

Time Series forecasting



Author: Amit KULKARNI

Table Contents

1	Project Objective.....	03
2	Assumptions.....	03
3	Exploratory Data Analysis.....	03
	A. Environmental Set up and Data Import.....	03
	B. Install packages and invoke Libraries.....	03
4	Trend Visualization.....	04
5	Seasonality.....	10
6	Stationary Test.....	13
7	Demand Forecasting.....	19
8	Conclusion.....	20

Project Objective: Demand for 2 products A and B are provided for period Jan 2002 till Jul 2017. We are required to check the trends, seasonality and perform best model fit of the time series and forecast the demands for both the products from Sep 2017 till Dec 2018

1. Assumptions: There are no specific assumptions made in the explanation of the solutions
2. Dataset: The provided data is in the .csv file and a sample format is provided below

	A	B	C	D	E	F
1	Year	Month	Item A	Item B	Diff A	Diff B
2	2002	1	1954	2585		
3	2002	2	2302	3368	348	783
4	2002	3	3054	3210	752	-158
5	2002	4	2414	3111	-640	-99
6	2002	5	2226	3756	-188	645
7	2002	6	2725	4216	499	460
8	2002	7	2589	5225	-136	1009
9	2002	8	3470	4426	881	-799
10	2002	9	2400	3932	-1070	-494
11	2002	10	3180	3816	780	-116
12	2002	11	4009	3661	829	-155
13	2002	12	3924	3795	-85	134
14	2003	1	2072	2285	-1852	-1510
15	2003	2	2434	2934	362	649
16	2003	3	2956	2985	522	51
17	2003	4	2828	3646	-128	661
18	2003	5	2687	4198	-141	552
19	2003	6	2629	4935	-58	737
20	2003	7	3150	5618	521	683
21	2003	8	4119	5454	969	-164
22	2003	9	3030	3624	-1089	-1830
23	2003	10	3055	2898	25	-726

3.

4. Exploratory Data Analysis

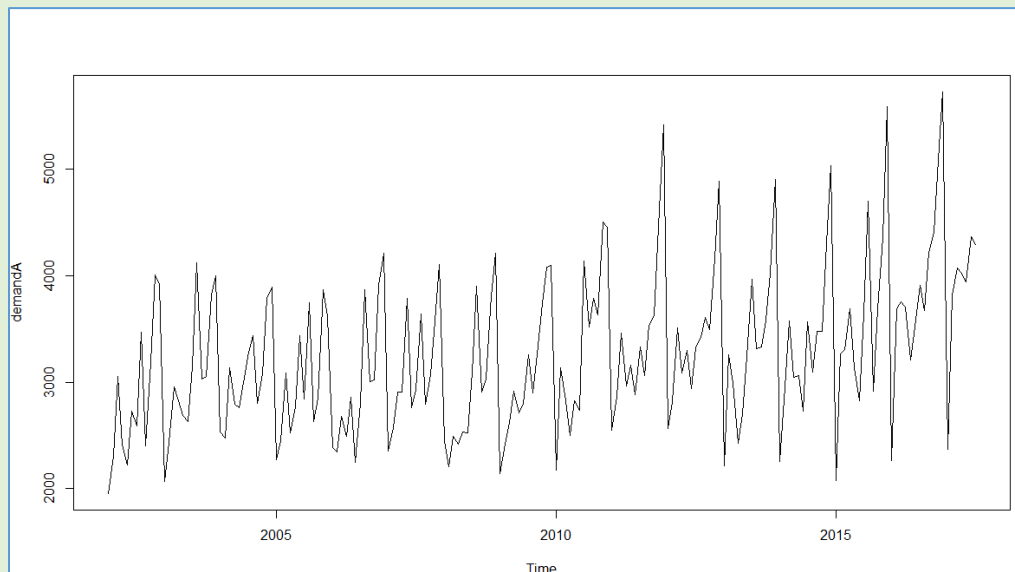
4.1. Environmental Set Up and Data Import: The provided dataset will be analyzed in R STUDIO by importing it using the `read.csv` function. Simple Time Series Forecasting methods like ARIMA etc. will be used to forecast the demands for the 2 products. Data will be divided in the train and test and will be validated using the MAPE technique

4.2. Install packages and invoke libraries: Variety of packages used and invoked libraries.

- 4.2.1. tseries
- 4.2.2. reshape2
- 4.2.3. fpp2
- 4.2.4. Forecast

Q1: Before a formal extraction of time series components is done, can you check for seasonal changes in the data for the two series separately? Particularly whether there are more variability in a season compared to the others, whether seasonal variations are changing across years etc. Compare the behavior of the two series

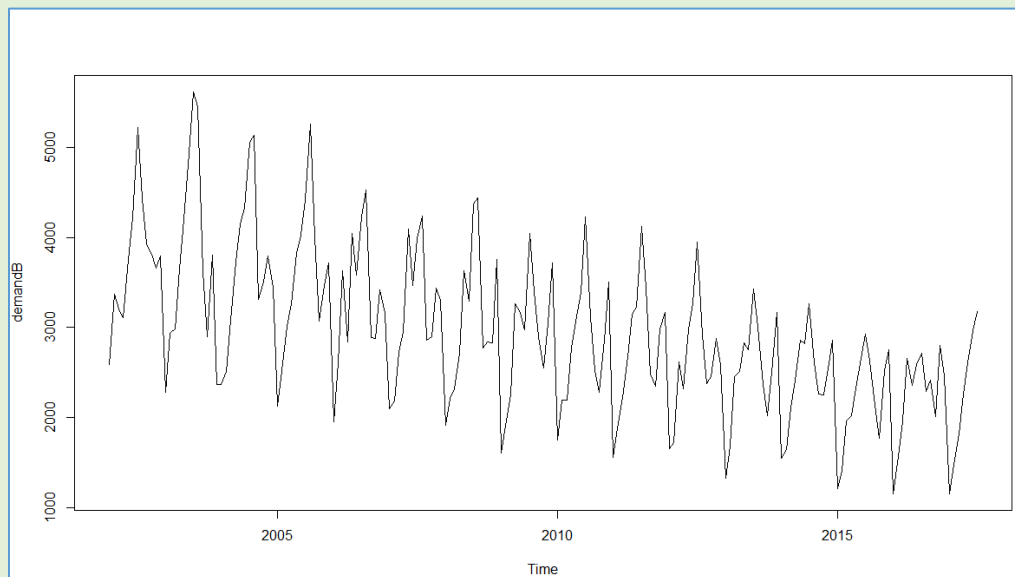
Visualization for Product A



Visualizing the data it can be seen that there exists no trend for product A as it shows steady demand for the product over the period and also no seasonality

Visualization for Product B:

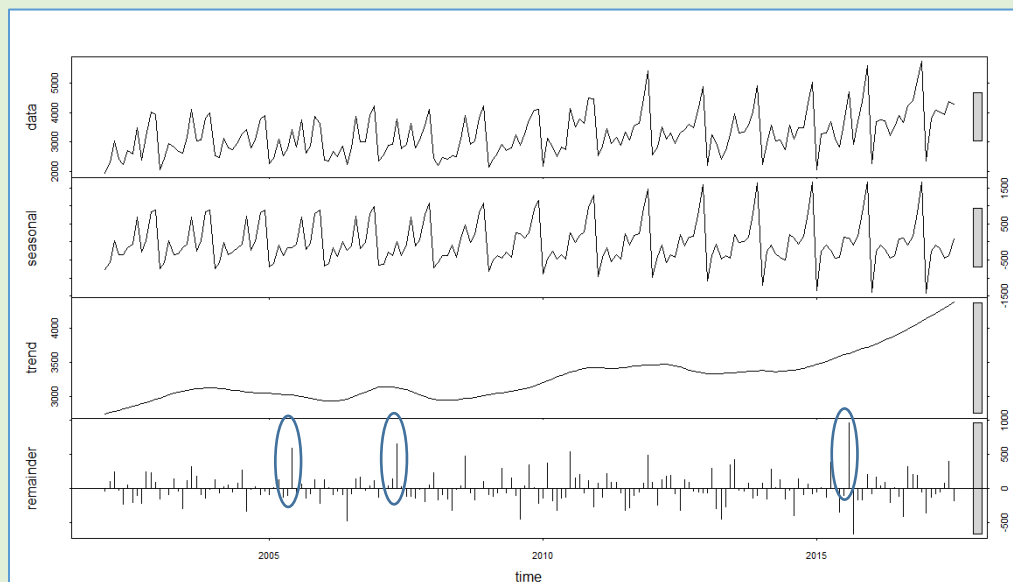
The below graph depicts that there is gradual declining trend for the product and seasonality also exists to a certain extent



Q2: Decompose each series to extract trend and seasonality, if there are any. Which seasonality is more appropriate – additive or multiplicative? Explain the seasonal indices. In which month(s) do you see higher sales and which month(s) you see lower sales? Any difference in the nature of demand of the two items?

```
# decomposition  
  
demandAst1 = stl(demandA, s.window = 7)  
plot(demandAst1)
```

Decomposition is done using the 'stl' function in R. It is assumed that there are no seasonal patterns for A, s.window has been used with number 7.



The decomposed graph is divided in 4 parts.

1st part is the actual data plotted

2nd part is the seasonal component

3rd part is the trend of the data. This part is very similar to the first part but only in a smoothed manner

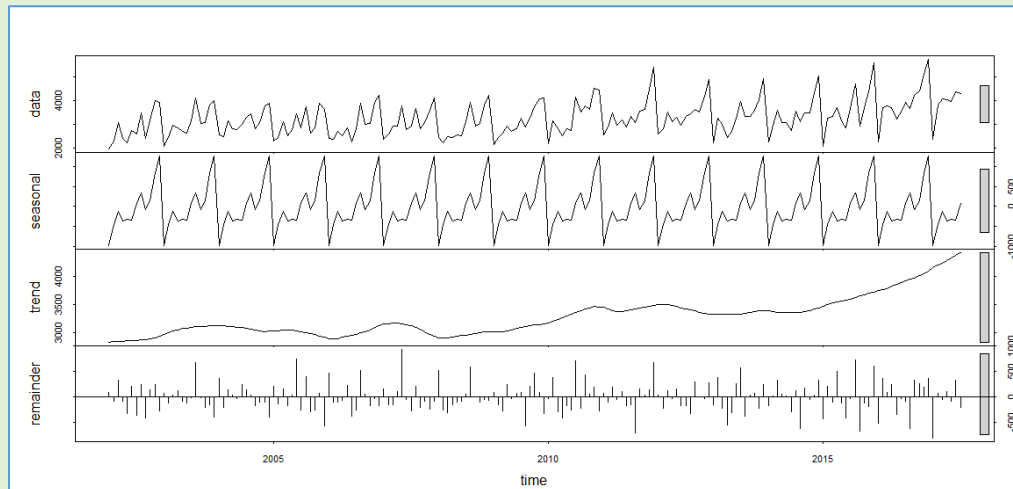
4th part is the unexplained or the remainder part which the decomposition could not explain. It can be observed that there are few occasions that the decomposition could not identify the spikes in the data (marked in circles). These can be investigated to see what could have been the reasons for the spikes. Since no information is available not much comments could be done.

Since we have assumed that there are no seasonality we considered a number while decomposing and we could see that the seasonality is not evenly spread. However if you

look at the grey bars at the right hand corner, its length of the bar decides the significance of either trend or seasonality.

Lesser the length of the bar the more significant that component and thus it can be seen that the seasonality is significant than the trend of the data.

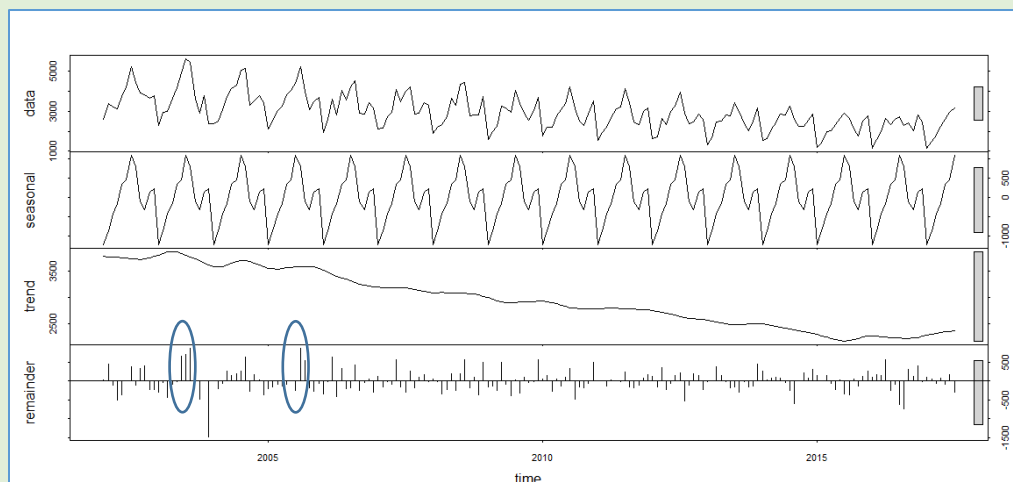
Let us also observe the data assuming there is seasonality and we use 'p' for the s.window



It can be observed that due to seasonality assumption the seasonal part of the graph is now much more aligned. It can also be seen that the grey line is still smaller than the trend and thus making seasonality significant. Rest of the graph remains the same.

Similar can be observed for demand for product B

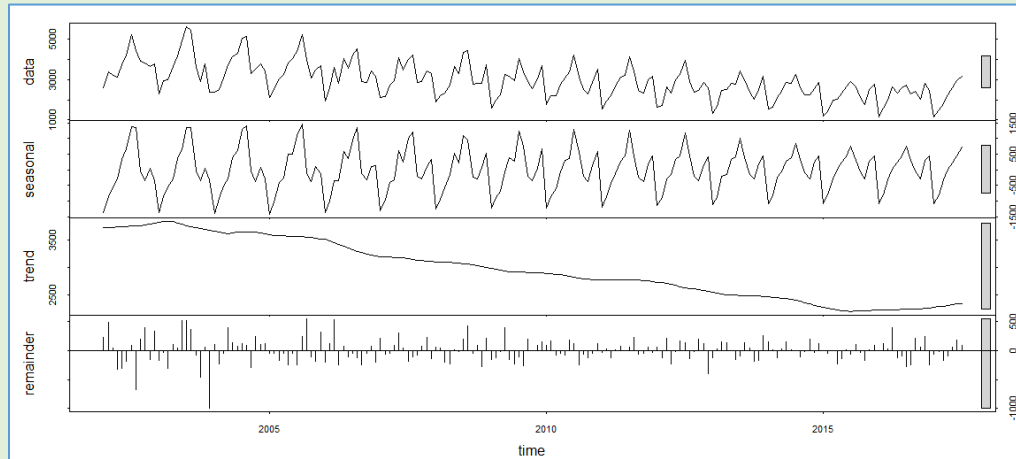
Assumption that there exists no seasonality and thus using s.window as 7



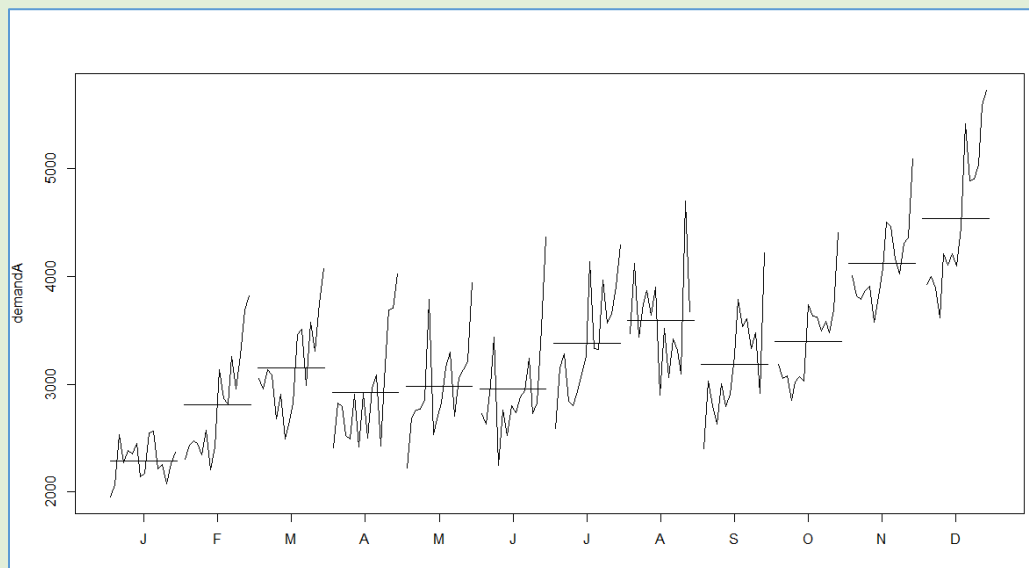
It can be observed that the seasonality pattern is much aligned and also the grey bar is also shorter than the trend.

There are again at times when the decomposition could not identify spikes as shown in the circles. The trend also shows that there is a declining trend for the product.

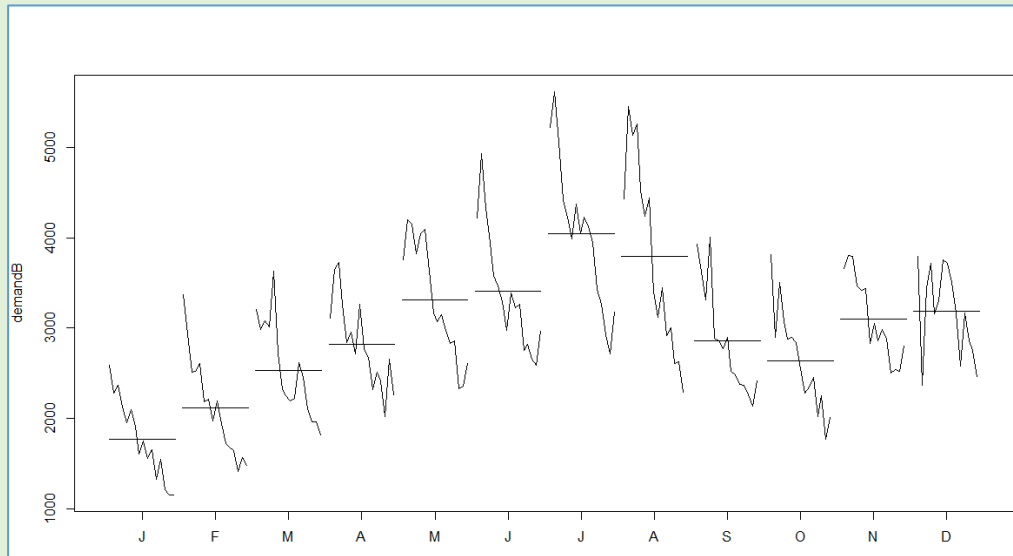
With seasonality assumed the significance still remains for seasonality. Rest everything remains the same.



Visualization using month plot for product A: With the below graph it is evident that the sales are high during the last 2 months i.e. Nov and Dec and lowest during Jan. The sales stabilize during Apr, May and Jun post which spikes are seen in the next couple of months only to witness reduced sales in Sep and gradually reach the peak



Month plot for product B: For Product B, it is evident that sales are high during the months of Jul. Lowest sales being in Jan and they gradually increase every month to reach peak in Jul and later dip further month on month except a small recovery during Nov and Dec



With the above product demands being broken down into monthly sales and considering the entire product demand analysis it can be said that

- i. The nature of the products are very different since the demand for product A is a steady one.
- ii. There can be some seasonality but overall the demand for the product A is steady.
- iii. The peak periods of the product A also is towards the end of the year as compared to B which is somewhere at the middle of the year.
- iv. Product B is on the declining trend. This could mean that the product is at its end of life cycle or it requires reinventing the wheel so as to sustain it into the market.

Q3. Can you extract the residuals for the two decomposition exercises and check if they form a stationary series? Do a formal test for stationarity writing down the null and alternative hypothesis. What is your conclusion in each case?

The products are defined in the below time series object in R

demandA and demandB

To perform an ARIMA model test, it is a mandatory requirement that the data has to be stationary. Typically data with trends or seasonality will not stationary series. However to test if the data is stationary or not Augmented Dickey Fuller test is done. In this test

The test is a Hypothesis testing done where

Null hypothesis is the data is not stationary

Alternative hypothesis is that the data is stationary

In our case the test is conducted on the time series defined objects demandA and demandB

```
# Doing the stationary test using the ADF test
adf.test(demandA)
adf.test(demandB)
```

The results for both the TS objects demandA and demandB are given below

```
Augmented Dickey-Fuller Test

data: demandA
Dickey-Fuller = -7.8632, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

```
Augmented Dickey-Fuller Test

data: demandB
Dickey-Fuller = -12.967, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Given the p value of 0.01 for both the time series objects, and considering the threshold value of p-value to be 0.05 it can be said that the alternative hypothesis will be accepted which means that the data is stationary and an ARIMA model can be built on this data

Q4. Before the final forecast is undertaken one would like to compare a few models. Use the last 21 months as hold-out sample fit a suitable exponential smoothing model to the rest of the data and calculate MAPE. What are the values of α , β and γ ? What role do they play in the modeling? For the same hold-out period compare forecast by decomposition and compute MAPE. Which model gives smaller MAPE? Give a comparison for the two demands

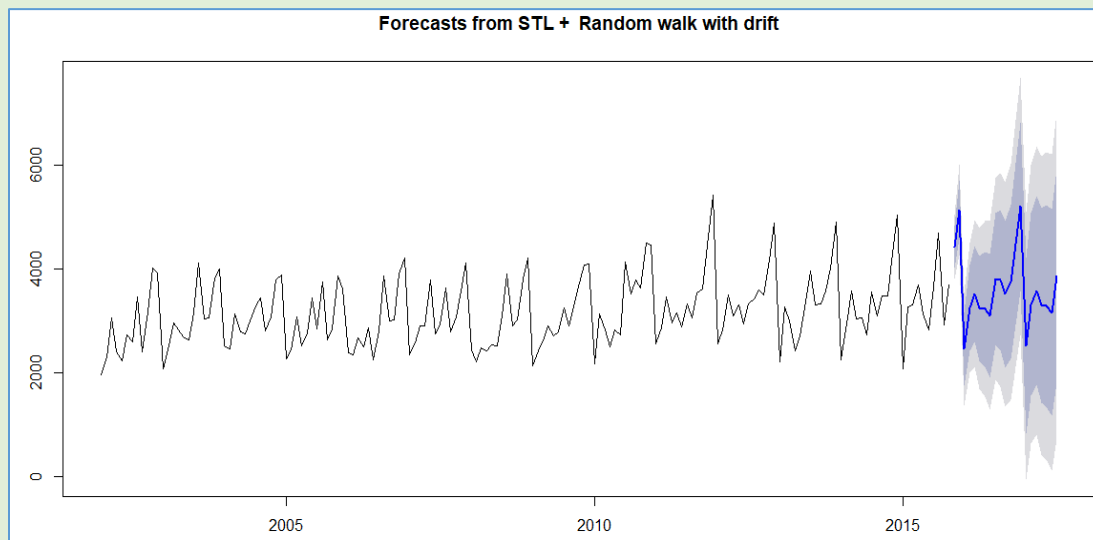
R Codes for doing an exponential smoothing forecast with MAPE

```
# Demand Forecasting and calculation of MAPE under exponential smoothing
forecast.demandA <- forecast(DmndTrnA, method="rwdrift", h=21)
plot(forecast.demandA)
forecast.demandA
forecast.demandB <- forecast(DmndTrnB, method="rwdrift", h=21)
plot(forecast.demandB)
forecast.demandB
VecA<- cbind(demandAtest,forecast.demandA$mean)
ts.plot(VecA, col=c("blue", "red"), main="Demand A vs Forecast")
MAPE <- mean(abs(VecA[,1]-VecA[,2])/VecA[,1])
MAPE
VecB<- cbind(demandBtest,forecast.demandB$mean)
ts.plot(VecB, col=c("blue", "red"), main="Demand A vs Forecast")
MAPE <- mean(abs(VecB[,1]-VecB[,2])/VecB[,1])
MAPE
# Forecast for simple exponential smoothing
fcdemandA = ses(demandAtest, h =5)
plot(fcdemandA)
fcdmndA = ses(demandA, h =3)
plot(fcdmndA)
ts.plot(demandA, fcdmndA$fitted, col=c("brown", "gold"))
round(accuracy(fcdmndA),2)
fcdmndA$model
fcdmndA$mean
fcdmndA2 = ses(demandA, alpha = 0.2)
fcdmndA2
fcdmndA5 = ses(demandA, alpha = 0.5)
fcdmndA5
fcdmndA8 = ses(demandA, alpha = 0.8)
fcdmndA8
ts.plot(demandA,fcdmndA2$fitted, fcdmndA5$fitted, fcdmndA8$fitted, col=c("black","red","blue"))
```

MAPE for both products

```
> VecA<- cbind(demandAtest,forecast.demandA$mean)
> ts.plot(VecA, col=c("blue", "red"), main="Demand A vs Forecast")
> MAPE <- mean(abs(VecA[,1]-VecA[,2])/VecA[,1])
> MAPE
[1] 0.105762
> VecB<- cbind(demandBtest,forecast.demandB$mean)
> ts.plot(VecB, col=c("blue", "red"), main="Demand A vs Forecast")
> MAPE <- mean(abs(VecB[,1]-VecB[,2])/VecB[,1])
> MAPE
[1] 0.1694894
```

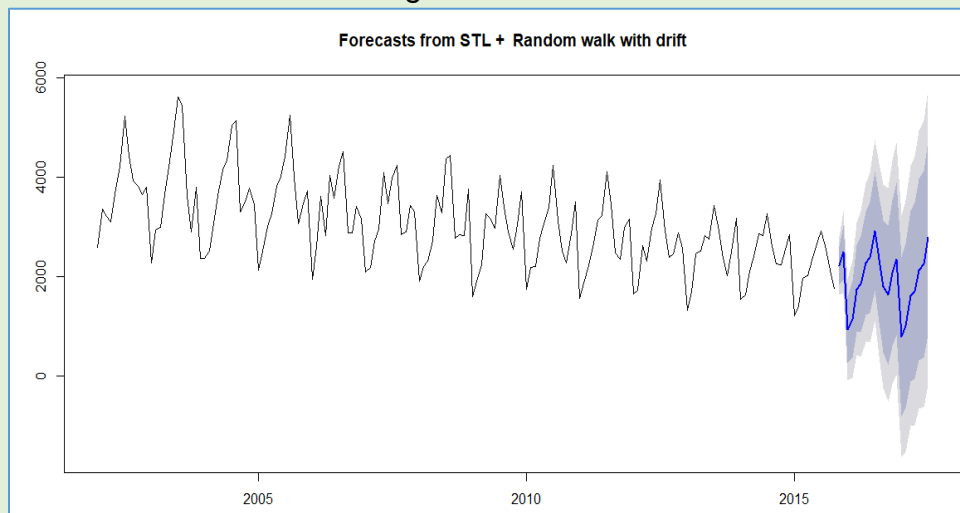
Demand Forecast for A using the random walk with drift



```
> forecast.demandA
```

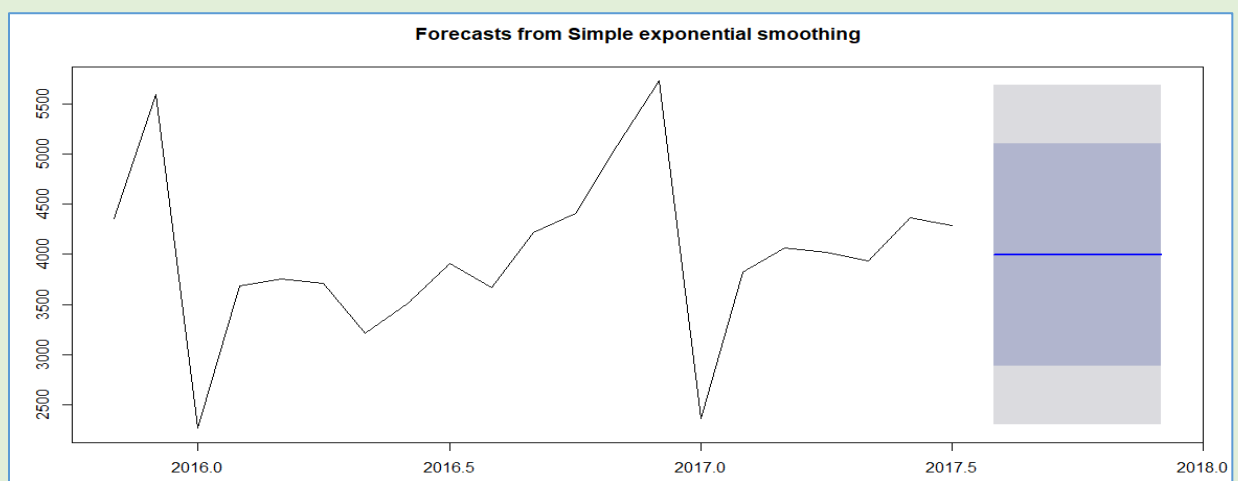
	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Nov	2015	4420.996	4018.426	4823.565	3805.3185	5036.673
Dec	2015	5138.832	4566.093	5711.571	4262.9033	6014.760
Jan	2016	2453.320	1747.698	3158.942	1374.1643	3532.475
Feb	2016	3252.907	2433.346	4072.468	1999.4968	4506.317
Mar	2016	3521.415	2599.807	4443.024	2111.9363	4930.894
Apr	2016	3240.289	2224.932	4255.646	1687.4340	4793.144
May	2016	3226.707	2123.782	4329.632	1539.9290	4913.485
Jun	2016	3100.183	1914.501	4285.865	1286.8389	4913.527
Jul	2016	3805.255	2540.682	5069.828	1871.2578	5739.253
Aug	2016	3785.818	2445.537	5126.099	1736.0353	5835.601
Sep	2016	3513.509	2100.192	4926.827	1352.0266	5674.992
Oct	2016	3748.572	2264.495	5232.650	1478.8720	6018.273
Nov	2016	4483.568	2930.697	6036.439	2108.6563	6858.479
Dec	2016	5201.404	3581.454	6821.354	2723.9044	7678.904
Jan	2017	2515.892	830.374	4201.410	-61.8855	5093.670
Feb	2017	3315.479	1565.733	5065.225	639.4737	5991.485
Mar	2017	3583.988	1771.212	5396.763	811.5868	6356.388
Apr	2017	3302.861	1428.134	5177.589	435.7127	6170.010
May	2017	3289.280	1353.573	5224.986	328.8724	6249.687
Jun	2017	3162.755	1166.955	5158.555	110.4424	6215.068
Jul	2017	3867.828	1812.741	5922.915	724.8432	7010.812

Demand Forecast for B using the random walk with drift



> forecast.demandB					
	Point	Forecast	Lo 80	Hi 80	Lo 95
Nov 2015	2222.7378	1842.98553	2602.490	1641.95680	
Dec 2015	2496.3945	1956.11808	3036.671	1670.11303	
Jan 2016	936.5148	270.88704	1602.143	-81.47503	
Feb 2016	1146.8242	373.71543	1919.933	-35.54362	
Mar 2016	1752.3607	882.98798	2621.733	422.76988	
Apr 2016	1853.1222	895.31451	2810.930	388.28174	
May 2016	2273.2389	1232.82680	3313.651	682.06591	
Jun 2016	2391.3840	1272.90557	3509.862	680.81885	
Jul 2016	2927.5475	1734.64951	4120.446	1103.16745	
Aug 2016	2344.5098	1080.19483	3608.825	410.90692	
Sep 2016	1808.2158	475.00392	3141.428	-230.75583	
Oct 2016	1634.3589	234.39799	3034.320	-506.69655	
Nov 2016	2085.0967	620.24105	3549.952	-155.20671	
Dec 2016	2358.7534	830.62095	3886.886	21.67650	
Jan 2017	798.8738	-791.11063	2388.858	-1632.79752	
Feb 2017	1009.1831	-641.38861	2659.755	-1515.14848	
Mar 2017	1614.7196	-95.30920	3324.748	-1000.54376	
Apr 2017	1715.4811	-52.98854	3483.951	-989.15984	
May 2017	2135.5978	309.60587	3961.590	-657.01586	
Jun 2017	2253.7429	371.06317	4136.423	-625.56727	
Jul 2017	2789.9064	851.30012	4728.513	-174.93607	

Under simple exponential smoothing, the value forecasted is a single value for any period as shown in the below graph



Using the Holt Winter Model

Forecast graphical representation for product A

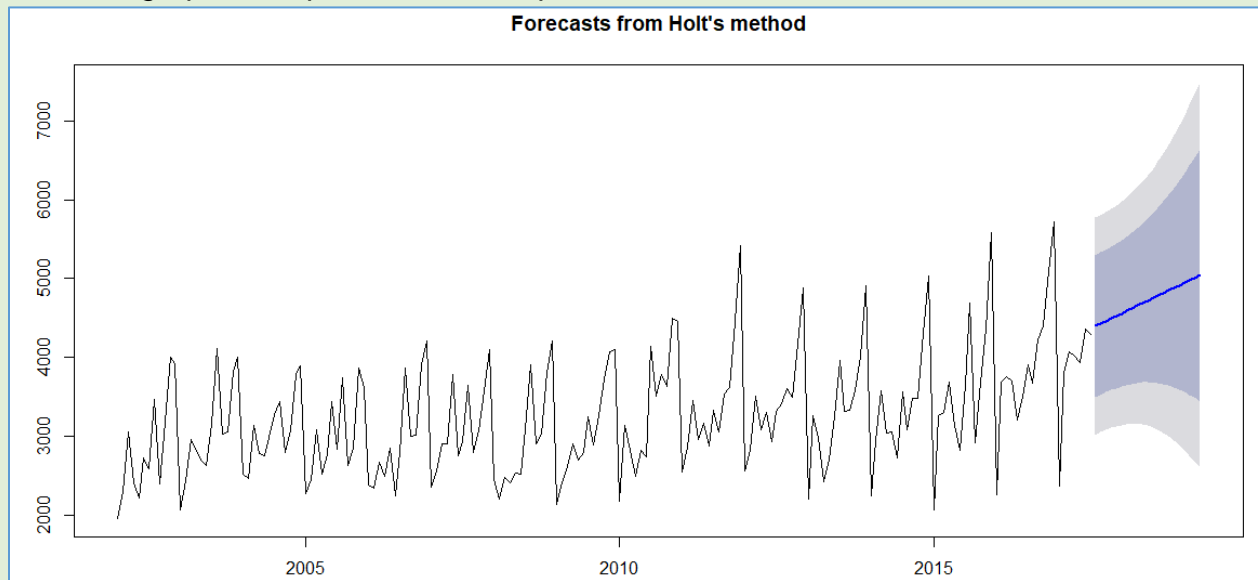


Table depicting the values forecasted

```
> fcdmnda_Holt
```

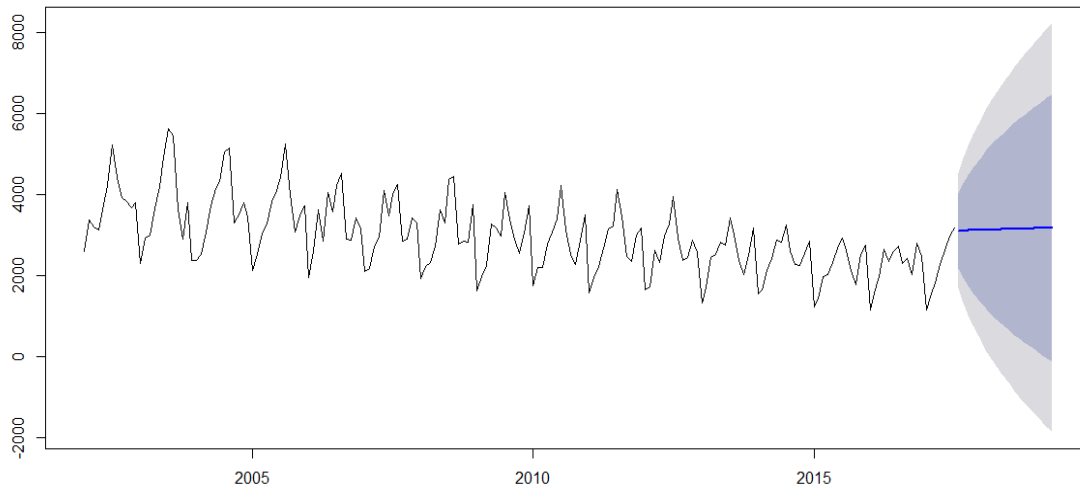
	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Aug 2017		4397.713	3498.450	5296.977	3022.409	5773.018
Sep 2017		4430.042	3529.612	5330.473	3052.953	5807.132
Oct 2017		4462.371	3559.320	5365.422	3081.274	5843.468
Nov 2017		4494.700	3587.009	5402.390	3106.507	5882.892
Dec 2017		4527.028	3612.136	5441.921	3127.821	5926.236
Jan 2018		4559.357	3634.192	5484.522	3144.439	5974.275
Feb 2018		4591.686	3652.719	5530.653	3155.660	6027.712
Mar 2018		4624.015	3667.321	5580.708	3160.878	6087.151
Apr 2018		4656.343	3677.675	5635.011	3159.600	6153.087
May 2018		4688.672	3683.537	5693.807	3151.451	6225.893
Jun 2018		4721.001	3684.745	5757.257	3136.184	6305.818
Jul 2018		4753.330	3681.213	5825.446	3113.669	6392.990
Aug 2018		4785.658	3672.929	5898.387	3083.886	6487.431
Sep 2018		4817.987	3659.939	5976.035	3046.906	6589.068
Oct 2018		4850.316	3642.338	6058.293	3002.874	6697.758
Nov 2018		4882.644	3620.258	6145.031	2951.991	6813.298
Dec 2018		4914.973	3593.854	6236.093	2894.495	6935.451
Jan 2019		4947.302	3563.297	6331.307	2830.649	7063.955
Feb 2019		4979.631	3528.766	6430.496	2760.724	7198.537
Mar 2019		5011.959	3490.439	6533.480	2684.994	7338.925
Apr 2019		5044.288	3448.491	6640.085	2603.727	7484.849

MAPE A

```
> MAPE_Holt
[1] 0.1733669
```

For Product B

Forecasts from Holt's method



MAPE B

```
> MAPE_Holt_B
[1] 0.2190072
```

```
> fcdmndB_Holt
      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
Aug 2017    3106.126 2177.34388 4034.908 1685.67627 4526.576
Sep 2017    3109.574 1942.32632 4276.822 1324.42265 4894.726
Oct 2017    3113.022 1748.31346 4477.731 1025.88032 5200.164
Nov 2017    3116.470 1579.41990 4653.520  765.75464 5467.186
Dec 2017    3119.918 1427.95228 4811.884  532.27965 5707.557
Jan 2018    3123.366 1289.48914 4957.243  318.69327 5928.039
Feb 2018    3126.814 1161.21178 5092.417  120.68471 6132.944
Mar 2018    3130.262 1041.19253 5219.332  -64.69416 6325.219
Apr 2018    3133.710  928.04415 5339.376 -239.56495 6506.985
May 2018    3137.158  820.72887 5453.588 -405.51477 6679.831
Jun 2018    3140.606  718.44634 5562.766 -563.76768 6844.980
Jul 2018    3144.054  620.56387 5667.545 -715.29127 7003.400
Aug 2018    3147.502  526.57107 5768.434 -860.86613 7155.871
Sep 2018    3150.950  436.04908 5865.852 -1001.13282 7303.033
Oct 2018    3154.398  348.64916 5960.148 -1136.62475 7445.421
Nov 2018    3157.846  264.07723 6051.615 -1267.79162 7583.484
Dec 2018    3161.294  182.08263 6140.506 -1395.01681 7717.606
Jan 2019    3164.742  102.44960 6227.035 -1518.63028 7848.115
Feb 2019    3168.190   24.99089 6311.390 -1638.91842 7975.299
Mar 2019    3171.638  -50.45730 6393.734 -1756.13176 8099.409
Apr 2019    3175.086 -124.03924 6474.212 -1870.49088 8220.664
```

Holt's method

```
Call:
holt(y = demandB, h = 21)
```

```
Smoothing parameters:
  alpha = 0.7611
  beta  = 1e-04
```

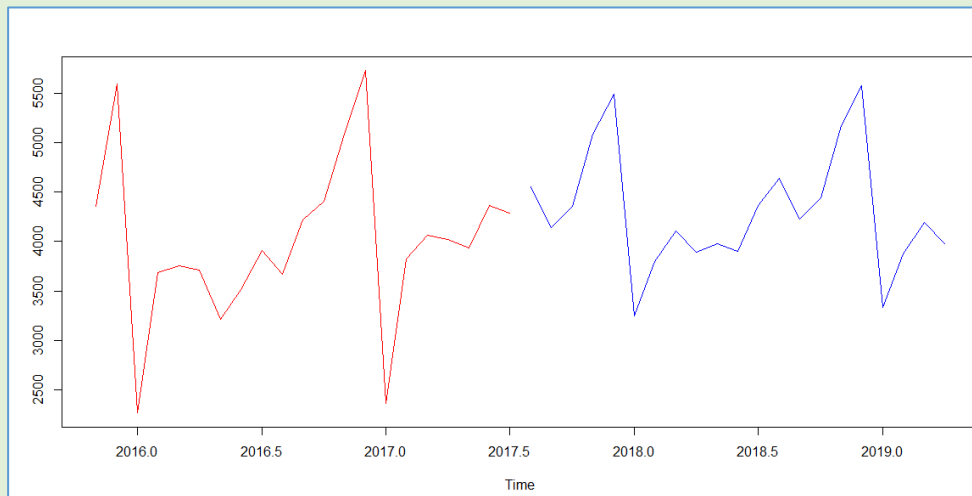
```
Initial states:
  l = 3478.2099
  b = 3.5827
```

```
sigma: 724.7326
```

```
      AIC      AICc      BIC
3447.264 3447.596 3463.420
```

Use of “hw” function in Holt winter model

For product A: Graphical representation of forecast trends colored in “Blue”



MAPE A

```
> MAPE_HW  
[1] 0.08203897
```

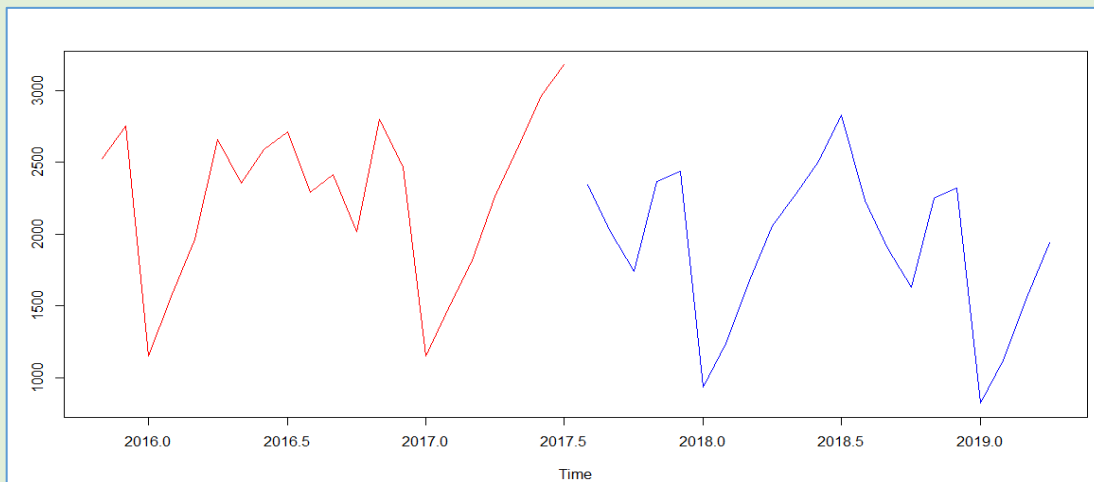
The Alpha, Beta and Gamma Values are as below for product A

```
Smoothing parameters:  
alpha = 0.1331  
beta  = 1e-04  
gamma = 4e-04
```

	Jan	Feb	Mar	Apr	May	Jun	Jul
2017							
2018	3250.725	3793.940	4109.453	3893.741	3977.923	3899.630	4361.673
2019	3334.871	3878.086	4193.599	3977.887			
	Aug	Sep	Oct	Nov	Dec		
2017	4552.034	4140.217	4354.567	5088.377	5488.777		
2018	4636.179	4224.362	4438.712	5172.522	5572.923		
2019							

Use of “hw” function in Holt winter model

For product B: Graphical representation of forecast trends colored in “Blue”



MAPE B

```
> MAPE_HW_B  
[1] 0.07989064
```

The Alpha, Beta and Gamma Values are as below for product B

```
Smoothing parameters:  
alpha = 1e-04  
beta  = 1e-04  
gamma = 0.3467
```

```
> demandBHW$mean
```

	Jan	Feb	Mar	Apr	May	Jun
2017						
2018	937.1552	1239.7282	1679.0992	2056.5112	2281.9294	2511.2094
2019	821.1228	1123.6958	1563.0669	1940.4788		
	Jul	Aug	Sep	Oct	Nov	Dec
2017		2346.8725	2016.8723	1744.5830	2366.0301	2436.7843
2018	2832.0218	2230.8402	1900.8400	1628.5507	2249.9978	2320.7520
2019						

Under the Auto Correlation Factor (ARIMA model)

Code sheet

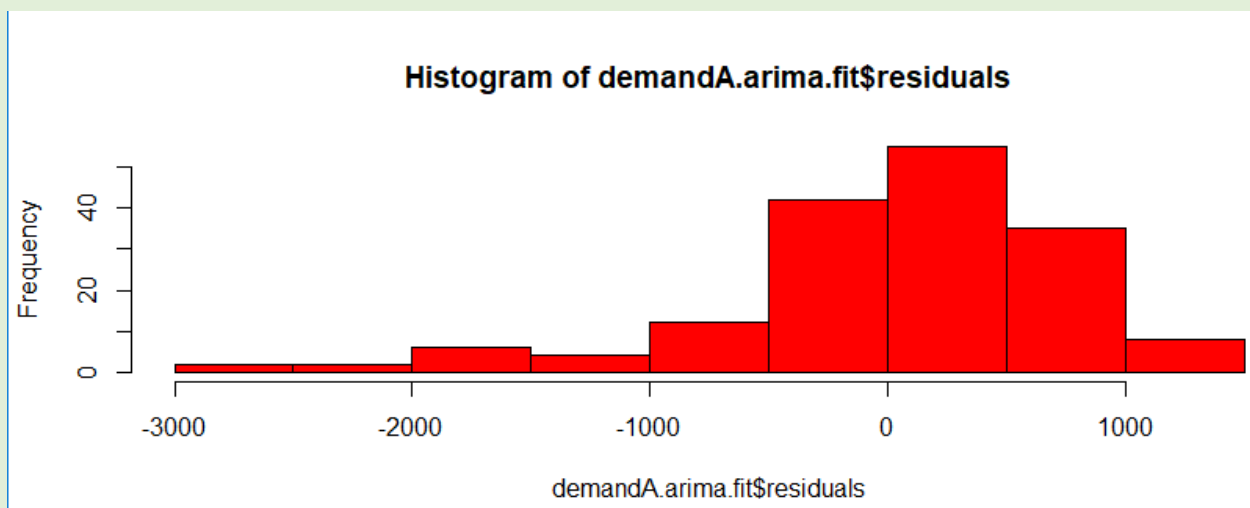
```
#AutoRegression ACF
plot(acf(demandAtrain, lag=50))
plot(acf(demandBtrain, lag=50))

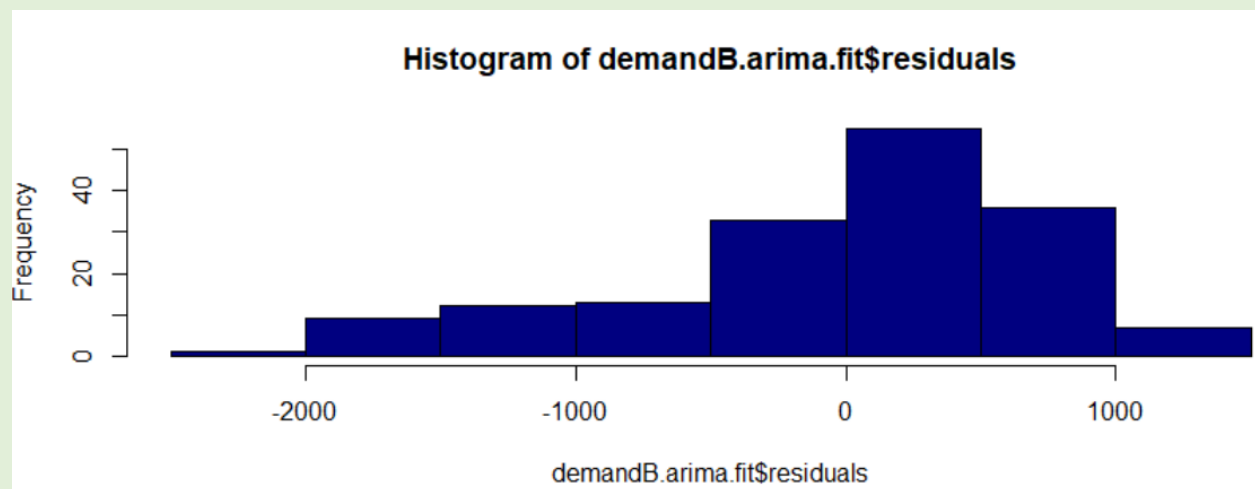
demandA.arima.fit = arima(demandAtrain, c(1,1,0))
demandA.arima.fit
hist(demandA.arima.fit$residuals, col="red")
demandB.arima.fit = arima(demandBtrain, c(1,1,0))
demandB.arima.fit
hist(demandB.arima.fit$residuals, col="navy")
Box.test(demandB.arima.fit$residuals, lag=30, type="Ljung-Box")
# setwd ("C:/Users/Amit Kulkarni/Documents/R Programming/packages")

demandA.arima.fit_FC = forecast(demandA.arima.fit, h=15)
plot(demandA.arima.fit_FC)
demandA.arima.fit_FC

demandB.arima.fit_FC = forecast(demandB.arima.fit, h=15)
plot(demandB.arima.fit_FC)
demandB.arima.fit_FC
```

Under the Auto regression model, the observed values are regressed amongst themselves. The 2nd one regressed with the 1st, 3rd with 2nd, 4th with 3rd so on so forth. This basically tries to identify that if there is any impact of the distant past observation that is having influence on the current one.





```

Coefficients:
      ar1
    -0.3167
s.e.    0.0738

sigma^2 estimated as 593831:  log likelihood = -1330.96,  aic = 2665.92

```

Portmanteau test using Box.Plot function for product B and A

```

> Box.test(demandB.arima.fit$residuals, lag=30, type="Ljung-Box")

      Box-Ljung test

data:  demandB.arima.fit$residuals
X-squared = 350.91, df = 30, p-value < 2.2e-16

```

```

> Box.test(demandA.arima.fit$residuals, lag=30, type="Ljung-Box")

      Box-Ljung test

data:  demandA.arima.fit$residuals
X-squared = 360.23, df = 30, p-value < 2.2e-16

```

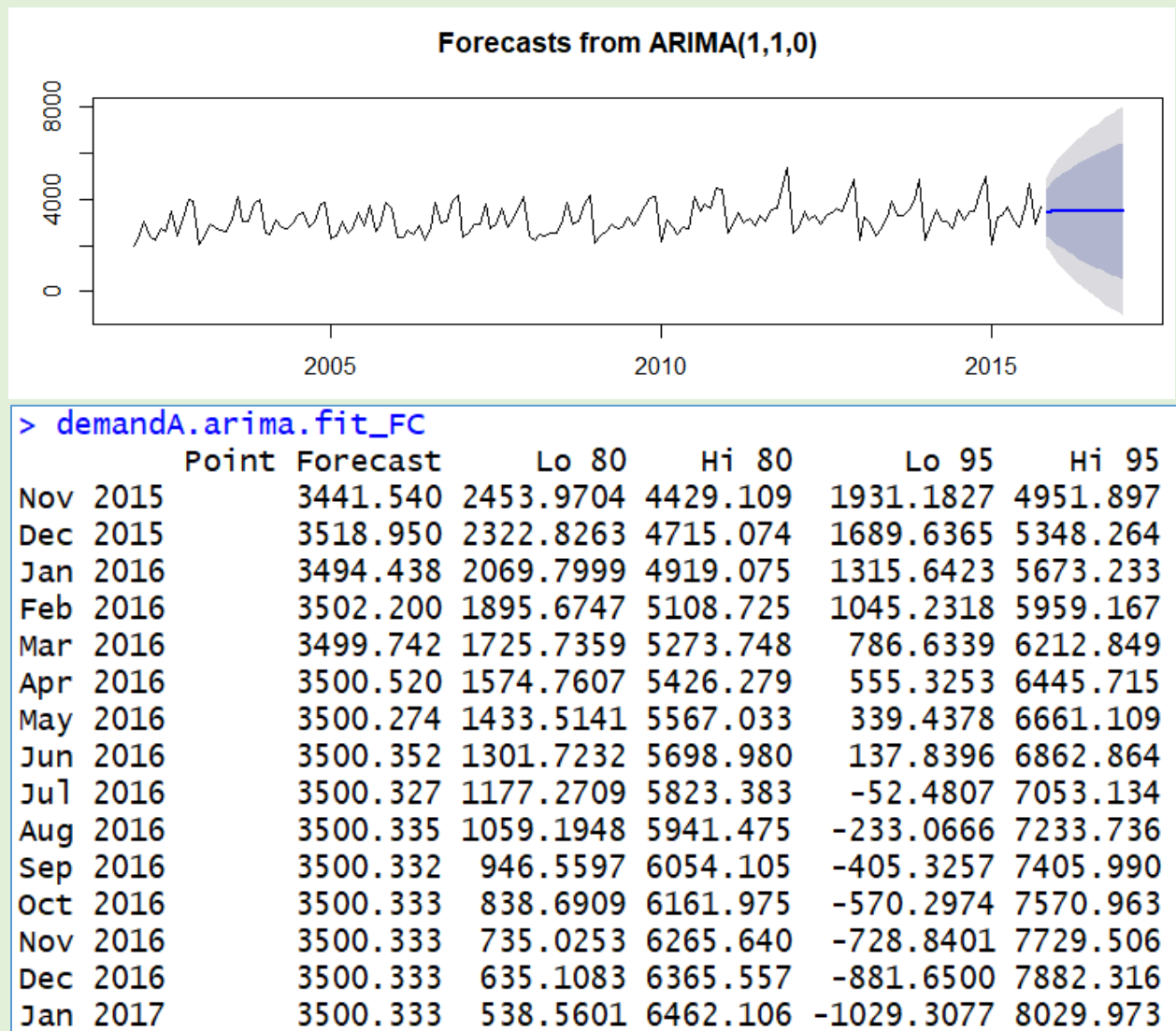
This is to test if the residuals are independent or not.

H_0 = the residual values are independent

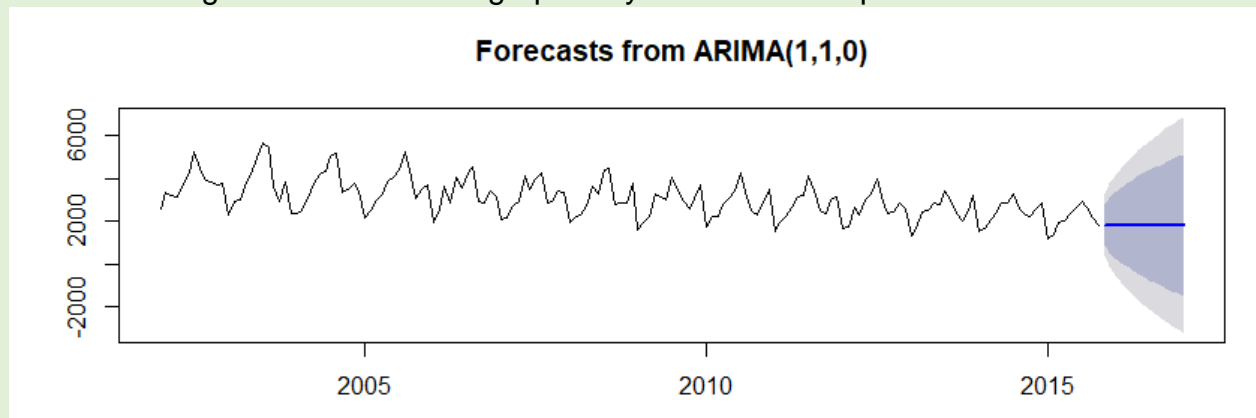
H_a = the residual values are not independent

Since the p values are very low the model selection is not the best and other models need to be evaluated

Forecast using the ARIMA model graphically and values for product A



Forecast using the ARIMA model graphically and values for product B



```
> demandB.arima.fit_FC
```

	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Nov 2015		1815.370	871.40154	2759.339	371.6947	3259.046
Dec 2015		1810.145	553.01152	3067.279	-112.4749	3732.766
Jan 2016		1810.775	296.63108	3324.918	-504.9082	4126.458
Feb 2016		1810.699	78.04386	3543.354	-839.1683	4460.566
Mar 2016		1810.708	-115.91881	3737.335	-1135.8135	4757.230
Apr 2016		1810.707	-292.06461	3913.479	-1405.2047	5026.619
May 2016		1810.707	-454.55455	4075.969	-1653.7116	5275.126
Jun 2016		1810.707	-606.14455	4227.559	-1885.5485	5506.963
Jul 2016		1810.707	-748.77203	4370.186	-2103.6784	5725.093
Aug 2016		1810.707	-883.86057	4505.275	-2310.2785	5931.693
Sep 2016		1810.707	-1012.49258	4633.907	-2507.0042	6128.418
Oct 2016		1810.707	-1135.51386	4756.928	-2695.1490	6316.563
Nov 2016		1810.707	-1253.60024	4875.014	-2875.7466	6497.161
Dec 2016		1810.707	-1367.30186	4988.716	-3049.6382	6671.052
Jan 2017		1810.707	-1477.07369	5098.488	-3217.5197	6838.934

Summary of MAPE

	MAPE		
Product	Exponential smoothing	Holt Winter model	Holt Winter HW function
A	0.1057	0.1733	0.08203897
B	0.1694	0.219	0.07989064

Referring to the table it can be seen that the Holt winter with hw function model has the least MAPE value making it an efficient model for both the products

Conclusion:

The typical trend that the products show are

For A: The demand is stable and with some intensity of seasonality. Typically this product does makes maximum sales during the year end, I will ensure that sufficient stocking of the product is done. In our models as well the product demand is in sync with the previous cycle and it will have its ups and downs. Stocking of the product is necessary during the peak business days. Since the product information is not provided, not much can be said but as a store's manager I will ensure that this product is stocked basis the seasonality

For B: It is very evident that the product is on a declining trend. There are spikes in sales during July and August of the year but the product sold is fairly less compared to the previous year. The forecast of the demand also in our model predicts that there will be a downward trend with respect to the sales within the stores.

The product could be an outdated or an outgoing or out of fashion so people tend to demand less but there is going to be some time that people will buy this product and thus the stocking of the product is done accordingly.

As a store's manager, my focus will be to ensure that I push the product to ultimate consumers unless it is out of vogue.