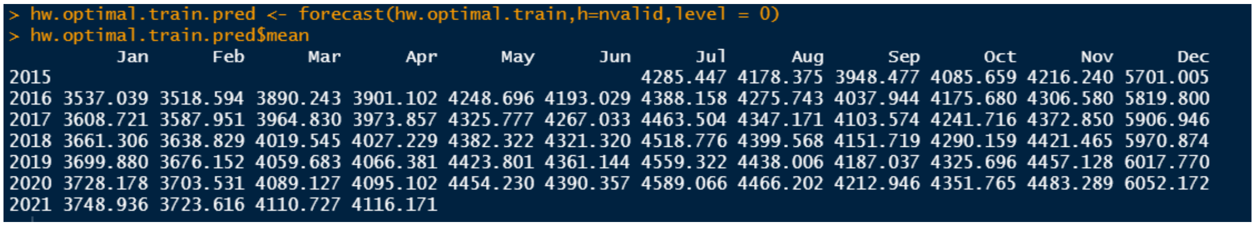
γ = 1e^-04 , smoothing constant for seasonality estimate

β = 0.0186 , smoothing constant for trend estimate

∅ = 0.9745 , damping parameter

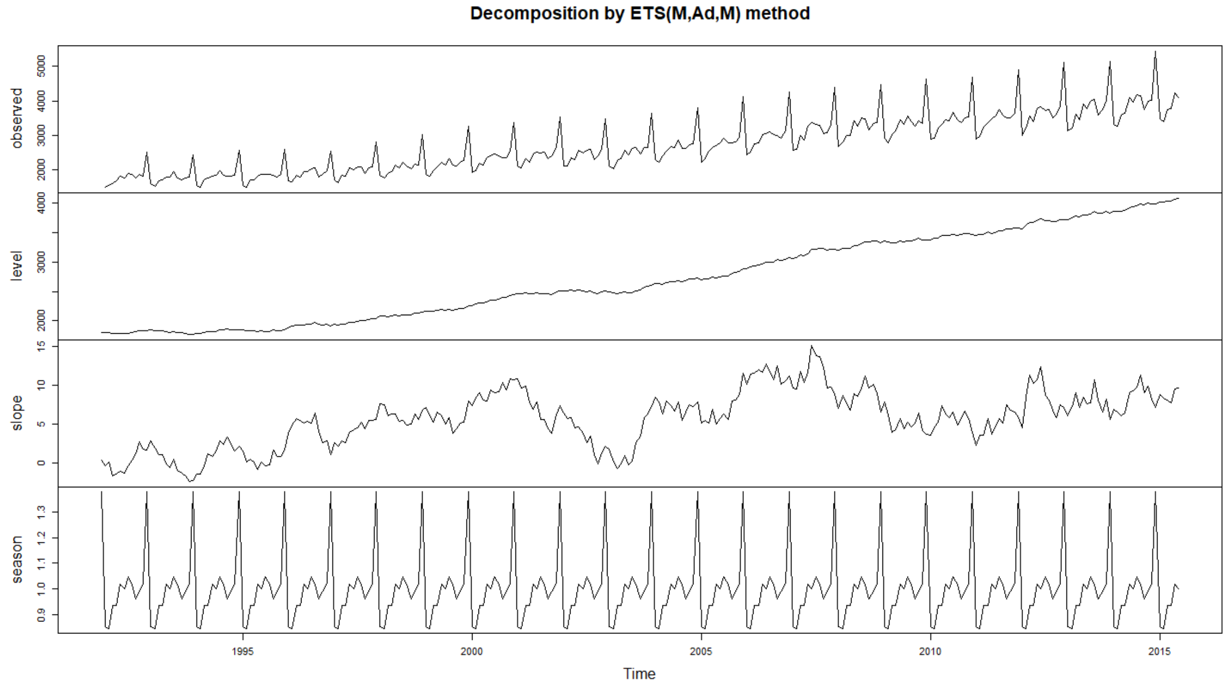
Holt-Winter's model will automatically assign the initial states for all components of time series which are used to calculate trend, level and seasonality for the data points at the beginning of the time series. As observed from the above summary l=1809.1807 is the initial state of level component , b = 0.3756 for trend component and s for seasonal component. Since data has monthly seasonality we have 12 initial values for seasonal component.

Below is the point forecasted values for validation period using training data and ETS(M,Ad,M).



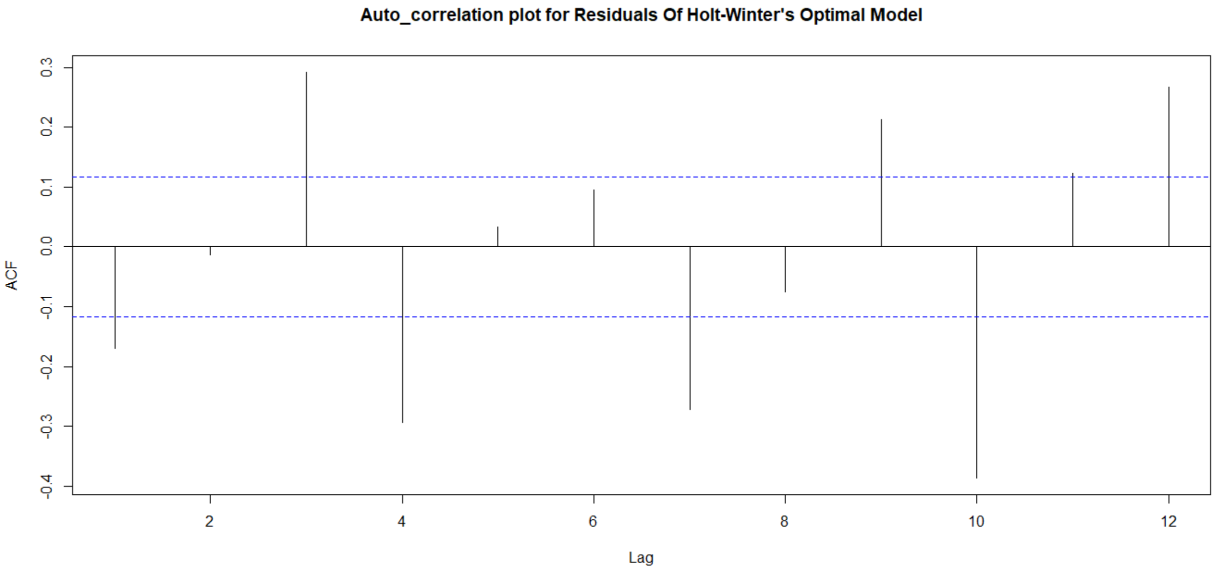
```
> hw.optimal.train.pred <- forecast(hw.optimal.train,h=nvalid,level = 0)
> hw.optimal.train.pred$mean
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015                                                        4285.447 4178.375 3948.477 4085.659 4216.240 5701.005
2016 3537.039 3518.594 3890.243 3901.102 4248.696 4193.029 4388.158 4275.743 4037.944 4175.680 4306.580 5819.800
2017 3608.721 3587.951 3964.830 3973.857 4325.777 4267.033 4463.504 4347.171 4103.574 4241.716 4372.850 5906.946
2018 3661.306 3638.829 4019.545 4027.229 4382.322 4321.320 4518.776 4399.568 4151.719 4290.159 4421.465 5970.874
2019 3699.880 3676.152 4059.683 4066.381 4423.801 4361.144 4559.322 4438.006 4187.037 4325.696 4457.128 6017.770
2020 3728.178 3703.531 4089.127 4095.102 4454.230 4390.357 4589.066 4466.202 4212.946 4351.765 4483.289 6052.172
2021 3748.936 3723.616 4110.727 4116.171
```

Components of ETS(M,Ad,M) Method-Training Data:



In the above plot level indicates the overall baseline without seasonal trends, slope indicates rate of change of level over time and season indicates seasonal trend of the data.

ACF plot for residuals:



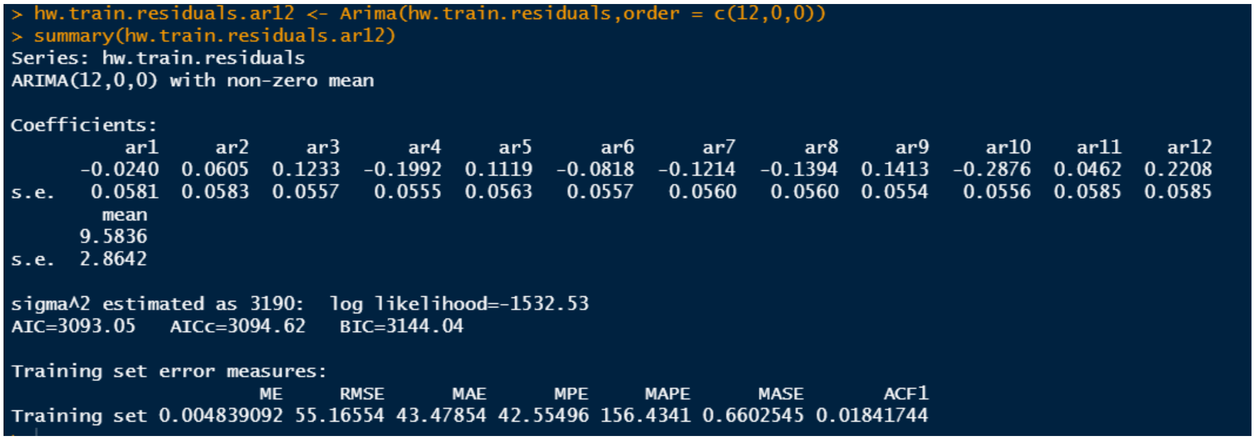
As we can observe from the above plot there are lot of significant relations in the residuals. So, to incorporate these relations or dependencies an AR() model is developed .

These residuals forecasted using AR() models is combined with holt-winter's forecast to form a two-level forecast model.

After testing various AR() models ,AR(12) is able to incorporate all the dependencies in the residuals of Holt-Winter's Automatic Model with optimal parameters.

AR Model for Holt-Winter's Residuals:

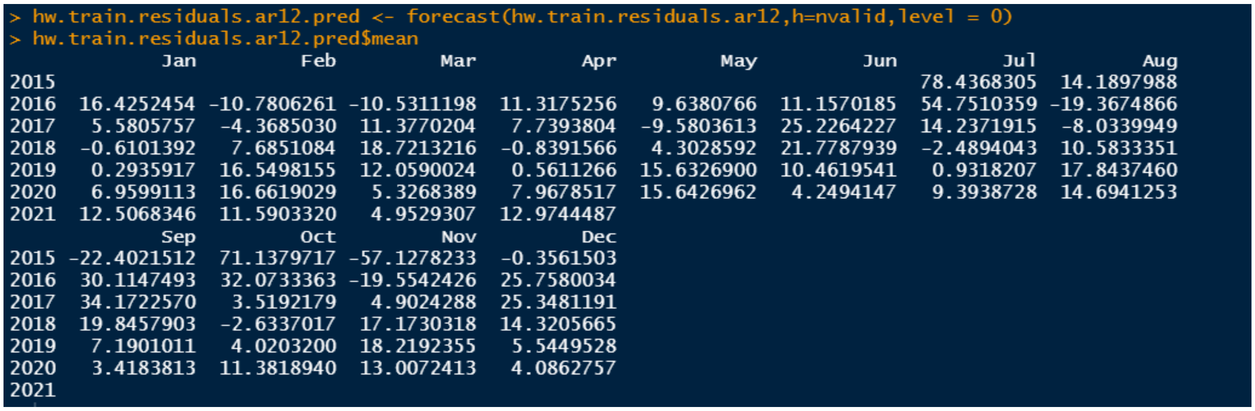
Below table shows summary of AR(12) model for Holt-Winter's model residuals:



```
> hw.train.residuals.ar12 <- Arima(hw.train.residuals,order = c(12,0,0))
> summary(hw.train.residuals.ar12)
Series: hw.train.residuals
ARIMA(12,0,0) with non-zero mean

Coefficients:
          ar1     ar2     ar3      ar4     ar5      ar6      ar7      ar8     ar9     ar10    ar11    ar12
      -0.0240  0.0605  0.1233  -0.1992  0.1119  -0.0818  -0.1214  -0.1394  0.1413  -0.2876  0.0462  0.2208
s.e.   0.0581  0.0583  0.0557   0.0555  0.0563   0.0557   0.0560   0.0560  0.0554   0.0556  0.0585  0.0585
        mean
      9.5836
s.e.  2.8642

sigma^2 estimated as 3190:  log likelihood=-1532.53
AIC=3093.05   AICc=3094.62   BIC=3144.04

Training set error measures:
                      ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 0.004839092 55.16554 43.47854 42.55496 156.4341 0.6602545 0.01841744
```

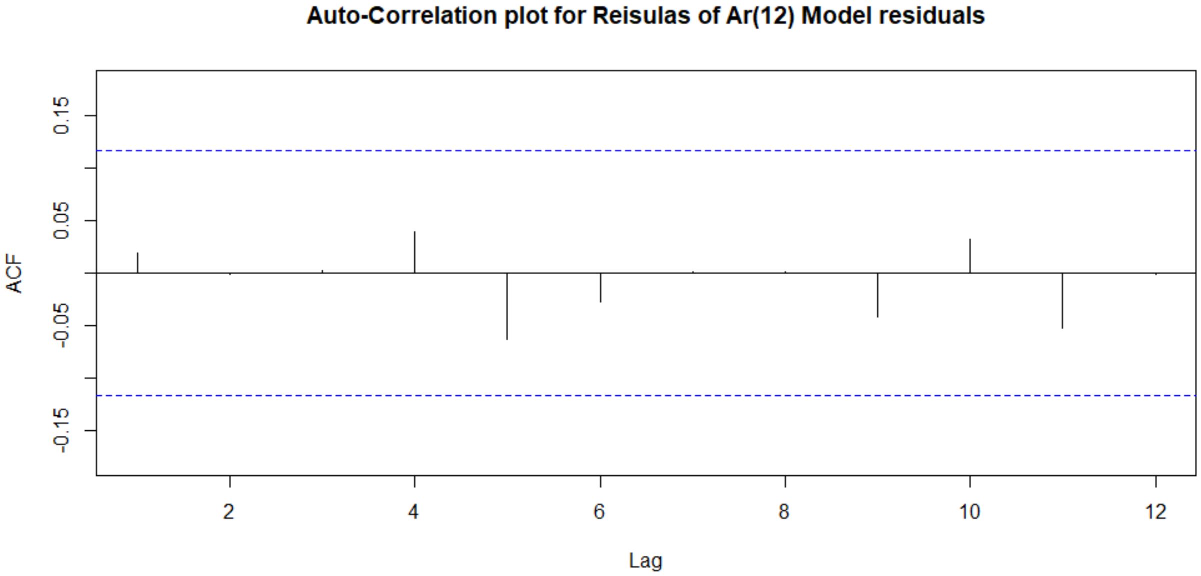
As we can observe we have 12 variables to form an AR equation. This AR model lagged 12 periods to incorporate all the dependencies in the residuals.

Below table represents point forecasted values of residuals for validation period using AR(12) model for Holt-Winter's Automatic Model with optimal parameters residuals.



```
> hw.train.residuals.ar12.pred <- forecast(hw.train.residuals.ar12,h=nvalid,level = 0)
> hw.train.residuals.ar12.pred$mean
             Jan         Feb         Mar         Apr         May         Jun         Jul         Aug
2015                                                                          78.4368305  14.1897988
2016  16.4252454 -10.7806261 -10.5311198  11.3175256   9.6380766  11.1570185  54.7510359 -19.3674866
2017   5.5805757  -4.3685030  11.3770204   7.7393804  -9.5803613  25.2264227  14.2371915  -8.0339949
2018  -0.6101392   7.6851084  18.7213216  -0.8391566   4.3028592  21.7787939  -2.4894043  10.5833351
2019   0.2935917  16.5498155  12.0590024   0.5611266  15.6326900  10.4619541   0.9318207  17.8437460
2020   6.9599113  16.6619029   5.3268389   7.9678517  15.6426962   4.2494147   9.3938728  14.6941253
2021  12.5068346  11.5903320   4.9529307  12.9744487
             Sep         Oct         Nov         Dec
2015 -22.4021512  71.1379717 -57.1278233  -0.3561503
2016  30.1147493  32.0733363 -19.5542426  25.7580034
2017  34.1722570   3.5192179   4.9024288  25.3481191
2018  19.8457903  -2.6337017  17.1730318  14.3205665
2019   7.1901011   4.0203200  18.2192355   5.5449528
2020   3.4183813  11.3818940  13.0072413   4.0862757
2021
```

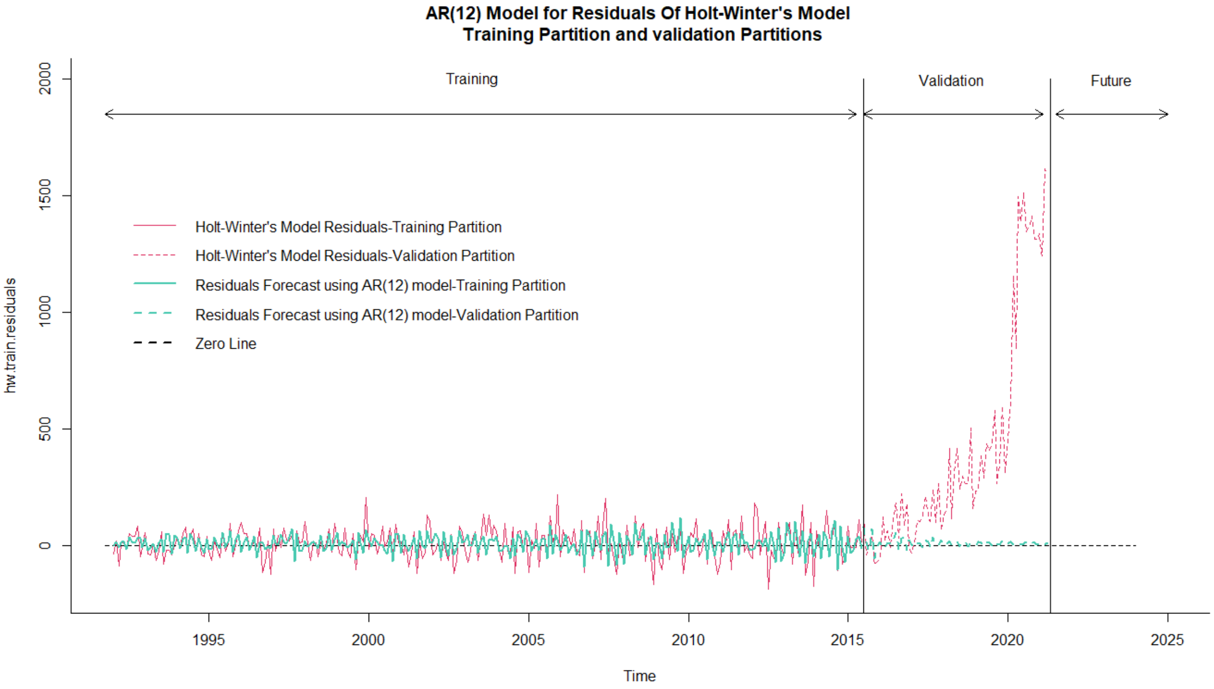
Below plot shows Autocorrelation plot for AR(12) model residuals i.e. residuals of residuals.



Since most of the dependencies are incorporated into the model we can combine AR(12) forecasted residuals with Holt-Winter's model's forecasted values in validation periods to form a two-level forecasted model.

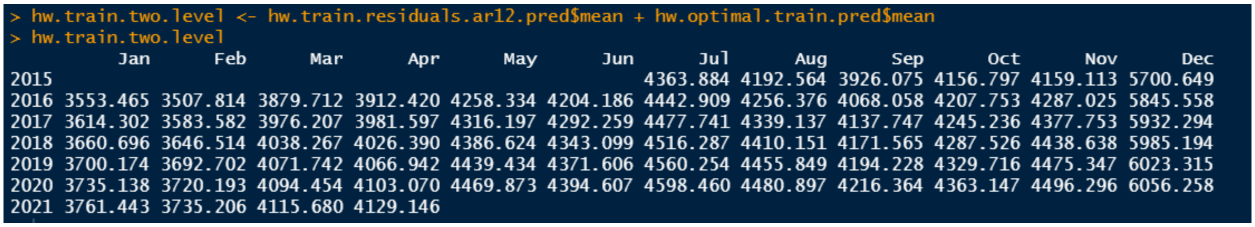
Visualizing AR(12) Model Forecast for training and validation:

Below plot show Holt-winter's residuals and 'AR(12) for residuals' model forecast for training and validation partitions.



Two-level Forecast with AR(12) Model for residuals – Validation Partition:

To develop a two-level forecast , AR(12) forecasted residuals for validation period and Holt Winter's forecasted values for validation period are combined to form a combined forecast. Below table shows forecasted values for validation period by two-level forecast.

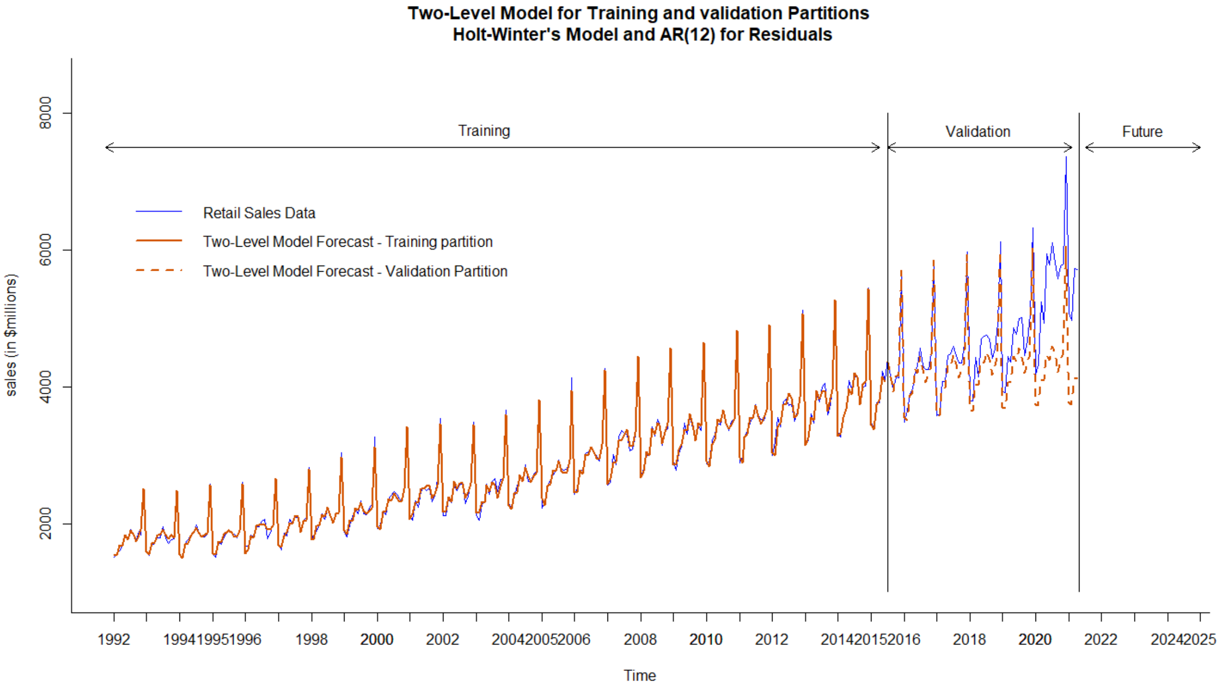


```
> hw.train.two.level <- hw.train.residuals.ar12.pred$mean + hw.optimal.train.pred$mean
> hw.train.two.level
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015                                                        4363.884 4192.564 3926.075 4156.797 4159.113 5700.649
2016 3553.465 3507.814 3879.712 3912.420 4258.334 4204.186 4442.909 4256.376 4068.058 4207.753 4287.025 5845.558
2017 3614.302 3583.582 3976.207 3981.597 4316.197 4292.259 4477.741 4339.137 4137.747 4245.236 4377.753 5932.294
2018 3660.696 3646.514 4038.267 4026.390 4386.624 4343.099 4516.287 4410.151 4171.565 4287.526 4438.638 5985.194
2019 3700.174 3692.702 4071.742 4066.942 4439.434 4371.606 4560.254 4455.849 4194.228 4329.716 4475.347 6023.315
2020 3735.138 3720.193 4094.454 4103.070 4469.873 4394.607 4598.460 4480.897 4216.364 4363.147 4496.296 6056.258
2021 3761.443 3735.206 4115.680 4129.146
```

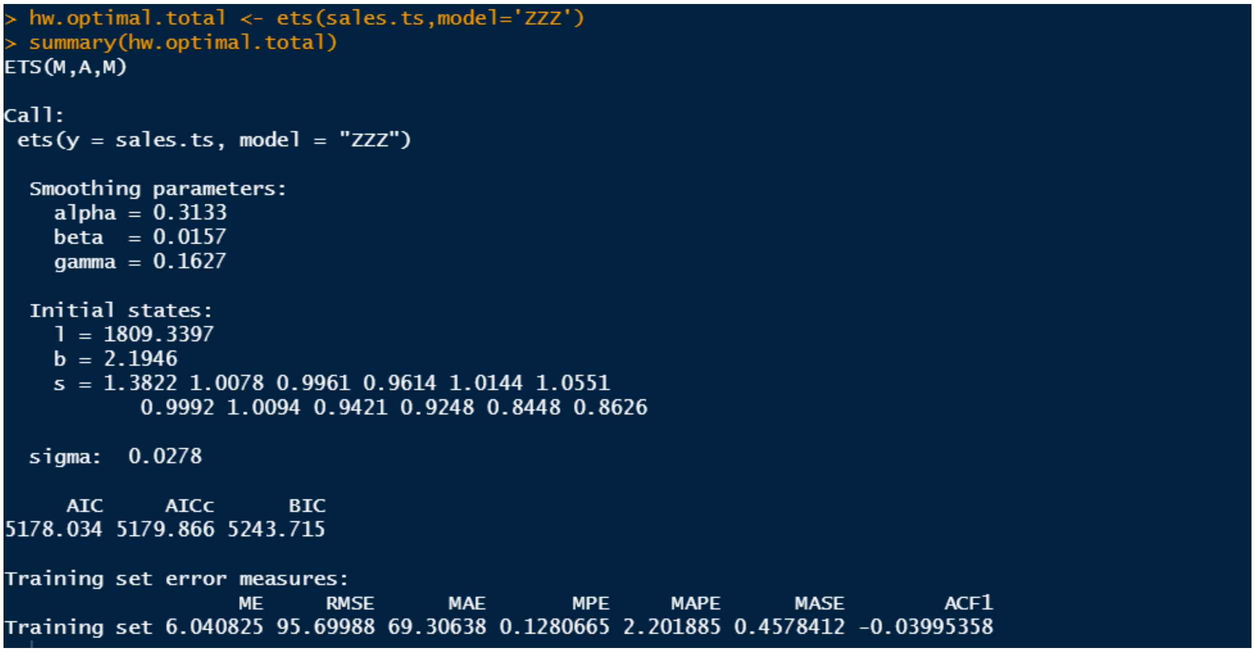
Below tables shows accuracy measures of Holt-Winter's model and two-level model (Holt-Winter's + AR(12) for residuals) for training and validation partitions.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Holt-Winter's Model	432.41	656.251	440.435	8.295	8.485	0.892	0.82
Two Level Forecast (Holt-Winters Model + AR(12) Model for residuals)	422.454	649.41	431.219	8.086	8.29	0.894	0.81

Visualizing retail sales using Two-Level model – Training & Validation Partition:



- **Holt-Winter's Automatic Model with optimal parameters –Entire Data:**



```
> hw.optimal.total <- ets(sales.ts,model='ZZZ')
> summary(hw.optimal.total)
ETS(M,A,M)

Call:
 ets(y = sales.ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.3133
    beta  = 0.0157
    gamma = 0.1627

  Initial states:
    l = 1809.3397
    b = 2.1946
    s = 1.3822 1.0078 0.9961 0.9614 1.0144 1.0551
           0.9992 1.0094 0.9421 0.9248 0.8448 0.8626

  sigma:  0.0278

     AIC     AICc      BIC
5178.034 5179.866 5243.715

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 6.040825 95.69988 69.30638 0.1280665 2.201885 0.4578412 -0.03995358
```

Above summary shows the model options and smoothing parameters provided by ets('ZZZ') for training partition. The model options are **ets(M, A, M)** i.e. Multiplicative error/level, Additive trend and Multiplicative seasonality and optimal smoothing parameters as below

α = 0.3133 , smoothing constant for exponential smoothing

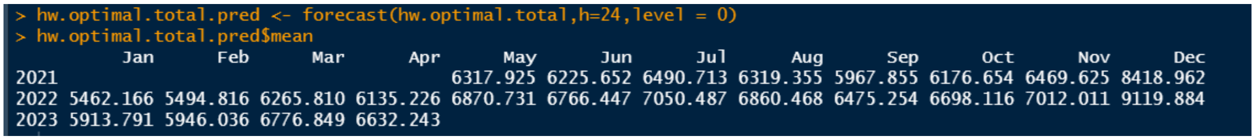
γ = 0.1627 , smoothing constant for seasonality estimate

β = 0.0157 , smoothing constant for trend estimate

∅ = 0, damping parameter

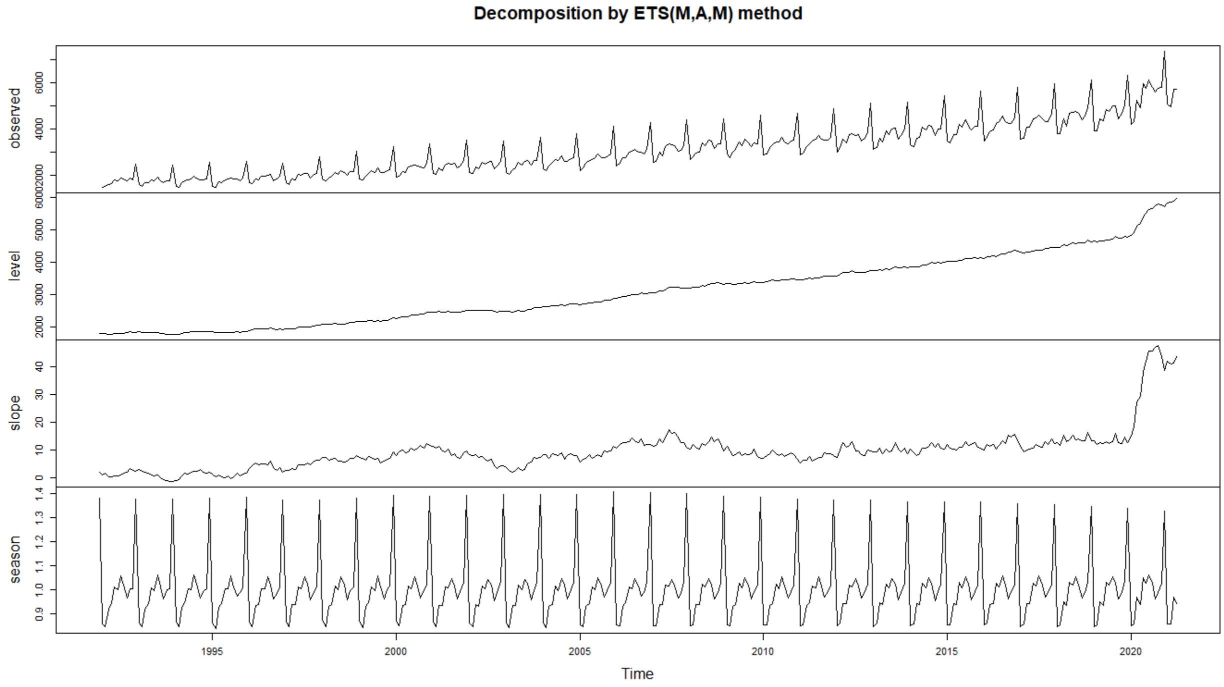
Holt-Winter's model will automatically assign the initial states for all components of time series which are used to calculate trend, level and seasonality for the data points at the beginning of the time series. As observed from the above summary l=1809.3397 is the initial state of level component , b = 2.1946 for trend component and s for seasonal component. Since data has monthly seasonality we have 12 initial values for seasonal component.

Below is the point forecasted values for future 24 periods using entire data and ETS('MAM').



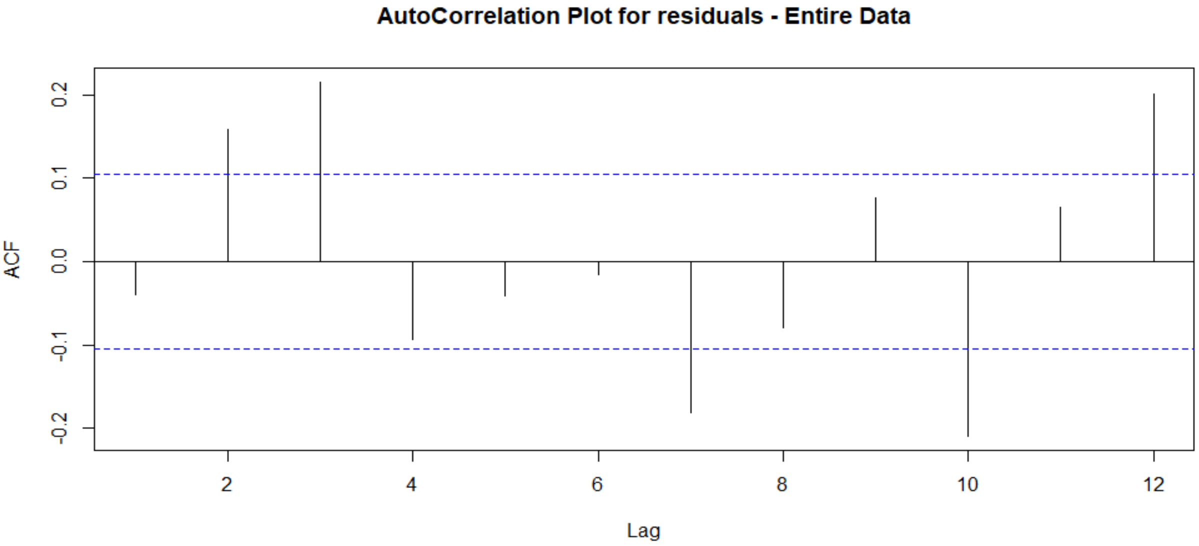
```
> hw.optimal.total.pred <- forecast(hw.optimal.total,h=24,level = 0)
> hw.optimal.total.pred$mean
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2021                                     6317.925 6225.652 6490.713 6319.355 5967.855 6176.654 6469.625 8418.962
2022 5462.166 5494.816 6265.810 6135.226 6870.731 6766.447 7050.487 6860.468 6475.254 6698.116 7012.011 9119.884
2023 5913.791 5946.036 6776.849 6632.243
```

Components of ETS(M,A,M) method - Entire Data:



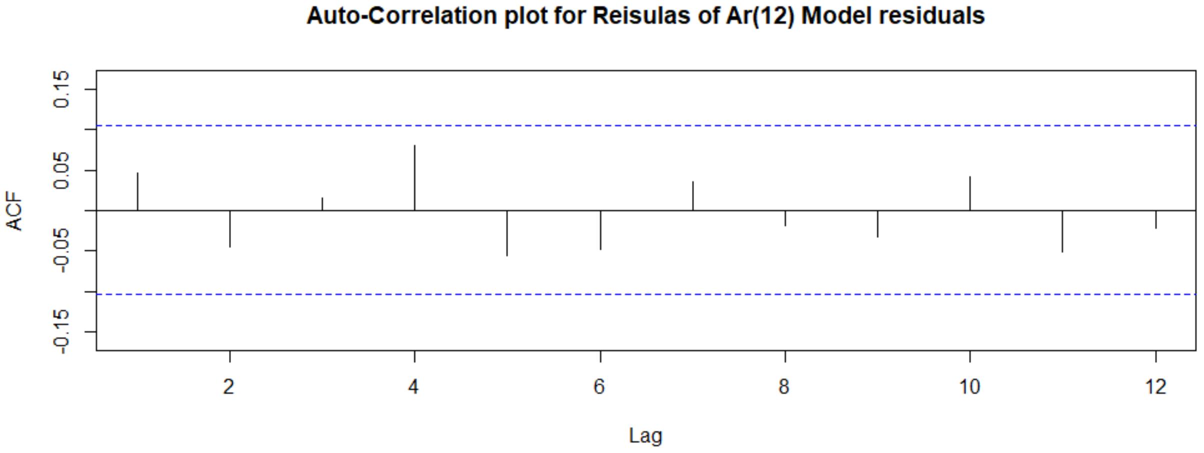
In the above plot level indicates the overall baseline without seasonal trends, slope indicates rate of change of level over time and season indicates seasonal trend of the data.

Two-level Forecast with AR(12) Model for residuals – Entire Data:



As we can observe from the above auto correlation plot there are significant relations still exists in the residuals of Holt-winter's model for entire data. So, to incorporate these AR(12) model is created with Holt-Winter's residuals and a two -level forecast is created to forecast the future values.

Below plot shows autocorrelation plot for residuals of AR(12) model for Holt-Winter's model for residuals for entire data. As we can observe there are no significant relations in the residuals of residuals.



To develop a two-level forecast , AR(12) forecasted residuals for future periods and Holt Winter's forecasted values for future periods are combined to form a combined forecast. Below table shows forecasted values for future 24 periods by two-level forecast.

Time	Holt-Winter's Forecast	AR(12) Forecast for residuals	Combined forecast
May-21	6317.925	185.333675	6503.258
June-21	6225.652	210.27018	6435.922
July-21	6490.713	181.481925	6672.195
August-21	6319.355	34.280103	6353.636
September-21	5967.855	126.636145	6094.491
October-21	6176.654	99.752503	6276.406
November-21	6469.625	-171.260693	6298.364
December-21	8418.962	-113.021195	8305.941
January-22	5462.166	-1.483266	5460.683
February-22	5494.816	-125.454506	5369.362
March-22	6265.81	-30.713697	6235.096
April-22	6135.226	-4.949783	6130.276
May-22	6870.731	24.16576	6894.897
June-22	6766.447	116.477349	6882.924
July-22	7050.487	75.590152	7126.077
August-22	6860.468	36.470186	6896.938
September-22	6475.254	132.936753	6608.191
October-22	6698.116	65.243766	6763.36
November-22	7012.011	-41.304053	6970.707
December-22	9119.884	12.477293	9132.361
January-23	5913.791	-25.108526	5888.682

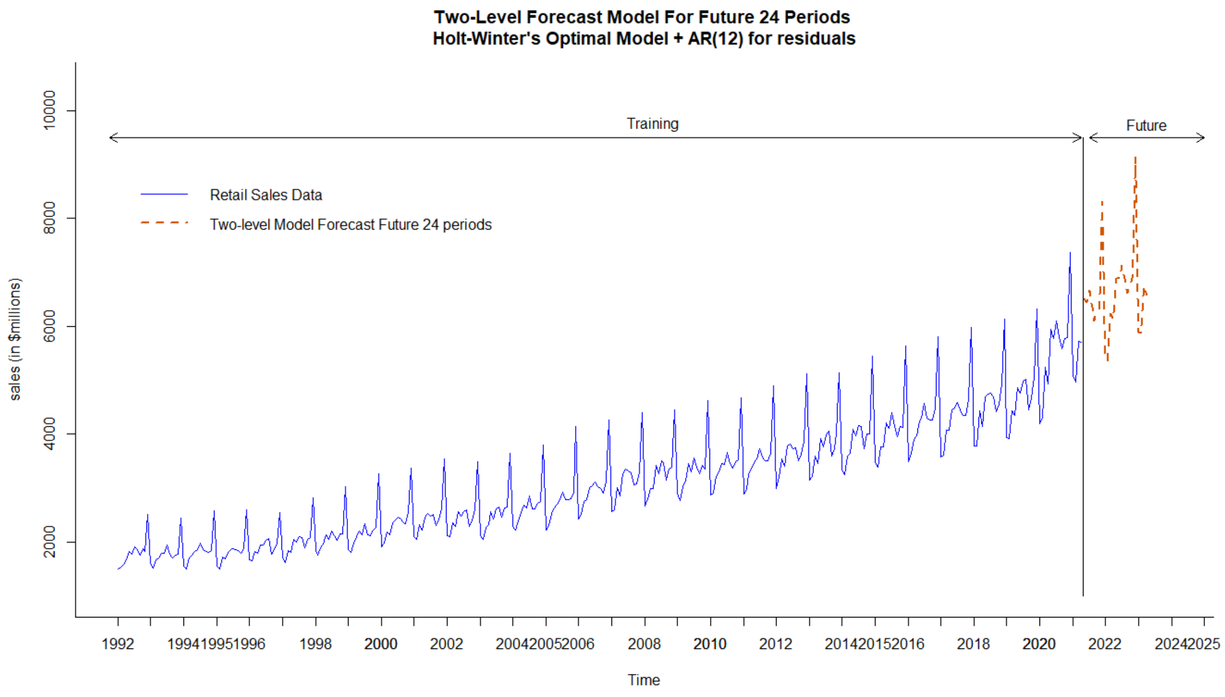
February-23	5946.036	-63.300716	5882.735
March-23	6776.849	-22.224803	6754.625
April-23	6632.243	-52.032802	6580.21

Below table show accuracies for both Holt-winters model and two-level model for entire data.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Two- level Forecast (Holt-Winters Model + AR(12) Model for residuals)	0.074	83.581	58.802	-0.076	1.874	0.046	0.154
Holt-Winter's Automatic Model with optimal parameters for entire data	6.041	95.7	69.306	0.128	2.202	-0.04	0.177

As we can observed Two level model performs better than Holt-Winter's Automatic Model with optimal parameters as it has less MAPE and RMSE values.

Visualizing retail sales using Two-Level model – Entire Data:



iii) Auto Regressive Integrated Moving Average(ARIMA):

Auto Regressive(**AR**)-Integrated(**I**)-Moving Average (**MA**) also referred as Box-Jenkins methodology or Box-Jenkins approach. This approach is capable of presenting every time series component like trend, seasonality and level as the approach can include up to 6 parameters .Non-seasonal ARIMA include three parts Auto Regressive (AR) ,Integrated (I) and Moving Average(MA) which only consider level and trend but not seasonality.

Auto Regressive(AR):

Auto-Regressive model is a type of model where it models the auto-correlation directly in regression model using past observations as predictors. The term auto-correlation indicates that it is a regression of the variable against itself. Auto-Regressive models can be built of any order depending on the autocorrelation in the data. Below are the equations and representation of various orders of AR model. It is represented as AR (p,0,0) where p is order of the model. p represents the lag order.

AR Model Equation of Order p:

$$\boldsymbol{Yt} \;=\; \beta 0 + \beta 1 * \boldsymbol{Yt - 1} \;+ \beta 2 * \boldsymbol{Yt - 2} \ldots\ldots + \beta p * \boldsymbol{Yt - p} + \boldsymbol{\varepsilon t}$$

Below is the example of auto regressive model on retail sales of order 2.

```
Series: sales.ts
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1     ar2       mean
      0.5553  0.3635  3179.0995
s.e.  0.0495  0.0500   341.2045

sigma^2 estimated as 295633:  log likelihood=-2715.87
AIC=5439.73   AICc=5439.85   BIC=5455.19

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE     MASE      ACF1
Training set 8.751821 541.3995 359.9102 -2.451123 11.61593 2.377584 -0.123233
```

From the above model summary, we can interpret that ar1 0.5553,ar2 0.3635 are co-efficients with mean as 3179.0995.

Yt = 3179.0995+ 0.5553 * Yt-1 + 0.3635 * Yt-2
Where Yt-1 and Yt-2 are preceding time period values

Moving Average(MA):

Moving average model works by analyzing the errors from the lagged observations i.e., residuals of AR model. The Moving average of order q can be represented as ARIMA(0,0,q).Below is the equation for MA of order q

$$\boldsymbol{Yt = c + \varepsilon t + \theta 1\ \varepsilon t - 1 + \theta 2\ \varepsilon t - 2} \ldots + \boldsymbol{\theta q\ \varepsilon t - q}$$

Where c= constant mean of MA model

$\boldsymbol{\varepsilon t}$ is error term (other coefficients are selected in a way to minimize this error)

$\boldsymbol{\varepsilon t - 1, \varepsilon t - 2,}$... $\boldsymbol{\varepsilon t - q}$ represents error terms of lagged time periods

$\boldsymbol{\theta 1}, \boldsymbol{\theta 2}, ..\boldsymbol{\theta q}$ represents coefficients of variables to be estimated

Below is the summary of ARIMA(0,0,1) that is order 1 moving average ,

```
Series: sales.ts
ARIMA(0,0,1) with non-zero mean


Coefficients:
       ma1       mean
    0.6698  3123.7317
s.e.  0.0313    72.6356


sigma^2 estimated as 671371:  log likelihood=-2860.17
AIC=5726.33   AICc=5726.4   BIC=5737.92


Training set error measures:
                  ME     RMSE      MAE       MPE     MAPE     MASE      ACF1
Training set 1.343633 817.0414 661.1787 -8.485985 23.36132 4.367777 0.3437518
```

From the model summary the equation for order 1 moving average is represented as below,

$$\mathrm{Yt} = 3123.7317 + 0.6698 \ * \varepsilon \mathrm{t} - 1$$

$\varepsilon \mathrm{t} - 1$ is the error term of first order autoregressive model at time $\mathrm{t} - 1$

Integrated (I):

Integrated means nothing but the difference between the values at lagged time periods (d). Differencing will help in stabilizing mean and will remove the trend from the data.

Typically, Auto-regressive and Moving average models works best with the data that has no trend or/and seasonality .So to remove the trend from the data and to stabilize the data around mean or to make stationary we introduce differencing into picture , which can be achieved using ARIMA(0,d,0) where d is level or order of differencing.

Below is the representation of how different level of differencing happened with value of d.

d = 0: no differencing (series does not have a trend), y_t

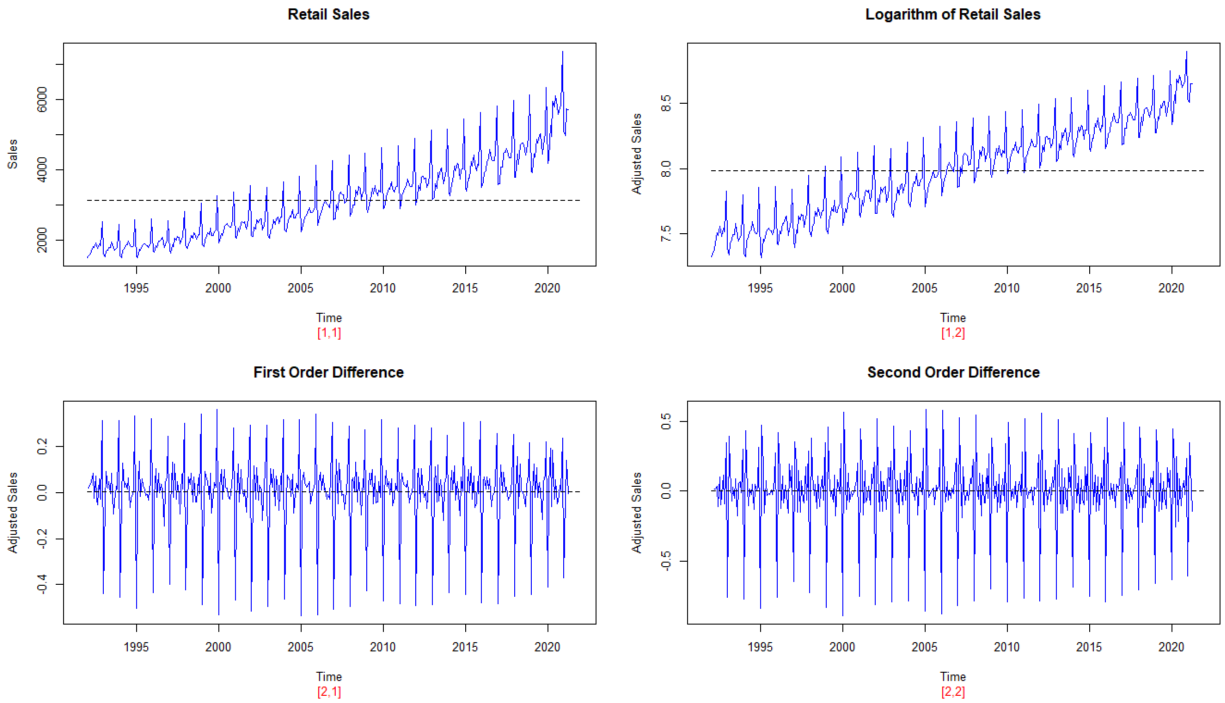
d = 1: difference the series once which can remove linear trend,

$y_t - y_{t-1}$, $y_2 - y_1$, $y_3 – y_2$, …

d = 2: difference the series twice, each time of lag-1 (first difference of the first difference),

$(y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t – 2\ y_{t-1} + y_{t-2}$, e.g., $y_3 – 2\ y_2 + y_1$, $y_4 – 2\ y_3 + y_2$, …

Visualizing Various orders of Differencing:



Adding log transformation$_{\mathbf{(plot\ \ 1,2)}}$ to the data will stabilize the variance which is one of the features of stationary data. As we can clearly observe from the above plots as we do various orders of difference on log transformed retail sales data, it removes the trend component from the original data .But still the data is not stationary as there is a cyclic behavior or seasonality in the data which can be removed by seasonal differencing which makes data more stationary. So, to overcome this Seasonal ARIMA is introduced.

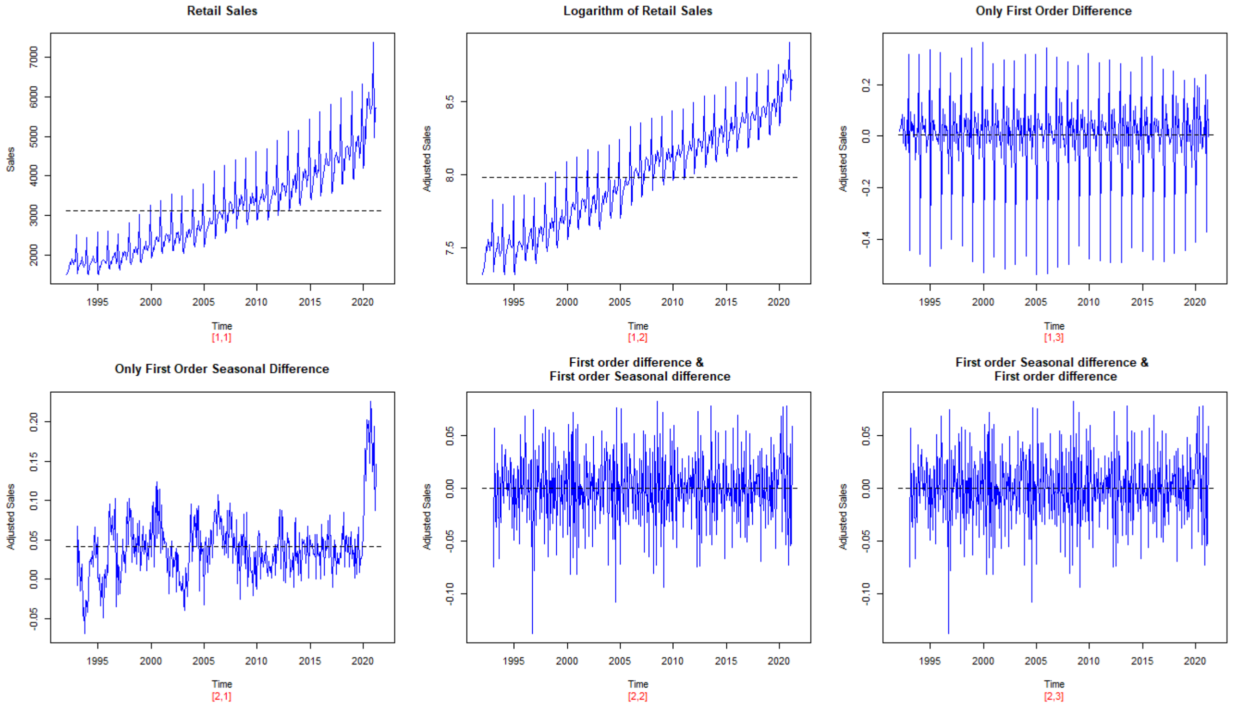
Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

Normal ARIMA (p,d,q) does not include seasonality which is why does not works best for a data that has seasonality. As observed in the above plots ,only first order differencing(i.e., removing trend) cannot make data stationary. We need to eliminates the seasonal patterns as well which makes data more stable or stationary. So, to overcome this few more parameters are introduced to ARIMA like P,D,Q,m .

Seasonal ARIMA (p,d,q) (P,D,Q)m

p, order p autoregressive model *AR(p)*
d , order d differencing to remove trend
q , order q moving average *MA(q)* for error lags
P , order P autoregressive model *AR(P) for seasonality*
D , order D differencing to remove seasonal patterns
Q , order Q moving average *MA(Q)* for error lags
m = 12, for monthly seasonality*4*, for quarterly seasonality

Below plot shows various transformations on retail sales data. As we can interpret first order difference**(plot [1,3])** cannot make data entirely stationary as seasonality still exists in the data. Sometimes applying only seasonal difference**(plot[2,1])** can make data stationary if the data is more seasonal. If it does not make data stationary then combination of differencing can be applied. So, a combination of first order difference and first order seasonal difference made data more stationary as observed from below plot**(plot [2,2][2,3])**. The order of applying the differencing also does not matter (i.e., applying first order and seasonal next or applying seasonal first and first order difference next) as we can observe from **plot[2,2][2,3]** both produced same outputs. Models with less complexity are preferred and as we increase order of difference it will be difficult to interpret the model and will become complex.

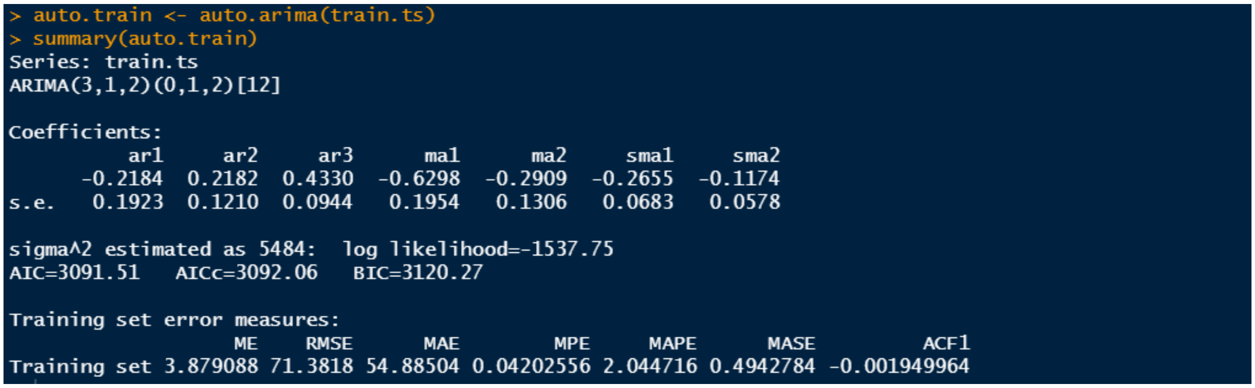


Now once the data becomes stationary after differencing(to remove trend and seasonality), various orders and combinations of auto regressive and moving average models can be applied on the differenced data to build a better forecasting model.

We can determine the values of these parameters by visualizing historical data or by ACF/PACF charts or various trail and errors etc. So, to determine optimal values for these components will be a hectic task so an automated model **Auto ARIMA** is introduced which will selected the parameter values based on many conditions such as AIC, AICc ,BIC, accuracy and log likelihood values. A model with less complexity or less AIC or BIC values with higher log likelihood is given preference as a best model.

- **<u>Auto Arima For Training Data:</u>**

Below is the model summary for auto arima model for training data ARIMA(3,1,2)(0,1,2)[12].



```
> auto.train <- auto.arima(train.ts)
> summary(auto.train)
Series: train.ts
ARIMA(3,1,2)(0,1,2)[12]

Coefficients:
          ar1     ar2     ar3      ma1      ma2     sma1     sma2
      -0.2184  0.2182  0.4330  -0.6298  -0.2909  -0.2655  -0.1174
s.e.   0.1923  0.1210  0.0944   0.1954   0.1306   0.0683   0.0578

sigma^2 estimated as 5484:  log likelihood=-1537.75
AIC=3091.51   AICc=3092.06   BIC=3120.27

Training set error measures:
                   ME     RMSE      MAE        MPE     MAPE      MASE          ACF1
Training set 3.879088 71.3818 54.88504 0.04202556 2.044716 0.4942784 -0.001949964
```

As we can interpret from the model summary the model it indicates we have first difference, first order seasonal difference ,third order auto regressive model, no auto regressive model for seasonality, non-seasonal second order MA for error lags and seasonal second order MA for error lags. Model equation can be represented as below ,

$$y_t - y_{t-1} = -0.2184\,(y_{t-1} - y_{t-2}) + 0.2182(y_{t-2} - y_{t-3}) + 0.4330(y_{t-3} - y_{t-4}) - 0.6298\ \varepsilon_{t-1} \\ -0.2909\ \varepsilon_{t-2} - 0.2655\ \rho_{t-1} - 0.1174\ \rho_{t-2}$$

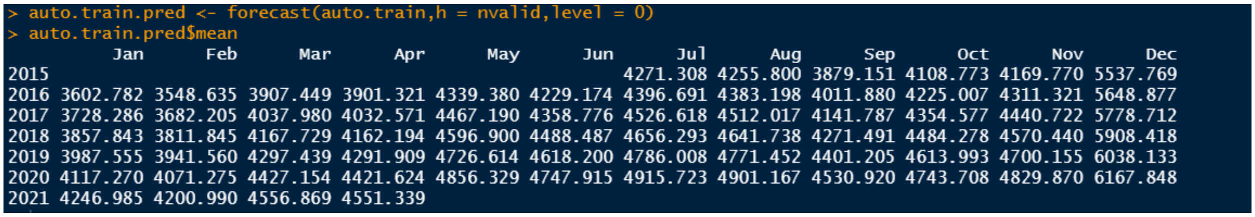
As we can interpret from the model equation it is first order differenced as we have yt-yt-1 on left side of the equation. -0.2184(ar1), 0.2182(ar2) and 0.4330(ar3) are the coefficients of third order auto regressive model, -0.6298(ma1) and -0.2909(ma2) are the coefficients of second order moving average for error lags. $y_{t-1} - y_{t-2}$, $y_{t-2} - y_{t-3}$, $y_{t-3} - y_{t-4}$ represents elements of the first order difference .ε_{t-1}, ε_{t-2} are error terms of second order auto regressive model. -0.2655(sma1),-0.1174(sma2) are the coefficients of seasonal second order moving average for error lags. ρ_{t-1}, ρ_{t-2} are error terms of second order seasonal auto regressive model.

The ARIMA(3,1,2)(0,1,2)[12] has a log likelihood of -1537.75,BIC as 3120.27 ,AICc as 3092.06 and AIC as 3091.51.These metrics can be used to compare with other models.

Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

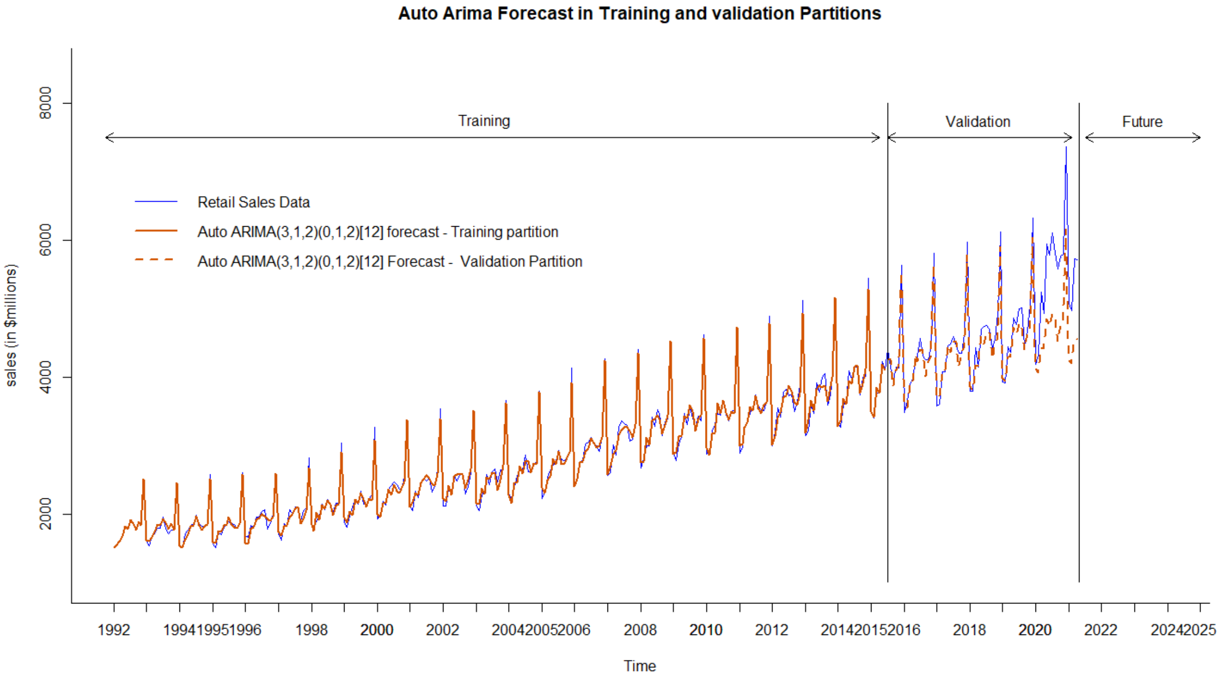
<u>Retail sales forecast for validation Period:</u>

Below table represents the point forecasted values for validation period using ARIMA(3,1,2)(0,1,2)[12].



```
> auto.train.pred <- forecast(auto.train,h = nvalid,level = 0)
> auto.train.pred$mean
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015                                                       4271.308 4255.800 3879.151 4108.773 4169.770 5537.769
2016 3602.782 3548.635 3907.449 3901.321 4339.380 4229.174 4396.691 4383.198 4011.880 4225.007 4311.321 5648.877
2017 3728.286 3682.205 4037.980 4032.571 4467.190 4358.776 4526.618 4512.017 4141.787 4354.577 4440.722 5778.712
2018 3857.843 3811.845 4167.729 4162.194 4596.900 4488.487 4656.293 4641.738 4271.491 4484.278 4570.440 5908.418
2019 3987.555 3941.560 4297.439 4291.909 4726.614 4618.200 4786.008 4771.452 4401.205 4613.993 4700.155 6038.133
2020 4117.270 4071.275 4427.154 4421.624 4856.329 4747.915 4915.723 4901.167 4530.920 4743.708 4829.870 6167.848
2021 4246.985 4200.990 4556.869 4551.339
```

<u>Visualize retail sales forecast using ARIMA(3,1,2)(0,1,2)[12]:</u>

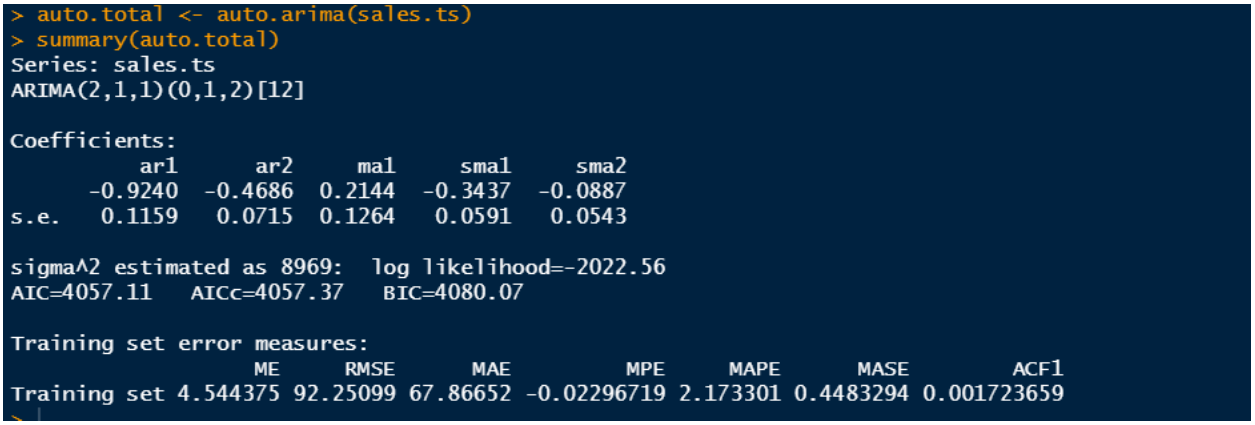


Below table represents accuracy for ARIMA(3,1,2)(0,1,2)[12] in the training and validation Partition,

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	3.879	71.382	54.885	0.042	2.045	0.494	-0.002	NA
Test set	266.606	466.377	293.282	4.843	5.528	2.641	0.823	0.575

- **Auto Arima For Entire Data:**

Below is the model summary for auto arima model for training data ARIMA(2,1,1)(0,1,2)[12].



```
> auto.total <- auto.arima(sales.ts)
> summary(auto.total)
Series: sales.ts
ARIMA(2,1,1)(0,1,2)[12]

Coefficients:
          ar1      ar2     ma1     sma1     sma2
      -0.9240  -0.4686  0.2144  -0.3437  -0.0887
s.e.   0.1159   0.0715  0.1264   0.0591   0.0543

sigma^2 estimated as 8969:  log likelihood=-2022.56
AIC=4057.11   AICc=4057.37   BIC=4080.07

Training set error measures:
                   ME     RMSE      MAE         MPE     MAPE      MASE        ACF1
Training set 4.544375 92.25099 67.86652 -0.02296719 2.173301 0.4483294 0.001723659
```

As we can interpret from the model summary the model it indicates we have first difference, first order seasonal difference ,second order auto regressive model, no auto regressive model for seasonality, non-seasonal first order MA for error lags and seasonal second order MA for error lags. Model equation can be represented as below ,

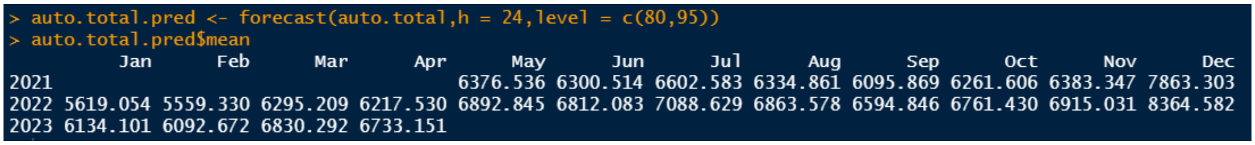
$$y_t - y_{t-1} = - 0.9240\ (y_{t-1} - y_{t-2})\ -0.4686(y_{t-2} - y_{t-3}) + 0.2144\ \varepsilon_{t-1}\ -0.3437\ \rho_{t-1}\ -0.0887\ \rho_{t-2}$$

As we can interpret from the model equation it is first order differenced as we have yt-yt-1 on left side of the equation. -0.9240(ar1), -0.4686(ar2) are the coefficients of second order auto regressive model, 0.2144 (ma1) is the coefficient of first order moving average for error lags. y_{t-1} -y_{t-2}, y_{t-2} -y_{t-3} represents elements of the first order difference .ε_{t-1} is error term of first order auto regressive model. -0.3437(sma1),-0.0887(sma2) are the coefficients of seasonal second order moving average for error lags. ρ_{t-1}, ρ_{t-2} are error terms of second order seasonal auto regressive model.

The ARIMA(2,1,1)(0,1,2)[12] has a log likelihood of -2022.56,BIC as 4080.07 ,AICc as 4057.37 and AIC as 4057.11.These metrics can be used to compare with other models with same differencing orders.

<u>Retail sales forecast for future 24 periods:</u>

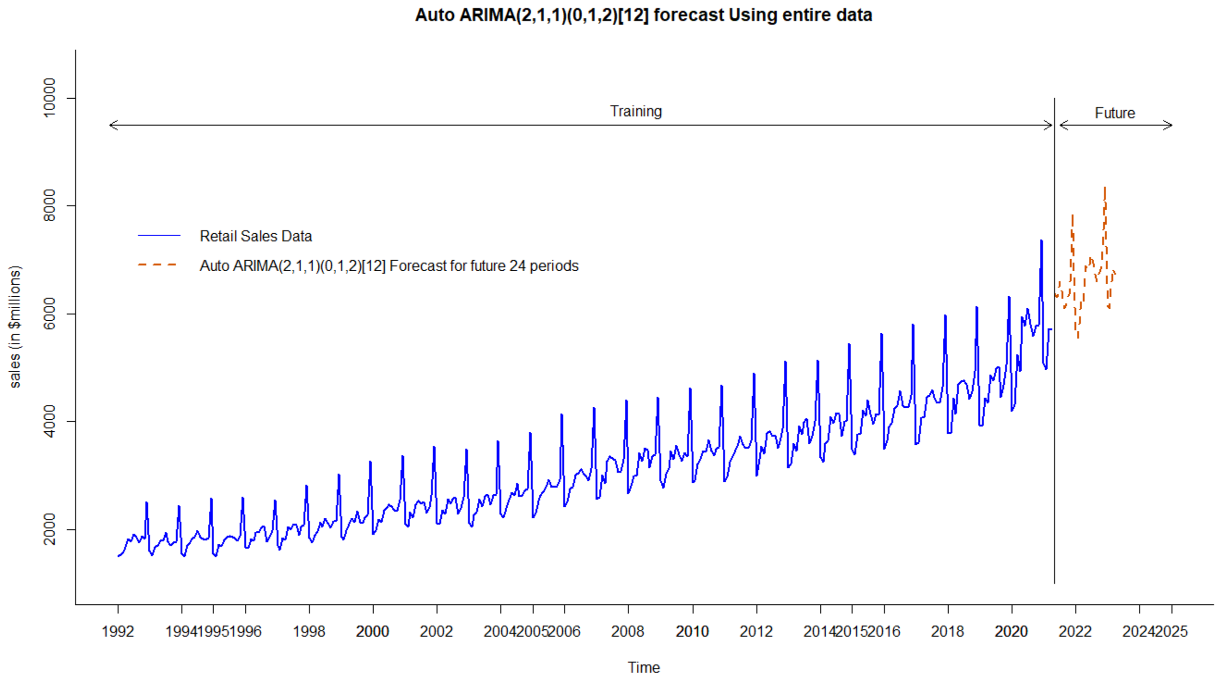
Below table represents the point forecasted values for future 24 periods using ARIMA(2,1,1)(0,1,2)[12].



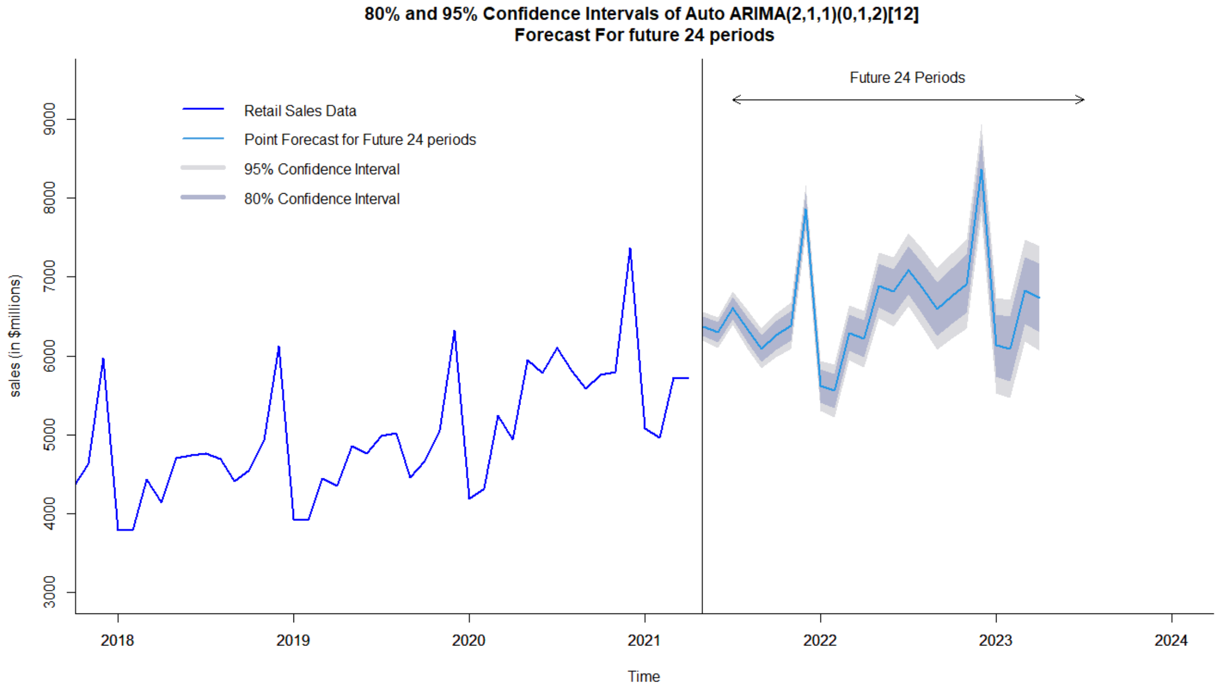
```
> auto.total.pred <- forecast(auto.total,h = 24,level = c(80,95))
> auto.total.pred$mean
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2021                                     6376.536 6300.514 6602.583 6334.861 6095.869 6261.606 6383.347 7863.303
2022 5619.054 5559.330 6295.209 6217.530 6892.845 6812.083 7088.629 6863.578 6594.846 6761.430 6915.031 8364.582
2023 6134.101 6092.672 6830.292 6733.151
```

Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

Visualizing retail sales forecast using ARIMA(2,1,1)(0,1,2)[12] for future 24 periods:



Visualizing Confidence intervals of future 24 periods:



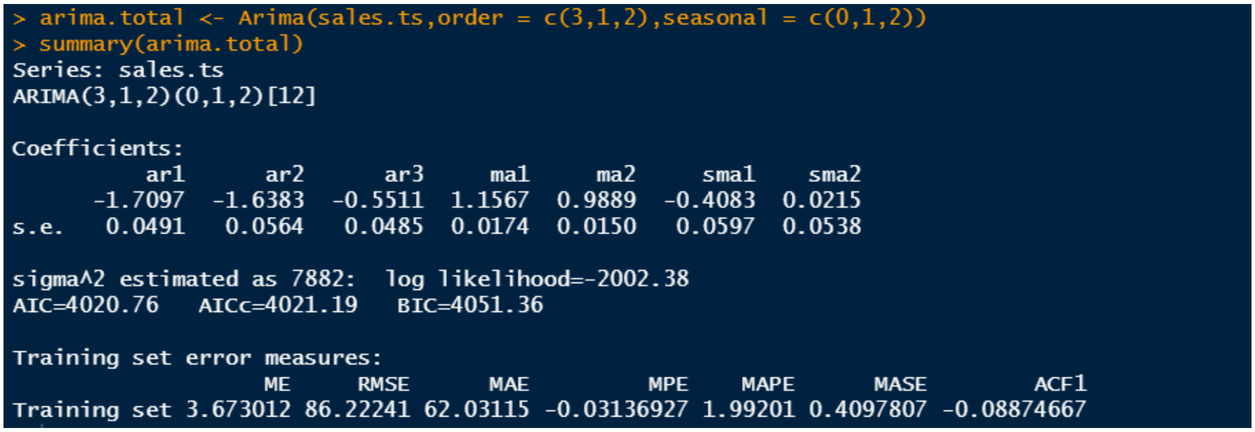
Below table shows accuracy for Arima(2,1,1)(0,1,2) for entire data set.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Arima(2,1,1)(0,1,2)	4.544	92.251	67.867	-0.023	2.173	0.002	0.179

- **<u>ARIMA(3,1,2)(0,1,2)[12] for Entire Data:</u>**

In this module we are trying to apply Arima(3,1,2)(0,1,2) (model which was chosen by auto-arima for training validation) on the entire data set.

Below is the model summary for ARIMA(3,1,2)(0,1,2)[12] model for entire data.



```
> arima.total <- Arima(sales.ts,order = c(3,1,2),seasonal = c(0,1,2))
> summary(arima.total)
Series: sales.ts
ARIMA(3,1,2)(0,1,2)[12]

Coefficients:
          ar1      ar2      ar3     ma1     ma2     sma1    sma2
      -1.7097  -1.6383  -0.5511  1.1567  0.9889  -0.4083  0.0215
s.e.   0.0491   0.0564   0.0485  0.0174  0.0150   0.0597  0.0538

sigma^2 estimated as 7882:  log likelihood=-2002.38
AIC=4020.76   AICc=4021.19   BIC=4051.36

Training set error measures:
                   ME     RMSE      MAE         MPE    MAPE      MASE        ACF1
Training set 3.673012 86.22241 62.03115 -0.03136927 1.99201 0.4097807 -0.08874667
```

As we can interpret from the model summary the model it indicates we have first difference, first order seasonal difference ,third order auto regressive model, no auto regressive model for seasonality, non-seasonal second order MA for error lags and seasonal second order MA for error lags. Model equation can be represented as below ,

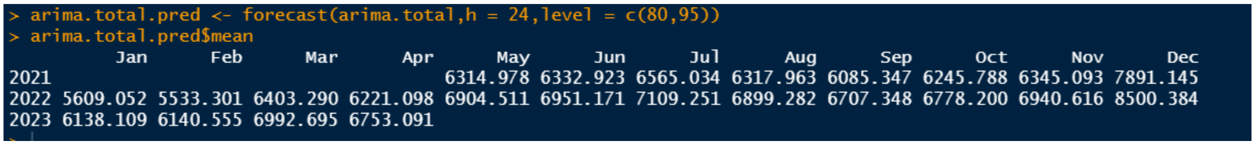
$$y_t - y_{t-1} = - 1.7097 (y_{t-1} - y_{t-2}) - 1.6383(y_{t-2} - y_{t-3}) - 0.5511(y_{t-3} - y_{t-4}) - 1.1567\ \varepsilon_{t-1}$$
$$+0.9889\ \varepsilon_{t-2} - 0.4083\ \rho_{t-1} + 0.0215\ \rho_{t-2}$$

As we can interpret from the model equation it is first order differenced as we have yt-yt-1 on left side of the equation. -1.7097(ar1), -1.6383(ar2) and -0.5511(ar3) are the coefficients of third order auto regressive model, -1.1567(ma1) and 0.9889(ma2) are the coefficients of second order moving average for error lags. y_{t-1} -y_{t-2}, y_{t-2} -y_{t-3}, y_{t-3} -y_{t-4} represents elements of the first order difference .ε_{t-1}, ε_{t-2} are error terms of second order auto regressive model. -0.4083(sma1),-0.0215(sma2) are the coefficients of seasonal second order moving average for error lags. $\rho_{t-1,}$ ρ_{t-2} are error terms of second order seasonal auto regressive model.

The ARIMA(3,1,2)(0,1,2)[12] has a log likelihood of -2002.38,BIC as 4051.36 ,AICc as 4021.19 and AIC as 4020.76.These metrics can be used to compare with other models with same differencing orders.

<u>Retail sales forecast for future 24 periods:</u>

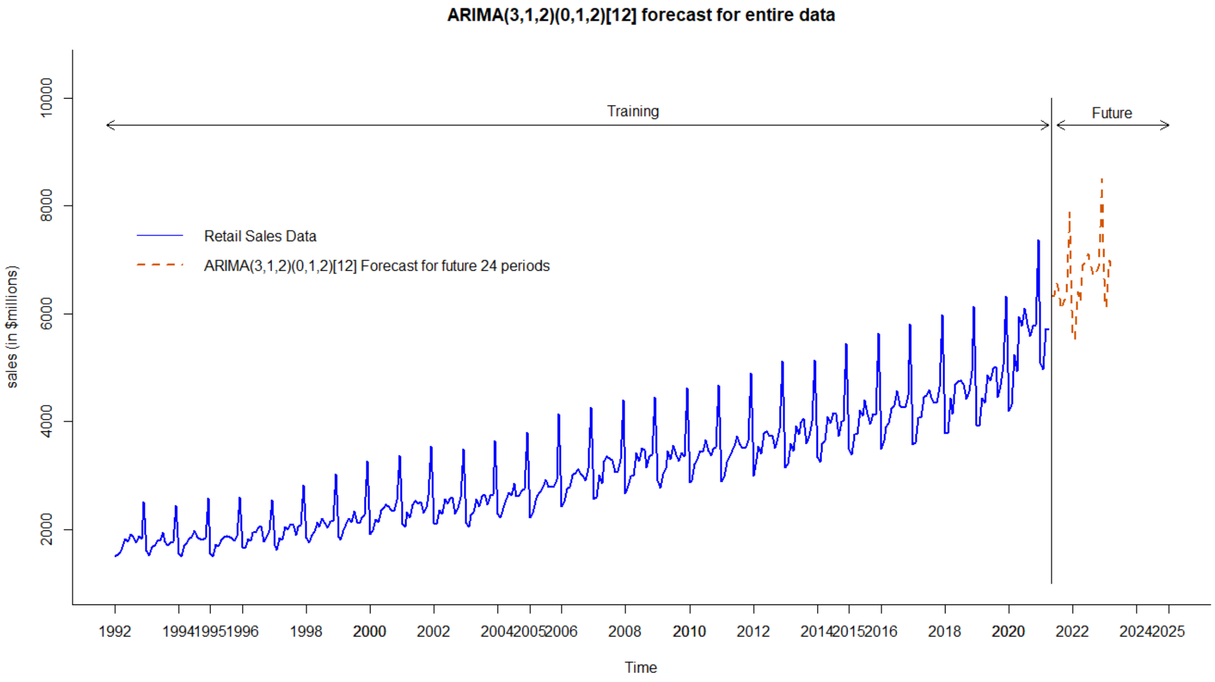
Below table represents the point forecasted values for future 24 periods using ARIMA(3,1,2)(0,1,2)[12].



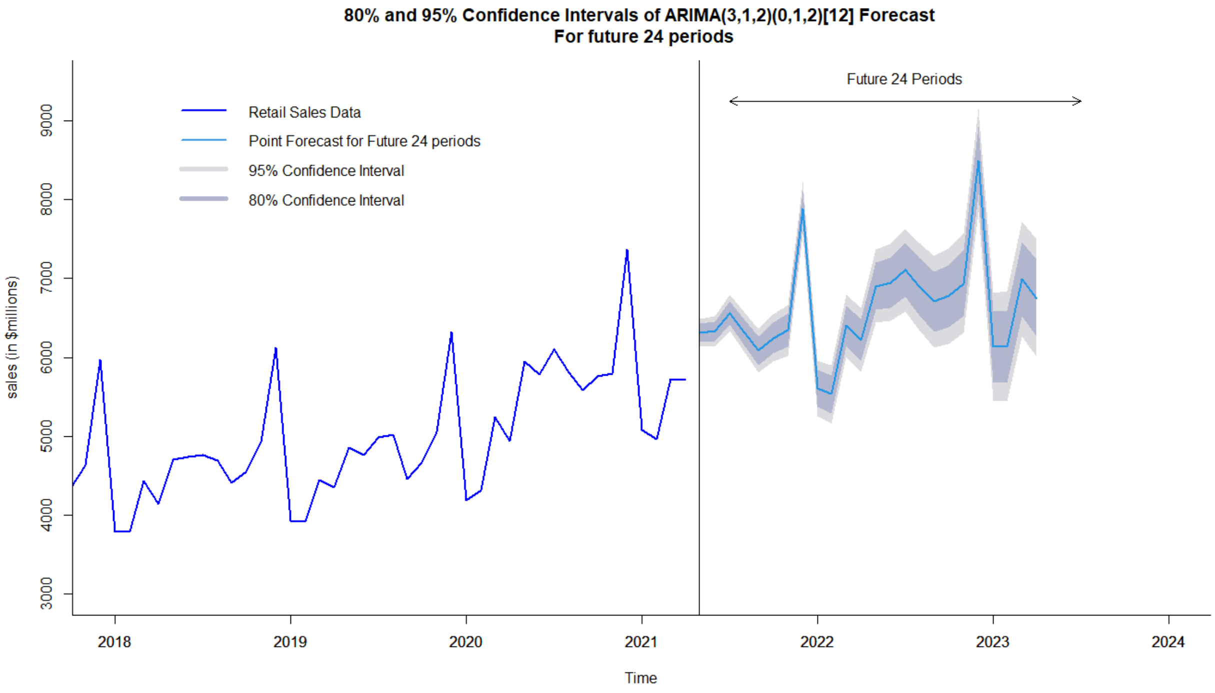
```
> arima.total.pred <- forecast(arima.total,h = 24,level = c(80,95))
> arima.total.pred$mean
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2021                                    6314.978 6332.923 6565.034 6317.963 6085.347 6245.788 6345.093 7891.145
2022 5609.052 5533.301 6403.290 6221.098 6904.511 6951.171 7109.251 6899.282 6707.348 6778.200 6940.616 8500.384
2023 6138.109 6140.555 6992.695 6753.091
```

Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

Visualizing retail sales forecast using ARIMA(3,1,2)(0,1,2)[12] for future 24 periods:



Visualizing Confidence intervals of future 24 periods:



Below table shows accuracy for Arima(3,1,2)(0,1,2) for entire data set.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Arima(3,1,2)(0,1,2)	3.673	86.222	62.031	-0.031	1.992	-0.089	0.166

7) Evaluate & Compare performance:

After developing various forecasting models, it is important to compare the accuracy measures of various model and select a model with best forecasting accuracy. The smaller the forecasting error the better the model.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Two-level Forecast (Linear + Trailing MA for regression residuals)	5.211	136.681	95.872	-0.024	3.174	0.145	0.287
Two-level Forecast (Quadratic + Trailing MA for regression residuals)	5.134	134.422	94.33	0.145	3.184	0.119	0.287
Holt-Winter’s Automatic Model with optimal parameters	6.041	95.7	69.306	0.128	2.202	-0.04	0.177
Two- level Forecast (Holt-Winters Model + AR(12) Model for residuals)	0.074	83.581	58.802	-0.076	1.874	0.046	0.154
Arima (3,1,2)(0,1,2)	3.673	86.222	62.031	-0.031	1.992	-0.089	0.166
Auto Arima (2,1,1)(0,1,2)	4.544	92.251	67.867	-0.023	2.173	0.002	0.179

As we can observe from above models two-level forecast (Holt-Winter’s Automatic Model with optimal parameters + AR(12) model for residuals) has less MAPE and RMSE values among all the models. Although two-level forecast (Holt-Winter’s Automatic Model with optimal parameters + AR(12) model for residuals) is best in terms of accuracy one should notice that AR(12) model is a complex model with 12 variables and an ensemble model will increase cost and computational time in real time. If complexity and computational time is not an issue, we can choose that two-level forecast (Holt-Winter’s Automatic Model with optimal parameters + AR(12) model for residuals) as best model for forecasting into future. Else Arima(3,1,2)(0,1,2) model can be chosen which closely follows two-level forecast in terms of forecasting accuracy.

8) IMPLEMENT FORECAST/SYSTEM:

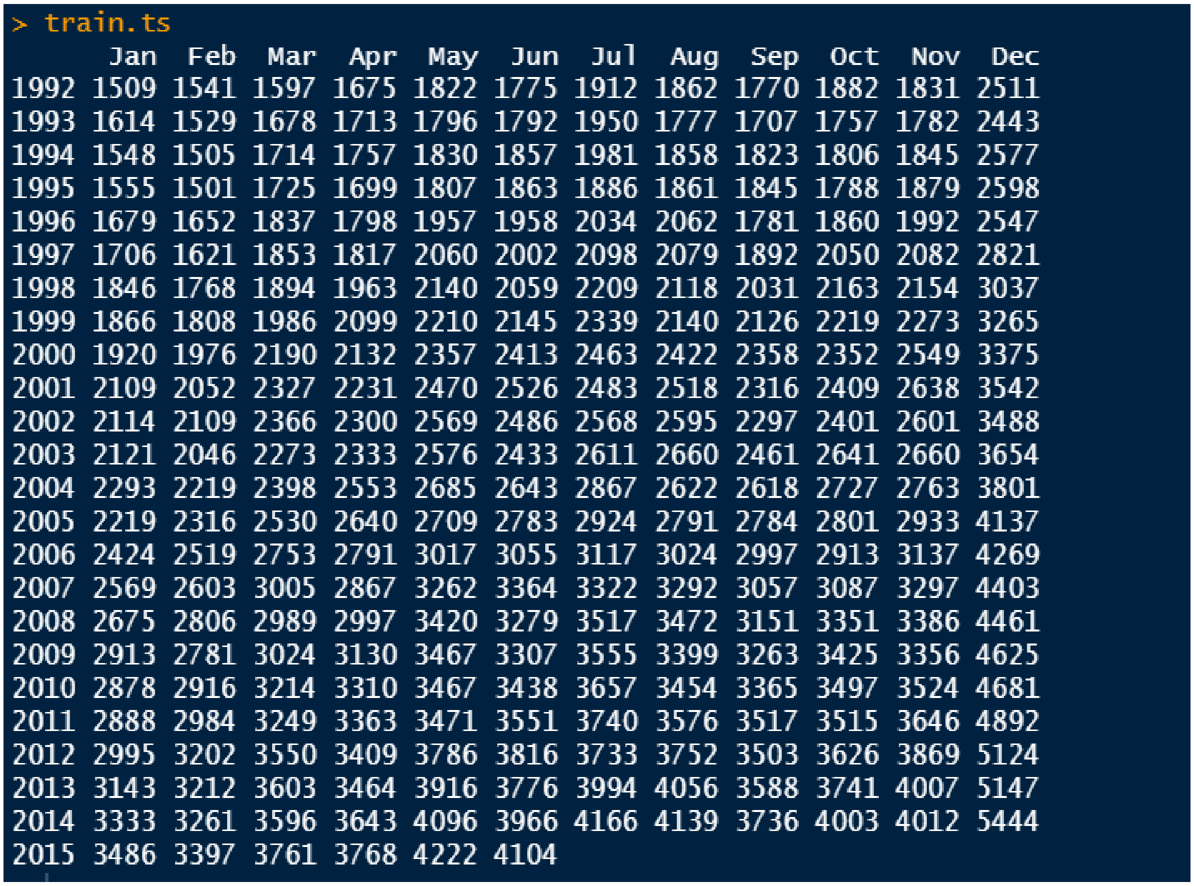
As we observed from the accuracy measures two-level forecast model (Holt-Winter's Automatic Model with optimal parameters + AR(12) model for residuals) is best forecasting model among all other models.

Once the best forecasting model is chosen it should be implemented in such a way that it accommodates new data as it comes each and every cycle. The model should be reevaluated at regular intervals as the new data comes in .In this case model should be reevaluated at least quarterly or semi – annually as we get additional data points every month. Models can also be automated so that it will be an ongoing forecasting with less manual intervention.

APPENDIX

Training Data :

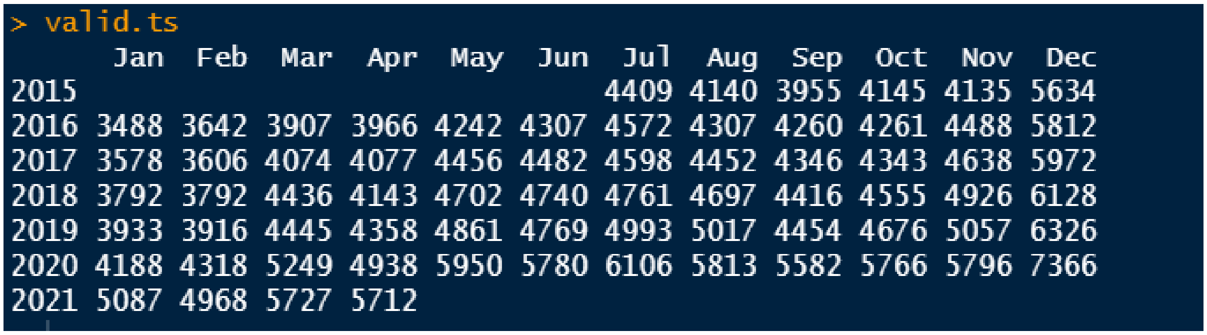
Below table shows training partition data utilized in the project.



```
> train.ts
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
1992 1509 1541 1597 1675 1822 1775 1912 1862 1770 1882 1831 2511
1993 1614 1529 1678 1713 1796 1792 1950 1777 1707 1757 1782 2443
1994 1548 1505 1714 1757 1830 1857 1981 1858 1823 1806 1845 2577
1995 1555 1501 1725 1699 1807 1863 1886 1861 1845 1788 1879 2598
1996 1679 1652 1837 1798 1957 1958 2034 2062 1781 1860 1992 2547
1997 1706 1621 1853 1817 2060 2002 2098 2079 1892 2050 2082 2821
1998 1846 1768 1894 1963 2140 2059 2209 2118 2031 2163 2154 3037
1999 1866 1808 1986 2099 2210 2145 2339 2140 2126 2219 2273 3265
2000 1920 1976 2190 2132 2357 2413 2463 2422 2358 2352 2549 3375
2001 2109 2052 2327 2231 2470 2526 2483 2518 2316 2409 2638 3542
2002 2114 2109 2366 2300 2569 2486 2568 2595 2297 2401 2601 3488
2003 2121 2046 2273 2333 2576 2433 2611 2660 2461 2641 2660 3654
2004 2293 2219 2398 2553 2685 2643 2867 2622 2618 2727 2763 3801
2005 2219 2316 2530 2640 2709 2783 2924 2791 2784 2801 2933 4137
2006 2424 2519 2753 2791 3017 3055 3117 3024 2997 2913 3137 4269
2007 2569 2603 3005 2867 3262 3364 3322 3292 3057 3087 3297 4403
2008 2675 2806 2989 2997 3420 3279 3517 3472 3151 3351 3386 4461
2009 2913 2781 3024 3130 3467 3307 3555 3399 3263 3425 3356 4625
2010 2878 2916 3214 3310 3467 3438 3657 3454 3365 3497 3524 4681
2011 2888 2984 3249 3363 3471 3551 3740 3576 3517 3515 3646 4892
2012 2995 3202 3550 3409 3786 3816 3733 3752 3503 3626 3869 5124
2013 3143 3212 3603 3464 3916 3776 3994 4056 3588 3741 4007 5147
2014 3333 3261 3596 3643 4096 3966 4166 4139 3736 4003 4012 5444
2015 3486 3397 3761 3768 4222 4104
```

Validation Data:

Below table shows validation partition data utilized in the project.



```
> valid.ts
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2015                               4409 4140 3955 4145 4135 5634
2016 3488 3642 3907 3966 4242 4307 4572 4307 4260 4261 4488 5812
2017 3578 3606 4074 4077 4456 4482 4598 4452 4346 4343 4638 5972
2018 3792 3792 4436 4143 4702 4740 4761 4697 4416 4555 4926 6128
2019 3933 3916 4445 4358 4861 4769 4993 5017 4454 4676 5057 6326
2020 4188 4318 5249 4938 5950 5780 6106 5813 5582 5766 5796 7366
2021 5087 4968 5727 5712
```

Retail Sales: Beer, Wine, and Liquor Stores – Time Series Analysis

Auto-Correlation :

Auto-Correlation represents the correlation between a random variable (time series data) itself and the same variable lagged one or more periods

$$\mathbf{r_k} = \frac{\sum_{t=k+1}^{n} (\mathbf{Y_t} - \overline{\mathbf{Y}})(\mathbf{Y_{t-k}} - \overline{\mathbf{Y}})}{\sum_{t=1}^{n} (\mathbf{Y_t} - \overline{\mathbf{Y}})^2}$$

where
r_k = autocorrelation coefficient for a lag of k periods ($k = 1, 2, 3, \ldots, 12, \ldots$)
= mean of the values of the series
Y_t= observation in time period t
Y_{t-k} = observation k time periods earlier or at time period t-k

MAPE:

Mean absolute percentage error gives an absolute percentage score of how forecast deviates (on the average) from actual values; useful for comparing performance across series of data that have different scales. The lower the MAPE the better the forecast of the model. The less MAPE also signify the less margin of error

$$MAPE = \frac{100}{v} \sum_{t=1}^{v} \left| \frac{e_t}{y_t} \right|$$

RMSE:

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are. In other words, it tells you how concentrated the data is around the line of best fit. Root mean square error measures the square root from the squared errors. The smaller the RSME values of any of the measures, the better the forecast i.e. errors are smaller the better.

$$RMSE = \sqrt{\frac{1}{v} \sum_{t=1}^{v} {e_t}^2}$$

REFERENCES

1. https://otexts.com/fpp2/intro.html
2. https://bootstrappers.umassmed.edu/bootstrappers-courses/pastCourses/rCourse_2016-04/Additional_Resources/Rcolorstyle.html
3. https://fred.stlouisfed.org/series/MRTSSM4453USN#0