Business Forecasting

Midterm Exam: Mac Sales

Ronak Parikh

Business Forecasting Mid-Term Exam

Introduction

Apple makes great consumer technology products. Even with cheaper options available, their market share and profits keep increasing. We will look at quarterly sales data for Mac to forecast future sales. Data has been provided to you as well as the commands to import the dataset as a time series.

Import Data

Please do the following steps once the csv file is on your desktop.

- library(readr)
- apple_data <- read_csv("C:/Users/rrparikh/Desktop/apple_data.csv")

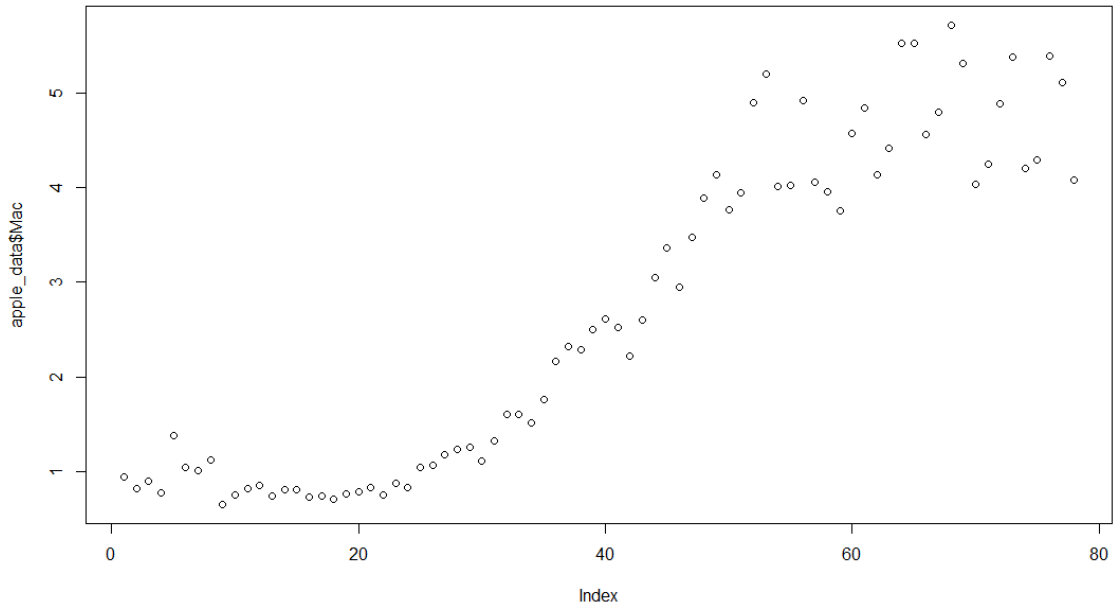
```
> library(readr)
Warning message:
package 'readr' was built under R version 3.4.4
> apple_data <- read_csv("C:/Users/asher/Desktop/Asher_MS/B
F/Midterm/apple_data.csv")
Parsed with column specification:
cols(
  Time = col_character(),
  Period = col_integer(),
  iPhone = col_double(),
  iPad = col_double(),
  iPod = col_double(),
  Mac = col_double()
)
```

- View(apple_data)

```
> View(apple_data)
```

- plot(mac_sales)

```
plot(apple_data$Mac)
```



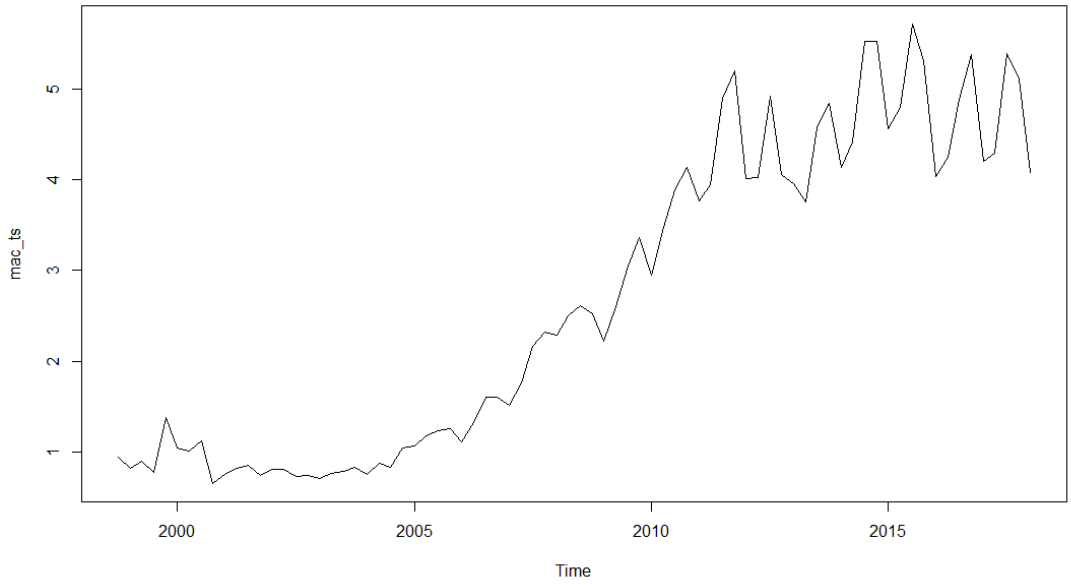
- mac_ts <- ts(mac_sales,start=c(1998,4),frequency = 4)

- plot(mac_ts)

```
mac_ts <- ts(apple_data$Mac,start=c(1998,4),frequency = 4)
plot(mac_ts)
```

Plot and Inference

- Show a time series plot.



- Please summaries your observations of the times series plot

    - We see a trend. It is overall upward rising trend.
    - There is sudden rise from 2009 to 2012
    - We have some cyclical seasonality too, with alternate rise and drop
    - Overall we see stability or slight drop after 2015

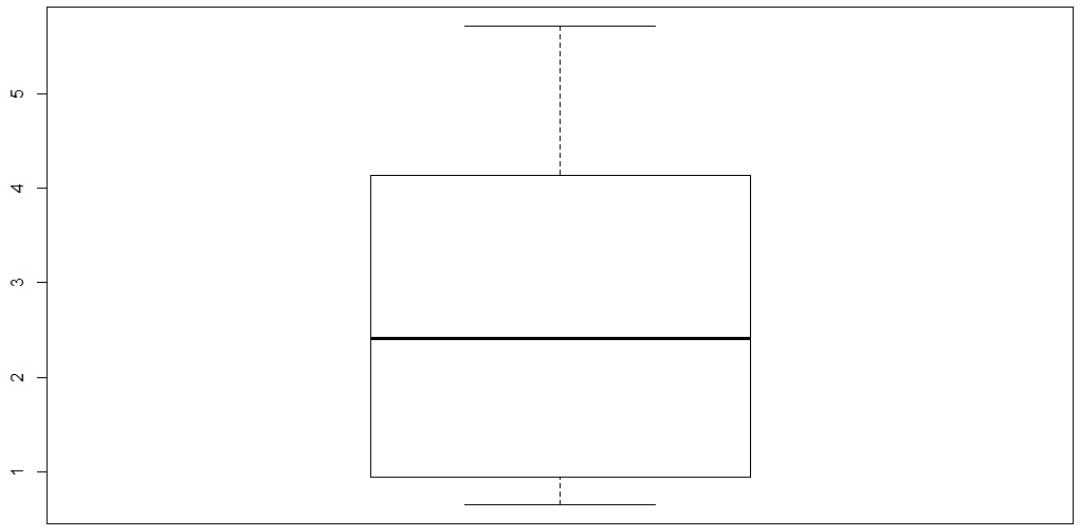
Central Tendency

- What are the min, max, mean, median, 1st and 3rd Quartile values of the times series?

```
> summary(mac_ts)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.659   0.962   2.408   2.677   4.136   5.710
> |
```

- Show the box plot.

```
> boxplot(mac_ts)
```



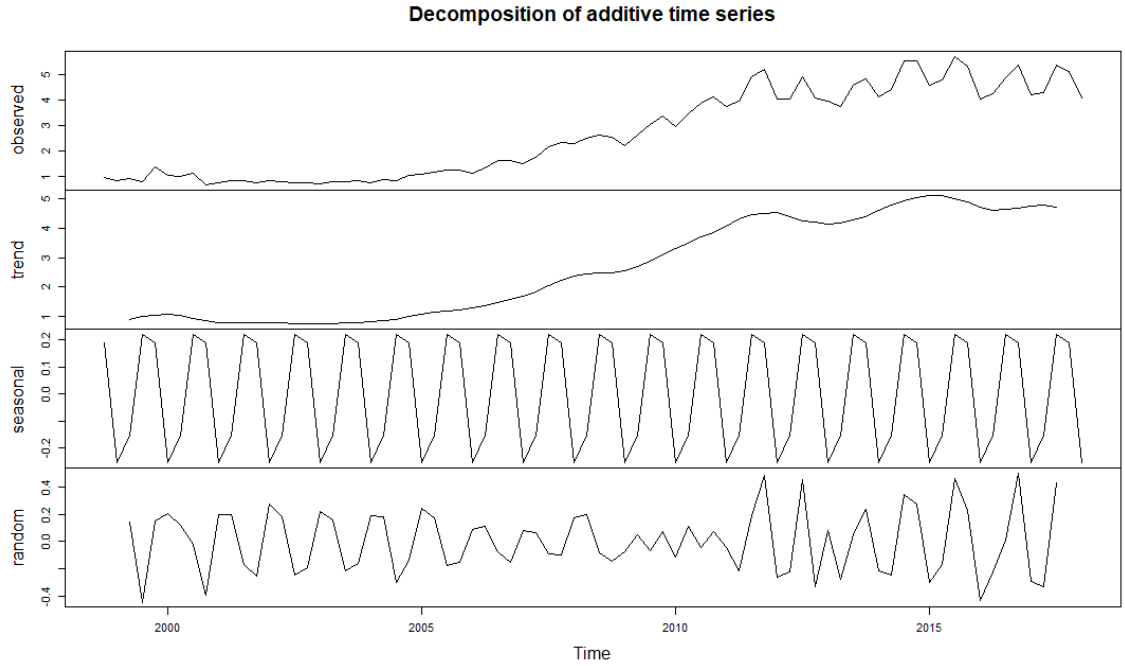
- Can you summarize your observation about the time series from the summary stats and box plot?

    The Median is about 2.408.
    The first quartile is the lowest line, 0.962
    The Mean is 2.677
    The third quartile is 4.136

Decomposition

- Plot the decomposition of the time series.

```
> decompo<-decompose(mac_ts)
>plot(decompo)
```



- Is the times series seasonal?
  Yes time series is seasonal as seen from the decomposition.

- Is the decomposition additive or multiplicative?
  It is additive

```
decompo$type
[1] "additive"
```

- If seasonal, what are the values of the seasonal monthly indices?

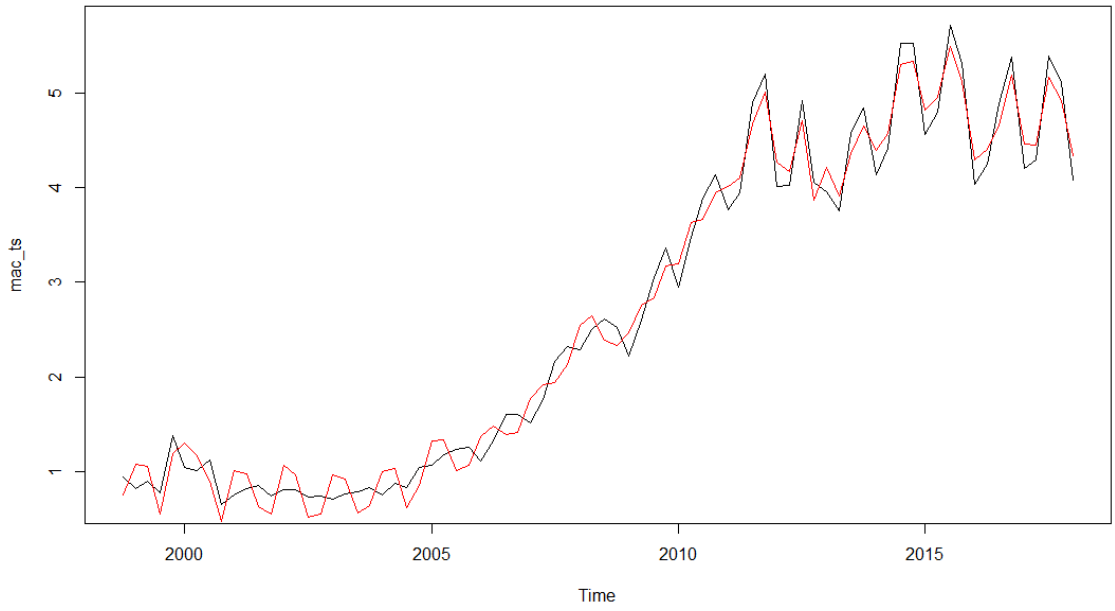
```
> decompo $ seasonal
           Qtr1       Qtr2       Qtr3       Qtr4
1998                                     0.1897854
1999 -0.2561660 -0.1537735  0.2201541  0.1897854
2000 -0.2561660 -0.1537735  0.2201541  0.1897854
2001 -0.2561660 -0.1537735  0.2201541  0.1897854
2002 -0.2561660 -0.1537735  0.2201541  0.1897854
2003 -0.2561660 -0.1537735  0.2201541  0.1897854
2004 -0.2561660 -0.1537735  0.2201541  0.1897854
2005 -0.2561660 -0.1537735  0.2201541  0.1897854
2006 -0.2561660 -0.1537735  0.2201541  0.1897854
2007 -0.2561660 -0.1537735  0.2201541  0.1897854
2008 -0.2561660 -0.1537735  0.2201541  0.1897854
2009 -0.2561660 -0.1537735  0.2201541  0.1897854
2010 -0.2561660 -0.1537735  0.2201541  0.1897854
2011 -0.2561660 -0.1537735  0.2201541  0.1897854
2012 -0.2561660 -0.1537735  0.2201541  0.1897854
2013 -0.2561660 -0.1537735  0.2201541  0.1897854
2014 -0.2561660 -0.1537735  0.2201541  0.1897854
2015 -0.2561660 -0.1537735  0.2201541  0.1897854
2016 -0.2561660 -0.1537735  0.2201541  0.1897854
2017 -0.2561660 -0.1537735  0.2201541  0.1897854
2018 -0.2561660
```

- For which month is the value of time series high and for which month is it low?
  It is low for first quarter and high for third quarter

- Can you think of the reason behind the value being high in those months and low in those months?
  Usually Apple launches new products in the third quarter, around September, so sales are high. It is also the festive time of the year.
  Sales are low in first two quarters as there are no new product launch and no festivities.

- Show the plot for time series adjusted for seasonality. Overlay this with the line for actual time series? Does seasonality have big fluctuations to the value of time series?

```
temp_sesadjust<-seasadj(decompo)
plot(mac_ts)
lines(temp_sesadjust,col='red')
```

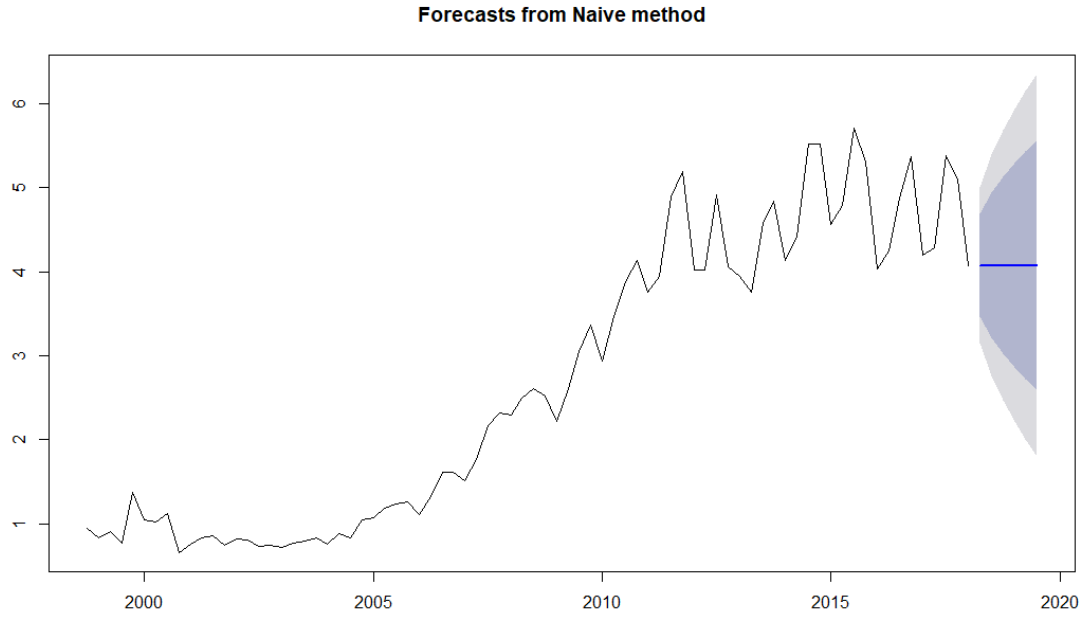


Big fluctuations are seen between 2000 and 2005 , but not after 2005.

Naïve Method

- Output

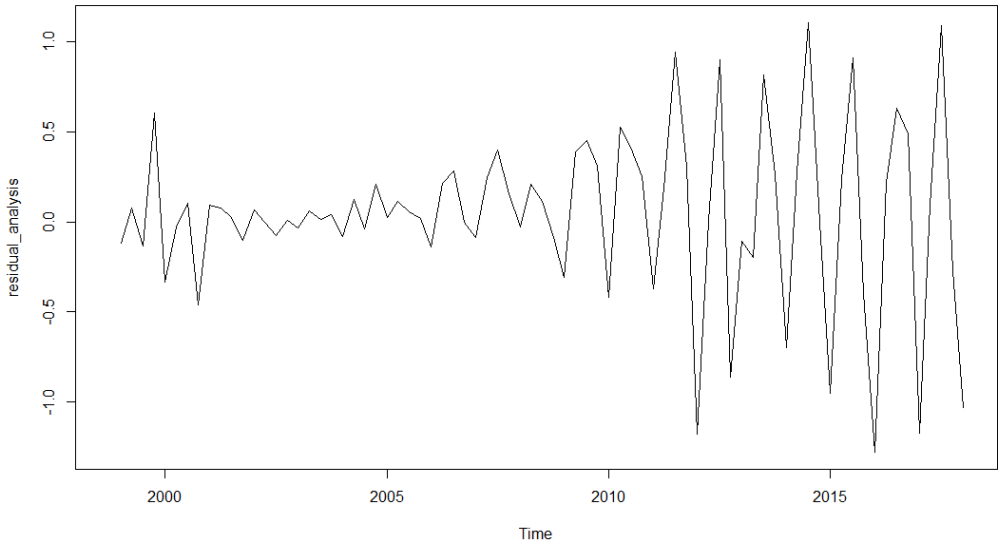
```
naive_forecast<-naive(mac_ts,6)
plot(naive_forecast)
```



- Perform Residual Analysis for this technique.

- Do a plot of residuals. What does the plot indicate?

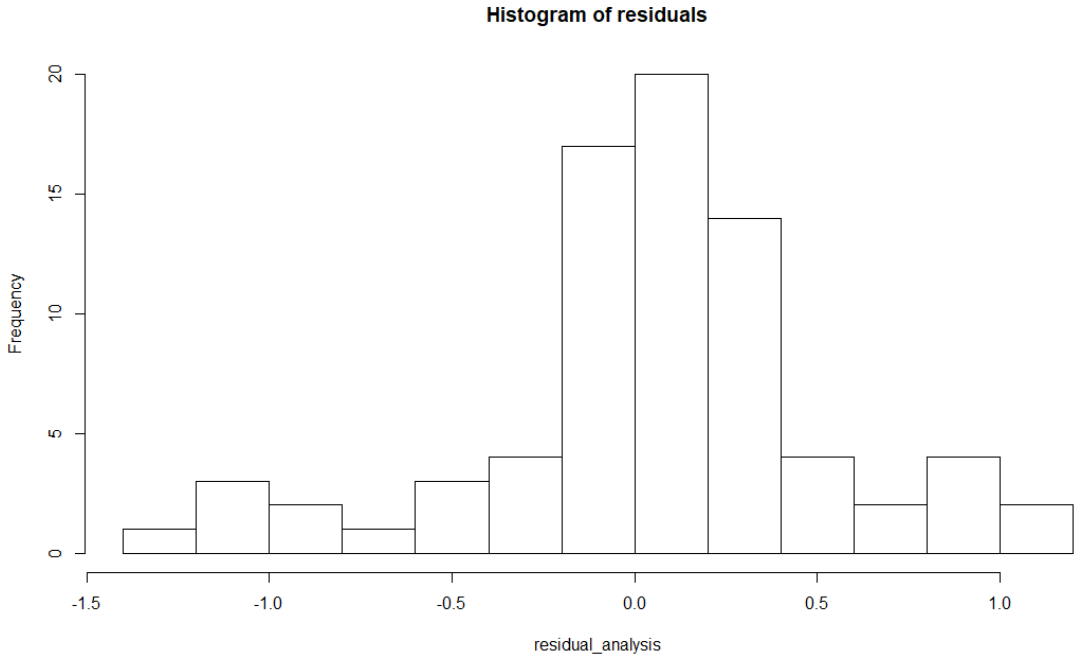
```
residual_analysis<-residuals(naive_forecast)
plot(residual_analysis)
```



The residual value is around zero between 2003 to 2005 and it highly fluctuates between 1 and -1 from 2012 onwards. Residuals should be close to zero which indicates highly significant values.

- Do a Histogram plot of residuals. What does the plot indicate?

```
histo<-hist(residual_analysis,breaks =10,main = "Histogram of residuals")
```

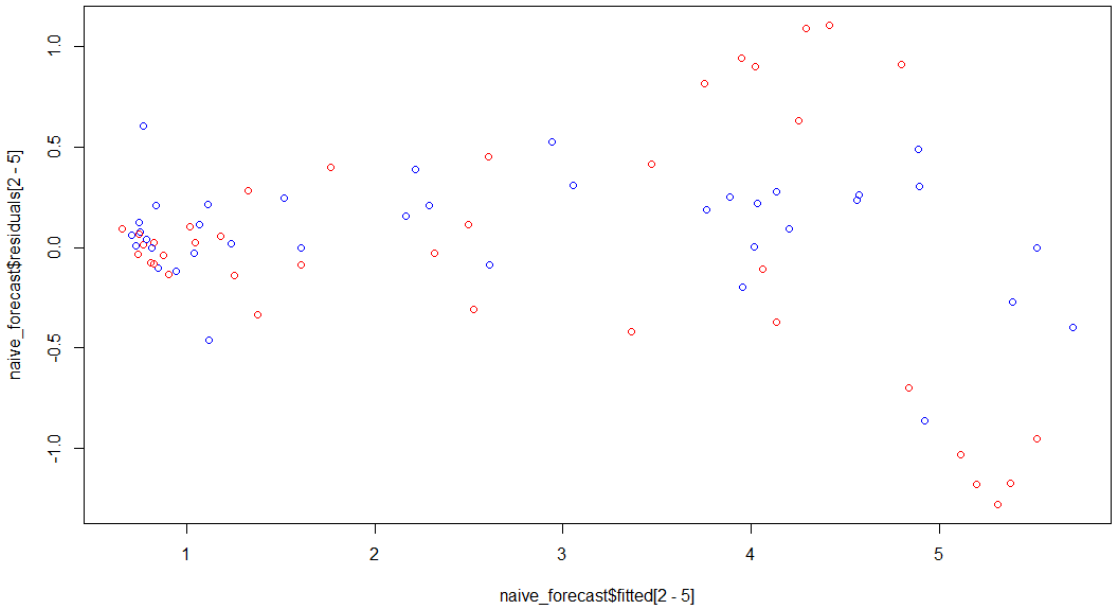


The residuals are left skewed.

- Do a plot of fitted values vs. residuals. What does the plot indicate?

```
plot(naive_forecast$fitted[2-5],naive_forecast$residuals[2-
5],col=c("red","blue"))
```

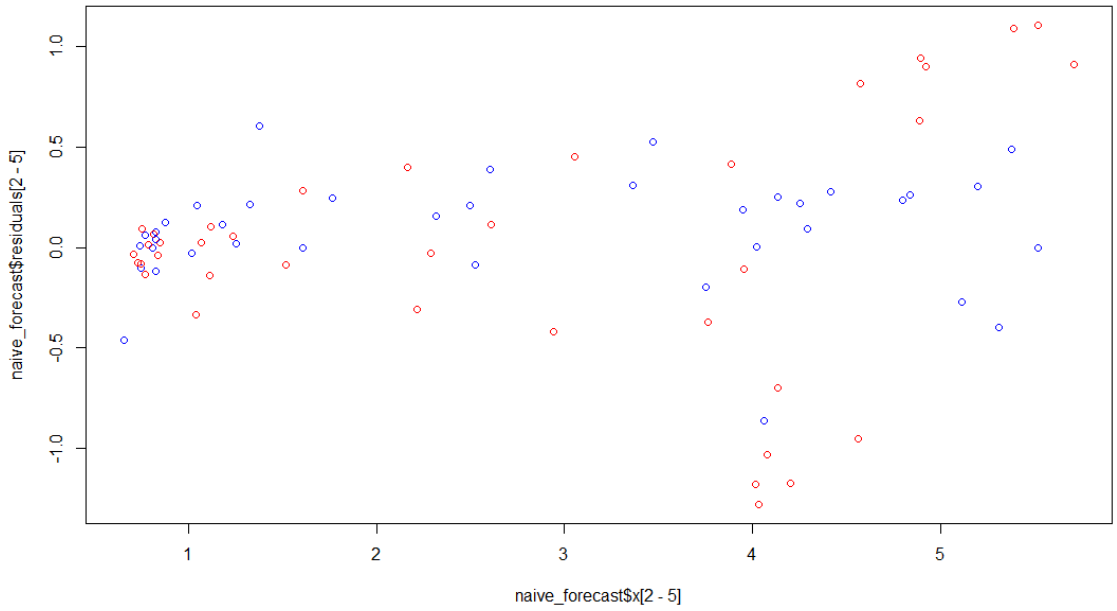
Upto 2005, the difference between values is less. Accuracy is high. It increases after 2005 and hence accuracy is less.



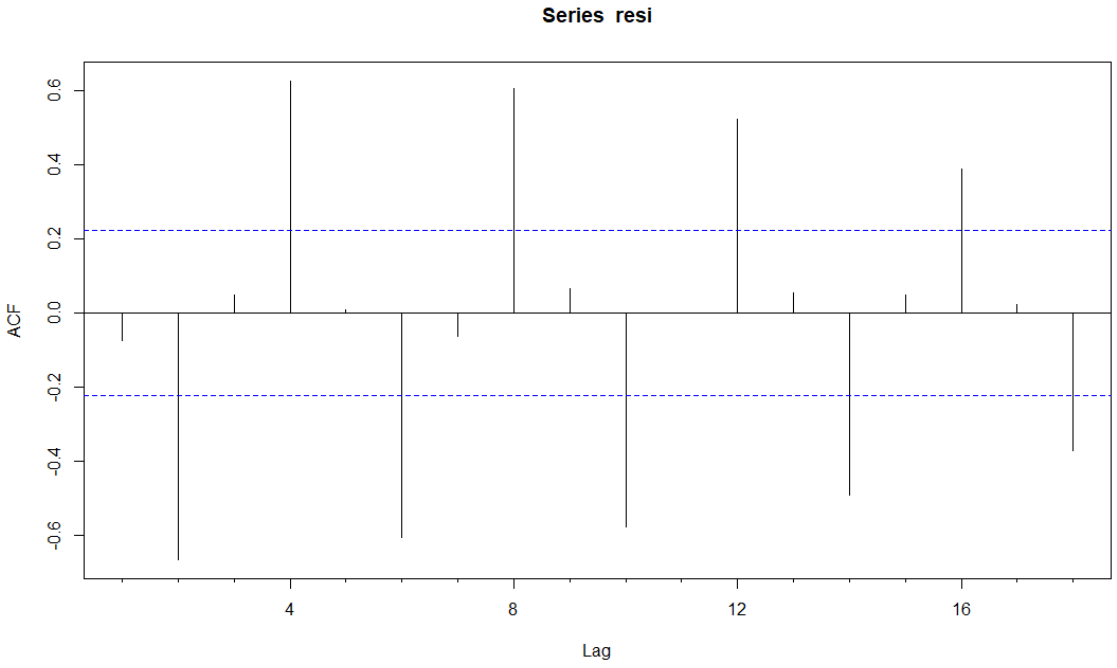
- Do a plot of actual values vs. residuals. What does the plot indicate?

Upto 2005, the difference between values is less. Accuracy is high. It increases after 2005 and hence accuracy is less.

```
plot(naive_forecast$x[2-5],naive_forecast$residuals[2-5],col=c("red","blue"))
```



- Do an ACF plot of the residuals? What does this plot indicate?
  >Acf(resi)



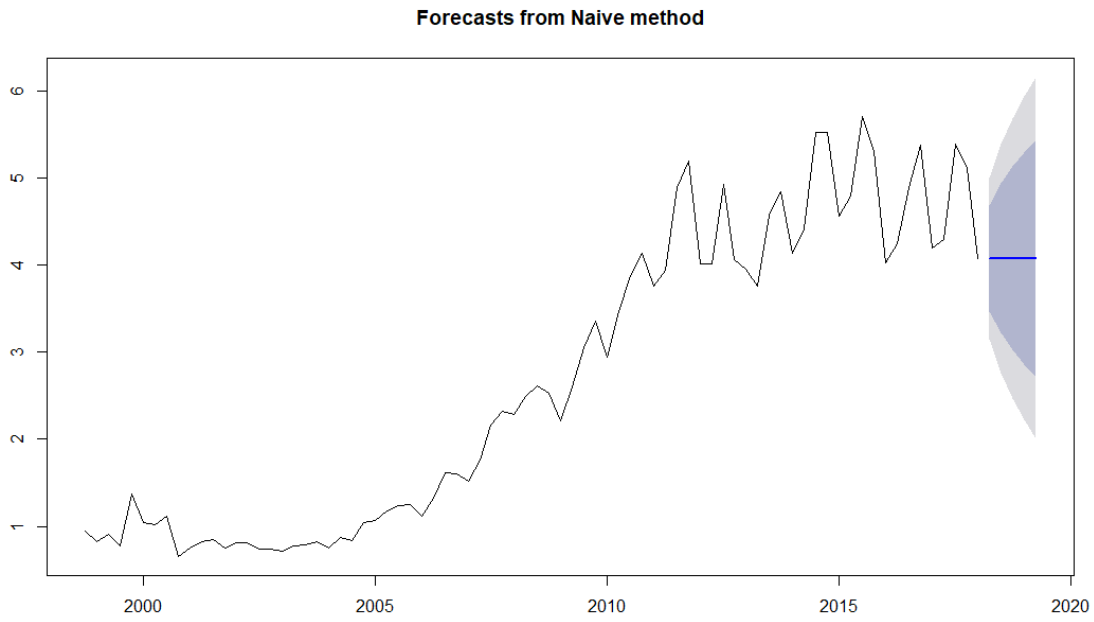
We can see that values are significant for alternate 2 quarters.

- Print the 5 measures of accuracy for this forecasting technique

```
> accuracy(naive_forecast)
                     ME      RMSE       MAE       MPE     MAPE      MASE        ACF1
Training set 0.0407013 0.4738365 0.3272727 0.6541332 12.01044 0.8748711 -0.07297452
> |
```

- Forecast
  - Time series value for next year. Show table and plot

```
> naive_fore<-forecast(naive_forecast, h=5)
> naive_fore
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q2          4.078 3.470754 4.685246 3.149298 5.006702
2018 Q3          4.078 3.219225 4.936775 2.764616 5.391384
2018 Q4          4.078 3.026219 5.129781 2.469440 5.686560
2019 Q1          4.078 2.863508 5.292492 2.220595 5.935405
2019 Q2          4.078 2.720157 5.435843 2.001358 6.154642
> plot(naive_fore)
```

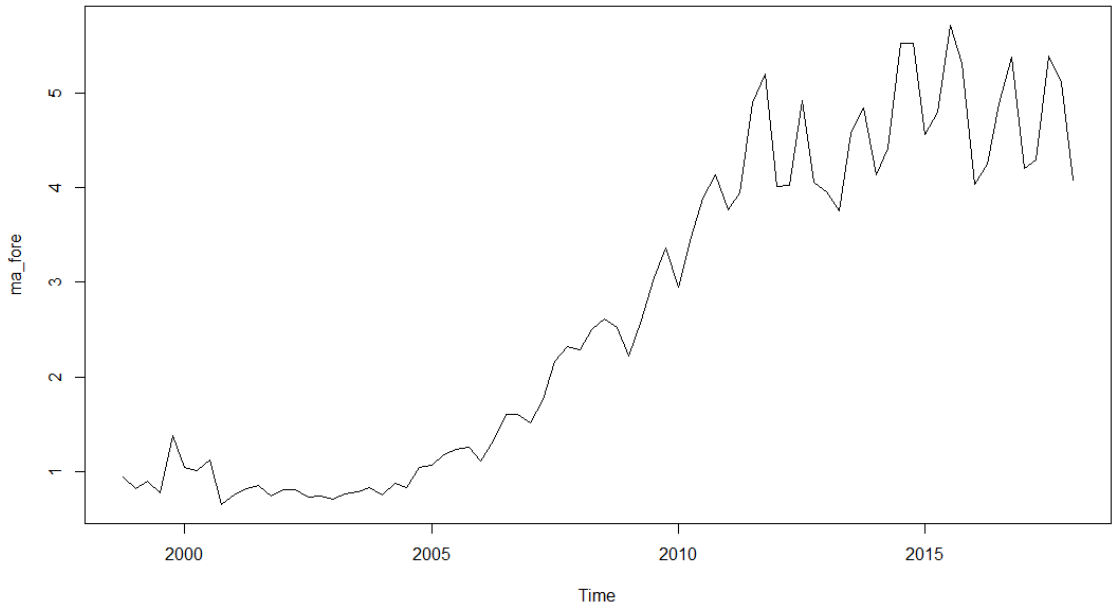


- Summarize this forecasting technique
  - How good is the accuracy?
    The RMSE is 0.47, which is low. Hence there is no much difference in actual and predicted values. Accuracy is high.
  - What does it predict the value of time series will be in one year?
    The predicted value is 4.078
  - Other observation
    For 2018 Q2, 95% confidence interval lies between 3,14 and 5

Simple Moving Averages

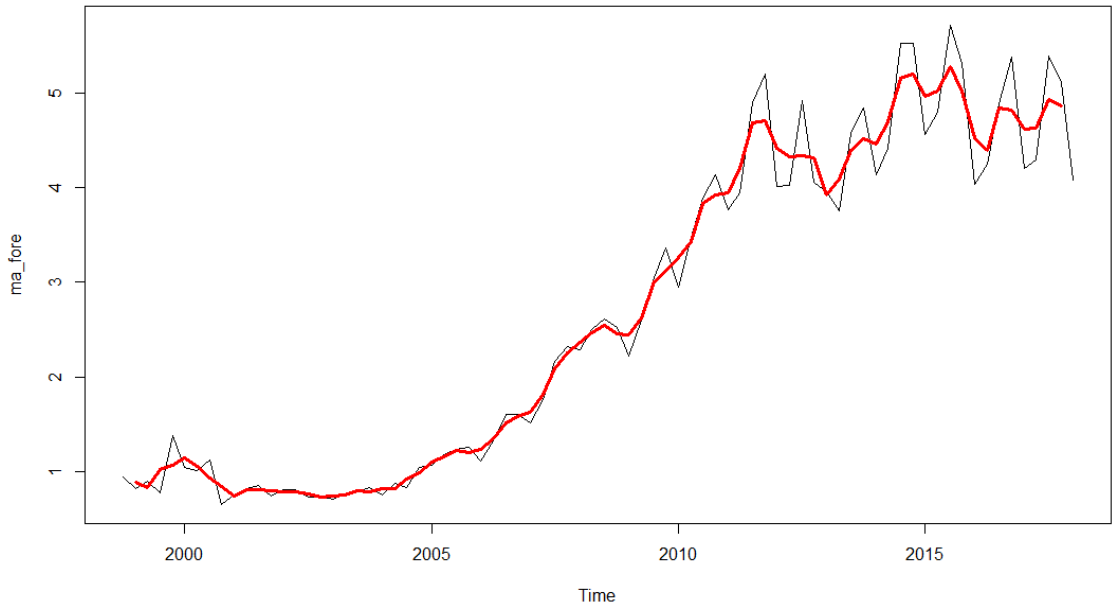
- Plot the graph for time series.

```
> ma_fore<-ma(mac_ts, order=1)
> plot(ma_fore)
```



- Show the Simple Moving average of order 3 on the plot above in Red

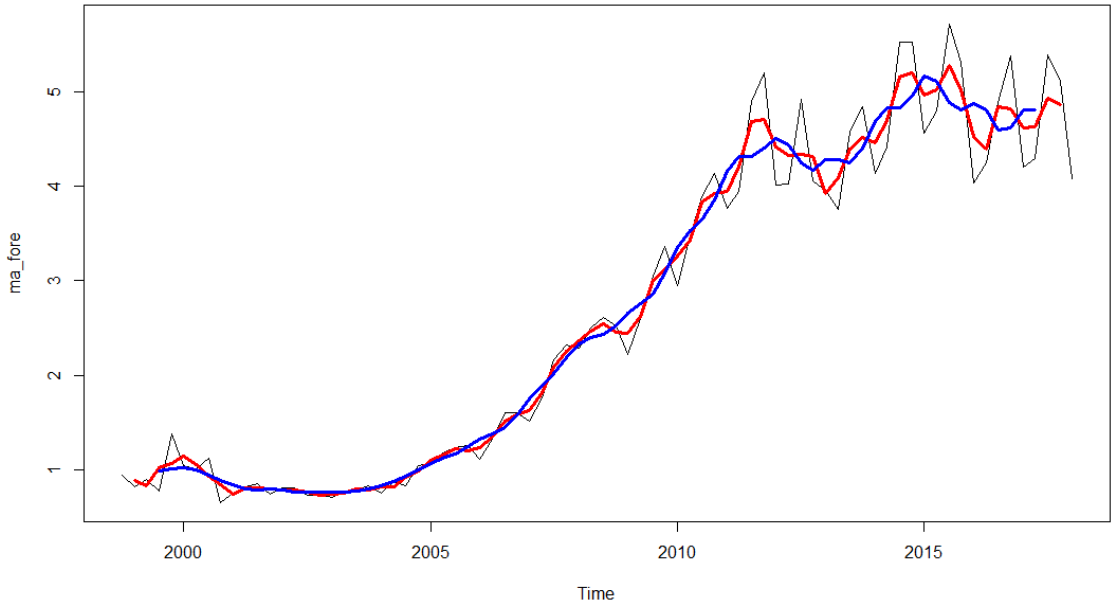
```
> ma3_fore<-ma(mac_ts, order=3)
> lines(ma3_fore,col="red",lwd=3)
```



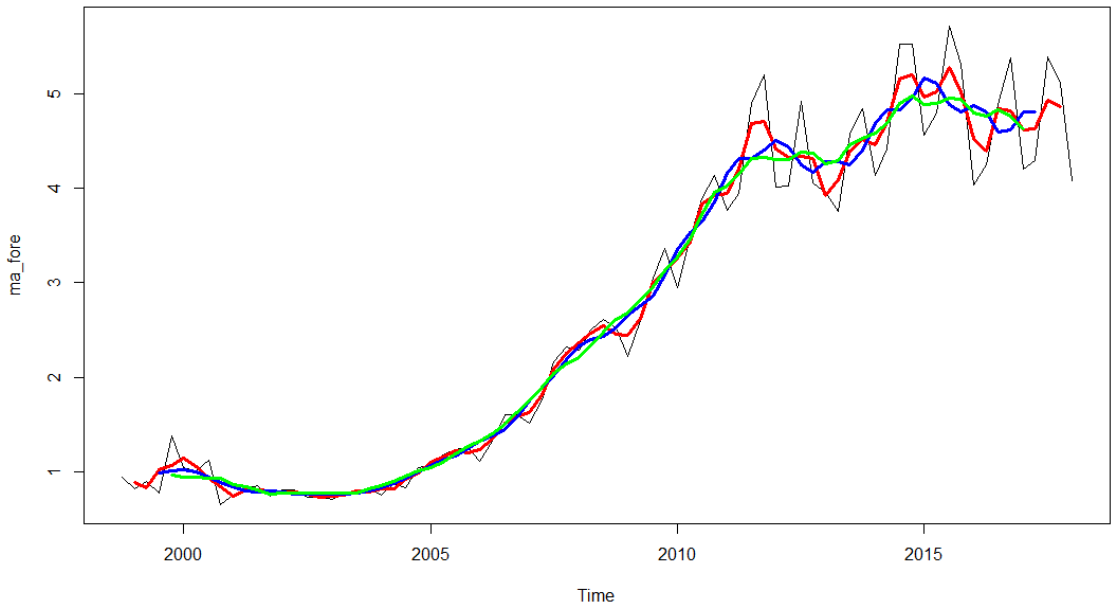
- Show the Simple Moving average of order 6 on the plot above in Blue

```
ma6_fore<-ma(mac_ts, order=6)
> lines(ma6_fore,col="blue",lwd=3)


>
```



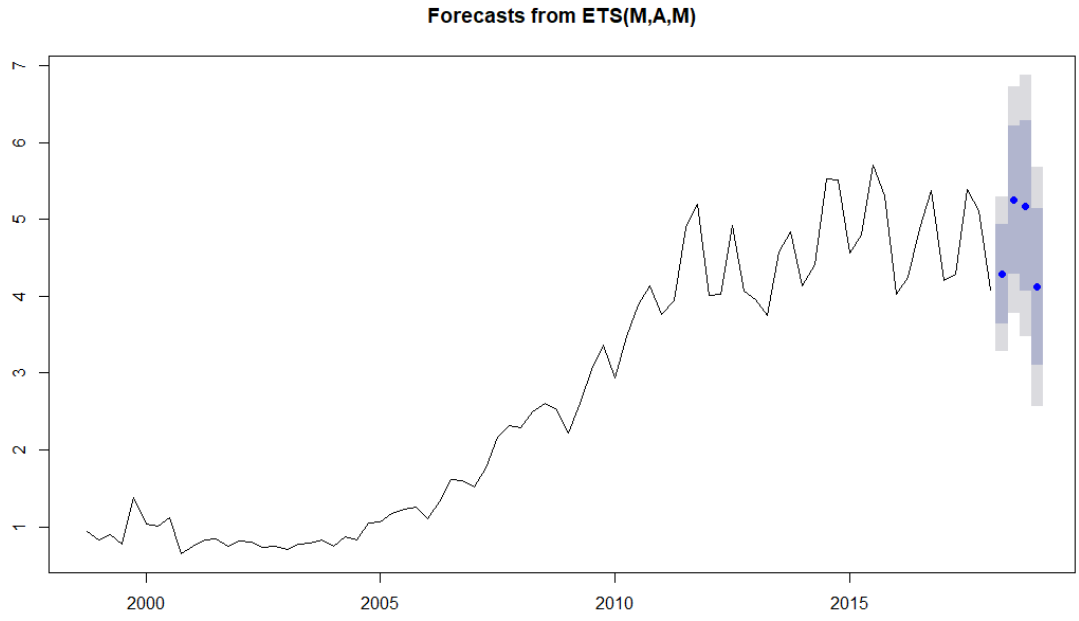
- Show the Simple Moving average of order 9 on the plot above in Green



```
> ma9_fore<-ma(mac_ts, order=9)
> lines(ma9_fore,col="green",lwd=3)
```

- (Bonus) show the forecast of next 12 months using one of the simple average order that you feel works best for time series

```
> ets_forecast <- ets(mac_ts)
> forecast.ets(ets_forecast, h=4)
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q2       4.287120 3.633379 4.940861 3.287310 5.286930
2018 Q3       5.245058 4.281225 6.208890 3.771002 6.719113
2018 Q4       5.172728 4.061757 6.283699 3.473644 6.871811
2019 Q1       4.122237 3.108422 5.136053 2.571740 5.672734
> fore_ets<-forecast.ets(ets_forecast, h=4)
> plot(fore_ets)
>
```



- What are your observations of the plot as the moving average order goes up?
  Error of prediction increases with it.

Simple Smoothing

- Perform a simple smoothing forecast for next 12 months for the time series.

```
> ses(mac_ts, h=4)
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q2       4.545093 3.963248 5.126938 3.655238 5.434948
2018 Q3       4.545093 3.893961 5.196225 3.549272 5.540913
2018 Q4       4.545093 3.831368 5.258817 3.453545 5.636640
2019 Q1       4.545093 3.773839 5.316346 3.365562 5.724623
>
```

  - What is the value of alpha? What does that value signify?

```
> summary(ses(mac_ts, h=4))

Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(y = mac_ts, h = 4)

  Smoothing parameters:
    alpha = 0.5023

  Initial states:
    l = 0.9143

  sigma:  0.454

     AIC     AICc      BIC
220.6160 220.9403 227.6861

Error measures:
                     ME      RMSE      MAE      MPE   MAPE      MASE        ACF1
Training set 0.09266392 0.4481574 0.326483 2.404932 11.762 0.8727599 0.08573371

Forecasts:
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q2       4.545093 3.963248 5.126938 3.655238 5.434948
2018 Q3       4.545093 3.893961 5.196225 3.549272 5.540913
2018 Q4       4.545093 3.831368 5.258817 3.453545 5.636640
2019 Q1       4.545093 3.773839 5.316346 3.365562 5.724623
>
```

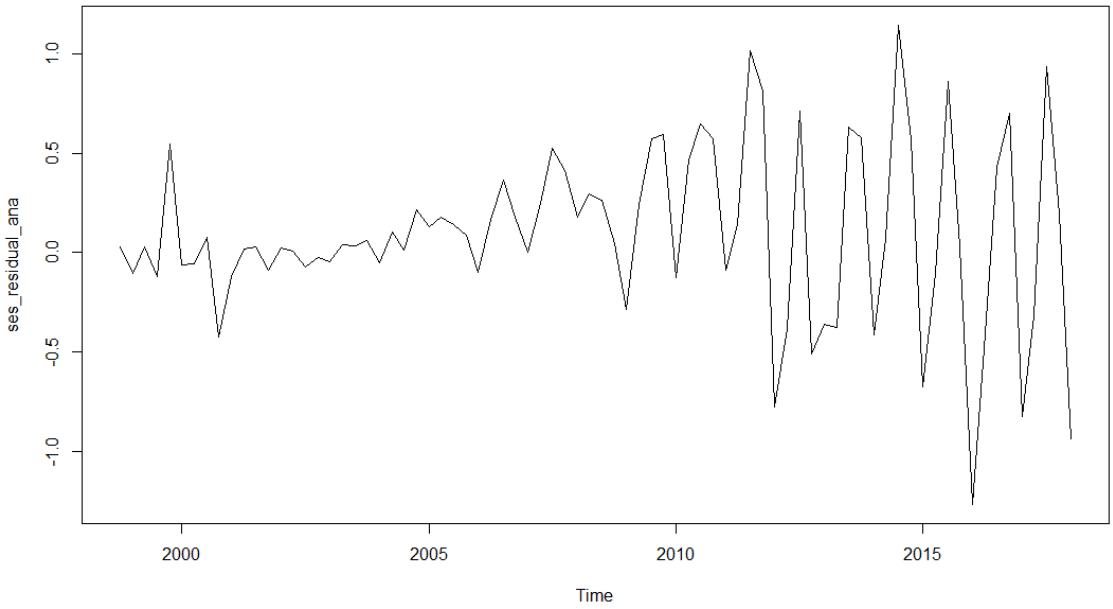
    Value of alpha is o.5023. The value of alpha signifies the optimal smoothing parameter for the model to get minimum error

  - What is the value of initial state?
    Value of initial state is 0.9143

  - What is the value of sigma? What does the sigma signify?
    Value of sigma is 0.454. It signifies that the residuals have more variation around the residual mean.

- Perform Residual Analysis for this technique.
  - Do a plot of residuals. What does the plot indicate?

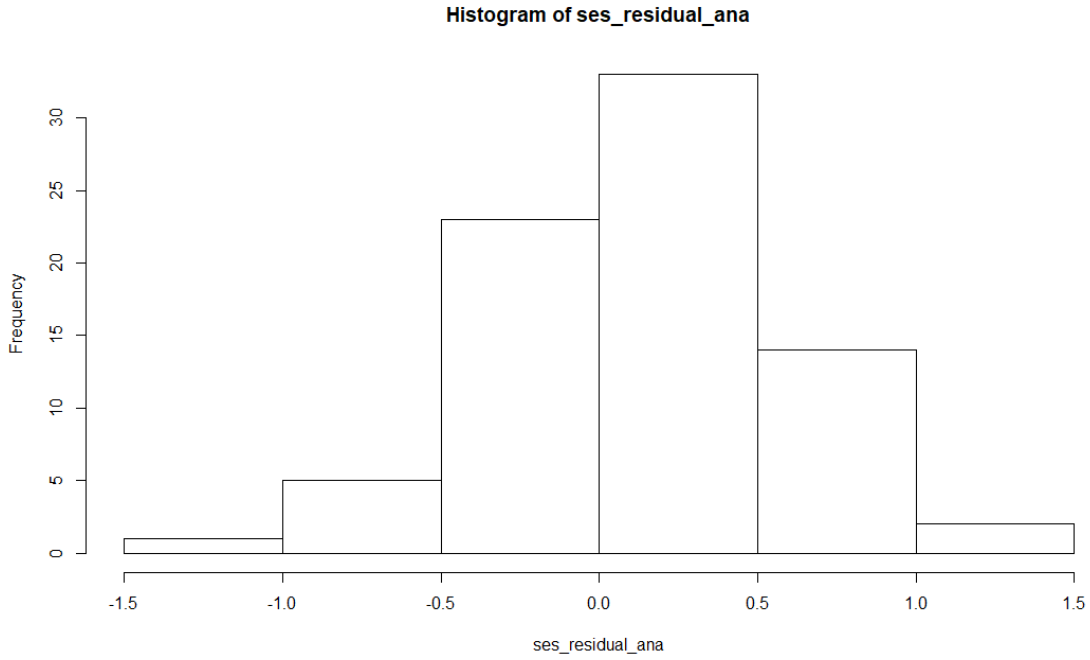
```
> ses_fore<-ses(mac_ts, h=4)
> ses_residual_ana<-residuals(ses_fore)
> plot(ses_residual_ana)
>
```



The variation in residuals is comparatively high after 2008.

- Do a Histogram plot of residuals. What does the plot indicate?

  hist(ses_residual_ana)



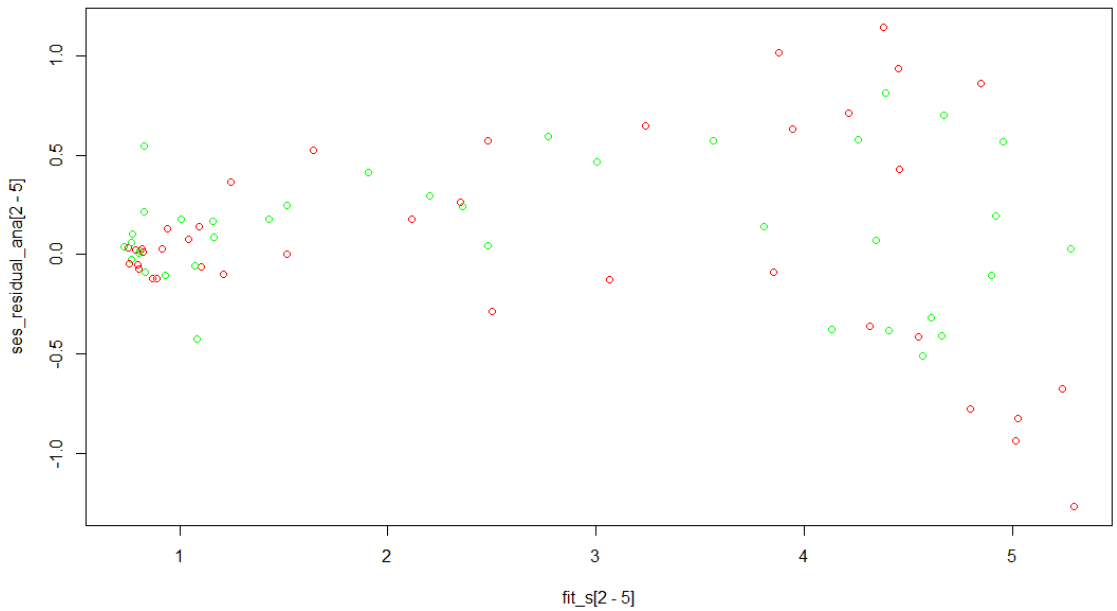
The residual data is slightly left skewed

- Do a plot of fitted values vs. residuals. What does the plot indicate?

Upto 2005, the difference between values is less. Accuracy is high. It increases after 2005 and hence accuracy is less.

```
fit_s<-ses_fore$fitted
> plot(fit_s[2-5],ses_residual_ana[2-5], col=c("Red","Green"))


>
```

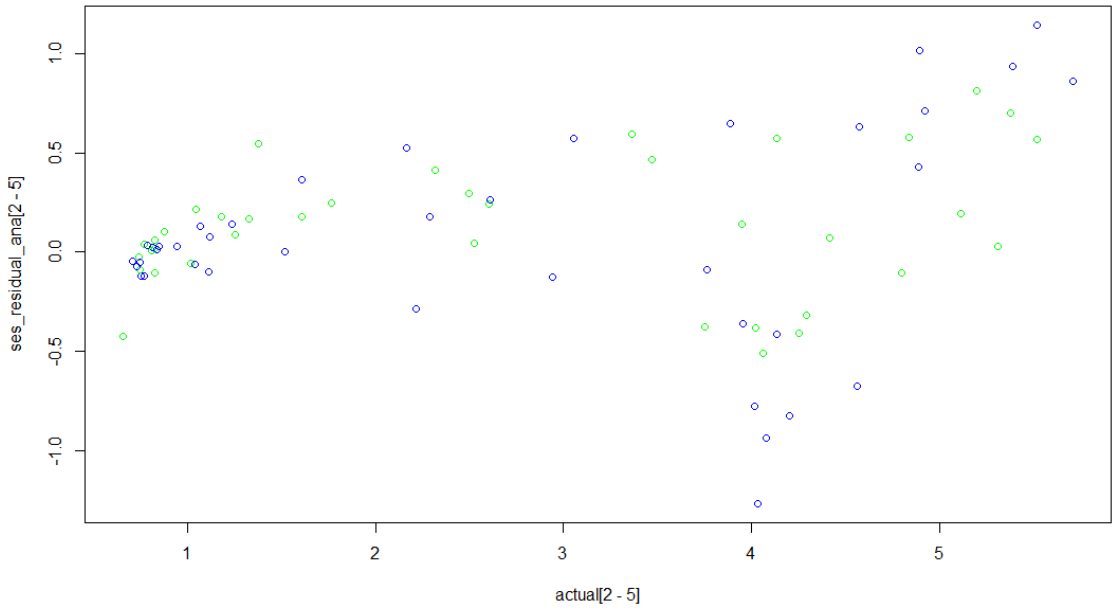


- Do a plot of actual values vs. residuals. What does the plot indicate?

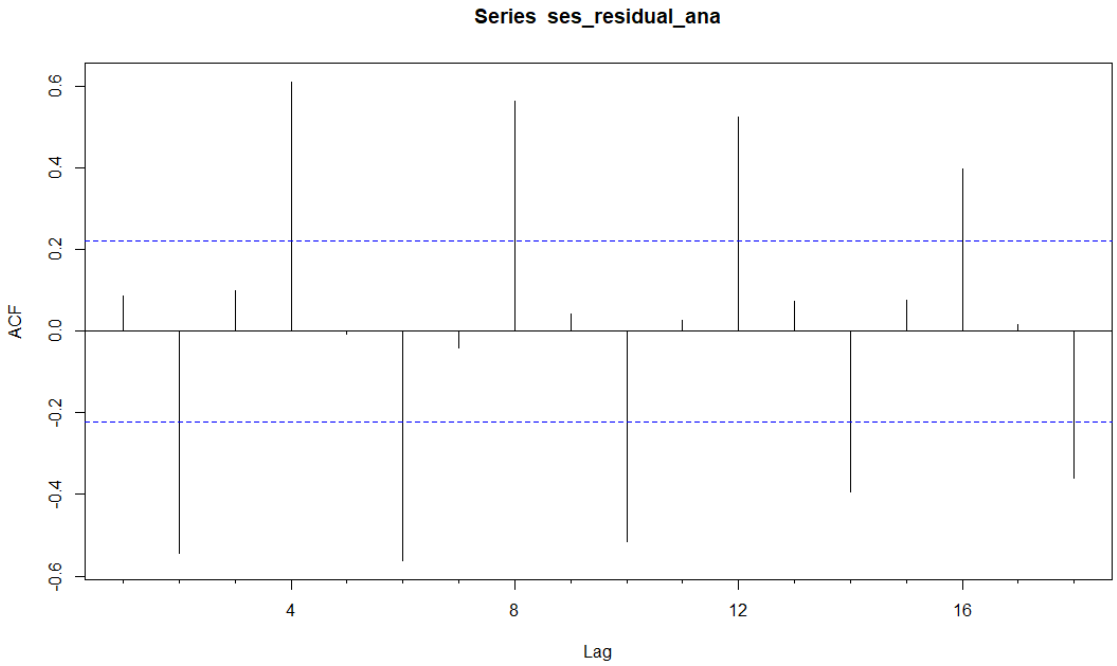
Upto 2005, the difference between values is less. Accuracy is high. It increases after 2005 and hence accuracy is less.

```
actual<-ses_fore$x

plot(actual[2-5],ses_residual_ana[2-5],col=c("Blue","Green"))
```



- Do an ACF plot of the residuals? What does this plot indicate?
  >Acf(ses_residual_ana)
  We can see that values are significant for alternate 2 quarters.

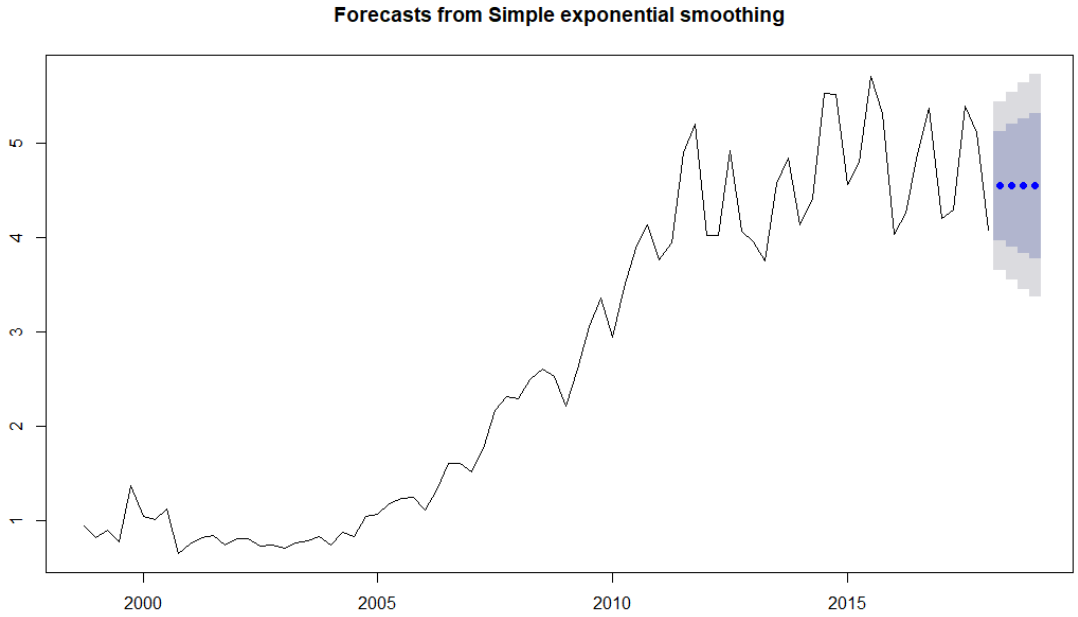


- Print the 5 measures of accuracy for this forecasting technique

```
> accuracy(ses_fore)
                     ME      RMSE       MAE       MPE    MAPE      MASE       ACF1
Training set 0.09266392 0.4481574 0.326483 2.404932 11.762 0.8727599 0.08573371
> |
```

- Forecast
  - Time series value for next year. Show table and plot

```
> forecast(ses_fore,h=4)
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q2       4.545093 3.963248 5.126938 3.655238 5.434948
2018 Q3       4.545093 3.893961 5.196225 3.549272 5.540913
2018 Q4       4.545093 3.831368 5.258817 3.453545 5.636640
2019 Q1       4.545093 3.773839 5.316346 3.365562 5.724623
> |
```



- Summarize this forecasting technique

```
> summary(ses_fore)

Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(y = mac_ts, h = 4)

  Smoothing parameters:
    alpha = 0.5023

  Initial states:
    l = 0.9143

  sigma:  0.454

     AIC     AICc      BIC
220.6160 220.9403 227.6861

Error measures:
                    ME      RMSE       MAE      MPE   MAPE      MASE        ACF1
Training set 0.09266392 0.4481574 0.326483 2.404932 11.762 0.8727599 0.08573371

Forecasts:
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q2       4.545093 3.963248 5.126938 3.655238 5.434948
2018 Q3       4.545093 3.893961 5.196225 3.549272 5.540913
2018 Q4       4.545093 3.831368 5.258817 3.453545 5.636640
2019 Q1       4.545093 3.773839 5.316346 3.365562 5.724623
```

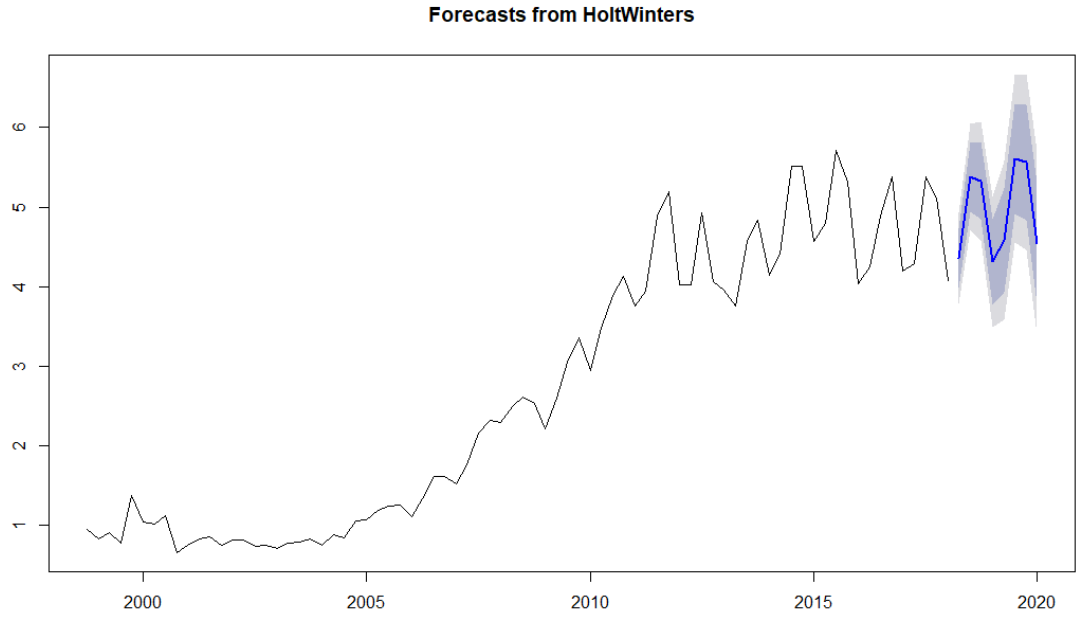
  - How good is the accuracy?
    Since MSE is close to zero, accuracy is high.
  - What does it predict the value of time series will be in one year?
    The predicted value in one year is 4.54

- Other observation
  We do not see a seasonal behavior in the predictions

Holt-Winters

- Perform Holt-Winters forecast for next 12 months for the time series.

```
> plot(forecast(ses_fore,h=4))
> holt<-HoltWinters(mac_ts)
> holt_forecast<-forecast(holt)
> plot(holt_forecast)
>
```




```
> summary(holt_forecast)

Forecast method: HoltWinters

Model Information:
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = mac_ts)

Smoothing parameters:
 alpha: 0.591872
 beta : 0
 gamma: 0.9719223

Coefficients:
         [,1]
a   4.7632368
b   0.0579750
s1 -0.4684586
s2  0.4965668
s3  0.3928622
s4 -0.6848847

Error measures:
                      ME      RMSE       MAE       MPE     MAPE      MASE        ACF1
Training set -0.0089909 0.2905301 0.2094488 -2.914616 9.993218 0.5599021 -0.01148062
```

```
Forecasts:
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q2       4.352753 3.978062 4.727445 3.779712 4.925794
2018 Q3       5.375754 4.940351 5.811156 4.709863 6.041644
2018 Q4       5.330024 4.841397 5.818652 4.582733 6.077315
2019 Q1       4.310252 3.773653 4.846851 3.489595 5.130909
2019 Q2       4.584653 3.932631 5.236675 3.587471 5.581835
2019 Q3       5.607654 4.918949 6.296359 4.554370 6.660937
2019 Q4       5.561924 4.838394 6.285454 4.455380 6.668468
2020 Q1       4.542152 3.785397 5.298907 3.384796 5.699509
```

- What is the value of alpha? What does that value signify?
  Value of alpha is 0.59 it signifies that the predictions are stable and random variations are smoothed.
- What is the value of beta? What does that value signify?
  Value of beta is 0. it signifies that trend completely depends on the previous period value.
- What is the value of gamma? What does that value signify?
  Value of Gamma is 0.971. it signifies that the seasonality repeats according to cycles at regular time intervals.
- What is the value of initial states for the level, trend and seasonality? What do these values signify?
  From the below
  a is Level,
  b is Trend ,
  s1 to s4 are the initial seasonality values for each quarter

```
> holt$coefficients
         a          b         s1         s2         s3         s4
4.7632368  0.0579750 -0.4684586  0.4965668  0.3928622 -0.6848847
```

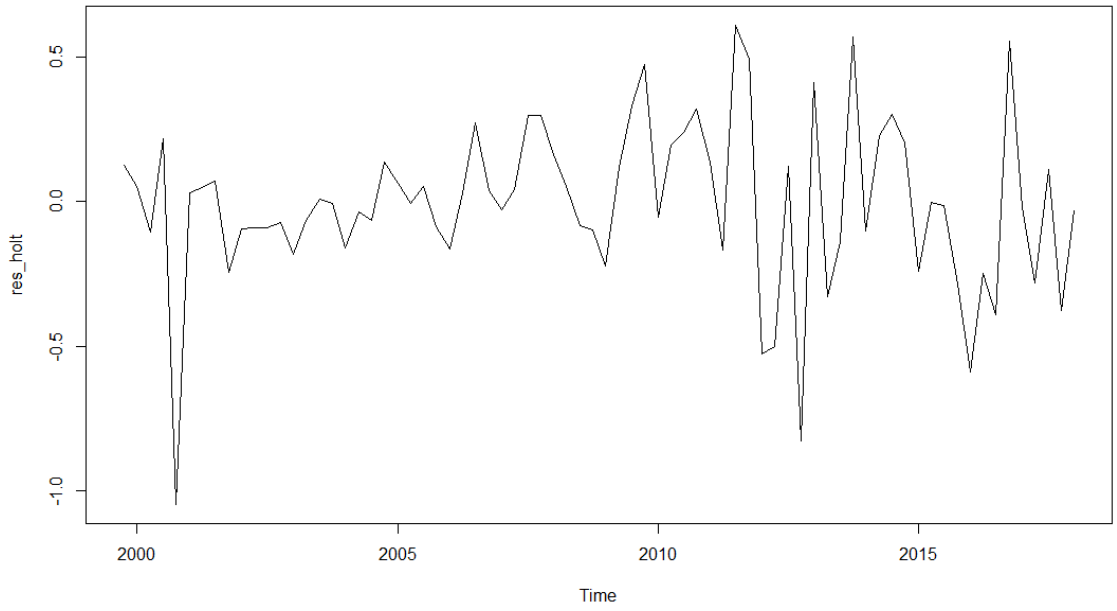
- What is the value of sigma? What does the sigma signify?

```
> sd(complete.cases(holt_forecast$residuals))
[1] 0.2220001
```

  The value of sigma signifies the value of standard deviation.

- Perform Residual Analysis for this technique.
  - Do a plot of residuals. What does the plot indicate?

```
> res_holt<-residuals(holt)
> plot(res_holt)
```

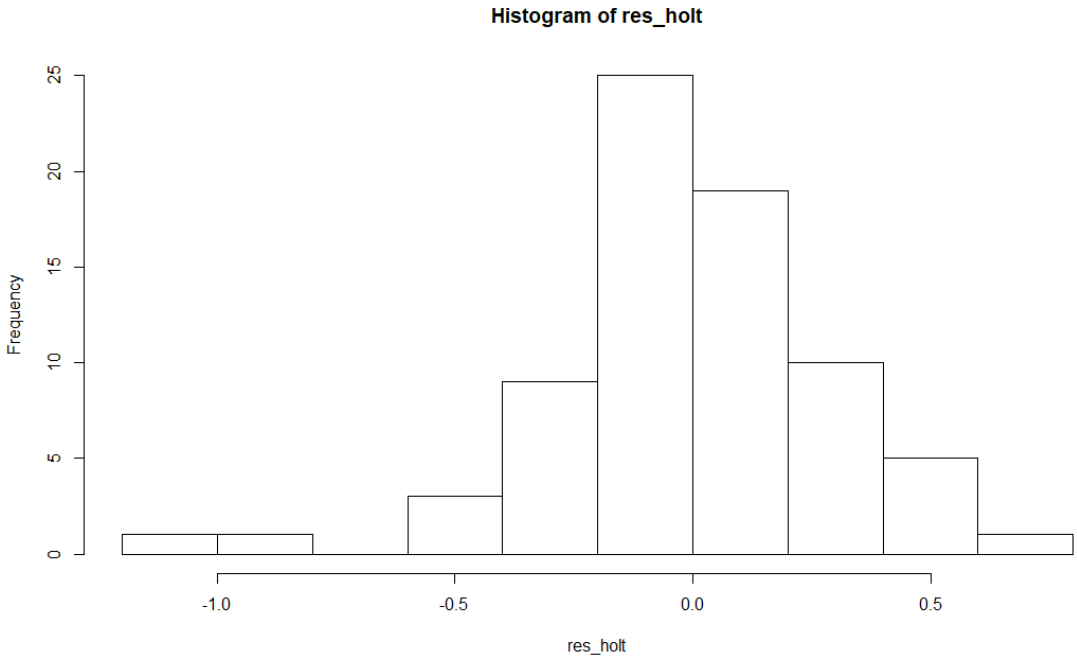


The variation in residuals is high.

- Do a Histogram plot of residuals. What does the plot indicate?

>hist(res_holt)

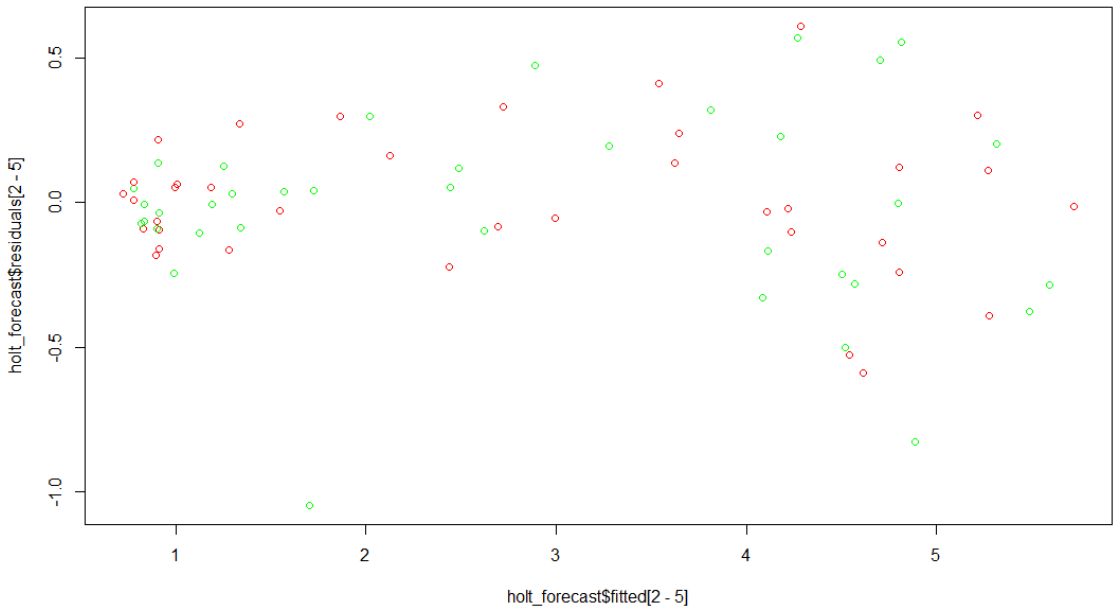
We can see that data is left skewed



- Do a plot of fitted values vs. residuals. What does the plot indicate?

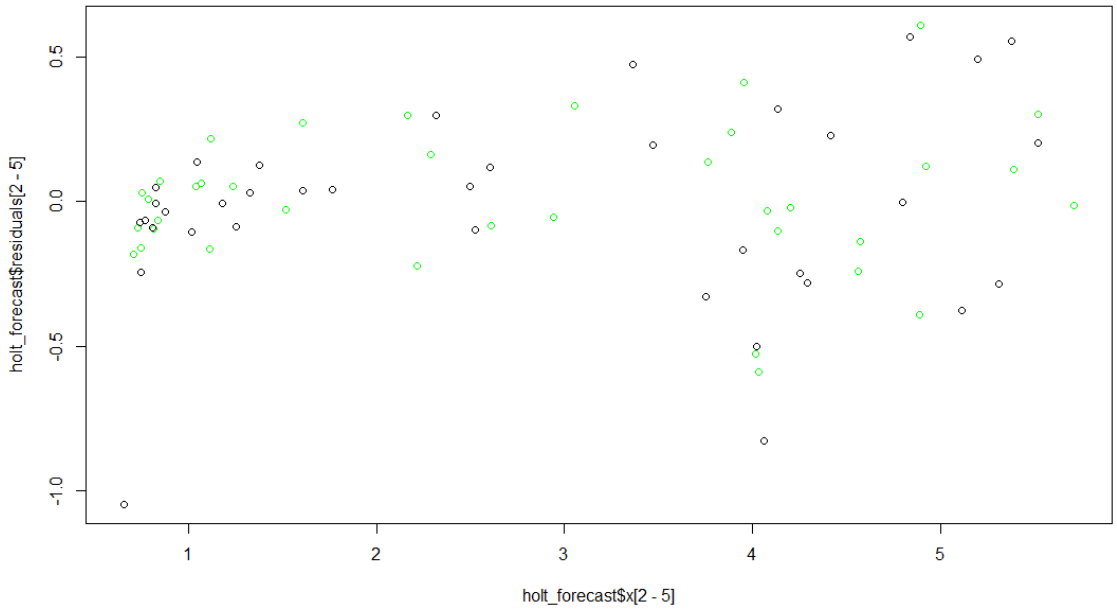
Upto 2005, the difference between values is less. Accuracy is high. It increases after 2005 and hence accuracy is less.

plot(holt_forecast$fitted[2-5],holt_forecast$residuals[2-5],col=c("red","green"))

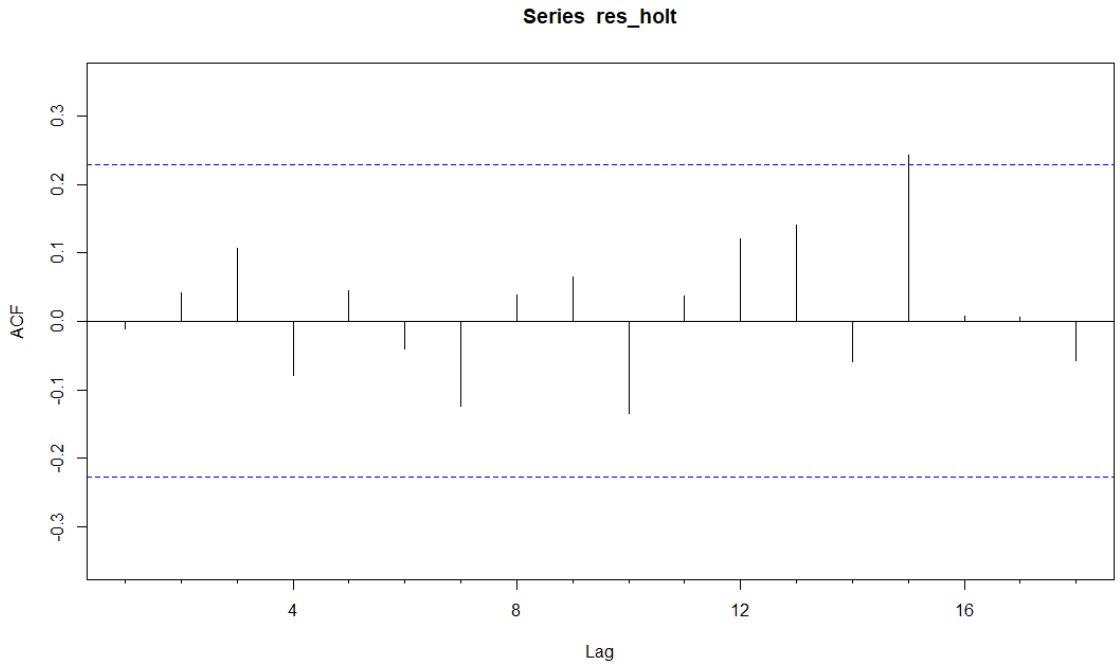


- Do a plot of actual values vs. residuals. What does the plot indicate?
  plot(holt_forecast$x[2-5],holt_forecast$residuals[2-5],col=c("Green","Black"))

  Upto 2005, the difference between values is less. Accuracy is high. It increases after 2005 and hence accuracy is less.



- Do an ACF plot of the residuals? What does this plot indicate?
  Acf(res_holt)
  We can see that values are not very significant



- Print the 5 measures of accuracy for this forecasting technique

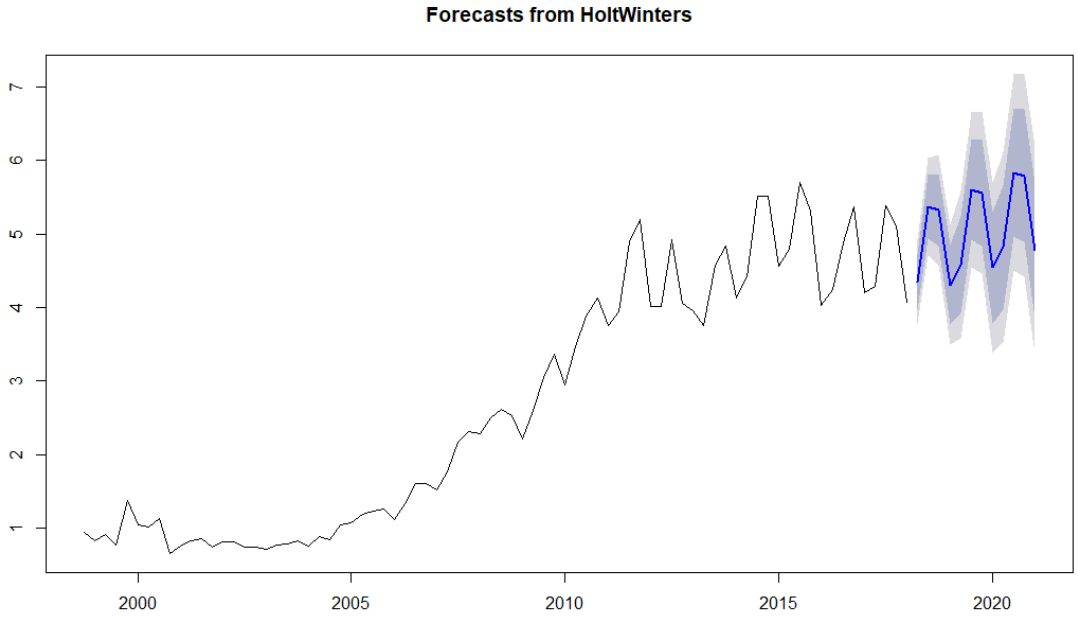
```
> accuracy(holt_forecast)
                     ME      RMSE       MAE       MPE     MAPE      MASE         ACF1
Training set -0.0089909 0.2905301 0.2094488 -2.914616 9.993218 0.5599021 -0.01148062


>
```

- Forecast
  - Time series value for next year. Show table and plot

```
> forecast(holt, h=12)
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q2       4.352753 3.978062 4.727445 3.779712 4.925794
2018 Q3       5.375754 4.940351 5.811156 4.709863 6.041644
2018 Q4       5.330024 4.841397 5.818652 4.582733 6.077315
2019 Q1       4.310252 3.773653 4.846851 3.489595 5.130909
2019 Q2       4.584653 3.932631 5.236675 3.587471 5.581835
2019 Q3       5.607654 4.918949 6.296359 4.554370 6.660937
2019 Q4       5.561924 4.838394 6.285454 4.455380 6.668468
2020 Q1       4.542152 3.785397 5.298907 3.384796 5.699509
2020 Q2       4.816553 3.974014 5.659092 3.528001 6.105106
2020 Q3       5.839554 4.968317 6.710791 4.507112 7.171996
2020 Q4       5.793824 4.894805 6.692843 4.418893 7.168755
2021 Q1       4.774052 3.848084 5.700020 3.357906 6.190198


>
```



- Summarize this forecasting technique
  - How good is the accuracy?
    MSE is low, hence accuracy is high.
  - What does it predict the value of time series will be in one year?
    The predicted value for Q2, 2019 is 4.58
  - Other observation

Accuracy Summary

- Show a table of all the forecast method above with their accuracy measures.

	NAIVE	SIMPLE-SMOOTHING	Holt-Winter
ME	0.0407013	0.09266392	-0.0089909
RMSE	0.4738365	0.4481574	0.2905301
MAE	0.3272727	0.326483	0.2094488
MPE	0.6541332	2.404932	-2.914616
MAPE	12.01044	11.762	9.993218
MASE	0.8748711	0.8727599	0.5599021

- Separately define each forecast method and why it is useful. Show the best and worst forecast method for each of the accuracy measures.

  For naïve forecasts, we simply set all forecasts to be the value of the last observation. This method works remarkably well for many economic and financial time series.

  The simple exponential smoothing method does not account for any trend or seasonal components, rather, it only uses the decreasing weights to forecast future results. This makes the method suitable only for time series without trend and seasonality

  Holt-Winters method, is one of the many methods or algorithms that can be used to forecast data points in a series, provided that the series is "seasonal", i.e. repetitive over some period.

  ME is best for Holts winters followed by Naive
  RMSE is best for Holts winters followed by simple smoothing
  MAE is best for Holts winters followed by simple smoothing
  MPE is best for Holts winters followed by Naive
  MAPE is best for Holts winters followed by simple smoothing
  MASE is best for Holts winters followed by simple smoothing

Conclusion

- Summarize your analysis of time series value over the time-period.

  Overall the trend is rising, with seasonal high sales in third and fourth quarter compared to first and second.

- Based on your analysis and forecast above, do you think the value of the time series will increase, decrease or stay flat over the next year? How about next 2 years?
  Based on the trend, seasonal pattern and our forecast, value will increase over the next year and next 2 years as well.

- Rank forecasting methods that best forecast for this time series based on historical values.
  1. Holts Winters
  2. Simple smoothing
  3. Naive