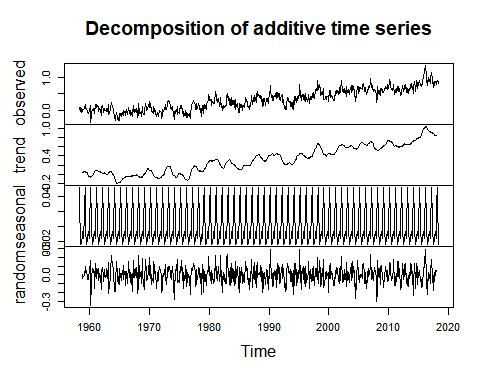
ECON 5337 Project

Load csv file

setwd("D:/ECON5337")  
library(urca)  
library(forecast)  
  
df\_monthly = read.csv("monthly\_csv.csv")  
Date = rev(df\_monthly$Date)  
Source = rev(df\_monthly$Source)  
Mean = rev(df\_monthly$Mean)  
df\_monthly = data.frame(Source, Date, Mean)  
  
df\_GISTEMP\_montly = df\_monthly[c(TRUE, FALSE),]  
df\_GCAG\_montly = df\_monthly[!c(TRUE, FALSE),]  
  
gisTEMP = ts(df\_GISTEMP\_montly$Mean, start=c(1880,1), frequency=12)  
  
gisFULL = window(gisTEMP, start=c(1958,3), end=c(2018,10))  
gisHOLD = window(gisTEMP, start=c(1958,3), end=c(2018,6)) # holding 4 observations

figure out it has seasonality

plot(decompose(gisHOLD))



BoxCox test

BoxCox.lambda(gisHOLD)

## [1] 0.9707678

no need to make log formation

Unit root test

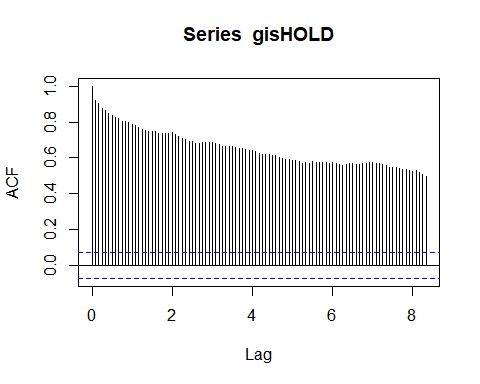
summary(ur.df(gisHOLD, type="trend", lags=30, selectlags = "AIC"))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.32126 -0.07562 0.00191 0.07056 0.38569   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.620e-02 1.091e-02 -4.235 2.59e-05 \*\*\*  
## z.lag.1 -2.379e-01 3.400e-02 -6.996 6.27e-12 \*\*\*  
## tt 3.351e-04 5.138e-05 6.522 1.34e-10 \*\*\*  
## z.diff.lag1 -2.816e-01 4.274e-02 -6.589 8.82e-11 \*\*\*  
## z.diff.lag2 -5.791e-02 4.263e-02 -1.358 0.1748   
## z.diff.lag3 -1.144e-01 3.785e-02 -3.022 0.0026 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1067 on 687 degrees of freedom  
## Multiple R-squared: 0.2418, Adjusted R-squared: 0.2362   
## F-statistic: 43.81 on 5 and 687 DF, p-value: < 2.2e-16  
##   
##   
## Value of test-statistic is: -6.9959 16.4266 24.4997   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -3.96 -3.41 -3.12  
## phi2 6.09 4.68 4.03  
## phi3 8.27 6.25 5.34

reject the null

however, when we look at the acf(gisHOLD),

acf(gisHOLD,100)

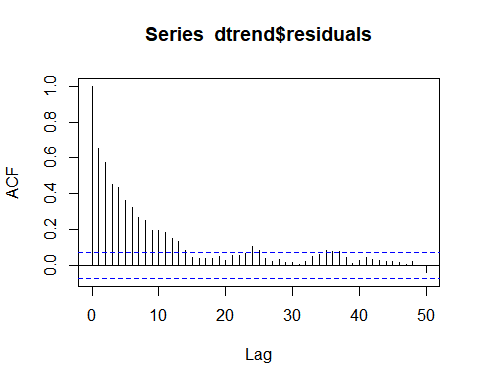
 it has very slow decay which means it might be non-stationary

Detrend the data

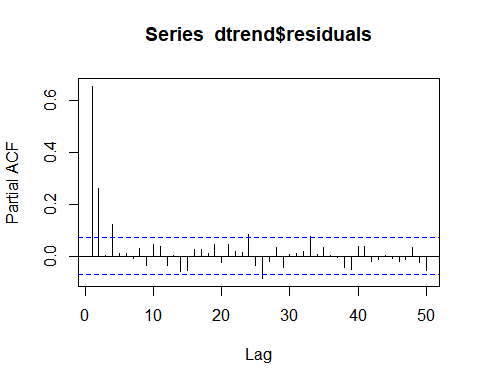
length(gisHOLD) #724

## [1] 724

t=1:724  
dtrend=lm(gisHOLD~t)  
acf(dtrend$residuals,50)



pacf(dtrend$residuals,50)

 seasonal lags are getting smaller, no need to take seasonal difference

estimating models

modelINIT = Arima(gisHOLD, order=c(1,1,0), seasonal=c(1,0,0), include.drift = TRUE) # initial  
modelINIT

## Series: gisHOLD   
## ARIMA(1,1,0)(1,0,0)[12] with drift   
##   
## Coefficients:  
## ar1 sar1 drift  
## -0.3878 0.0086 0.0011  
## s.e. 0.0343 0.0371 0.0030  
##   
## sigma^2 estimated as 0.01268: log likelihood=554.4  
## AIC=-1100.8 AICc=-1100.74 BIC=-1082.47

Arima(gisHOLD, order=c(1,1,1), seasonal=c(1,0,0), include.drift = TRUE) # trial 1

## Series: gisHOLD   
## ARIMA(1,1,1)(1,0,0)[12] with drift   
##   
## Coefficients:  
## ar1 ma1 sar1 drift  
## 0.3028 -0.7650 -0.0260 0.0011  
## s.e. 0.1010 0.0805 0.0387 0.0013  
##   
## sigma^2 estimated as 0.0121: log likelihood=571.86  
## AIC=-1133.72 AICc=-1133.64 BIC=-1110.81

Arima(gisHOLD, order=c(1,1,2), seasonal=c(1,0,0), include.drift = TRUE) # trial 2

## Series: gisHOLD   
## ARIMA(1,1,2)(1,0,0)[12] with drift   
##   
## Coefficients:  
## ar1 ma1 ma2 sar1 drift  
## 0.8642 -1.3875 0.3875 0.0229 0.0013  
## s.e. 0.0269 0.0529 0.0494 0.0373 0.0001  
##   
## sigma^2 estimated as 0.01157: log likelihood=586.52  
## AIC=-1161.04 AICc=-1160.93 BIC=-1133.54

Arima(gisHOLD, order=c(2,1,1), seasonal=c(1,0,0), include.drift = TRUE) # trial 3

## Series: gisHOLD   
## ARIMA(2,1,1)(1,0,0)[12] with drift   
##   
## Coefficients:  
## ar1 ar2 ma1 sar1 drift  
## 0.4668 0.2448 -0.9828 0.0171 0.0013  
## s.e. 0.0384 0.0382 0.0121 0.0379 0.0003  
##   
## sigma^2 estimated as 0.01159: log likelihood=587.4  
## AIC=-1162.81 AICc=-1162.69 BIC=-1135.31

Arima(gisHOLD, order=c(3,1,1), seasonal=c(1,0,0), include.drift = TRUE) # trial 4

## Series: gisHOLD   
## ARIMA(3,1,1)(1,0,0)[12] with drift   
##   
## Coefficients:  
## ar1 ar2 ar3 ma1 sar1 drift  
## 0.4675 0.2468 -0.0054 -0.9823 0.0165 0.0013  
## s.e. 0.0389 0.0409 0.0386 0.0127 0.0382 0.0003  
##   
## sigma^2 estimated as 0.0116: log likelihood=587.41  
## AIC=-1160.83 AICc=-1160.67 BIC=-1128.74

Arima(gisHOLD, order=c(2,1,1), seasonal=c(1,0,1), include.drift = TRUE) # trial 5

## Series: gisHOLD   
## ARIMA(2,1,1)(1,0,1)[12] with drift   
##   
## Coefficients:  
## ar1 ar2 ma1 sar1 sma1 drift  
## 0.4680 0.2430 -0.9808 -0.8162 0.7854 0.0013  
## s.e. 0.0389 0.0384 0.0129 0.2036 0.2175 0.0003  
##   
## sigma^2 estimated as 0.01157: log likelihood=588.33  
## AIC=-1162.65 AICc=-1162.5 BIC=-1130.57

Arima(gisHOLD, order=c(2,1,1), seasonal=c(1,0,2), include.drift = TRUE) # trial 6

## Series: gisHOLD   
## ARIMA(2,1,1)(1,0,2)[12] with drift   
##   
## Coefficients:  
## ar1 ar2 ma1 sar1 sma1 sma2 drift  
## 0.4662 0.2663 -0.9896 0.4808 -0.4665 0.1077 0.0013  
## s.e. 0.0376 0.0381 0.0107 0.2595 0.2587 0.0422 0.0002  
##   
## sigma^2 estimated as 0.01144: log likelihood=592.85  
## AIC=-1169.69 AICc=-1169.49 BIC=-1133.03

Arima(gisHOLD, order=c(2,1,1), seasonal=c(1,0,3), include.drift = TRUE) # trial 7

## Series: gisHOLD   
## ARIMA(2,1,1)(1,0,3)[12] with drift   
##   
## Coefficients:  
## ar1 ar2 ma1 sar1 sma1 sma2 sma3 drift  
## 0.4662 0.2679 -0.9896 0.8245 -0.8116 0.1037 -0.0548 0.0013  
## s.e. 0.0377 0.0380 0.0108 0.2244 0.2295 0.0485 0.0540 0.0002  
##   
## sigma^2 estimated as 0.01145: log likelihood=593.01  
## AIC=-1168.03 AICc=-1167.77 BIC=-1126.78

Arima(gisHOLD, order=c(2,1,1), seasonal=c(2,0,2), include.drift = TRUE) # trial 8

## Series: gisHOLD   
## ARIMA(2,1,1)(2,0,2)[12] with drift   
##   
## Coefficients:  
## ar1 ar2 ma1 sar1 sar2 sma1 sma2 drift  
## 0.4676 0.2675 -0.9897 0.0971 0.5183 -0.0799 -0.4066 0.0013  
## s.e. 0.0377 0.0380 0.0109 0.2379 0.2410 0.2523 0.2550 0.0002  
##   
## sigma^2 estimated as 0.01143: log likelihood=593.39  
## AIC=-1168.77 AICc=-1168.52 BIC=-1127.52

modelFINAL = Arima(gisHOLD, order=c(2,1,1), seasonal=c(1,0,2), include.drift = TRUE) # final  
modelFINAL

## Series: gisHOLD   
## ARIMA(2,1,1)(1,0,2)[12] with drift   
##   
## Coefficients:  
## ar1 ar2 ma1 sar1 sma1 sma2 drift  
## 0.4662 0.2663 -0.9896 0.4808 -0.4665 0.1077 0.0013  
## s.e. 0.0376 0.0381 0.0107 0.2595 0.2587 0.0422 0.0002  
##   
## sigma^2 estimated as 0.01144: log likelihood=592.85  
## AIC=-1169.69 AICc=-1169.49 BIC=-1133.03

Ljung-Box test

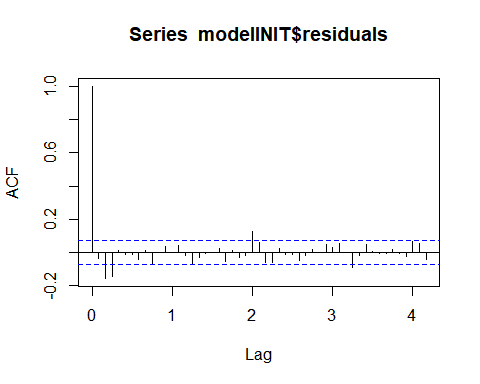
Box.test(modelFINAL$residuals,30,type="Ljung-Box")

##   
## Box-Ljung test  
##   
## data: modelFINAL$residuals  
## X-squared = 25.027, df = 30, p-value = 0.7237

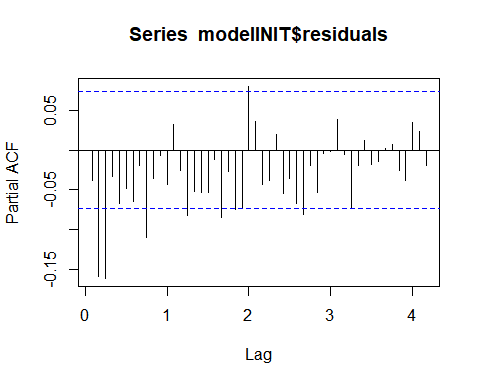
tells us …….

acf and pacf graph

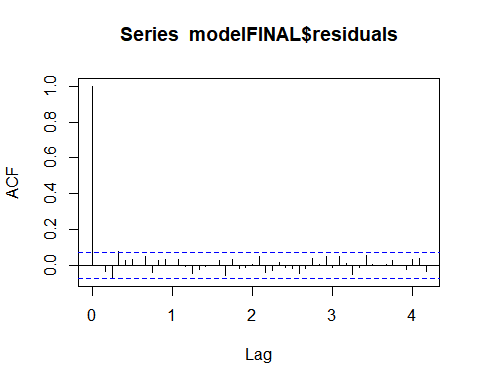
#Initial model  
acf(modelINIT$residuals,50)



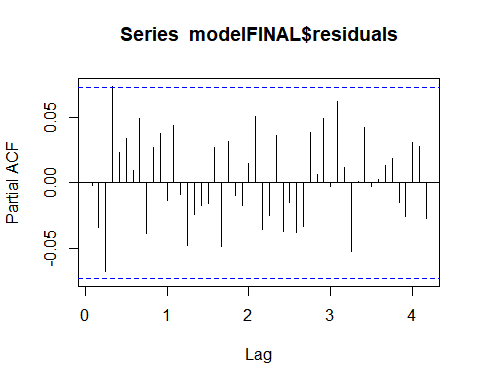
pacf(modelINIT$residuals,50)



#Final model  
acf(modelFINAL$residuals,50)

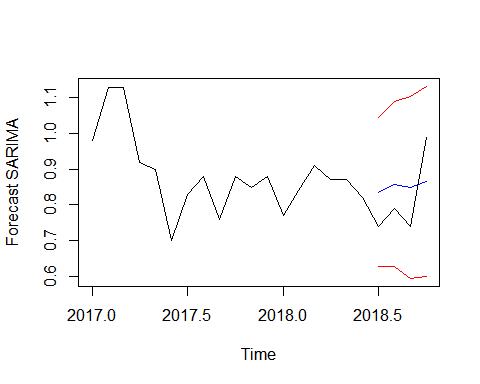


pacf(modelFINAL$residuals,50)

 pretty good.

forecasting

fore=forecast(modelFINAL, h=4)  
foreSARIMA=fore$mean  
upper = ts(fore$upper[,2], start=c(2018,7), frequency=12)  
lower = ts(fore$lower[,2], start=c(2018,7), frequency=12)  
windowDATA=window(gisFULL, start=c(2017,1))  
plot(cbind(fore$mean, upper, lower, windowDATA), plot.type="single", col=c("BLUE", "RED","RED", "BLACK"), ylab="Forecast SARIMA")



VAR gisTEMP & CO2

Unit root test for CO2

co2TEMP = read.csv("pset3.csv")  
co2FULL = ts(co2TEMP$C02, start=c(1958,3), end=c(2018,10), frequency=12)  
co2HOLD = ts(co2TEMP$C02, start=c(1958,3), end=c(2018,6), frequency=12)  
  
summary(ur.df(co2HOLD, type="trend", lags=30, selectlags = "AIC"))

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression trend   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.92161 -0.21070 -0.00959 0.20633 0.99026   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.7276605 1.1907977 0.611 0.541362   
## z.lag.1 -0.0019092 0.0039115 -0.488 0.625637   
## tt 0.0007160 0.0005117 1.399 0.162210   
## z.diff.lag1 -0.3259246 0.0383101 -8.508 < 2e-16 \*\*\*  
## z.diff.lag2 -0.2282662 0.0395648 -5.769 1.22e-08 \*\*\*  
## z.diff.lag3 -0.1927406 0.0406640 -4.740 2.62e-06 \*\*\*  
## z.diff.lag4 -0.1704475 0.0410283 -4.154 3.69e-05 \*\*\*  
## z.diff.lag5 -0.0827992 0.0410042 -2.019 0.043858 \*   
## z.diff.lag6 -0.0578561 0.0406777 -1.422 0.155407   
## z.diff.lag7 -0.0728818 0.0402361 -1.811 0.070537 .   
## z.diff.lag8 -0.0853308 0.0402647 -2.119 0.034439 \*   
## z.diff.lag9 -0.0299558 0.0401815 -0.746 0.456225   
## z.diff.lag10 -0.1225889 0.0401914 -3.050 0.002378 \*\*   
## z.diff.lag11 0.0394545 0.0403791 0.977 0.328874   
## z.diff.lag12 0.2488098 0.0404094 6.157 1.28e-09 \*\*\*  
## z.diff.lag13 0.1460304 0.0411364 3.550 0.000413 \*\*\*  
## z.diff.lag14 0.0316914 0.0405139 0.782 0.434355   
## z.diff.lag15 -0.0191960 0.0405457 -0.473 0.636054   
## z.diff.lag16 -0.0790883 0.0402530 -1.965 0.049855 \*   
## z.diff.lag17 -0.1319867 0.0403422 -3.272 0.001124 \*\*   
## z.diff.lag18 -0.0723487 0.0405364 -1.785 0.074752 .   
## z.diff.lag19 -0.1538910 0.0405226 -3.798 0.000159 \*\*\*  
## z.diff.lag20 -0.1456479 0.0408823 -3.563 0.000393 \*\*\*  
## z.diff.lag21 -0.1758914 0.0411255 -4.277 2.17e-05 \*\*\*  
## z.diff.lag22 -0.1275877 0.0411008 -3.104 0.001989 \*\*   
## z.diff.lag23 0.0146585 0.0405942 0.361 0.718140   
## z.diff.lag24 0.1858696 0.0397723 4.673 3.59e-06 \*\*\*  
## z.diff.lag25 0.1700082 0.0384427 4.422 1.14e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.32 on 665 degrees of freedom  
## Multiple R-squared: 0.9348, Adjusted R-squared: 0.9321   
## F-statistic: 352.9 on 27 and 665 DF, p-value: < 2.2e-16  
##   
##   
## Value of test-statistic is: -0.4881 11.7937 10.9755   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau3 -3.96 -3.41 -3.12  
## phi2 6.09 4.68 4.03  
## phi3 8.27 6.25 5.34

fail to reject

find model

library(vars)

## Loading required package: MASS

## Loading required package: strucchange

## Loading required package: zoo

##   
## Attaching package: 'zoo'

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

## Loading required package: sandwich

## Loading required package: lmtest

X=cbind(diff(gisHOLD), diff(co2HOLD))  
VARselect(X, lag.max=20, type="const")

## $selection  
## AIC(n) HQ(n) SC(n) FPE(n)   
## 18 14 13 18   
##   
## $criteria  
## 1 2 3 4 5  
## AIC(n) -4.61086443 -4.89977396 -5.041973594 -5.046907297 -5.063424633  
## HQ(n) -4.59583827 -4.87473036 -5.006912553 -5.001828816 -5.008328712  
## SC(n) -4.57198514 -4.83497514 -4.951255249 -4.930269425 -4.920867234  
## FPE(n) 0.00994322 0.00744827 0.006460993 0.006429204 0.006323898  
## 6 7 8 9 10  
## AIC(n) -5.244543677 -5.591863933 -5.911651467 -6.056566931 -6.139838093  
## HQ(n) -5.179430315 -5.516733131 -5.826503225 -5.961401249 -6.034654971  
## SC(n) -5.076066750 -5.397467479 -5.691335486 -5.810331423 -5.867683059  
## FPE(n) 0.005276273 0.003728121 0.002707763 0.002342491 0.002155349  
## 11 12 13 14 15  
## AIC(n) -6.159952391 -6.40705189 -6.51718862 -6.528869437 -6.537606638  
## HQ(n) -6.044751829 -6.28183389 -6.38195318 -6.383616554 -6.382336315  
## SC(n) -5.861877830 -6.08305780 -6.16727500 -6.153036294 -6.135853968  
## FPE(n) 0.002112453 0.00164998 0.00147793 0.001460793 0.001448116  
## 16 17 18 19 20  
## AIC(n) -6.53984633 -6.548634685 -6.557883519 -6.548251403 -6.542523878  
## HQ(n) -6.37455857 -6.373329481 -6.372560874 -6.352911318 -6.337166353  
## SC(n) -6.11217413 -6.095042961 -6.078372267 -6.042820624 -6.011173572  
## FPE(n) 0.00144491 0.001432305 0.001419162 0.001432945 0.001441229

BIC: lag 13 AIC: lag 18

VAR lags 13

VAR13=VAR(X, p=13, type="const")  
foreVAR13=forecast(VAR13, h=6)  
foreVAR13

## diff.gisHOLD.  
## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jul 2018 0.002717295 -0.1364465 0.1418811 -0.2101154 0.2155500  
## Aug 2018 0.010646203 -0.1449567 0.1662491 -0.2273279 0.2486204  
## Sep 2018 0.016395762 -0.1393188 0.1721103 -0.2217491 0.2545406  
## Oct 2018 0.004035923 -0.1530830 0.1611548 -0.2362567 0.2443285  
## Nov 2018 -0.008700766 -0.1661813 0.1487798 -0.2495465 0.2321450  
## Dec 2018 -0.008314879 -0.1659642 0.1493345 -0.2494187 0.2327890  
##   
## diff.co2HOLD.  
## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jul 2018 -1.8865671 -2.3237683 -1.4493660 -2.5552086 -1.2179256  
## Aug 2018 -2.1060039 -2.5522927 -1.6597152 -2.7885437 -1.4234642  
## Sep 2018 -1.5462984 -1.9947993 -1.0977975 -2.2322213 -0.8603755  
## Oct 2018 0.1347446 -0.3163572 0.5858464 -0.5551560 0.8246453  
## Nov 2018 1.4690307 1.0132756 1.9247857 0.7720134 2.1660479  
## Dec 2018 1.6794814 1.2219468 2.1370160 0.9797426 2.3792202

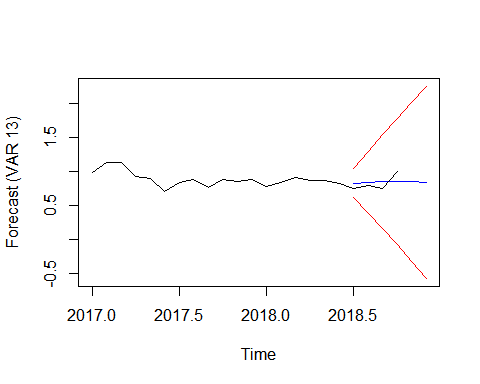
foreVAR13$forecast$diff.gisHOLD.$mean

## Jul Aug Sep Oct Nov  
## 2018 0.002717295 0.010646203 0.016395762 0.004035923 -0.008700766  
## Dec  
## 2018 -0.008314879

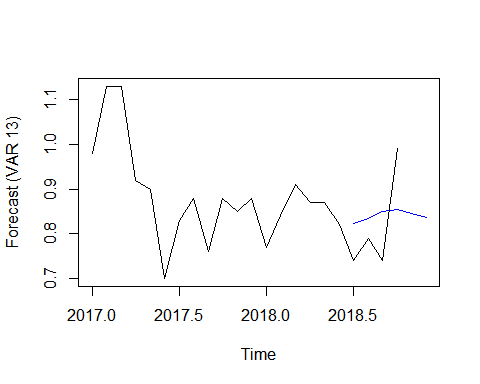
tail(gisHOLD)

## Jan Feb Mar Apr May Jun  
## 2018 0.77 0.84 0.91 0.87 0.87 0.82

tempFORE13=diffinv(foreVAR13$forecast$diff.gisHOLD.$mean, xi=0.82)  
meanVAR13=ts(tempFORE13[2:7], start=c(2018,7), frequency=12)  
tempFORE13=diffinv(foreVAR13$forecast$diff.gisHOLD.$upper[,2], xi=0.82)  
upperVAR13=ts(tempFORE13[2:7], start=c(2018,7), frequency=12)  
tempFORE13=diffinv(foreVAR13$forecast$diff.gisHOLD.$lower[,2], xi=0.82)  
lowerVAR13=ts(tempFORE13[2:7], start=c(2018,7), frequency=12)  
plot(cbind(meanVAR13, upperVAR13, lowerVAR13, windowDATA), plot.type="single", col=c("BLUE", "RED","RED", "BLACK"), ylab="Forecast (VAR 13)")



plot(cbind(meanVAR13, windowDATA), plot.type="single", col=c("BLUE", "BLACK"), ylab="Forecast (VAR 13)")



VAR lags 18

VAR18=VAR(X, p=18, type="const")  
foreVAR18=forecast(VAR18, h=6)  
foreVAR18

## diff.gisHOLD.  
## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jul 2018 0.031472997 -0.1066601 0.1696061 -0.1797833 0.2427293  
## Aug 2018 0.015995946 -0.1393340 0.1713259 -0.2215606 0.2535525  
## Sep 2018 0.013882720 -0.1414651 0.1692305 -0.2237012 0.2514667  
## Oct 2018 0.026852088 -0.1300030 0.1837072 -0.2130371 0.2667413  
## Nov 2018 -0.005381664 -0.1625490 0.1517857 -0.2457484 0.2349851  
## Dec 2018 -0.017630720 -0.1750179 0.1397565 -0.2583337 0.2230722  
##   
## diff.co2HOLD.  
## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jul 2018 -1.8240860 -2.2507575 -1.3974146 -2.4766238 -1.1715483  
## Aug 2018 -2.2418839 -2.6854568 -1.7983109 -2.9202702 -1.5634976  
## Sep 2018 -1.5866303 -2.0371984 -1.1360621 -2.2757148 -0.8975457  
## Oct 2018 0.2092663 -0.2461632 0.6646957 -0.4872530 0.9057856  
## Nov 2018 1.4878368 1.0295882 1.9460855 0.7870060 2.1886677  
## Dec 2018 1.6534141 1.1928356 2.1139926 0.9490201 2.3578081

foreVAR18$forecast$diff.gisHOLD.$mean

## Jul Aug Sep Oct Nov  
## 2018 0.031472997 0.015995946 0.013882720 0.026852088 -0.005381664  
## Dec  
## 2018 -0.017630720

tail(gisHOLD)

## Jan Feb Mar Apr May Jun  
## 2018 0.77 0.84 0.91 0.87 0.87 0.82

tempFORE18=diffinv(foreVAR18$forecast$diff.gisHOLD.$mean, xi=0.82)  
meanVAR18=ts(tempFORE18[2:7], start=c(2018,7), frequency=12)

compare SARIMA, VAR forecasting to real data

realDATA=window(gisFULL, start=c(2018,7))  
  
realDATA

## Jul Aug Sep Oct  
## 2018 0.74 0.79 0.74 0.99

foreSARIMA

## Jul Aug Sep Oct  
## 2018 0.8354674 0.8581591 0.8485555 0.8656921

meanVAR13

## Jul Aug Sep Oct Nov Dec  
## 2018 0.8227173 0.8333635 0.8497593 0.8537952 0.8450944 0.8367795

meanVAR18

## Jul Aug Sep Oct Nov Dec  
## 2018 0.8514730 0.8674689 0.8813517 0.9082038 0.9028221 0.8851914

calculate accuracy

accuSARIMA = accuracy(realDATA, foreSARIMA[1:4])  
accuVAR13 = accuracy(realDATA, meanVAR13[1:4])  
accuVAR18 = accuracy(realDATA, meanVAR18[1:4])  
  
accuSARIMA

## ME RMSE MAE MPE MAPE  
## Test set 0.0369685 0.1012379 0.09912246 4.450726 11.63041

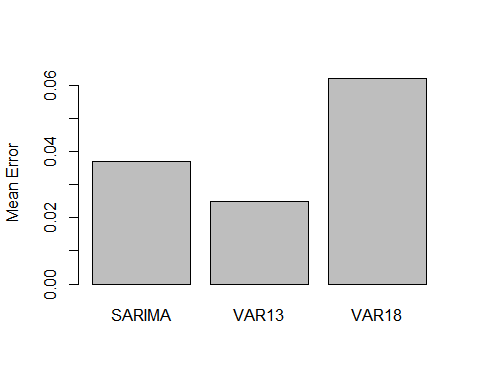
accuVAR13

## ME RMSE MAE MPE MAPE  
## Test set 0.02490881 0.09914811 0.09301122 3.055308 11.03174

accuVAR18

## ME RMSE MAE MPE MAPE  
## Test set 0.06212434 0.1061821 0.1030225 7.26348 11.76667

barplot(c(accuSARIMA[1],accuVAR13[1],accuVAR18[1]), names.arg = c("SARIMA", "VAR13", "VAR18"),ylab = "Mean Error")



head & tail of the data

head(gisFULL)

## Mar Apr May Jun Jul Aug  
## 1958 0.10 0.01 0.08 -0.05 0.06 -0.06

tail(gisFULL)

## May Jun Jul Aug Sep Oct  
## 2018 0.87 0.82 0.74 0.79 0.74 0.99

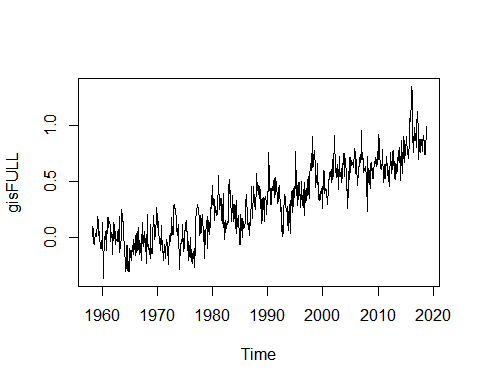
head(co2FULL)

## Mar Apr May Jun Jul Aug  
## 1958 315.71 317.45 317.50 317.10 315.86 314.93

tail(co2FULL)

## May Jun Jul Aug Sep Oct  
## 2018 411.24 410.79 408.71 406.99 405.51 406.00

plot(gisFULL)



plot(co2FULL)

