Atajan Abdyyev

2022-08-07

# Time Series and Forecasting

# Author: Atajan Abdyyev  
  
#1 Check your working directory  
#getwd()

#2 Set your working directory to "ANLY 565/RScript".   
  
setwd('C:/Users/ataja/OneDrive/Desktop/ANLY 565/RScript')  
  
#3 Download "ffrategdp.xls" data file and set the "observation\_date"   
# variable to the date format and the "FEDFUNDS" and "GDPC1" variables to the numeric format.  
# The "FEDFUNDS" variable represents the effective federal funds rate,  
# which indicates the interest rate at which depository institutions trade federal funds   
# (balances held at Federal Reserve Banks) with each other overnight.   
# The "GDPC1" variable represents real gross domestic product.  
  
library(readxl)  
ffrategdp = read\_excel("ffrategdp.xls", col\_types =c("date","numeric", "numeric"))  
head(ffrategdp)

## # A tibble: 6 x 3  
## observation\_date FEDFUNDS GDPC1  
## <dttm> <dbl> <dbl>  
## 1 1954-07-01 00:00:00 1.03 2683.  
## 2 1954-10-01 00:00:00 0.987 2735.  
## 3 1955-01-01 00:00:00 1.34 2813.  
## 4 1955-04-01 00:00:00 1.5 2859.  
## 5 1955-07-01 00:00:00 1.94 2898.  
## 6 1955-10-01 00:00:00 2.36 2915.

str(ffrategdp)

## tibble [260 x 3] (S3: tbl\_df/tbl/data.frame)  
## $ observation\_date: POSIXct[1:260], format: "1954-07-01" "1954-10-01" ...  
## $ FEDFUNDS : num [1:260] 1.027 0.987 1.343 1.5 1.94 ...  
## $ GDPC1 : num [1:260] 2683 2735 2813 2859 2898 ...

#4 By using ts() function create a time series object that contains two variables: "FEDFUNDS" and "GDPC1".  
# Label it as "ffrategdpts".  
range(ffrategdp$observation\_date)

## [1] "1954-07-01 UTC" "2019-04-01 UTC"

ffrategdpts = ts(ffrategdp[-1],start=c(1954,3), freq = 4)  
head(ffrategdpts)

## FEDFUNDS GDPC1  
## [1,] 1.0266667 2682.601  
## [2,] 0.9866667 2735.091  
## [3,] 1.3433333 2813.212  
## [4,] 1.5000000 2858.988  
## [5,] 1.9400000 2897.598  
## [6,] 2.3566667 2914.993

str(ffrategdpts)

## Time-Series [1:260, 1:2] from 1954 to 2019: 1.027 0.987 1.343 1.5 1.94 ...  
## - attr(\*, "dimnames")=List of 2  
## ..$ : NULL  
## ..$ : chr [1:2] "FEDFUNDS" "GDPC1"

#5 Create two stand alone variables "fedrate" and "gdp" that take on values of the "FEDFUNDS" and "GDPC1"  
# variables from the "ffrategdpts" data set  
fedrate = ffrategdpts[,1]  
gdp =ffrategdpts[,2]  
  
  
  
#6 When federal funds rate goes down, the commercial loan interest rates go down too.  
# This means that people can borrow cheaply and invest in their businesses,   
# which will result in higher gross domestic output.   
# Therefore, you suspect that the federal funds rate has a negative correlation with GDP.  
# To test this hypothesis you decide to use lm() function to estimate the coefficients of   
# a linear regression model in which "gdp" is a dependent variable and "fedrate"   
# is as an independent variable.   
# Save the estimated model as gdpfr.lm  
# Based on the results of this model can you make any conclusions about the nature of the   
# relationship between the gdp and the federal funds rate?  
  
gdpfr.lm = lm(gdp ~ fedrate)  
summary(gdpfr.lm)

##   
## Call:  
## lm(formula = gdp ~ fedrate)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8680.0 -3843.8 526.3 3522.0 8398.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11915.26 467.51 25.487 < 2e-16 \*\*\*  
## fedrate -538.30 78.18 -6.885 4.38e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4504 on 258 degrees of freedom  
## Multiple R-squared: 0.1552, Adjusted R-squared: 0.1519   
## F-statistic: 47.4 on 1 and 258 DF, p-value: 4.382e-11

##answer: for every gdp value of 1 % going up, the fed rate goes down by 538.3. We can see that   
#fedrate and gdp are negatively associated and fedrate will go do if gdp goes up.   
## results are significant in pvalue but R-squared being close to 0 at 0.15 suggests this is a bad model in prediction using.  
  
#7 You have suspected that the "gdp" variable may contain a unit root.  
# By using Augmented Dickey Fuller method test "gdp" variable for the   
# the presence of unit root.   
# Does "gdp" variable contain a unit root?  
# Is "gdp" variable stationary?  
  
library(tseries)

## Warning: package 'tseries' was built under R version 4.1.3

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

adf.test(gdp)

##   
## Augmented Dickey-Fuller Test  
##   
## data: gdp  
## Dickey-Fuller = -1.5023, Lag order = 6, p-value = 0.7855  
## alternative hypothesis: stationary

#answer: we see that adf test shows pvalue at 0.7855 and we failed to reject null hypotheses, I think that gdp data is nonstationary.   
#failed to reject null hypothesis unit root presents.   
  
#8 By using Augmented Dickey Fuller method test "fedrate" variable for the   
# the presence of unit root.   
# Does "fedrate" variable contain a unit root?  
# Is "fedrate" variable stationary?  
  
adf.test(fedrate)

##   
## Augmented Dickey-Fuller Test  
##   
## data: fedrate  
## Dickey-Fuller = -2.783, Lag order = 6, p-value = 0.2462  
## alternative hypothesis: stationary

#answer: we see that adf test shows pvalue at 0.2462 and we failed to reject null hypotheses, I think that gdp data is nonstationary.   
#failed to reject null hypothesis unit root presents.   
  
#9 The Phillips-Ouliaris test shows whether there is evidence that the series are  
# cointegrated, which justifies the use of a regression model.   
# Are "gdp" and "fedrate" variables cointegrated?  
# Is "gdpfr.lm" a suitable model to explore the relationship between "gdp" and "fedrate"?  
  
po.test(cbind(fedrate, gdp))

## Warning in po.test(cbind(fedrate, gdp)): p-value greater than printed p-value

##   
## Phillips-Ouliaris Cointegration Test  
##   
## data: cbind(fedrate, gdp)  
## Phillips-Ouliaris demeaned = -13.028, Truncation lag parameter = 2,  
## p-value = 0.15

#answer: p-value = 0.15 we can assume that series are not cointegrated (larger than null hypothesis = cointegrated). gdpfr.lm is a suitable model.  
  
#10 Create the following 2 new variables:  
# "gdpgrowth" - that represents quarterly percentage change in GDP  
# "fedratediff" - that represents quarterly difference in the federal funds rate (simple difference)  
# To each of the variables add "NA" as the first observation .  
# This will ensure that the new variables are of the same length as the existing variables.  
gdpgrowth <- log(gdp[-1]/gdp[-length(gdp)])\*100   
head(gdpgrowth)

## [1] 1.9377861 2.8162183 1.6140825 1.3414402 0.5985301 -0.3891620

fedratediff=diff(fedrate)  
head(fedratediff)

## [1] -0.0400000 0.3566667 0.1566667 0.4400000 0.4166667 0.1266667

#11 By using ts() and cbind() functions add "gdpgrowth" and "fedratediff" variables   
# to the "ffrategdpts" data set.   
ffrategdpts<-(ts(cbind(gdpgrowth,fedratediff),start=c(1954,3), freq = 4))  
head(ffrategdpts)

## gdpgrowth fedratediff  
## [1,] 1.9377861 -0.0400000  
## [2,] 2.8162183 0.3566667  
## [3,] 1.6140825 0.1566667  
## [4,] 1.3414402 0.4400000  
## [5,] 0.5985301 0.4166667  
## [6,] -0.3891620 0.1266667

#12 Use na.omit() function to get rid of the missing values in the "ffrategdpts" data set.   
# Save the new data set as "ffrategdptscc".   
  
ffrategdptscc = na.omit(ffrategdpts)  
head(ffrategdptscc)

## gdpgrowth fedratediff  
## [1,] 1.9377861 -0.0400000  
## [2,] 2.8162183 0.3566667  
## [3,] 1.6140825 0.1566667  
## [4,] 1.3414402 0.4400000  
## [5,] 0.5985301 0.4166667  
## [6,] -0.3891620 0.1266667

#13 Create 2 new variables:   
# "ggdp" - takes on values of the "gdpgrowth" from the "ffrategdptscc"  
# "dfrate" - takes on values of the "fedratediff" from the "ffrategdptscc"  
  
ggdp = ffrategdpts[,1]  
dfrate = ffrategdpts[,2]  
  
head(ggdp)

## [1] 1.9377861 2.8162183 1.6140825 1.3414402 0.5985301 -0.3891620

head(dfrate)

## [1] -0.0400000 0.3566667 0.1566667 0.4400000 0.4166667 0.1266667

#14 Use to Augmented Dickey-Fuller test to determine whether "ggdp" and "dfrate"  
# are stationary or not.   
# Does "ggdp" contain a unit root? Is "ggdp" stationary?   
# Does "dfrate" contain a unit root? Is "dfrate" stationary?  
  
adf.test(ggdp)

## Warning in adf.test(ggdp): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: ggdp  
## Dickey-Fuller = -6.051, Lag order = 6, p-value = 0.01  
## alternative hypothesis: stationary

adf.test(dfrate)

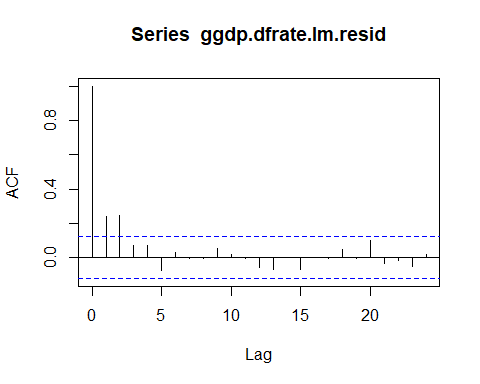
## Warning in adf.test(dfrate): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: dfrate  
## Dickey-Fuller = -7.2109, Lag order = 6, p-value = 0.01  
## alternative hypothesis: stationary

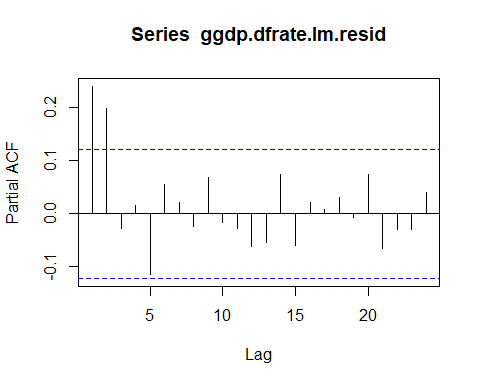
#answer: both gdp and dfrate differences have pvalue below 0.05 (0.01) meaning data is stationary and unit root presents in differences.  
  
#15 Use lm() function to estimate the coefficients of a linear regression model   
# in which "ggdp" is a dependent variable and "dfrate" is as an independent variable.  
# Lable these estimates as "ggdp.dfrate.lm".  
# Based on the findings of the linear regression model what is the nature of the relationship   
# between the growth rate of real gdp and difference in federal funds rate?  
  
ggdp.dfrate.lm = lm(ggdp~dfrate)  
summary(ggdp.dfrate.lm)

##   
## Call:  
## lm(formula = ggdp ~ dfrate)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.9319 -0.3987 0.0024 0.4209 2.8653   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.75461 0.05081 14.85 < 2e-16 \*\*\*  
## dfrate 0.32580 0.05902 5.52 8.27e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8177 on 257 degrees of freedom  
## Multiple R-squared: 0.106, Adjusted R-squared: 0.1025   
## F-statistic: 30.47 on 1 and 257 DF, p-value: 8.269e-08

#answer: data is significant with low pvalue, ad adjusted r is still 0.10, but we can see that dfrate is positive at 0.32. SO each GDP difference going up by 1  
## should result with rate difference to go up as well, now however the differences show a positive correlation with change of 0.32580 for the rate.  
  
#16 Create a variable called "ggdp.dfrate.lm.resid" that represents the residual series obtained   
# from the "ggdp.dfrate.lm" regression  
  
ggdp.dfrate.lm.resid = ggdp.dfrate.lm$residuals  
  
#17 Construct acf and pacf functions for "ggdp.dfrate.lm.resid".  
# What can you say about the goodness of the fit of the model?  
  
acf(ggdp.dfrate.lm.resid)



pacf(ggdp.dfrate.lm.resid)



##answer: there are residuals that are autocorrelated, so gdp and rate difference linear model fit is not the best  
  
#18 Maybe vector autoregression model would prove a better fit.   
# Upload "vars" library that contains VAR() function  
  
#install.packages('vars')  
library(vars)

## Warning: package 'vars' was built under R version 4.1.3

## Loading required package: MASS

## Loading required package: strucchange

## Warning: package 'strucchange' was built under R version 4.1.3

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 4.1.3

## Loading required package: urca

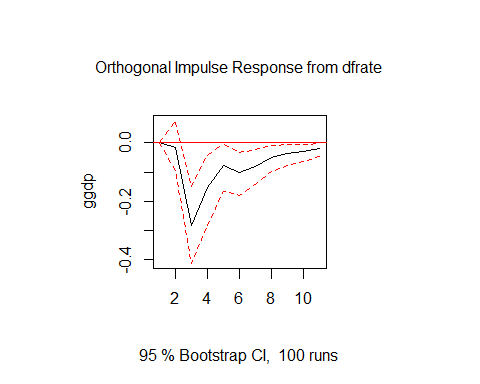
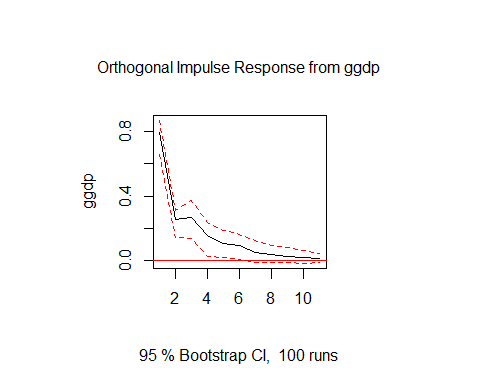
## Warning: package 'urca' was built under R version 4.1.3

## Loading required package: lmtest

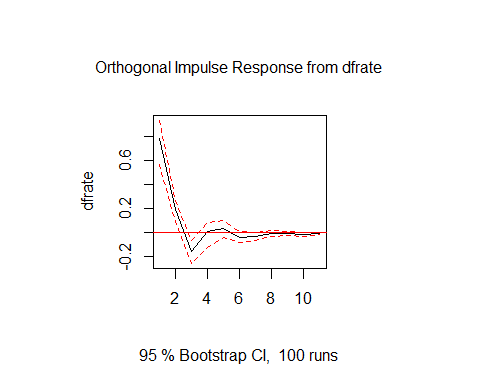
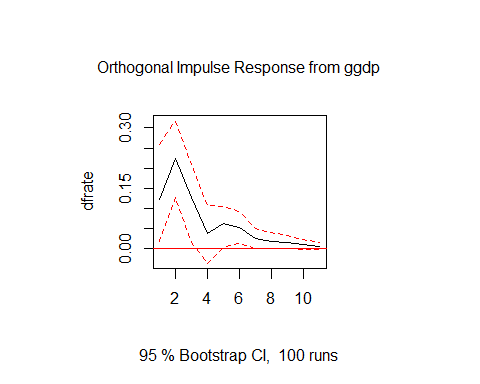
#19 Estimate a VAR model for the "ggdp" and "dfrate" variables.  
# In this model include 3 lags of each variable.   
# Save the estimates of the var model as "ggdp.dfrate.var"  
  
ggdp.dfrate.var =VAR(cbind(ggdp, dfrate), p = 3, type = "trend")   
ggdp.dfrate.var

##   
## VAR Estimation Results:  
## =======================   
##   
## Estimated coefficients for equation ggdp:   
## =========================================   
## Call:  
## ggdp = ggdp.l1 + dfrate.l1 + ggdp.l2 + dfrate.l2 + ggdp.l3 + dfrate.l3 + trend   
##   
## ggdp.l1 dfrate.l1 ggdp.l2 dfrate.l2 ggdp.l3 dfrate.l3   
## 0.324868938 -0.019580982 0.294905407 -0.351969101 0.092804853 0.007567916   
## trend   
## 0.001058961   
##   
##   
## Estimated coefficients for equation dfrate:   
## ===========================================   
## Call:  
## dfrate = ggdp.l1 + dfrate.l1 + ggdp.l2 + dfrate.l2 + ggdp.l3 + dfrate.l3 + trend   
##   
## ggdp.l1 dfrate.l1 ggdp.l2 dfrate.l2 ggdp.l3 dfrate.l3   
## 0.246345325 0.243252362 0.050724711 -0.255140685 -0.051840367 0.211175296   
## trend   
## -0.001032565

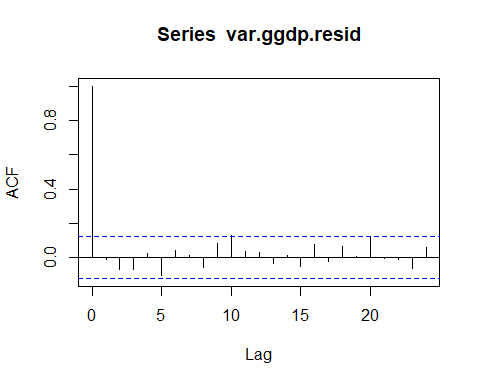
#20 Use plot() and irf() functions to obtain and plot impulse response functions for each variable.  
# IRF illustrates the behavior of a variable in response to one standard deviation shock   
# in its own value and in the value of the other variable.  
# Based on the these graph what conclusions can you draw about the nature of the relationship between   
# the growth rate of gdp and the difference in federal funds rate?   
# Any potential explanations?  
plot(irf(ggdp.dfrate.var, response="ggdp", boot=T, nsteps = 4))



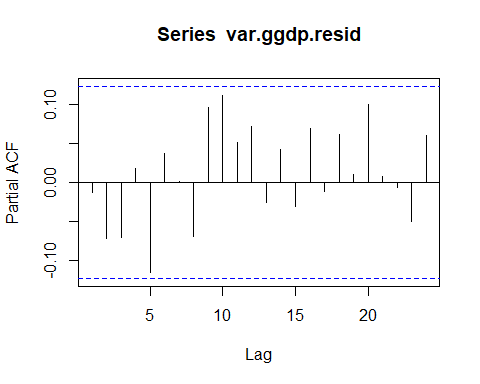
plot(irf(ggdp.dfrate.var, response="dfrate", boot=T, nsteps = 4))



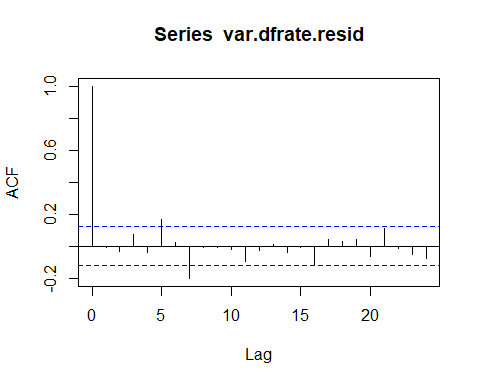
#answer: We see that GDP growth affects up to lag 3 in fed fund rate and when we get to lag 3 growth rate inverses.  
# we can assume that fed fund rate does change after 3 lags of following gdp growth change only to reverse its own trend.  
  
#21 Use resid() function to obtain the residuals from the ggdp equation of the "ggdp.dfrate.var" model.  
# Save this residual series as "var.ggdp.resid".  
var.ggdp.resid = (resid(ggdp.dfrate.var)[, 1])  
  
  
#22 #17 Use resid() function to obtain the residuals from the dfrate equation of the "ggdp.dfrate.var" model.  
# Save this residual series as "var.dfrate.resid".  
var.dfrate.resid = (resid(ggdp.dfrate.var)[, 2])  
  
  
#24 Plot acf and pacf functions for the 'var.ggdp.resid".   
# Does "ggdp.dfrate.var" model provide a good fit to explain growth rate of gdp?  
  
acf(var.ggdp.resid)



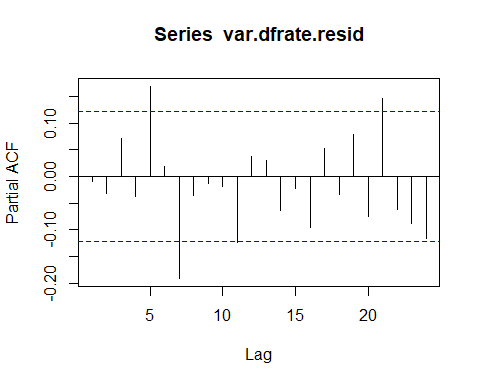
pacf(var.ggdp.resid)



#answer, there are no residuals autocorrelated making this VAR model a good fit to explain growth rate of gdp  
  
#25 Plot acf and pacf functions for the 'var.dfrate.resid".   
# Does "ggdp.dfrate.var" model provide a good fit to explain the difference in federal funds rate?  
acf(var.dfrate.resid)



pacf(var.dfrate.resid)



#answer, there are no residuals autocorrelated making this VAR model a good fit to explain growth rate of ratedifference  
  
  
#26 Use "ggdp.dfrate.var" model and predict() function to forecast growth rate of gdp and   
# change in federal funds rate over the upcoming year.   
# Save the predicted values as "VAR.pred"  
  
VAR.pred=predict(ggdp.dfrate.var,n.ahead= 4)  
VAR.pred

## $ggdp  
## fcst lower upper CI  
## [1,] 0.6284083 -0.9218272 2.178644 1.550236  
## [2,] 0.7070941 -0.9217267 2.335915 1.628821  
## [3,] 0.7785642 -1.0213790 2.578507 1.799943  
## [4,] 0.8392158 -1.0112413 2.689673 1.850457  
##   
## $dfrate  
## fcst lower upper CI  
## [1,] -0.10426595 -1.656041 1.447509 1.551775  
## [2,] -0.11335170 -1.768923 1.542220 1.655572  
## [3,] -0.09321943 -1.795246 1.608808 1.702027  
## [4,] -0.09225239 -1.795915 1.611410 1.703662

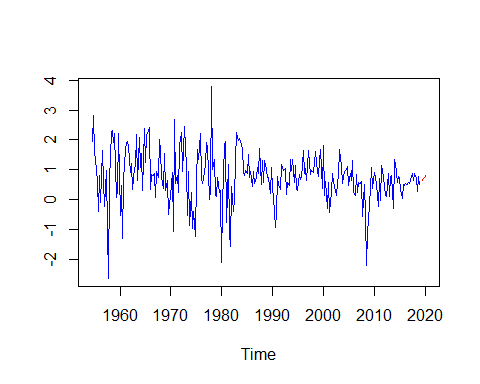
#27 Use ts() function and VAR.pred forecast to create a new variable "ggdp.pred".  
# It should contain the forcasted values of the growth rate of gdp over the next 4 quarters.  
  
ggdp.pred = ts(VAR.pred$fcst$ggdp[, 1],start = c(2019,3),fr= 4)  
ggdp.pred

## Qtr1 Qtr2 Qtr3 Qtr4  
## 2019 0.6284083 0.7070941  
## 2020 0.7785642 0.8392158

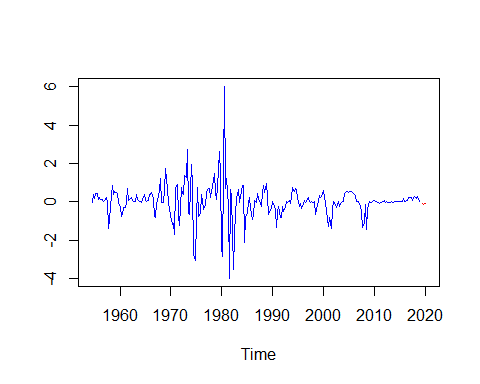
#28 Use ts() function and VAR.pred forecast to create a new variable "dfrate.pred".  
# It should contain the prediction of the change in the federal funds rate over the next 4 quarters.  
dfrate.pred =ts(VAR.pred$fcst$dfrate[, 1], start = c(2019,3),fr=4)  
dfrate.pred

## Qtr1 Qtr2 Qtr3 Qtr4  
## 2019 -0.10426595 -0.11335170  
## 2020 -0.09321943 -0.09225239

#29 Plot the times series graph of the past growth rates of gdp alongside   
# its future forecasted values. Do you expect the gdp to grow over the next 4 quarters?  
  
ts.plot(cbind(ggdp, ggdp.pred), lty = 1:2,col=c("blue","red"))



#answer: Yes, we expect data to go up over the next 4 quarters   
  
#30 Plot the times series graph of the past changes of the federal funds rate alongside   
# its future forecasted values.   
# Do you expect the federal funds rate to increase over the next 4 quarters?  
# Should one take out a loan now?  
  
ts.plot(cbind(dfrate, dfrate.pred),lty = 1:2,col=c("blue","red"))



#answer: no, we expect rates to go down and stay flat over the next 4 quarters, it might be good start taking loans over the next quarter as it will   
# stay flat for a while.