Hw 10

Colin White

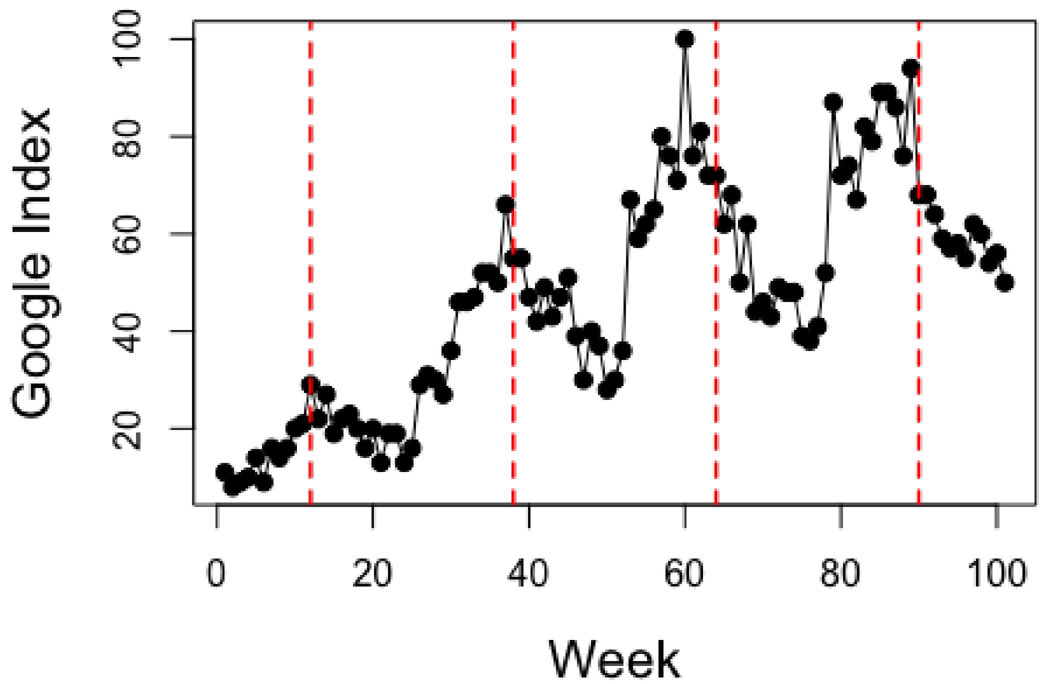
4/14/2021

```
library(forecast)

## Warning: package 'forecast' was built under R version 4.0.2

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

studioC <- read.table("~/Desktop/1A School/1A Winter 2021/STAT330/HW10/Studio
C.csv", sep =",",header = TRUE)
#attach(StudioC)
```

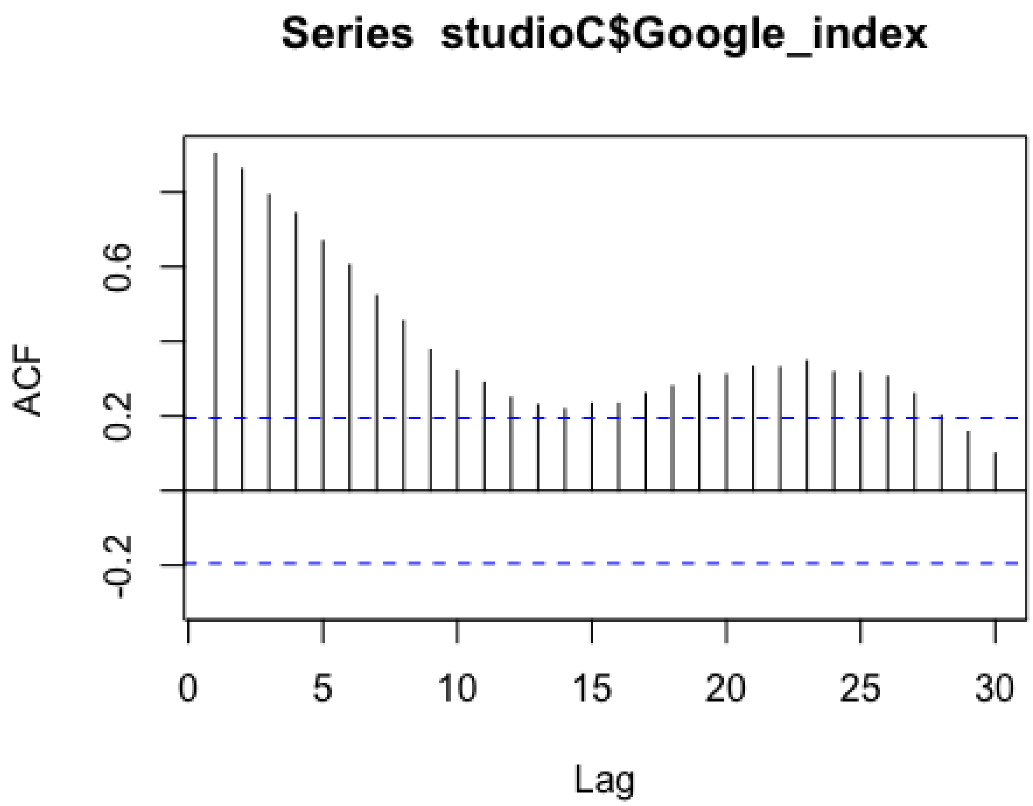
#1 *We want to use past measures of popularity (Google index) to predict future popularity of studio C.* We will use time series models becasue we beleive that we can leverage temporal (time) autocorelation to make better predictions.

#2 * There is a strong autocorrelation, and it seems to have a seasonality *Multiple regression methods that assume independance are likeluy not appropriate because there is autocorrelaiton.

```
plot(studioC$Week, studioC$Google_index, type = "o", xlab="Week", ylab="Googl
e Index", cex.lab = 1.4, pch = 19)
abline(v = (1:4)*26 - 14, col = "red", lwd = 1.5, lty = 2)
```




```
Acf(studioC$Google_index,lag.max = 30)
```



#3 * Below, we fit all the models * We use AIC becasue we are using the model for prediction and AIC is derived with prediciotn as the goal.

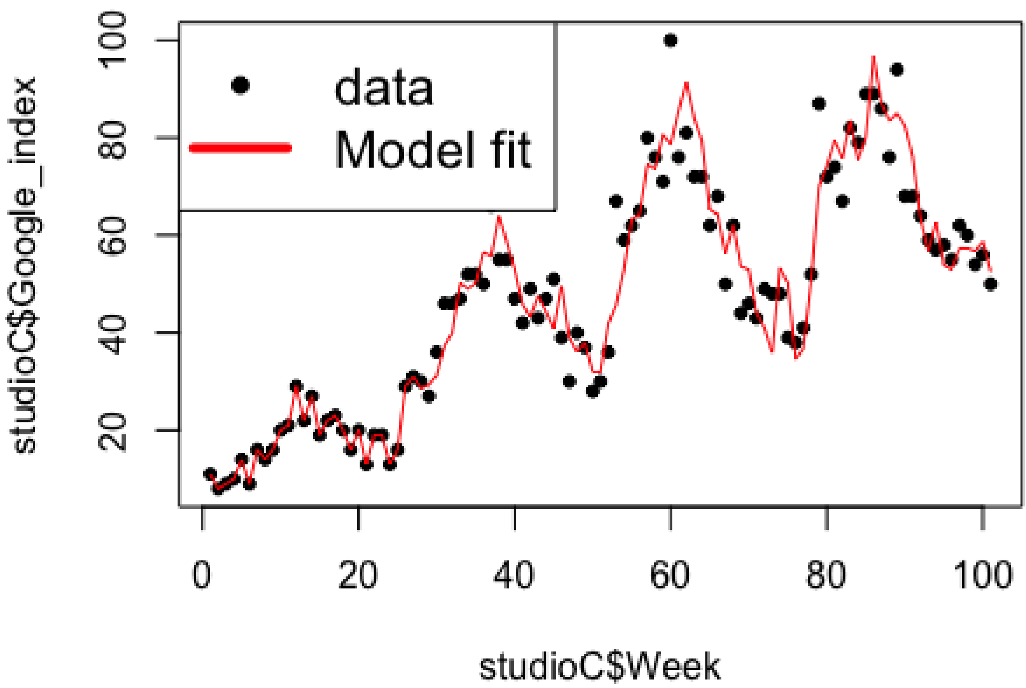
```
Google_ts = ts(studioC$Google_index, frequency = 26)

ar1 = forecast::Arima(studioC$Google_index, order = c(1,0,0))
ma1 = forecast::Arima(studioC$Google_index, order = c(0,0,1))
arima111 = forecast::Arima(studioC$Google_index, order = c (1,1,1))




sarima111 = forecast::Arima(studioC$Google_index, order = c (1,1,1), seasonal
= list(order = c (1,1,1), period = 26))
```

- AR(1) AIC = 745.6181621
- MA(1) AIC = 854.7395218
- ARMA(1,1) AIC = 727.4900944
- SARIMA$(1,1,1)\times(1,1,1)_{26}$ AIC = 531.6329

#4

The fit looks good.

```r
plot(studioC$Week, studioC$Google_index, pch = 20)
lines(studioC$Week, sarima111$fitted, col = "red")
legend("topleft", c("data", "Model fit"), col = c("black", "red"), lwd = c(NA
, 4), pch = c(20,NA), cex = 1.4)
```

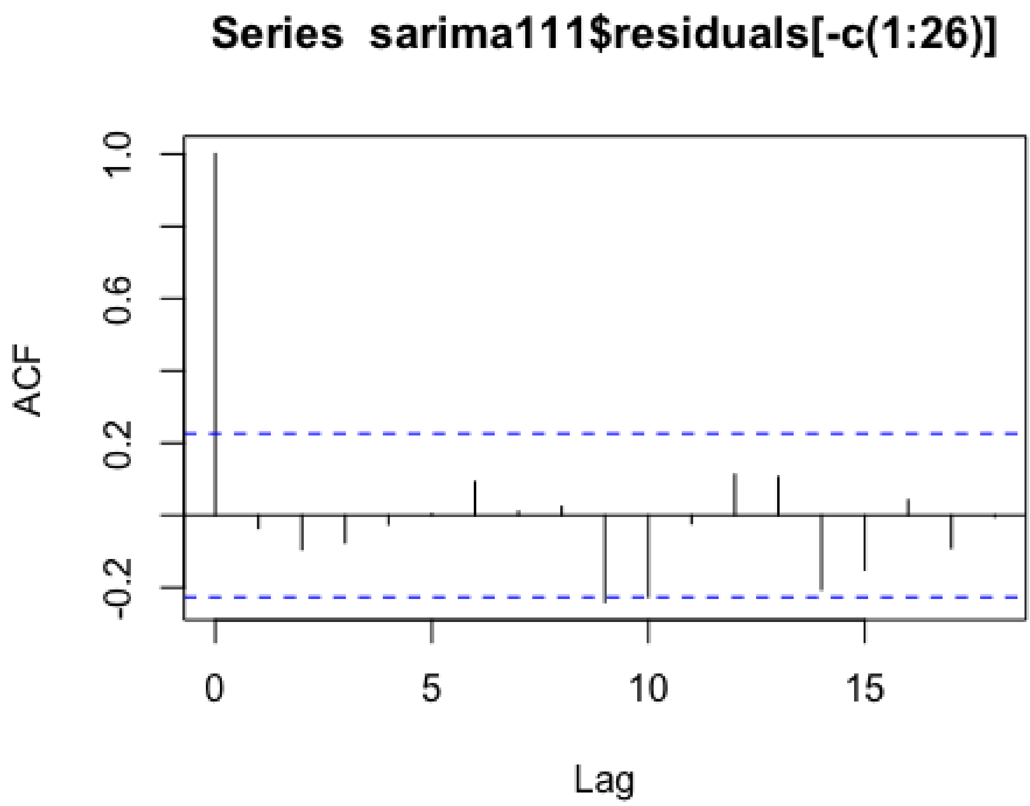


#5

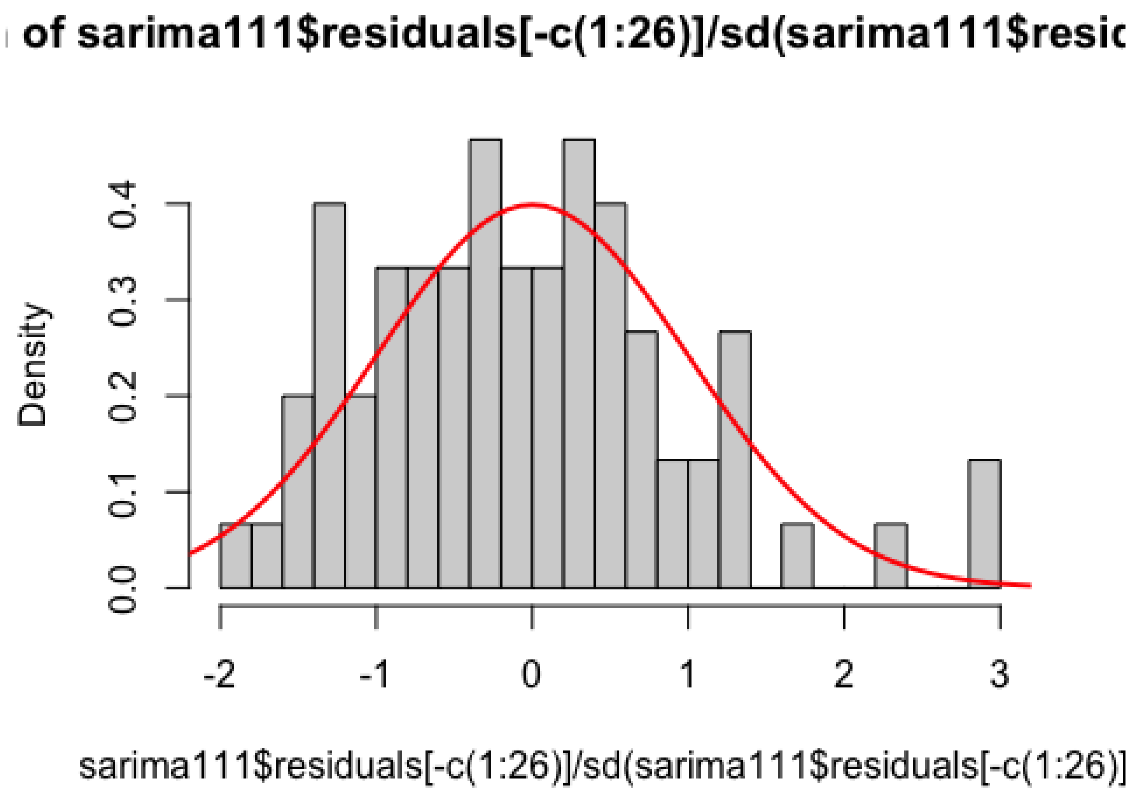
- l - we dont have to worry about linearity becasue we dont have any quantitaitve (or any) covariates.
- I - Independence - after time series terms are included we assume that ϵ_t are all independant. This is justified by the ACF plot, even though there are a couple ACF peaks just outside the significance bounds.
- N - Normality - we argue that this looks ok based on the histogram of residuals
- E - Equal variance - we thing that the fitted vs residuals look oh, so I argue that the equal variabce assumption is met.

Becasue of Seasonal lags, the model isnt fully initialized untill after the first season.

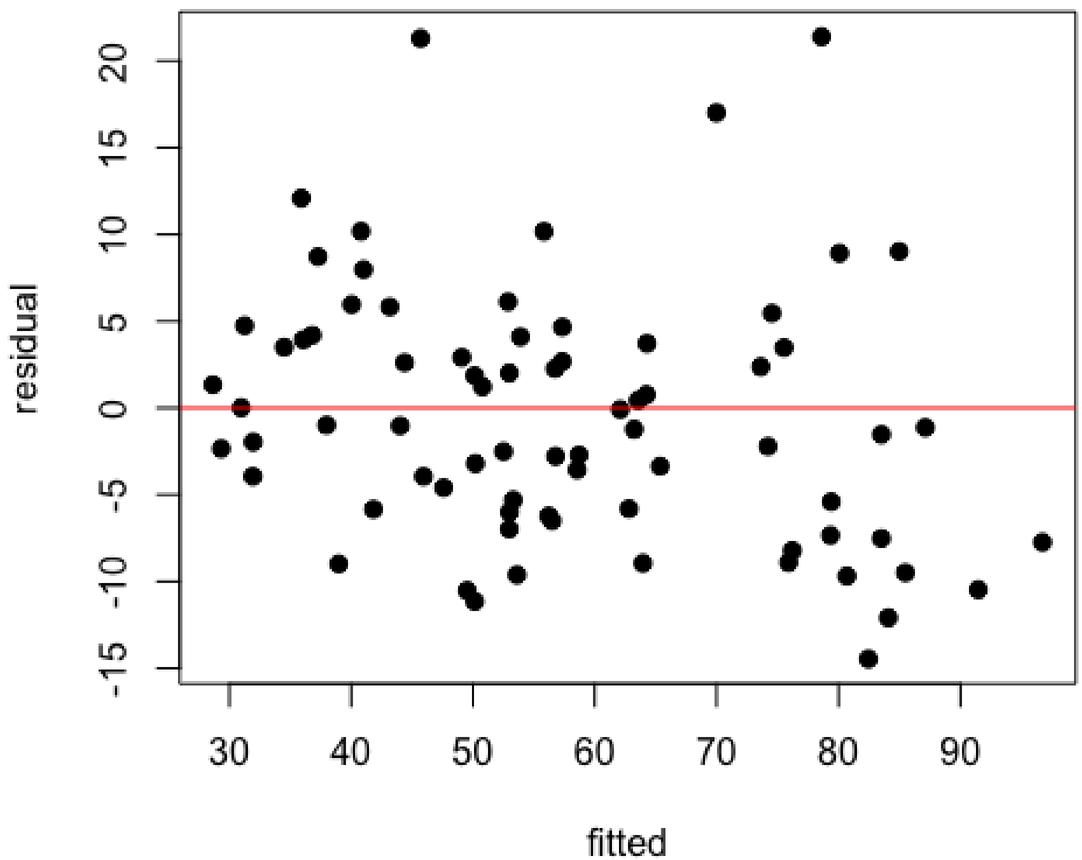
```r
acf(sarima111$residuals[-c(1:26)])
```




```
#hist
hist(sarima111$residuals[-c(1:26)] / sd(sarima111$residuals[-c(1:26)]), break
s = 20, freq = FALSE)
curve(dnorm(x), from = -4, to =4, add = TRUE, col = "red", lwd = 2)
```

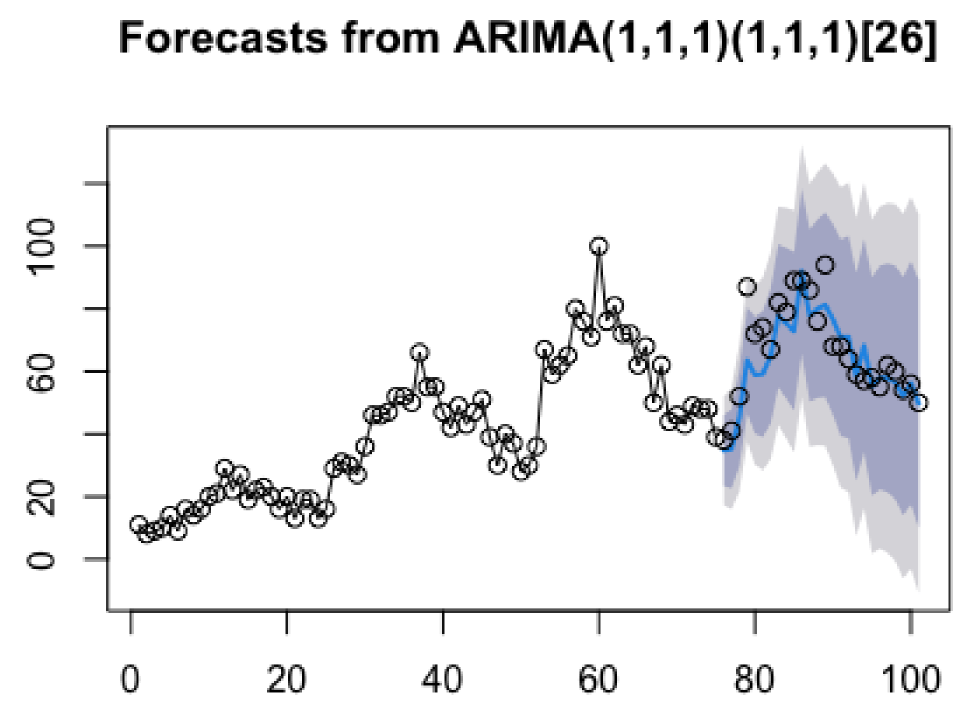



```r
#fitted vs r

par(mar = c(4,4,1,1))
plot(c(sarima111$fitted[-(1:26)]),c(sarima111$residuals[-(1:26)]), xlab = "fi
tted", ylab ="residual", pch = 19)

abline(h = 0, col = "red")
```



#6

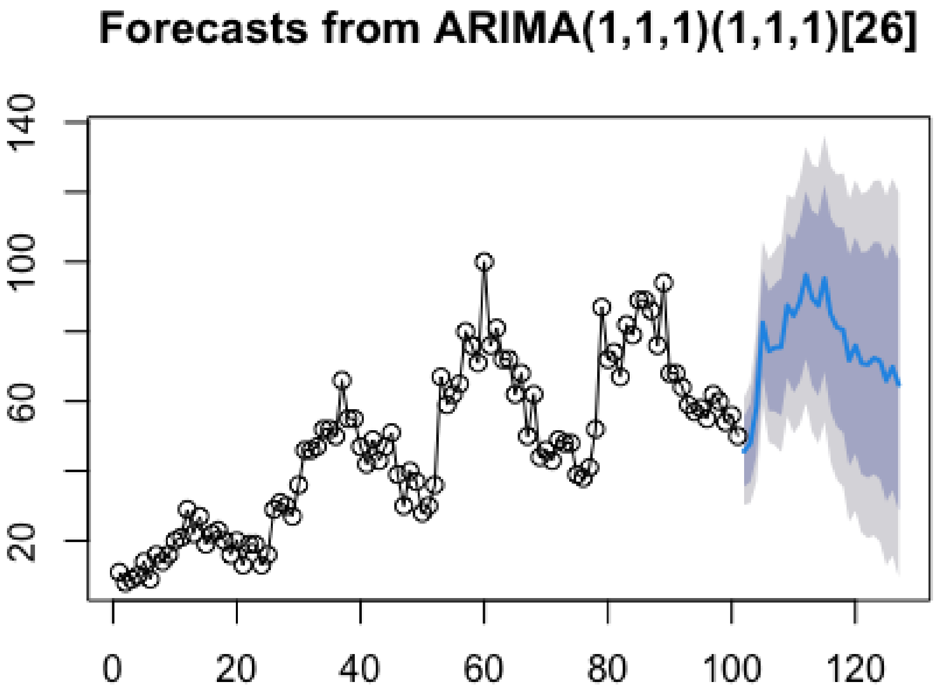
```
test.set = studioC$Google_index[76:length(studioC$Google_index)]
train.set = studioC$Google_index[-c(76:length(studioC$Google_index))]


sarima_train = Arima(train.set, order = c(1,1,1), season = list(order = c(1,1
,1), period = 26))

sarima_test_pred = forecast(sarima_train, h = 26)

plot(sarima_test_pred)
points(studioC$Week, studioC$Google_index)
```




```
bias <- mean(sarima_test_pred$mean - test.set)
bias

## [1] -3.37408

rpmse <- sqrt(mean((test.set-sarima_test_pred$mean)^2))
rpmse

## [1] 8.527953

diff(range(studioC$Google_index))

## [1] 92

sd(studioC$Google_index)

## [1] 23.03799
```

#7

```
season_five = forecast(sarima111, h = 26)
plot(season_five)
points(studioC$Week, studioC$Google_index)
```


Forecasts from ARIMA(1,1,1)(1,1,1)[26]
20
60
100
140
0
20
40
60
80
100
120