Quiz 10

Nicholas Colonna

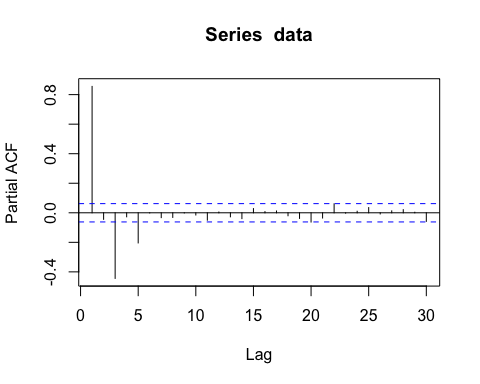
4/18/2018

library("stats")  
library("fBasics")

## Loading required package: timeDate

## Loading required package: timeSeries

sample\_data <- read.csv("data4Quiz10.csv", header=T)  
data <- sample\_data$x  
  
**#1. Use PACF to determine the order of AR model**.  
s2 <- pacf(data)



s2$acf

## , , 1  
##   
## [,1]  
## [1,] 0.8556810816  
## [2,] -0.0443614305  
## [3,] -0.4442316569  
## [4,] -0.0272523416  
## [5,] -0.2041017499  
## [6,] -0.0010467982  
## [7,] -0.0328661584  
## [8,] -0.0313604036  
## [9,] -0.0009317692  
## [10,] -0.0149577148  
## [11,] -0.0499551580  
## [12,] 0.0050273953  
## [13,] -0.0273898314  
## [14,] -0.0395844864  
## [15,] 0.0283575550  
## [16,] 0.0078725928  
## [17,] 0.0123286965  
## [18,] -0.0200926436  
## [19,] -0.0379855534  
## [20,] -0.0626654592  
## [21,] -0.0354469796  
## [22,] 0.0608222106  
## [23,] -0.0024594000  
## [24,] 0.0103759364  
## [25,] 0.0351455941  
## [26,] -0.0077949298  
## [27,] 0.0130659175  
## [28,] 0.0211395975  
## [29,] 0.0036269476  
## [30,] -0.0581316621

#Discovered that order=5 since the acf is significant at that lag  
  
**#2. Use ar() to build a model and name it as AR1**AR1 <- ar(data, method="mle")  
AR1

##   
## Call:  
## ar(x = data, method = "mle")  
##   
## Coefficients:  
## 1 2 3 4 5   
## 0.8546 0.2904 -0.3523 0.1419 -0.2014   
##   
## Order selected 5 sigma^2 estimated as 1.019

AR1$order #Reconfirming that order=5

## [1] 5

**#3. Use the order your select from question 1, build another model AR2 using arima()**AR2 <- arima(data, order=c(5,0,0))  
AR2

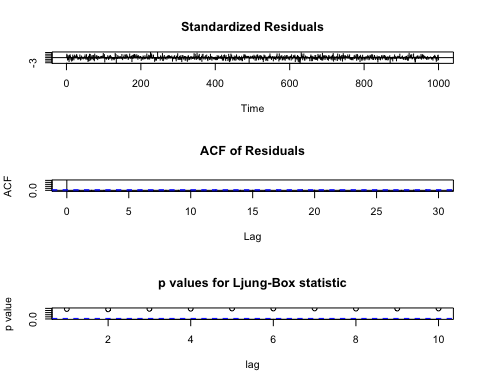
##   
## Call:  
## arima(x = data, order = c(5, 0, 0))  
##   
## Coefficients:  
## ar1 ar2 ar3 ar4 ar5 intercept  
## 0.8546 0.2904 -0.3523 0.1419 -0.2014 0.0925  
## s.e. 0.0310 0.0411 0.0405 0.0411 0.0312 0.1197  
##   
## sigma^2 estimated as 1.019: log likelihood = -1429.66, aic = 2873.33

**#4. Obtain the AIC values from AR1, what's your observation?**  
AR1\_aic <- AR1$aic  
AR1\_aic

## 0 1 2 3 4 5   
## 1593.450301 266.187487 266.083918 37.874201 39.215469 0.000000   
## 6 7 8 9 10 11   
## 1.992968 3.022340 4.042964 6.042435 7.919770 7.507738   
## 12   
## 9.473264

# The values shown show the differences in AIC between each model and the best-fitting model. Therefore, the order that has aic=0 is the "best" model that the rest are compared to. Through my observation of these values, you can see that when the order is 5, we get an aic value of 5, which matches the value we selected for order in the questions above. You will also notice that the aic values are very large for orders 0 through 4

**#5. Use tsdiag() on AR2, interpret your results.**  
tsdiag(AR2)



# The results of tsdiag() on AR2 show the following:

# First, the standardized residuals chart is comparing our observed frequency to the expected frequency. As you can see, most of the time it stays between -2 and 2, which means it is right around the expected frequency. However, there are times when it goes below -2, which means the observed frequency is less than the expected frequency, and above 2, which means the observed frequency is greater than the expected frequency. Since a majority of our data falls between -2 and 2, you can say the data appears to be normally distributed.

# Second, the ACF plot is showing the correlation between the various lags. Since our model is order 5, none of the values should be significant. Lag 19 looks very close to exceeding the threshold to be significant, however, it is not.

# Finally, we have the p-values for the Ljung-Box statistic, which has a null hypothesis that all the autocorrelations equal 0. Since for all lags it appears that the values are significant, we can determine that the values are not random and independent over time.