BUDA530Module5

## Problem 1 :

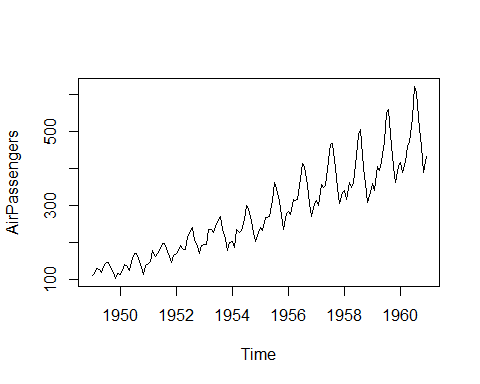
Using the AirPassengers data use the techniques you’ve learned in this module to forecast the next 10 time points. Give an interpretation to the next time point and discuss the intervals for prediction. You should at least have two comparisons. Discuss any thoughts you have on how you might select between the two models.

library(forecast)

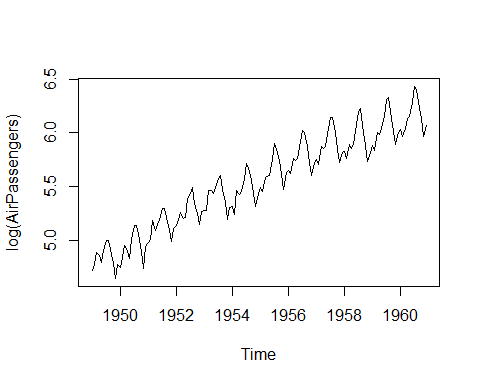
## Warning: package 'forecast' was built under R version 3.6.2

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

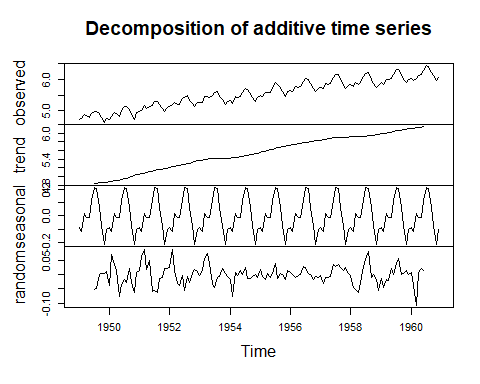
#alpha : the estimates of the level at the current time point  
#beta : slope b of the trend component at the current time point  
#gamma : the seasonal #component at the current time point  
#The parameters alpha, beta and gamma all have values between 0 and 1, and values that are close to 0 mean that relatively little weight is placed on the most recent observations when making forecasts of future values.  
  
#head(AirPassengers)  
#summary(AirPassengers)  
plot(AirPassengers)



plot(log(AirPassengers))



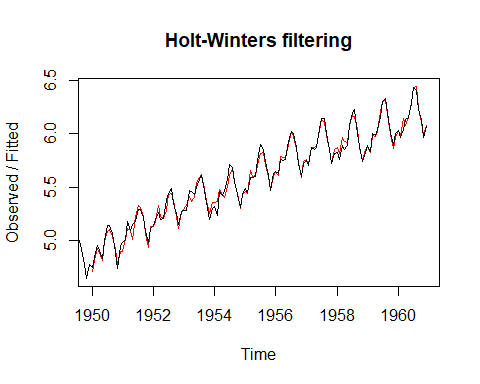
airPassengers\_dc <- decompose(log(AirPassengers))  
plot(airPassengers\_dc)



#airPassengers\_dc$seasonal  
  
airpassengersH <- HoltWinters(log(AirPassengers))  
airpassengersH

## Holt-Winters exponential smoothing with trend and additive seasonal component.  
##   
## Call:  
## HoltWinters(x = log(AirPassengers))  
##   
## Smoothing parameters:  
## alpha: 0.3266015  
## beta : 0.005744138  
## gamma: 0.8206654  
##   
## Coefficients:  
## [,1]  
## a 6.172308435  
## b 0.008981893  
## s1 -0.073201087  
## s2 -0.140973564  
## s3 -0.036703294  
## s4 0.014522733  
## s5 0.032554237  
## s6 0.154873570  
## s7 0.294317062  
## s8 0.276063997  
## s9 0.088237657  
## s10 -0.032657089  
## s11 -0.198012716  
## s12 -0.102863837

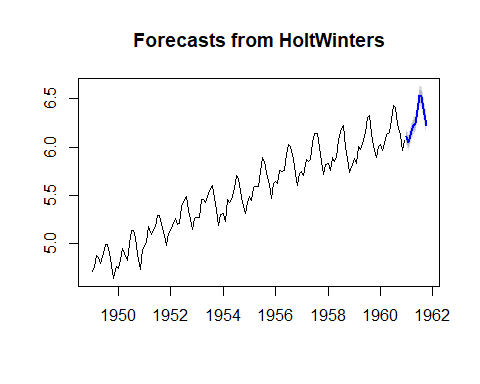
plot(airpassengersH)



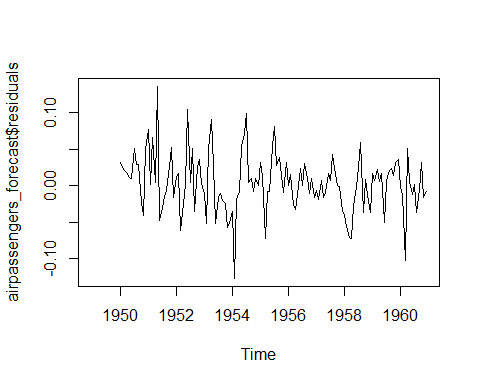
airpassengers\_forecast <- forecast(airpassengersH, h=10)  
airpassengers\_forecast

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jan 1961 6.108089 6.057867 6.158312 6.031281 6.184898  
## Feb 1961 6.049299 5.996436 6.102161 5.968452 6.130145  
## Mar 1961 6.162551 6.107146 6.217956 6.077816 6.247285  
## Apr 1961 6.222759 6.164896 6.280622 6.134265 6.311252  
## May 1961 6.249772 6.189525 6.310019 6.157633 6.341912  
## Jun 1961 6.381073 6.318508 6.443638 6.285388 6.476758  
## Jul 1961 6.529499 6.464674 6.594324 6.430357 6.628640  
## Aug 1961 6.520228 6.453195 6.587261 6.417709 6.622746  
## Sep 1961 6.341383 6.272189 6.410577 6.235560 6.447206  
## Oct 1961 6.229470 6.158159 6.300782 6.120408 6.338532

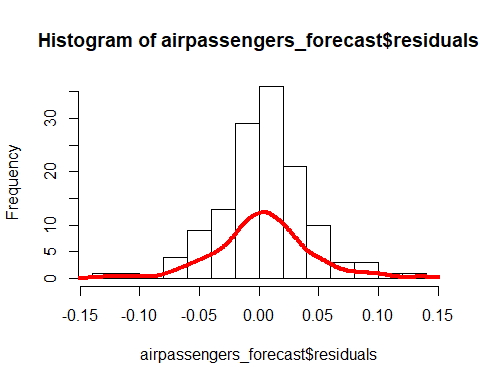
plot(airpassengers\_forecast)



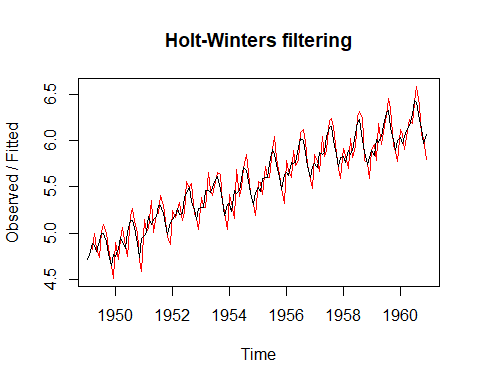
plot(airpassengers\_forecast$residuals)



hist(airpassengers\_forecast$residuals,breaks=10)  
lines(density(airpassengers\_forecast$residuals, na.rm=T), col="red",lwd=4)



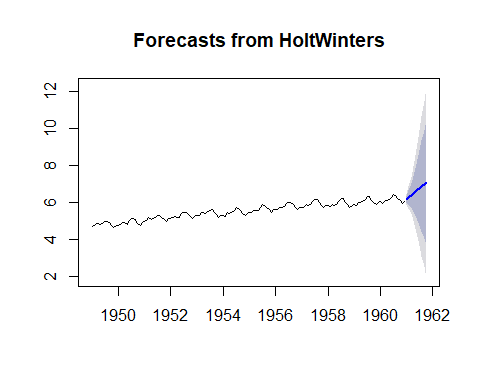
HoltWinters\_airpassengers1 <- HoltWinters(log(AirPassengers), beta = TRUE, gamma=FALSE, alpha=TRUE)  
#HoltWinters\_airpassengers1 <- HoltWinters(log(AirPassengers), beta = FALSE, gamma=TRUE, alpha=TRUE)  
plot(HoltWinters\_airpassengers1)



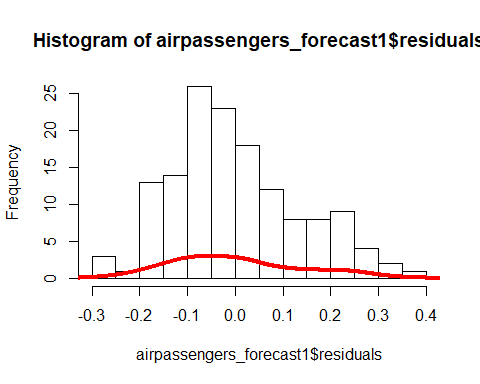
airpassengers\_forecast1 <- forecast(HoltWinters\_airpassengers1,h=10)  
forecast(HoltWinters\_airpassengers1,h=10)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jan 1961 6.170704 5.997685 6.343724 5.906094 6.435315  
## Feb 1961 6.272983 5.886101 6.659866 5.681297 6.864669  
## Mar 1961 6.375262 5.727883 7.022641 5.385182 7.365342  
## Apr 1961 6.477541 5.529876 7.425206 5.028212 7.926870  
## May 1961 6.579820 5.296675 7.862965 4.617419 8.542221  
## Jun 1961 6.682099 5.031600 8.332597 4.157880 9.206318  
## Jul 1961 6.784378 4.737186 8.831569 3.653469 9.915286  
## Aug 1961 6.886656 4.415448 9.357865 3.107269 10.666044  
## Sep 1961 6.988935 4.068035 9.909836 2.521803 11.456067  
## Oct 1961 7.091214 3.696332 10.486096 1.899190 12.283239

plot(forecast(airpassengers\_forecast1, h= 10))



hist(airpassengers\_forecast1$residuals,breaks=10)  
lines(density(airpassengers\_forecast1$residuals, na.rm=T), col="red",lwd=4)



### Interpretation :

It is seen in the plot of log of AirPassengers that the fluctuations in the time series are roughly constant in size over time. THe decomposed plot shows the level, trend and seasonality in the time series of AirPassengers data. The estimated seasonal factors are for the months January to December, and are the same for each year. The largest seasonal factor is for July of about 0.21 , and the lowest is for November of about -0.21 , indicating that there seems to be a peak in July and less passengers in November each year.

Using HoltWinters model without explicitly selecting values for alpha, beta and gamma, forecast for 10 next points which is Jan 1961 - Oct 1961 is done. The estimated values of alpha, beta and gamma are 0.33, 0.01, and 0.82, respectively. The forecasted passengers for Jan 1961 is about 6.11 , with a 95% prediction interval of 6.031 to 6.185. We see from the forecast plot that the prediction of the seasonal peaks, which occur roughly in July and November every year is constant for each year.The forecasts values are shown as a blue line along with the grey shades of 80% and 95% prediction intervals, respectively. The predicted intervals for the next 10 points looks consistent in range. From the histogram of forecast model, it seems plausible that the forecast errors are normally distributed with mean zero.

While comparing the models, model with beta = TRUE, gamma=FALSE, alpha=TRUE is less accurate in forecasting the next points. The plot of the model shows the variance in actual values and predicted values (red line). The forecasted points have large predicted intervals which shows large and inconsistent variance. HIstogram plot do not show normal distrubution. Forecasted plot shows a straigh blue line and do not show constant variance. Therefore, Model that selects alpha, gamma and beta by itself is good fit for forecasting airpassengers data.

## Problem 2 :

The wineind data in R is an interesting data set for time series analysis. Use the methods associated with ARIMA such as differences and autocorrelation to detect if there is a lag structure. Then compare a Holt Winters Model to that of an ARIMA model. Hint in the auto.arima function in R there is a seasonal aspect that adds seasonal functionality to arima. It will return what you need. In the arima function you can supply the argument seasonal = c(p,d,q) with the recommendation from auto.arima.

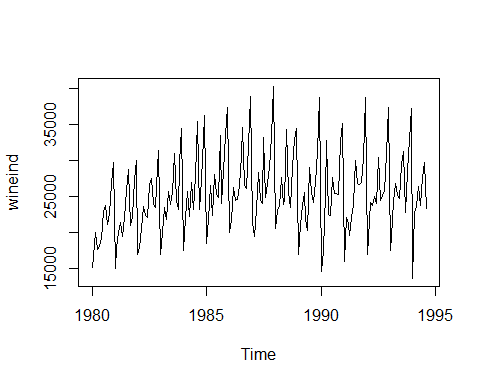
library(forecast)  
library(tseries)

## Warning: package 'tseries' was built under R version 3.6.2

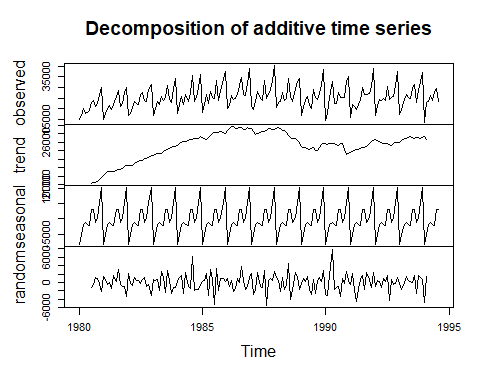
library(TTR)

## Warning: package 'TTR' was built under R version 3.6.2

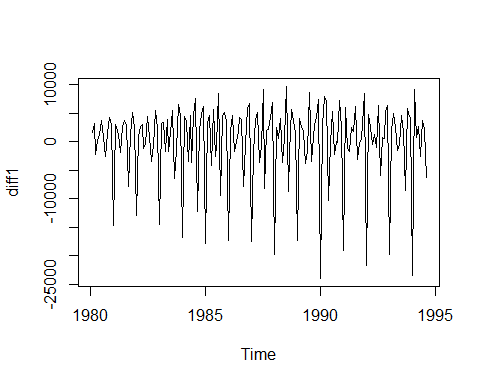
#wineind  
#summary(wineind)  
plot(wineind)



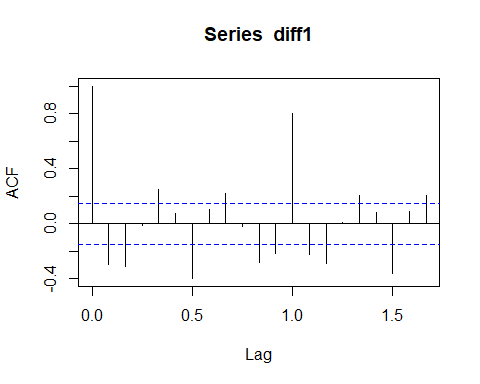
wineind\_ts <- ts(wineind, start=c(1980), frequency = 12)  
  
  
wineind\_dc <- decompose(wineind\_ts)  
plot(wineind\_dc)



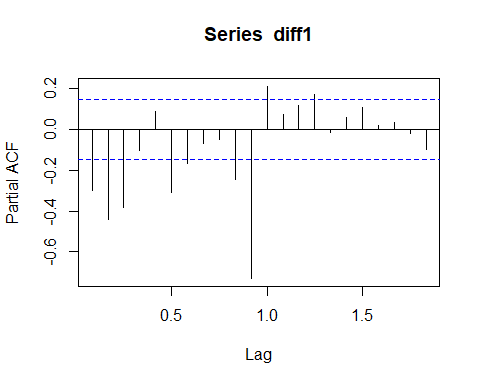
#wineind\_dc$seasonal  
  
diff1 <- diff(wineind\_ts, lag = 1)  
#diff1 <- diff(wineind,lag = 2)  
plot(diff1)



acf(diff1, lag.max = 20)



pacf(diff1)



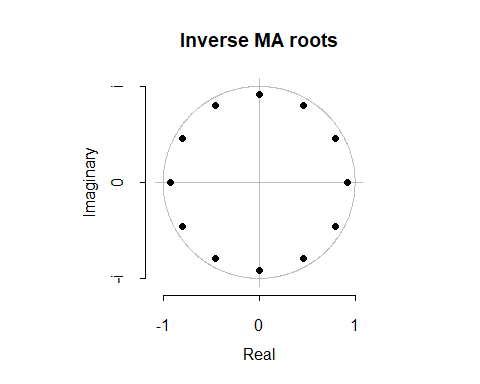
auto.arima(wineind\_ts)

## Series: wineind\_ts   
## ARIMA(1,1,2)(0,1,1)[12]   
##   
## Coefficients:  
## ar1 ma1 ma2 sma1  
## 0.4299 -1.4673 0.5339 -0.6600  
## s.e. 0.2984 0.2658 0.2340 0.0799  
##   
## sigma^2 estimated as 5399312: log likelihood=-1497.05  
## AIC=3004.1 AICc=3004.48 BIC=3019.57

arimaModel <- Arima(wineind\_ts, seasonal = c(0,1,1), include.drift = T)  
arimaModel

## Series: wineind\_ts   
## ARIMA(0,0,0)(0,1,1)[12] with drift   
##   
## Coefficients:  
## sma1 drift  
## -0.3671 27.6784  
## s.e. 0.0831 10.8321  
##   
## sigma^2 estimated as 6369104: log likelihood=-1517.26  
## AIC=3040.52 AICc=3040.67 BIC=3049.82

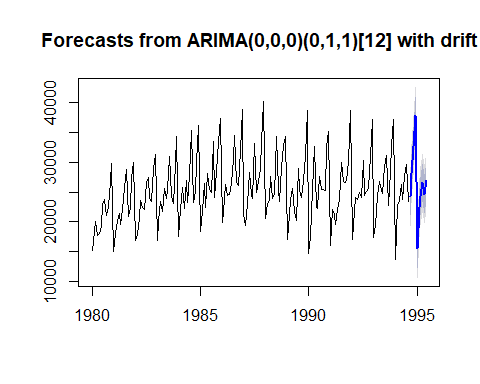
plot(arimaModel)



arimaforecast <- forecast(arimaModel, h=10)  
forecast(arimaModel, h=10)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Sep 1994 24260.73 21026.47 27495.00 19314.36 29207.11  
## Oct 1994 28112.86 24878.59 31347.12 23166.48 33059.23  
## Nov 1994 32678.77 29444.51 35913.03 27732.39 37625.15  
## Dec 1994 37804.31 34570.05 41038.57 32857.93 42750.69  
## Jan 1995 15464.18 12229.91 18698.44 10517.80 20410.55  
## Feb 1995 22355.30 19121.04 25589.56 17408.93 27301.68  
## Mar 1995 24273.96 21039.70 27508.22 19327.58 29220.34  
## Apr 1995 26600.63 23366.37 29834.89 21654.25 31547.01  
## May 1995 24671.97 21437.71 27906.23 19725.59 29618.35  
## Jun 1995 26903.99 23669.73 30138.26 21957.62 31850.37

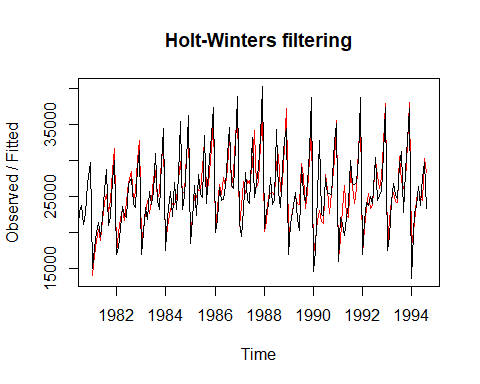
plot(arimaforecast)



#auto.arima(wineind\_ts, trace = T)  
  
holtwintersModel <- HoltWinters(wineind\_ts)  
holtwintersModel

## Holt-Winters exponential smoothing with trend and additive seasonal component.  
##   
## Call:  
## HoltWinters(x = wineind\_ts)  
##   
## Smoothing parameters:  
## alpha: 0.01655901  
## beta : 0.812572  
## gamma: 0.2759254  
##   
## Coefficients:  
## [,1]  
## a 26828.30329  
## b -61.97017  
## s1 -1756.05105  
## s2 667.28919  
## s3 5392.28635  
## s4 10730.34747  
## s5 -10141.59571  
## s6 -5212.79108  
## s7 -2617.86356  
## s8 -767.74288  
## s9 -2443.95319  
## s10 -1587.83526  
## s11 3193.31754  
## s12 71.02039

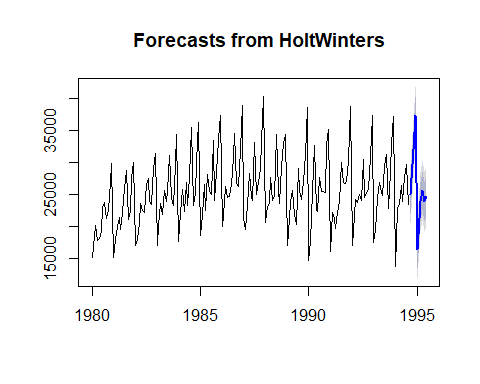
plot(holtwintersModel)



holtwintersforecast <- forecast(holtwintersModel, h=10)  
forecast(holtwintersModel, h=10)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Sep 1994 25010.28 22040.01 27980.55 20467.64 29552.92  
## Oct 1994 27371.65 24400.04 30343.26 22826.97 31916.34  
## Nov 1994 32034.68 29060.27 35009.09 27485.71 36583.65  
## Dec 1994 37310.77 34331.56 40289.98 32754.45 41867.09  
## Jan 1995 16376.86 13390.32 19363.40 11809.34 20944.38  
## Feb 1995 21243.69 18246.79 24240.59 16660.32 25827.06  
## Mar 1995 23776.65 20765.84 26787.45 19172.02 28381.28  
## Apr 1995 25564.80 22536.08 28593.52 20932.77 30196.83  
## May 1995 23826.62 20775.51 26877.72 19160.36 28492.88  
## Jun 1995 24620.77 21542.39 27699.15 19912.79 29328.74

plot(holtwintersforecast)

 ### Interpretation :

Plot wineind show the almost equal variance in the trend. DIfferences plot show the steady trend. Wine sales is in peak in November and december months and least is in January.

When comparing the holtwinters and arima models, forecasting of arima is better than holtswinter. The Confidence intervals are better for arima model compared to holtswinter model.