EY Casestudy

Ratul Wadhwa

2022-11-11

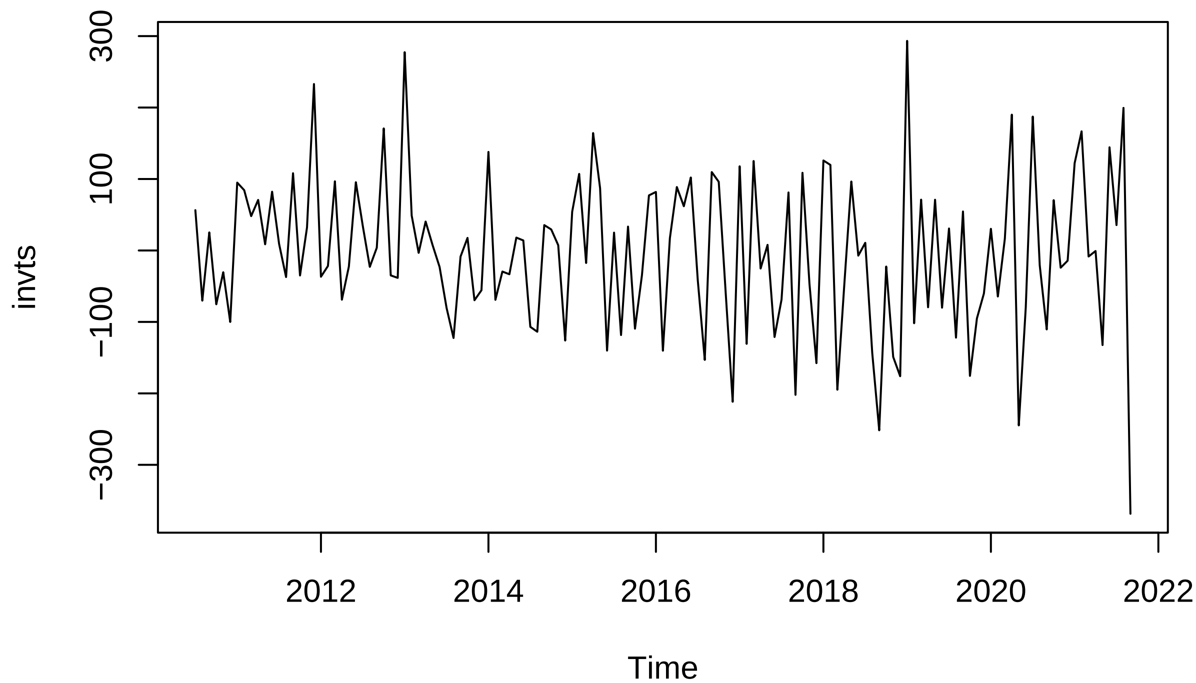
Contents

Objective

The objective of this case study is to produce forecasts of the total automotive gasoline inventory on a month-to-month basis, using a combination of econometrics and arithmetic functions in R.

The total automotive gasoline inventory can be calculated by inflow-outflow. Production and imports are considered as the inputs here and sales and exports are the outputs. **production+imports-sale-expor** gives us the total inventory of the automotive gasoline. The negative values here indicates shortage, outflow more than inflow.

Time Series Plot



Unit root tests

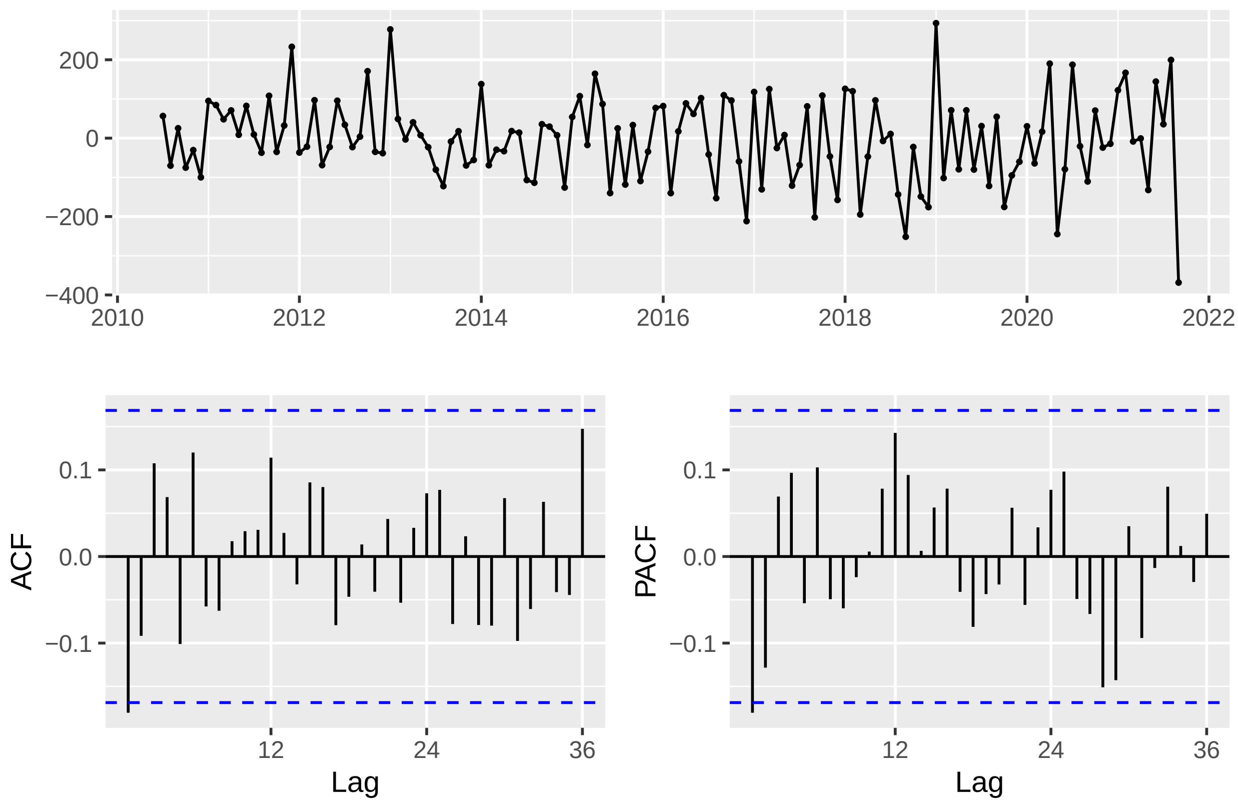
```
##
##  Box-Ljung test
##
## data:  invts
## X-squared = 4.4912, df = 1, p-value = 0.03407

## [1] 0

## [1] 0
```

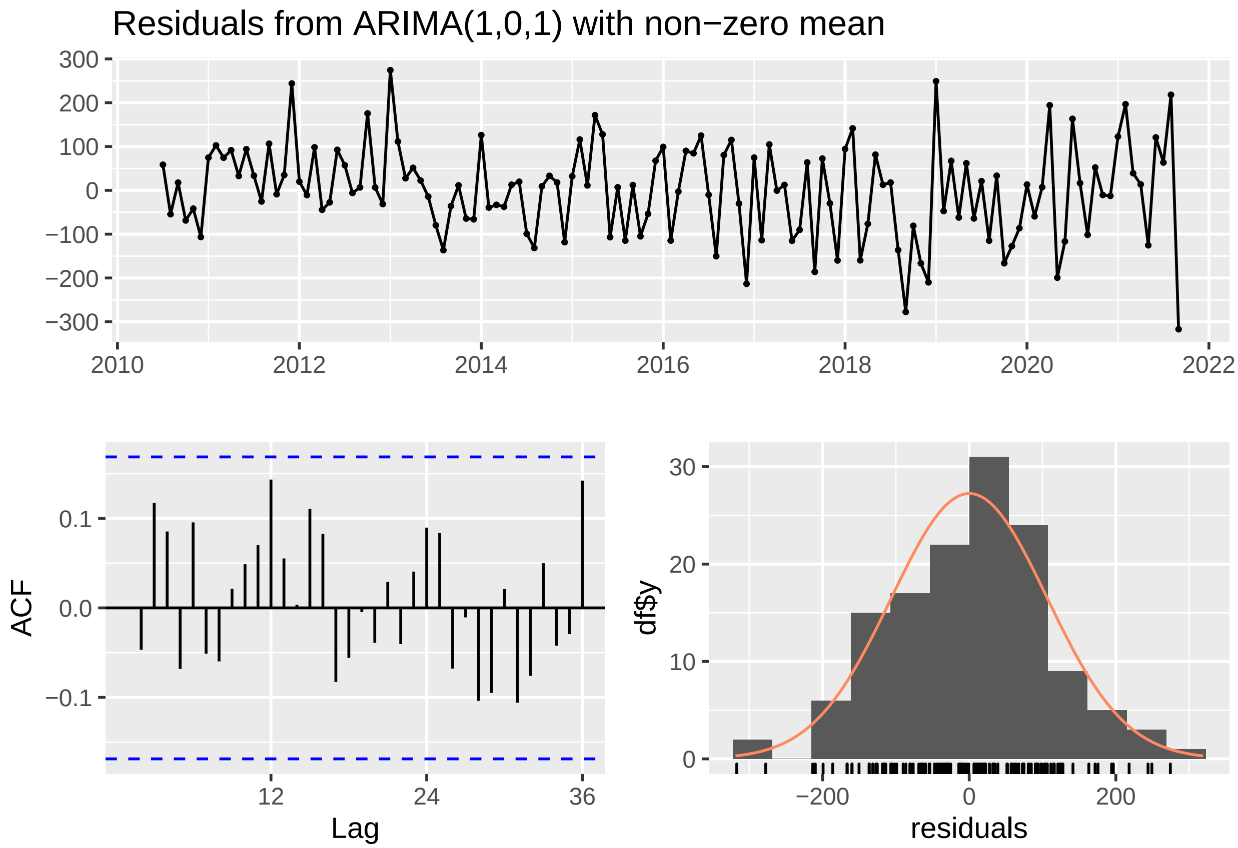
The test statistic is tiny, and well within the range we would expect for stationary data. So we can conclude that the data is stationary. To double check that I used nsdiffs() and ndiffs() functions and as they returns 0, it suggest we do not need seasonal difference and a first difference.

Choosing an appropriate ARIMA MODEL



Reading the ACF and the PACF plots above helps to identify the appropriate ARIMA models. The ACF plots helps to identify the non stationary part of the ARIMA model.

```
##
## Call:
## arima(x = (invts), order = c(1, 0, 1))
##
## Coefficients:
##          ar1      ma1  intercept
##       0.0582  -0.2747    -3.3877
## s.e.  0.2719   0.2511     7.0362
##
## sigma^2 estimated as 11211:  log likelihood = -821,  aic = 1649.99
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,1) with non-zero mean
## Q* = 17.627, df = 22, p-value = 0.7278
##
## Model df: 2.   Total lags used: 24
```

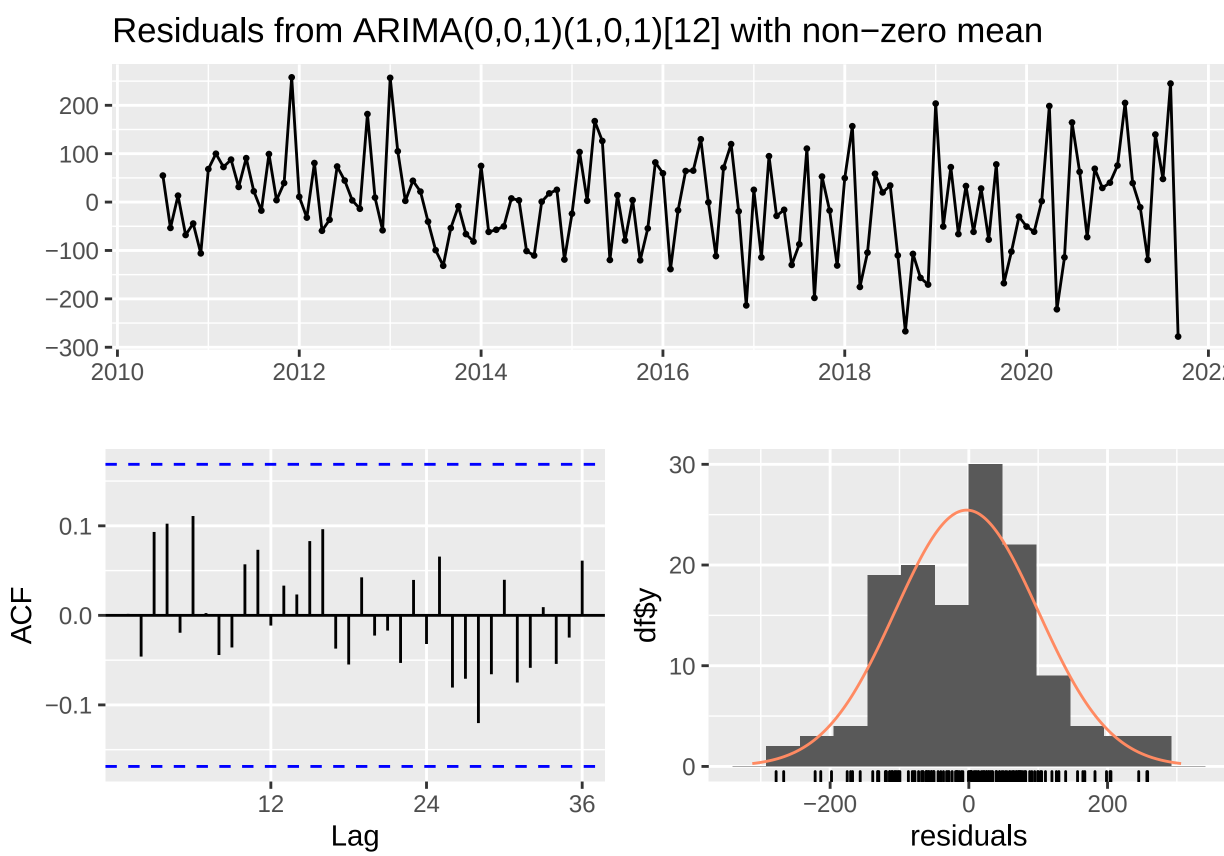
he p-value of the test is 0.772, which is much larger than 0.05. it fits the data reasonably well. The ARIMA model does well in capturing all the dynamics in the data as the residuals seem to be white noise.

Fitting auto.arima

```
##
##  ARIMA(2,0,2)(1,0,1)[12] with non-zero mean : 1650.511
##  ARIMA(0,0,0)            with non-zero mean : 1651.946
##  ARIMA(1,0,0)(1,0,0)[12] with non-zero mean : 1648.384
##  ARIMA(0,0,1)(0,0,1)[12] with non-zero mean : 1647.545
##  ARIMA(0,0,0)            with zero mean     : 1650.055
##  ARIMA(0,0,1)            with non-zero mean : 1648.22
##  ARIMA(0,0,1)(1,0,1)[12] with non-zero mean : 1645.723
##  ARIMA(0,0,1)(1,0,0)[12] with non-zero mean : 1647.115
##  ARIMA(0,0,1)(2,0,1)[12] with non-zero mean : 1647.825
##  ARIMA(0,0,1)(1,0,2)[12] with non-zero mean : Inf
##  ARIMA(0,0,1)(0,0,2)[12] with non-zero mean : 1649.213
##  ARIMA(0,0,1)(2,0,0)[12] with non-zero mean : 1648.386
##  ARIMA(0,0,1)(2,0,2)[12] with non-zero mean : Inf
```

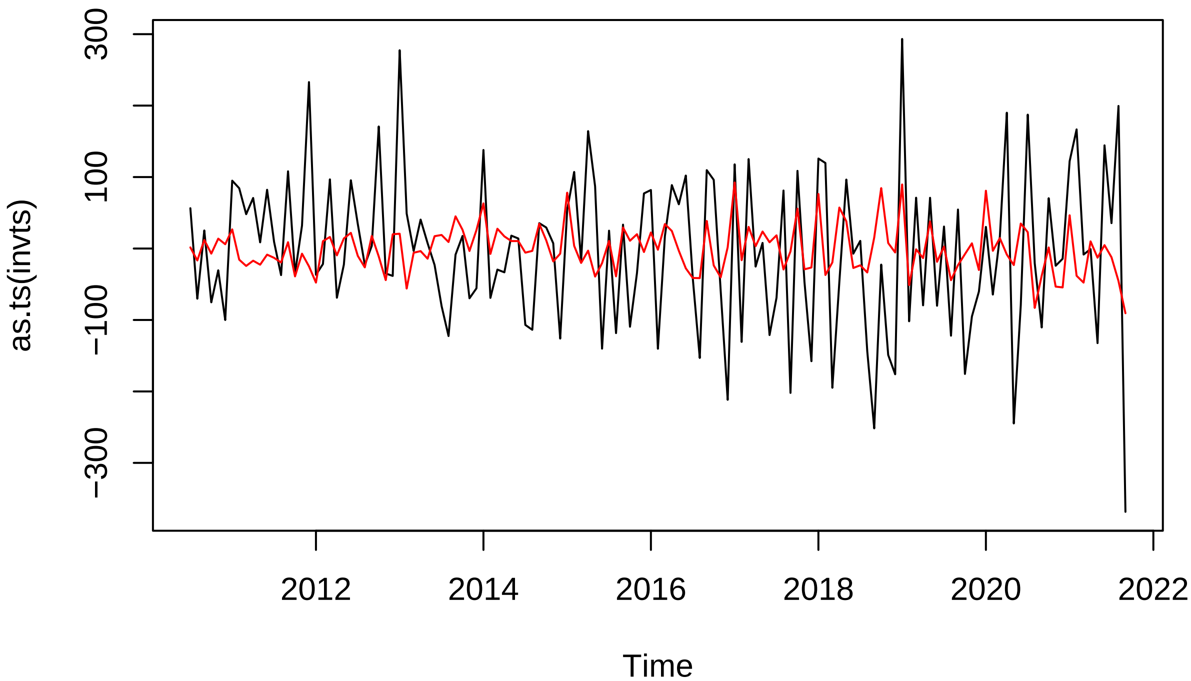
```
##  ARIMA(0,0,0)(1,0,1)[12] with non-zero mean : 1650.522
##  ARIMA(1,0,1)(1,0,1)[12] with non-zero mean : 1647.958
##  ARIMA(0,0,2)(1,0,1)[12] with non-zero mean : 1647.877
##  ARIMA(1,0,0)(1,0,1)[12] with non-zero mean : 1646.76
##  ARIMA(1,0,2)(1,0,1)[12] with non-zero mean : 1649.874
##  ARIMA(0,0,1)(1,0,1)[12] with zero mean     : 1643.595
##  ARIMA(0,0,1)(0,0,1)[12] with zero mean     : 1645.552
##  ARIMA(0,0,1)(1,0,0)[12] with zero mean     : 1645.098
##  ARIMA(0,0,1)(2,0,1)[12] with zero mean     : 1645.659
##  ARIMA(0,0,1)(1,0,2)[12] with zero mean     : Inf
##  ARIMA(0,0,1)            with zero mean     : 1646.355
##  ARIMA(0,0,1)(0,0,2)[12] with zero mean     : 1647.163
##  ARIMA(0,0,1)(2,0,0)[12] with zero mean     : 1646.293
##  ARIMA(0,0,1)(2,0,2)[12] with zero mean     : Inf
##  ARIMA(0,0,0)(1,0,1)[12] with zero mean     : 1648.432
##  ARIMA(1,0,1)(1,0,1)[12] with zero mean     : Inf
##  ARIMA(0,0,2)(1,0,1)[12] with zero mean     : 1645.716
##  ARIMA(1,0,0)(1,0,1)[12] with zero mean     : 1644.629
##  ARIMA(1,0,2)(1,0,1)[12] with zero mean     : 1647.68
##
##  Best model: ARIMA(0,0,1)(1,0,1)[12] with zero mean
```

The models with the lowest AICc is considered as the best model. For our case it is Best model: *ARIMA(0,0,1)(1,0,1)[12] with zero mean*

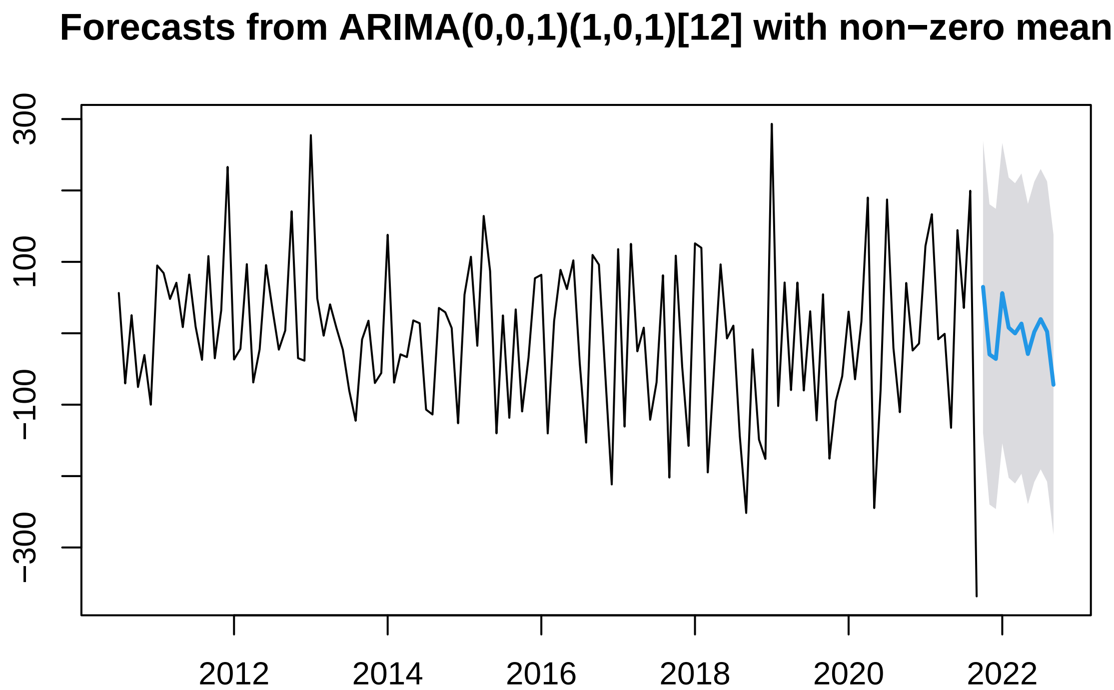



```
##
##  Ljung-Box test
```

```
##
## data:  Residuals from ARIMA(0,0,1)(1,0,1)[12] with non-zero mean
## Q* = 11.336, df = 21, p-value = 0.9558
##
## Model df: 3.   Total lags used: 24
```



Forecasting next 12 months



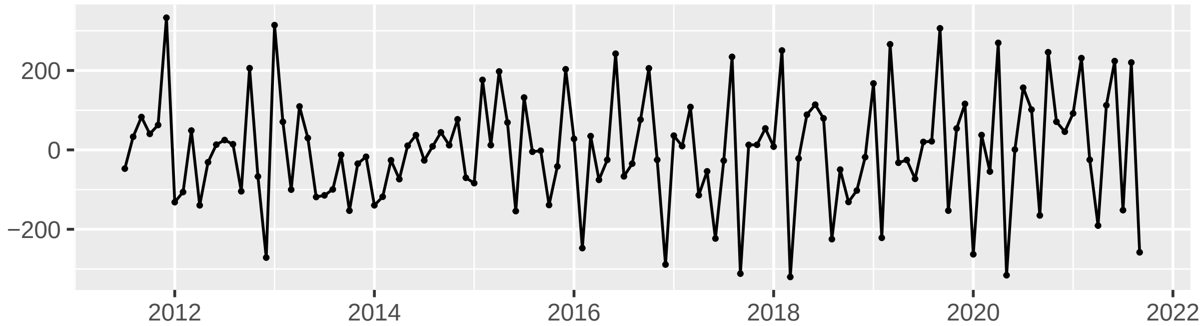
Values

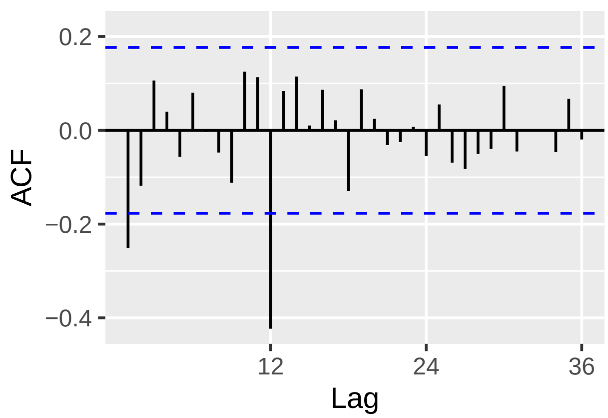
Values of the total automotive gasoline inventory forecasted for the next 10 months.

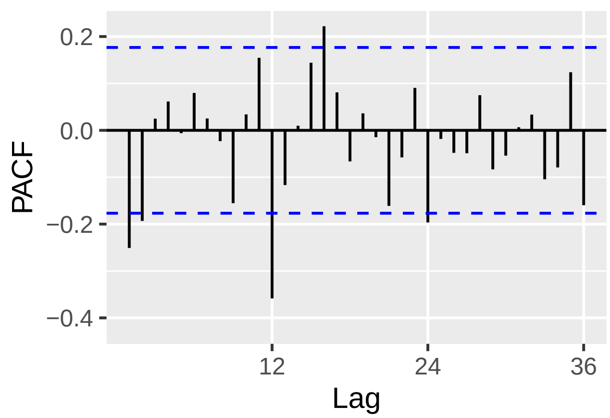
```
##              Jan         Feb         Mar         Apr         May         Jun
## 2021
## 2022  56.1700243   7.9440768  -0.1189176  13.5246087 -29.0508874   1.9352541
##              Jul         Aug         Sep         Oct         Nov         Dec
## 2021                                      64.9115631 -29.6207445 -35.9148091
## 2022  19.7943819   2.3421517 -72.0884939
```

Differencing the Data

Although the tests suggest that there is no differecing required as the data is stationary. There are very high fluctuations in the dataset as inventory is not a macroeconomic data. It is based on so many other factors, hence I conducted another forecast with keeping econometric point of view in mind and adding a differencing in the Seasonal part of the model. If we would have considered more variables in the dataset apart from production, sales, imports and exports, we might not need the differencing.



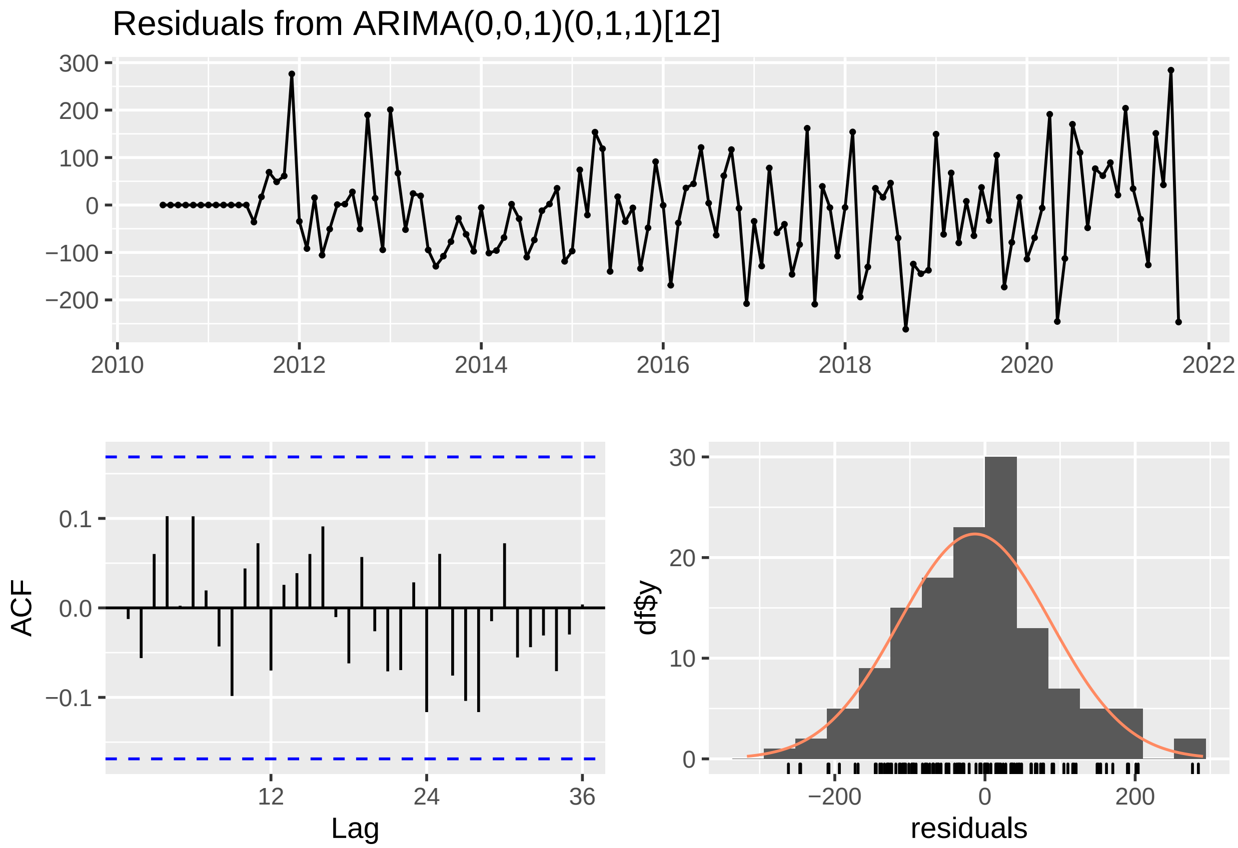




ARIMA (p,d,q) Non seasonal part (P,D,Q) Seasonal part

- Seasonal part- ACF lag at 12 means P=0 Q = 1
- Non seasonal part- exponential curve in PACF hence, ACF p = 0 q = 1
- Non seasonal part- sine curve in ACF Hence, PACF lag at p = 1 q = 0

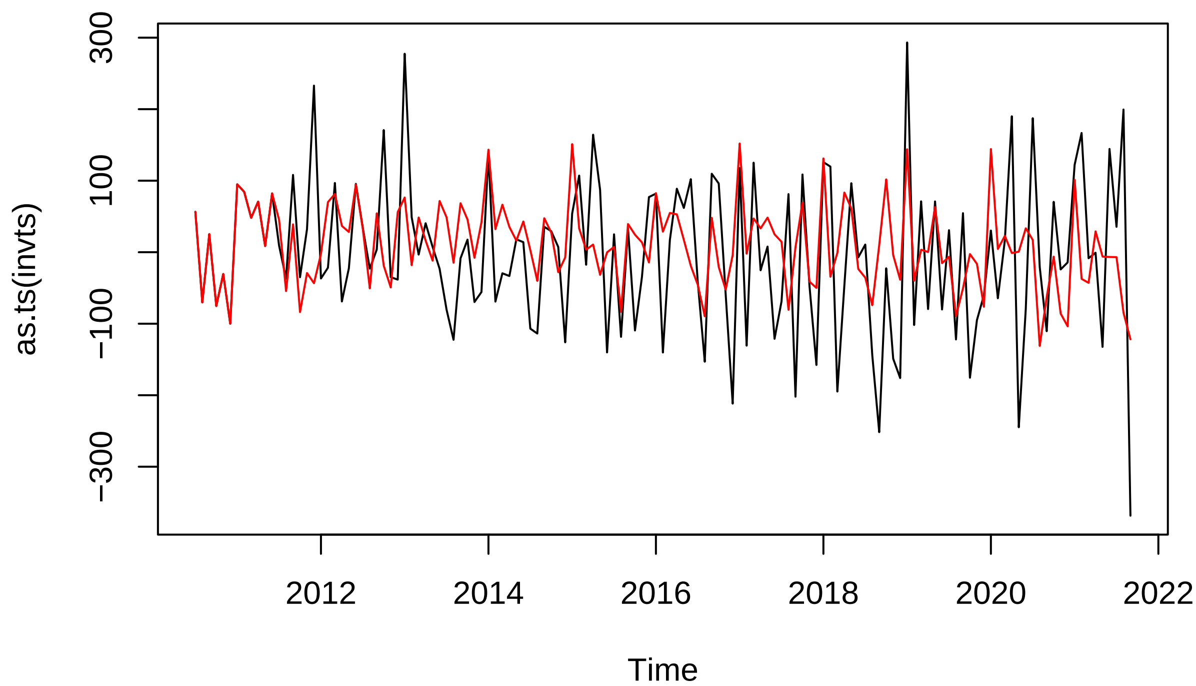
```
## Series: .
## ARIMA(0,0,1)(0,1,1)[12]
##
## Coefficients:
##           ma1     sma1
##       -0.2489  -0.7917
## s.e.   0.0858   0.0972
##
## sigma^2 = 11547:  log likelihood = -754.72
## AIC=1515.43   AICc=1515.63   BIC=1523.87
```

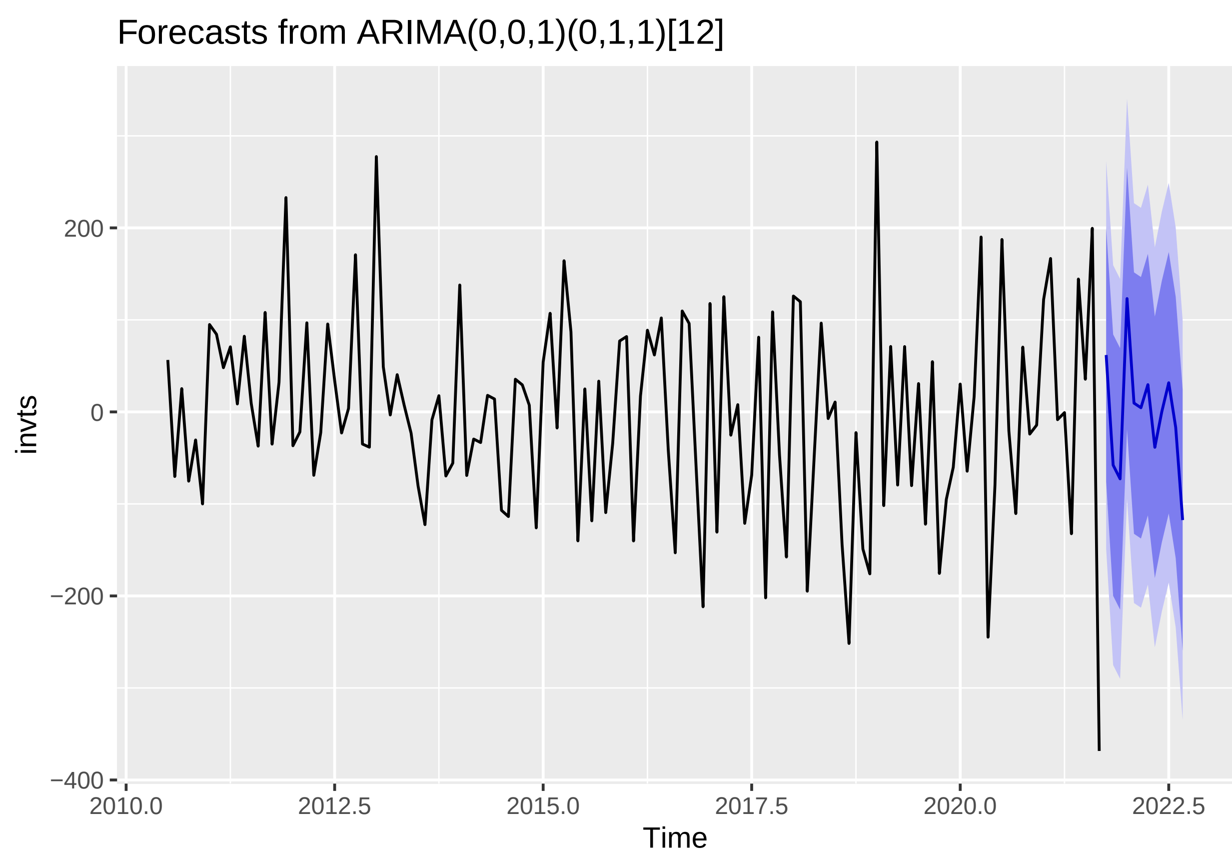



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,1)(0,1,1)[12]
## Q* = 14.928, df = 22, p-value = 0.8653
##
## Model df: 2.   Total lags used: 24
```

The ARIMA model does well in capturing all the dynamics in the data as the residuals seem to be white noise.

Forecasting next 12 months





Conclusion

```
##                      ME     RMSE      MAE      MPE     MAPE      MASE
## Training set -3.078576 102.7165 80.86974 89.00925 126.7029 0.7403807
##                     ACF1
## Training set 0.001665076
```

```
##                      ME    RMSE      MAE     MPE     MAPE      MASE        ACF1
## Training set -13.32024 101.731 76.50035 87.8092 175.6032 0.7003779 -0.01239161
```

This captures the fluctuations a little better than the last forecast ARIMA model. The P value for the model with differencing is higher (0.8653) than the p value without differencing (0.7278) which also suggests that the second model might be better. More over the RMSE value for the model with differencing is less than that of without differencing.