MATH_4070_Final_Project

Carlos Kelaidis

11/18/2020

```
crude.oil <- read.csv("~/Documents/Undergraduate Courses/MATH 6070:4070/Final project/dataset/CL=F.csv")
```

```
which(crude.oil$Open=="null")
```

```
##   [1]    4    9   15   21   27   32   37   43   49   50   55   61   67   73   74
##  [16]   79   85   91   97  103  108  114  120  126  132  138  144  150  156  162
##  [31]  167  173  179  185  191  196  202  208  214  220  226  232  238  244  249
##  [46]  255  261  267  273  278  284  290  296  301  307  312  318  324  330  336
##  [61]  341  346  352  357  363  369  375  381  386  392  398  404  410  416  422
##  [76]  427  433  439  445  451  457  463  468  474  480  486  492  493  497  503
##  [91]  509  515  521  527  533  539  545  550  556  562  568  574  580  586  592
## [106]  598  604  610  614  615  621  627  633  639  644  649  655  660  666  672
## [121]  678  684  689  695  701  707  713  718  724  730  736  742  748  754  760
## [136]  766  771  777  783  789  795  797  800  806  812  818  824  830  836  842
## [151]  848  853  859  865  871  877  883  889  895  901  907  913  917  918  924
## [166]  929  935  941  942  946  951  957  963  968  974  980  986  991  997 1003
## [181] 1009 1015 1021 1027 1033 1038 1044 1050 1056 1062 1068 1073 1079 1085 1091
## [196] 1097 1100 1102 1108 1114 1120 1126 1132 1138 1144 1150 1155 1161 1167 1173
## [211] 1179 1185 1191 1197 1203 1209 1215 1221 1225 1226 1232 1238 1244 1246 1249
## [226] 1254 1260 1266 1271 1277 1283 1289 1294 1300 1330 1459 1464 1470 1476 1482
## [241] 1488 1494 1500 1506 1512 1518
```

Removing all "null" rows.

```
for (i in 1:nrow(crude.oil)) {
  oil.data<-crude.oil[-which(crude.oil$Open=="null"),]
}
```

Create pre and post covid data

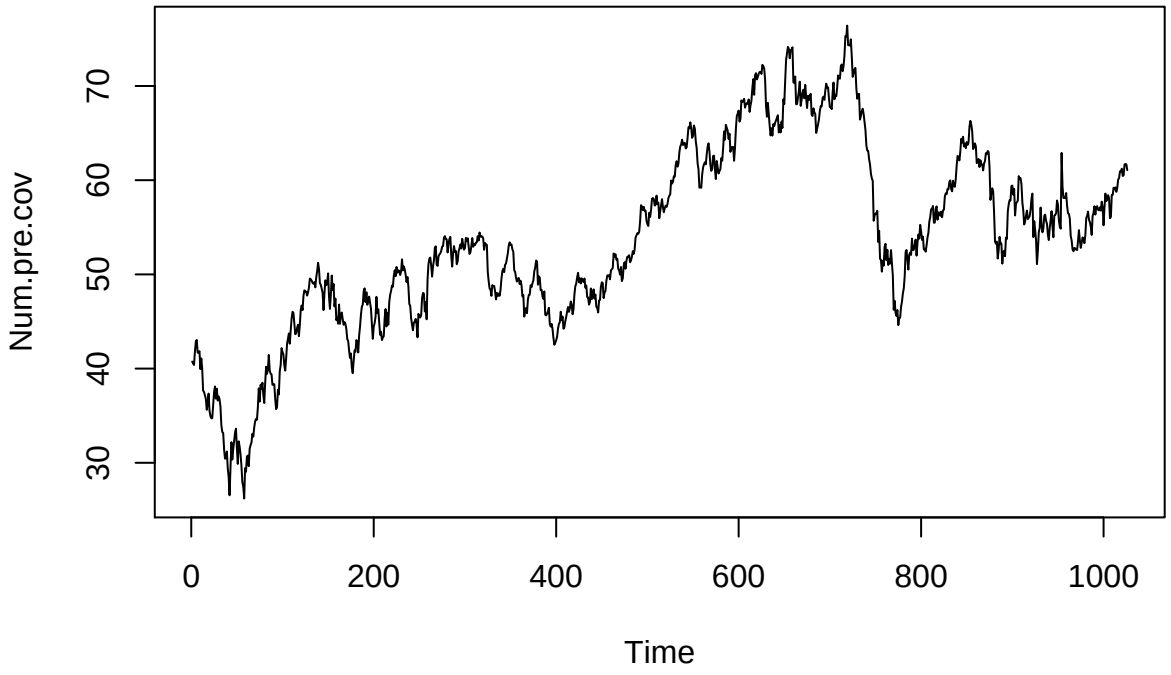
```
oil.data$Index<-cbind(1:nrow(oil.data))
pre_cov<-oil.data$Adj.Close[oil.data$Index[1:1026]]
post_cov<-oil.data$Adj.Close[oil.data$Index[1142:1274]]
```

Change data to numeric

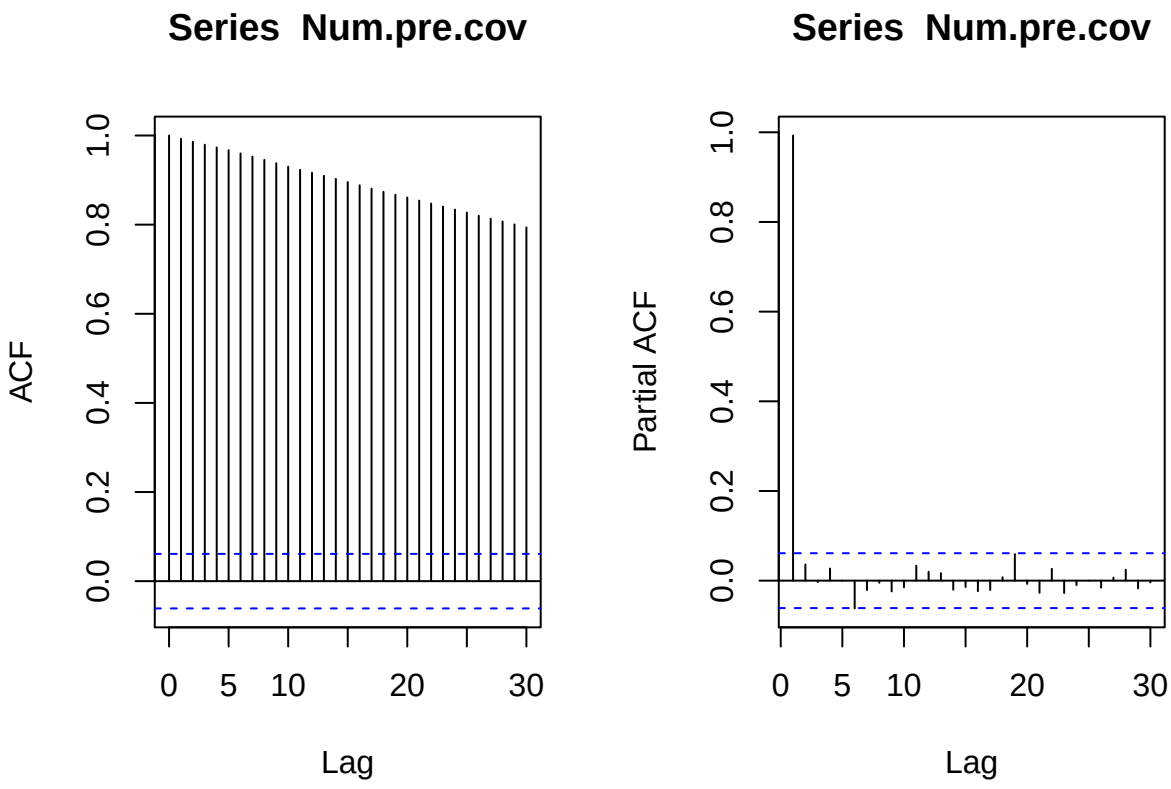
```
Num.pre.cov<-as.numeric(as.character(pre_cov))
Num.post.cov<-as.numeric(as.character(post_cov))
```

Start by analyzing *Adjusted close* prices of pre_cov

```
ts.plot(Num.pre.cov)
```



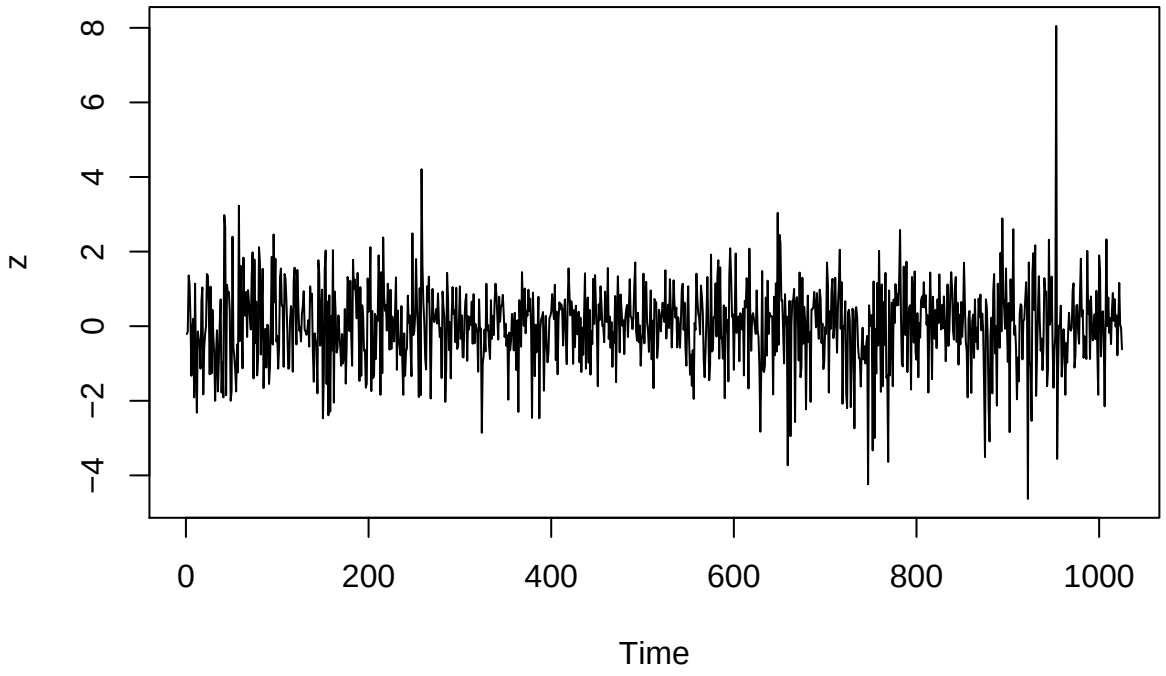

```
par(mfrow=c(1,2))
acf(Num.pre.cov)
pacf(Num.pre.cov)
```



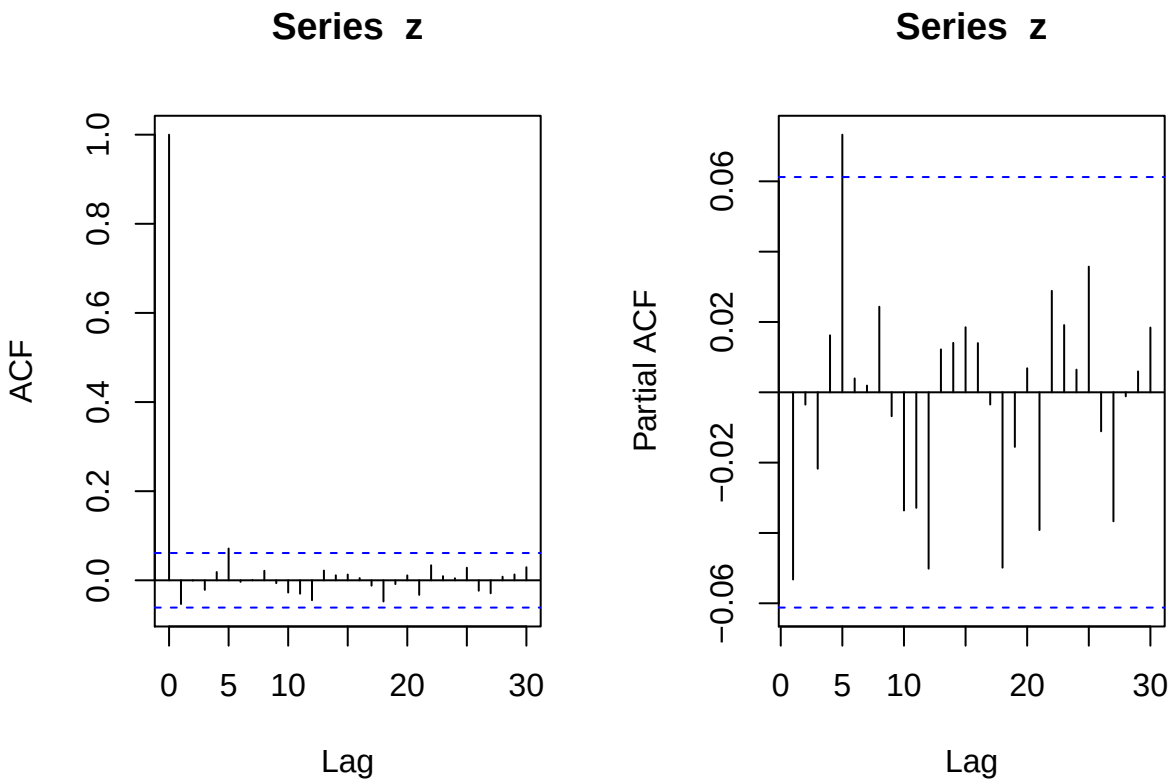
Gradual decrease in ACF indicates non-constant variance and non-stationarity.

Let's take first diff of pre_cov

```
z<-diff(Num.pre.cov)
ts.plot(z)
```




```
par(mfrow=c(1,2))
acf(z)
pacf(z)
```



After taking 1st diff, we see constant variance and no trend indicating stationarity.

Looking at PACF an AR(5) and MA(5) from ACF.

Finally try fit with one seasonal AR and MA at lag 5.

Pre_covid models:

AR(5)

```
t<-1:length(Num.pre.cov)
M<-cbind(t)
pre_ar5<-arima(Num.pre.cov, order=c(5,1,0), fixed=c(0,0,0,0,NA))
```

```
## Warning in arima(Num.pre.cov, order = c(5, 1, 0), fixed = c(0, 0, 0, 0, : some
## AR parameters were fixed: setting transform.pars = FALSE
```

```
Box.test(pre_ar5$residuals, lag=25, fitdf=5, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  pre_ar5$residuals
## X-squared = 16.309, df = 20, p-value = 0.6973
```

```
pre_ar5
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(5, 1, 0), fixed = c(0, 0, 0, 0, NA))
##
```

```
## Coefficients:
##       ar1  ar2  ar3  ar4     ar5
##         0    0    0    0  0.0715
## s.e.    0    0    0    0  0.0311
##
## sigma^2 estimated as 1.174:  log likelihood = -1536.8,  aic = 3077.59
```

```
BIC(pre_ar5)
```

```
## [1] 3087.456
```

AR(1) with seasonal AR at lag 5

```
pre_ar1sar5<-arima(Num.pre.cov, order=c(1,1,0),seasonal=list(order=c(1,0,0),period=5))
Box.test(pre_ar1sar5$residuals, lag=25, fitdf=2, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  pre_ar1sar5$residuals
## X-squared = 13.806, df = 23, p-value = 0.9323
```

```
pre_ar1sar5
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(1, 1, 0), seasonal = list(order = c(1, 0, 0),
##     period = 5))
##
## Coefficients:
##           ar1    sar1
##       -0.0545  0.0728
## s.e.   0.0312  0.0311
##
## sigma^2 estimated as 1.171:  log likelihood = -1535.27,  aic = 3076.54
```

```
BIC(pre_ar1sar5)
```

```
## [1] 3091.334
```

seasonal ma at lag 4

```
pre_sma4<-arima(Num.pre.cov,order=c(0,1,0),seasonal=list(order=c(0,0,1),period=4))
pre_sma4
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 0), seasonal = list(order = c(0, 0, 1),
##     period = 4))
##
## Coefficients:
##         sma1
##       0.0179
## s.e.  0.0305
##
## sigma^2 estimated as 1.18:  log likelihood = -1539.25,  aic = 3082.5
```

seasonal ar at lag 5

```
pre_sar5<-arima(Num.pre.cov, order=c(0,1,0),seasonal=list(order=c(1,0,0),period=5))
Box.test(pre_sar5$residuals, lag=25, fitdf=1, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  pre_sar5$residuals
## X-squared = 16.309, df = 24, p-value = 0.8766
```

```
pre_sar5
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 0), seasonal = list(order = c(1, 0, 0),
##     period = 5))
##
## Coefficients:
##         sar1
##       0.0715
## s.e.  0.0311
##
## sigma^2 estimated as 1.174:  log likelihood = -1536.8,  aic = 3077.59
```

```
BIC(pre_sar5)
```

```
## [1] 3087.456
```

seasonal ma at lag 5

```
pre_sma5<-arima(Num.pre.cov,order=c(0,1,0),seasonal=list(order=c(0,0,1),period=5))
Box.test(pre_sma5$residuals, lag=25, fitdf=1, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  pre_sma5$residuals
## X-squared = 16.088, df = 24, p-value = 0.8849
```

```
pre_sma5
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 0), seasonal = list(order = c(0, 0, 1),
##     period = 5))
##
## Coefficients:
##         sma1
##       0.0763
## s.e.  0.0321
##
## sigma^2 estimated as 1.174:  log likelihood = -1536.62,  aic = 3077.24
```

```
BIC(pre_sma5)
```

```
## [1] 3087.11
```

MA(5)

```
pre_ma5<-arima(Num.pre.cov, order=c(0,1,5), fixed = c(0,0,0,0,NA))
Box.test(pre_ma5$residuals, lag=25, fitdf=5, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  pre_ma5$residuals
## X-squared = 16.088, df = 20, p-value = 0.7112
```

```
pre_ma5
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 5), fixed = c(0, 0, 0, 0, NA))
##
## Coefficients:
##       ma1  ma2  ma3  ma4     ma5
##         0    0    0    0  0.0763
## s.e.    0    0    0    0  0.0321
##
## sigma^2 estimated as 1.174:  log likelihood = -1536.62,  aic = 3077.24
```

```
BIC(pre_ma5)
```

```
## [1] 3087.11
```

seasonal MA and AR at lag 5

```
pre_sarma5<-arima(Num.pre.cov, order=c(0,1,0), seasonal=list(order=c(1,0,1), period=5))
```

```
## Warning in arima(Num.pre.cov, order = c(0, 1, 0), seasonal = list(order = c(1, :
## possible convergence problem: optim gave code = 1
```

```
Box.test(pre_sarma5$residuals, lag=25, fitdf=2, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  pre_sarma5$residuals
## X-squared = 17.289, df = 23, p-value = 0.7948
```

```
pre_sarma5
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 0), seasonal = list(order = c(1, 0, 1),
##     period = 5))
##
## Coefficients:
##          sar1    sma1
##       -0.9455  0.9754
## s.e.   0.0732  0.0547
##
## sigma^2 estimated as 1.171:  log likelihood = -1535.51,  aic = 3077.02
```

```
BIC(pre_sarma5)
```

```
## [1] 3091.814
```

Try AR(8) with fixed 0

```
pre_ar8<-arima(Num.pre.cov, order=c(8,1,0),xreg=M,fixed=c(0,0,0,0,NA,0,0,0,0))
```

```
## Warning in arima(Num.pre.cov, order = c(8, 1, 0), xreg = M, fixed = c(0, : some
## AR parameters were fixed: setting transform.pars = FALSE
```

```
pre_ar8
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(8, 1, 0), xreg = M, fixed = c(0, 0, 0, 0, NA,
##     0, 0, 0, 0))
##
## Coefficients:
##       ar1  ar2  ar3  ar4     ar5  ar6  ar7  ar8  t
##         0    0    0    0  0.0715    0    0    0  0
## s.e.    0    0    0    0  0.0311    0    0    0  0
##
## sigma^2 estimated as 1.174:  log likelihood = -1536.8,  aic = 3077.59
```

```
BIC(pre_ar8)
```

```
## [1] 3087.456
```

seaosnal MA at lag 4

```
pre_sma4<-arima(Num.pre.cov,order=c(0,1,0),seasonal=list(order=c(0,0,1),period=4))
pre_sma4
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 0), seasonal = list(order = c(0, 0, 1),
##     period = 4))
##
## Coefficients:
##         sma1
##       0.0179
## s.e.  0.0305
##
## sigma^2 estimated as 1.18:  log likelihood = -1539.25,  aic = 3082.5
```

seasonal MA at lag 5 WE PICK THIS ONE!

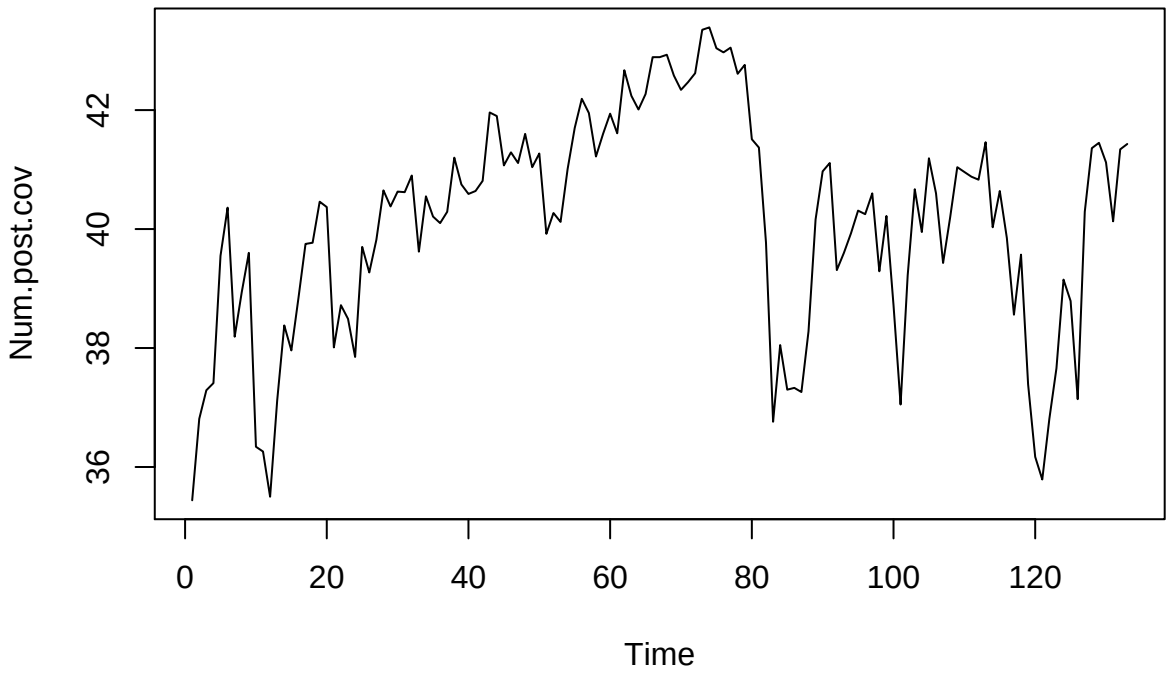
```
pre_sma5
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 0), seasonal = list(order = c(0, 0, 1),
```

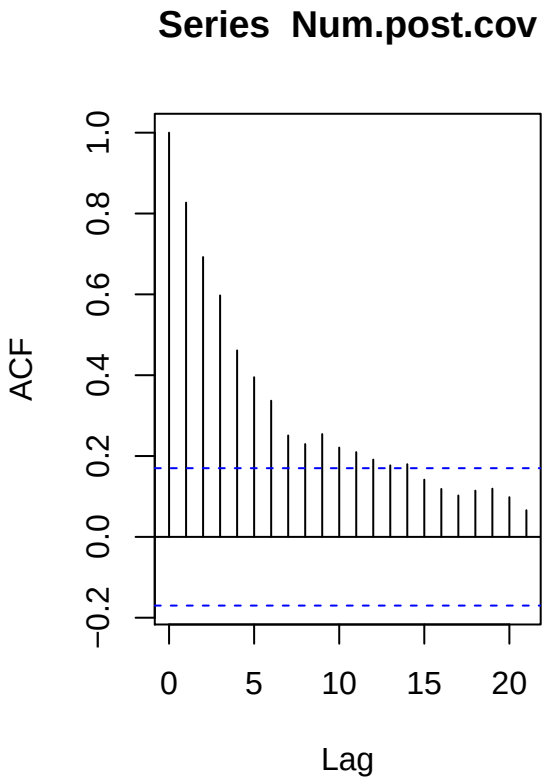
```
##     period = 5))
##
## Coefficients:
##         sma1
##       0.0763
## s.e.  0.0321
##
## sigma^2 estimated as 1.174:  log likelihood = -1536.62,  aic = 3077.24
```

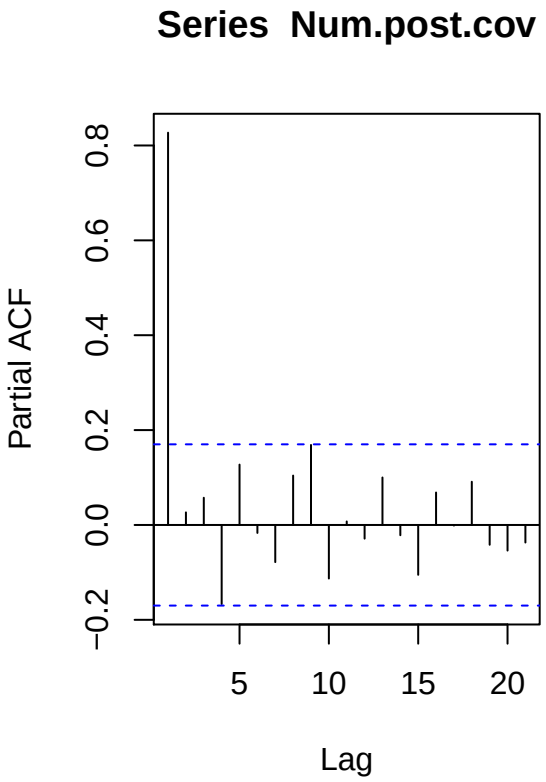
Post covid data

```
ts.plot(Num.post.cov)
```



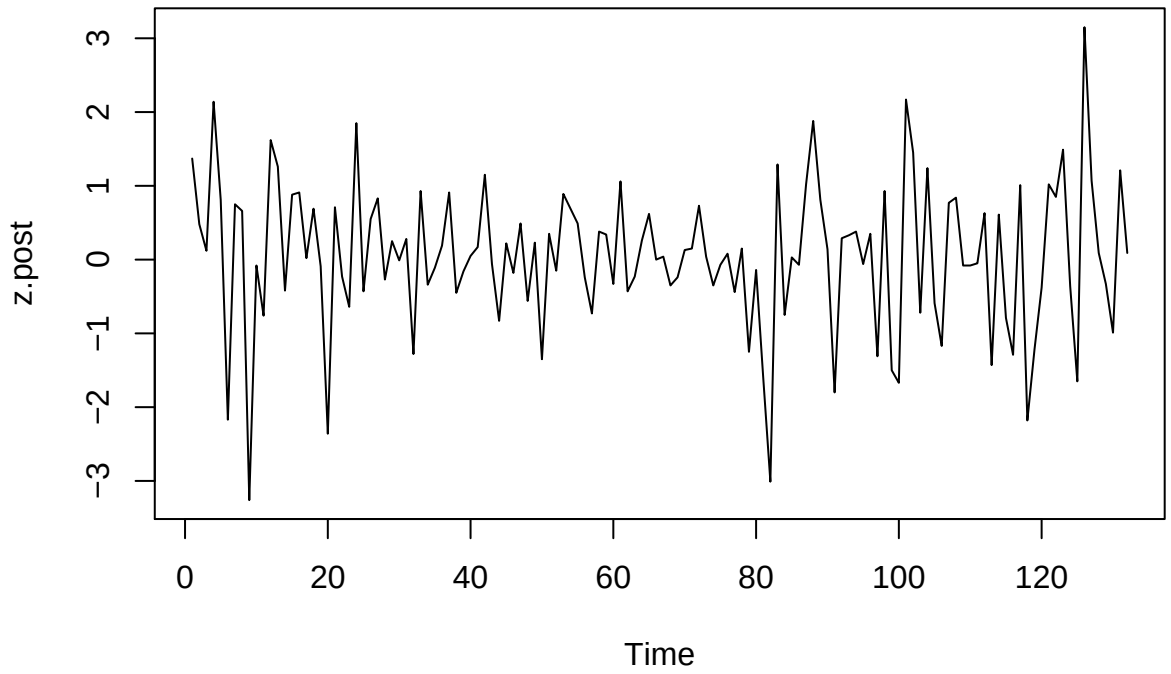

```
par(mfrow=c(1,2))
acf(Num.post.cov)
pacf(Num.post.cov)
```



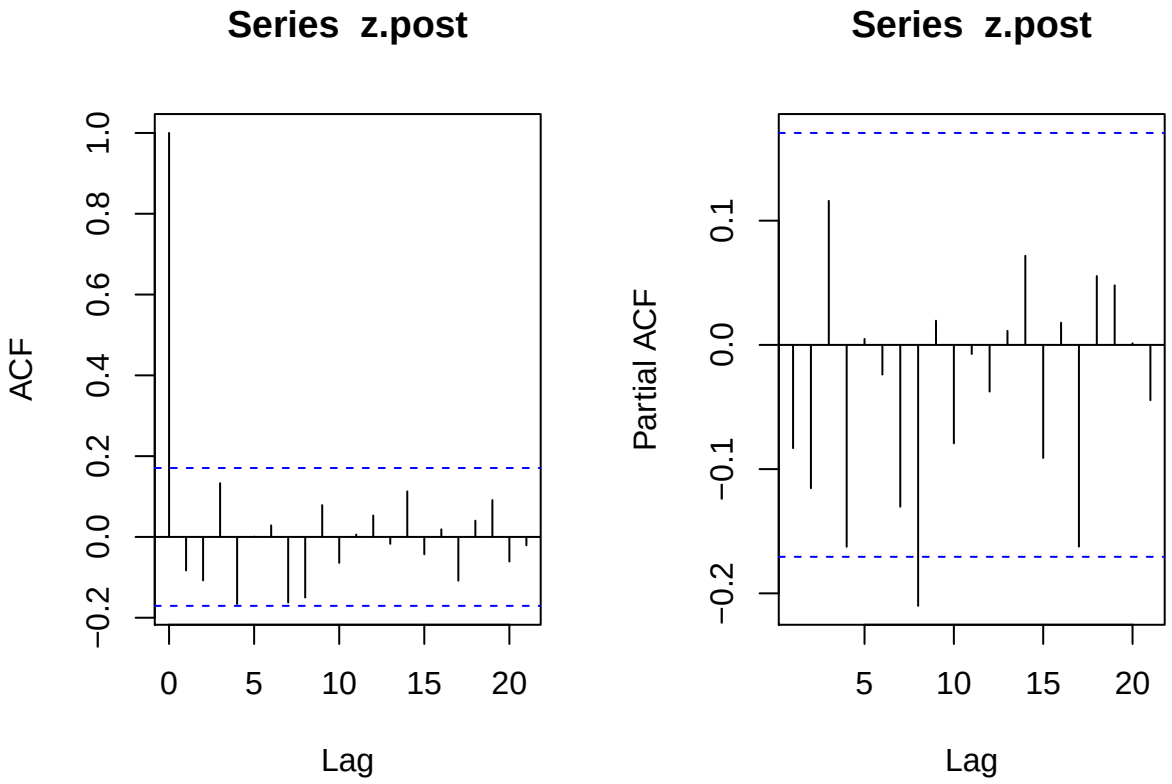


Take post 1st diff

```
z.post<-diff(Num.post.cov)
ts.plot(z.post)
```



```
par(mfrow=c(1,2))
acf(z.post)
pacf(z.post)
```



Try AR(8) WE PICK THIS ONE!

```
X<-cbind(1:length(Num.post.cov))
post_ar8<-arima(Num.post.cov,order=c(8,1,0))
Box.test(post_ar8$residuals, lag=25, fitdf=8, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  post_ar8$residuals
## X-squared = 9.6291, df = 17, p-value = 0.9184
```

```
post_ar8
```

```
##
## Call:
## arima(x = Num.post.cov, order = c(8, 1, 0))
##
## Coefficients:
##           ar1      ar2     ar3      ar4      ar5      ar6      ar7      ar8
##       -0.1042  -0.1369  0.0684  -0.1986  -0.0038  -0.0543  -0.1749  -0.2503
## s.e.   0.0850   0.0844  0.0850   0.0870   0.0862   0.0873   0.0903   0.0919
##
## sigma^2 estimated as 0.8699:  log likelihood = -178.55,  aic = 375.09
```

```
BIC(post_ar8)
```

```
## [1] 401.0393
```

Try seasonal MA at lag 4 !!!

```
post_sma4<-arima(Num.post.cov,order=c(0,1,0),seasonal=list(order=c(0,0,1),period=4))
post_sma4
```

```
##
## Call:
## arima(x = Num.post.cov, order = c(0, 1, 0), seasonal = list(order = c(0, 0,
##     1), period = 4))
##
## Coefficients:
##          sma1
##       -0.2606
## s.e.   0.1042
##
## sigma^2 estimated as 0.9607:  log likelihood = -184.79,  aic = 373.59
```

```
BIC(post_sma4)
```

```
## [1] 379.3532
```

For post we could use either MA or AR.

We wish to predict the next 24 values using the pre-cov model and plot:

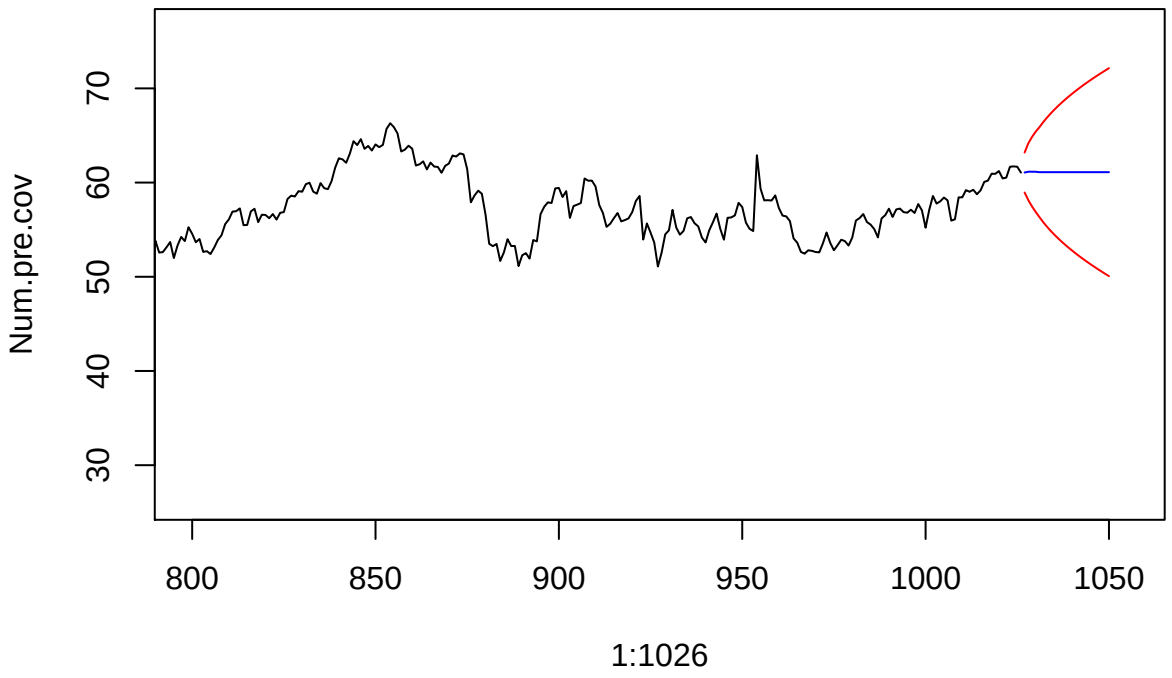
```
pre_sma5 # Model used to predict
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 0), seasonal = list(order = c(0, 0, 1),
##     period = 5))
##
## Coefficients:
##         sma1
##       0.0763
## s.e.  0.0321
##
## sigma^2 estimated as 1.174:  log likelihood = -1536.62,  aic = 3077.24
```

```
newX<-1027:1038
pre_cov.fore.24<-predict(pre_sma5, n.ahead=24)
U<-pre_cov.fore.24$pred+1.96*pre_cov.fore.24$se
L<-pre_cov.fore.24$pred-1.96*pre_cov.fore.24$se

#par(mfrow=c(1,1))
plot(1:1026, Num.pre.cov,type="l", xlim=c(800,1055))+
  lines(1027:1050, pre_cov.fore.24$pred, col="blue")+
  lines(1027:1050, U, col="red")+
```

```
  lines(1027:1050, L, col="red")
```




```
## integer(0)
```

Still predicting pre-covid data but this time using the MA(5)

```
pre_ma5 # Model used
```

```
##
## Call:
## arima(x = Num.pre.cov, order = c(0, 1, 5), fixed = c(0, 0, 0, 0, NA))
##
## Coefficients:
##       ma1  ma2  ma3  ma4     ma5
##         0    0    0    0  0.0763
## s.e.    0    0    0    0  0.0321
##
## sigma^2 estimated as 1.174:  log likelihood = -1536.62,  aic = 3077.24
```

Let's predict our post_covid data using the post_covid model:

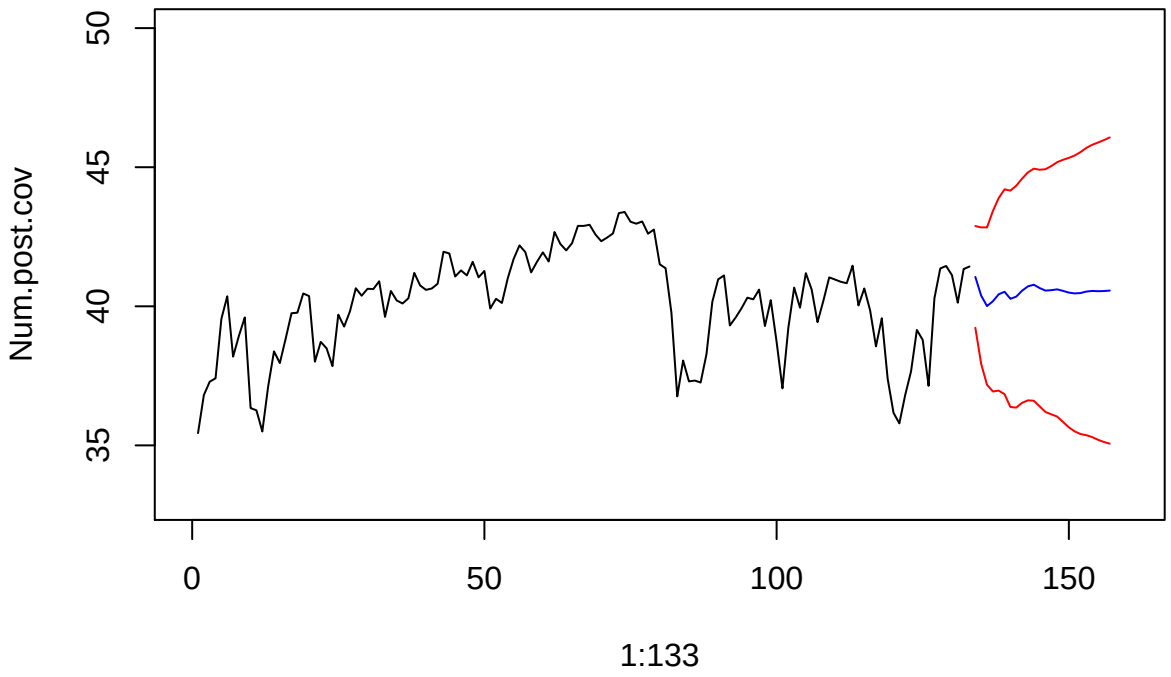
```
post_ar8 # Model used
```

```
##
## Call:
## arima(x = Num.post.cov, order = c(8, 1, 0))
##
## Coefficients:
##           ar1      ar2     ar3      ar4      ar5      ar6      ar7      ar8
##       -0.1042  -0.1369  0.0684  -0.1986  -0.0038  -0.0543  -0.1749  -0.2503
## s.e.   0.0850   0.0844  0.0850   0.0870   0.0862   0.0873   0.0903   0.0919
```

```
##
## sigma^2 estimated as 0.8699:  log likelihood = -178.55,  aic = 375.09
```

```
post_cov.fore.24<-predict(post_ar8, n.ahead=24)
post.U<-post_cov.fore.24$pred+1.96*post_cov.fore.24$se
post.L<-post_cov.fore.24$pred-1.96*post_cov.fore.24$se

plot(1:133, Num.post.cov, type="l", ylim=c(33, 50), xlim=c(0,160))+
  lines(134:157, post_cov.fore.24$pred, col="blue")+
  lines(134:157, post.U, col="red")+
  lines(134:157, post.L, col="red")
```




```
## integer(0)
```

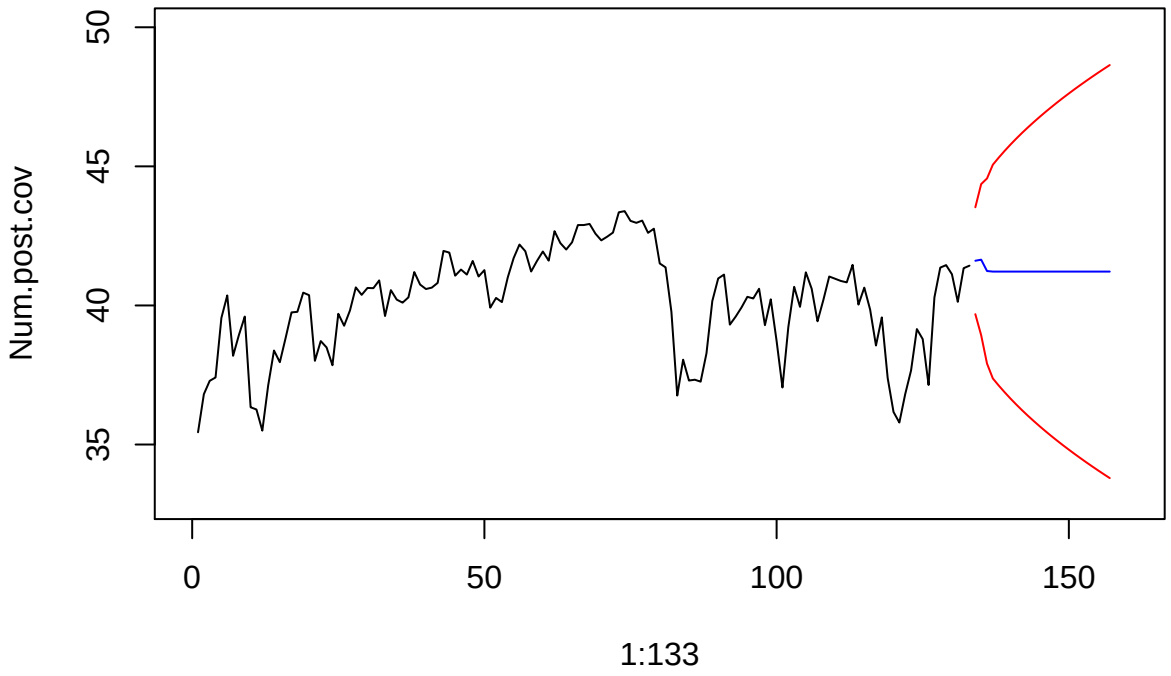
Predciting pos_covid using seasonal ma at lag 4:

```
post_sma4 # Model used
```

```
##
## Call:
## arima(x = Num.post.cov, order = c(0, 1, 0), seasonal = list(order = c(0, 0,
##     1), period = 4))
##
## Coefficients:
##          sma1
##       -0.2606
## s.e.   0.1042
##
## sigma^2 estimated as 0.9607:  log likelihood = -184.79,  aic = 373.59
```

```
post_cov.fore<-predict(post_sma4, n.ahead=24)
post_cov.U<-post_cov.fore$pred+1.96*post_cov.fore$se
post_cov.L<-post_cov.fore$pred-1.96*post_cov.fore$se
```

```
plot(1:133, Num.post.cov, type="l", xlim=c(0,160), ylim=c(33,50))+
  lines(134:157, post_cov.fore$pred, col="blue")+
  lines(134:157, post_cov.U, col="red")+
  lines(134:157, post_cov.L, col="red")
```




```
## integer(0)
```