2017

PSTAT 174 Time Series Final Project

BY: ANREW DELOS SANTOS PERM: 4830675
PROFESSOR: DR. RAYA FELDMAN
TA: NHAN HUYNH M: 7:00-7:50 PM

Abstract

My time series final project is based on monthly robberies that occurred in Boston during the years of 1966 to 1975 and being able to forecast the number of robberies from June to October. The main questions I wanted to address from this data set that would help me forecast were:

- Can I create a time series model that could accurately model this behavior?
- Can my model forecast a close estimate of the amount of robberies that would occur in the next five months?

Therefore, in order to conduct this model, I implemented the necessary steps to build my time series model. In order to come up with this time series model, I obtained the data set and removed the last five months to see if my model was valid. I then observed three different transformations which were, Box Cox, log, and square root. After observing the trend and non-constant variance features on the graph of the Box-Coxed data, I differenced it. After, I came up with a model of ARIMA(0,1,1) and tested the residuals for normality by conducting Box test, Mcleod test, and Shapiro test. I obtained positive results from these tests and as a result I was able to forecast the next five months. The key result I obtained was that my model indeed followed an ARIMA(0,1,1) for both the transformed data and the original data because the forecasted values were within the 95% confidence interval along with the actual data.

Introduction

The focus of this project was to forecast how many robberies would occur in 1975 from June to October in Boston. This particular data set is important because robberies are always present in the world. One particular example is a personal experience of mine that happened last year here in Santa Barbara when someone broke into my house. Therefore, this data set on robberies is of the utmost importance for me to analyze because I would like to predict how many people may get robbed, and as a result, be able to enforce more law enforcement to be on patrol during those months. In order to solve the problem of being able to forecast the robberies during June to October, there were two main questions I wanted to address. Based on knowing how many robberies occurred each month, could I create a model and could I forecast the robberies that would specifically happen within the next five months? The data set of the robberies I obtained for my project is from Data Market and the data itself had monthly data from 1966 to 1975 with 118 values. In order to address this problem, I used R statistical software and the techniques of transformations, such as Box Cox. I then differenced, and built the model from the differenced data by analyzing the acf and pacf. From there, I ran AICc tests to see which model yielded the lowest AICc value and would be my best model. I then analyzed the residuals, and forecasted the next five months from June to October. The results were positive due to the fact that I was able to predict the values in those months and they were all contained in the 95% confidence interval with 95% confidence that the true value was in it. The model in both the original data and transformed data that yielded these results was an ARIMA(0,1,1).

Time Series Plot

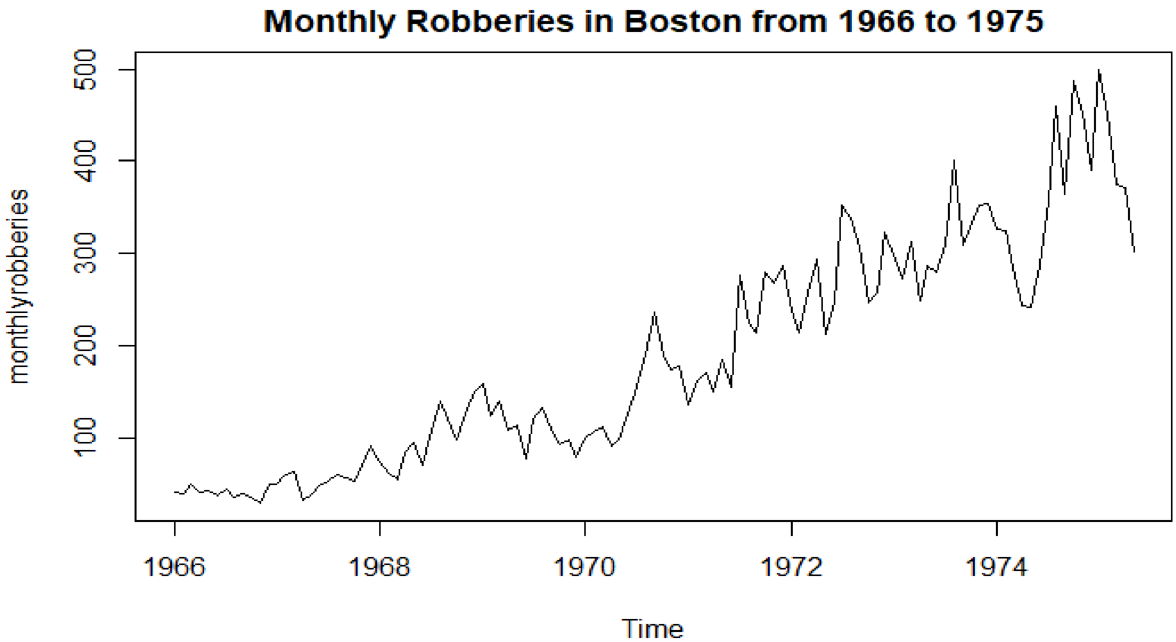


Figure 1

From this particular time series, one can see that there is an increasing trend involved of the monthly robberies throughout the plot as the years increase. Also there is non-constant variance throughout this particular plot due to the difference in heights throughout the years. However, even though there is an increasing trend and non-constant variance present, there is no apparent seasonality and there are some changes but not extremely sharp changes.

Transformation

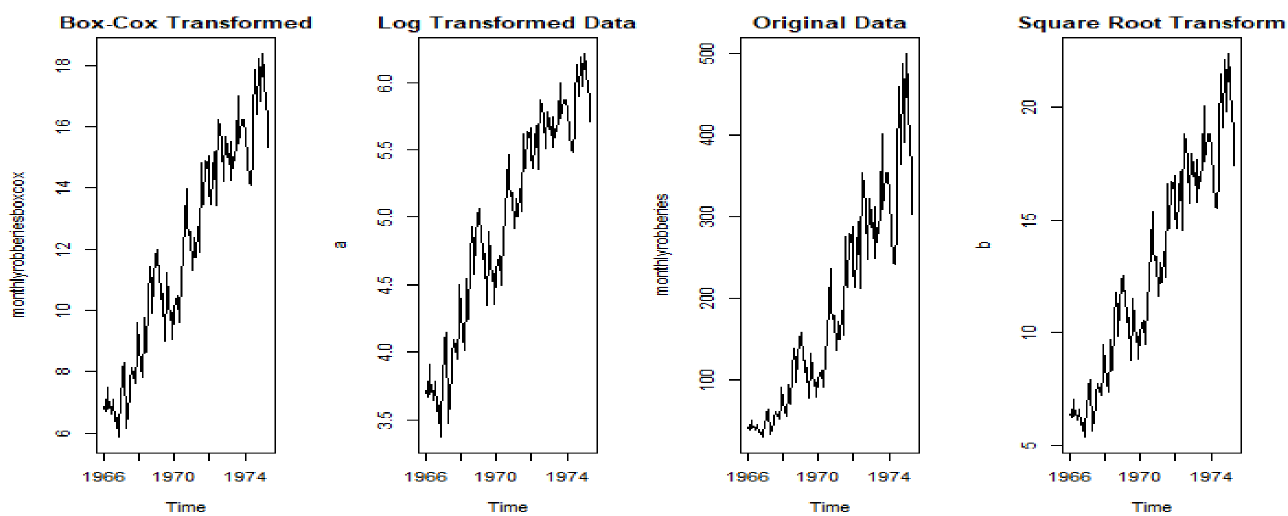


Figure 2

As seen from figure 2, I plotted three transformations in order to transform the data. The reason for implementing a transformation is because based on my original time series data, the series itself did not appear stationary and the variance was very high. Therefore, to stabilize the variance and make the data more normal, transformations are necessary. I plotted a log transformation, square root transformation, and a Box-Cox transformation alongside the original plot. In order to distinguish which transformation is the best, one needs to analyze the transformations and see which transformation created a more stabilized variance. From the graphs above, they all appear to yield the same characteristics of an upward trend and a low variance after implementing the transformation. Therefore, the transformation I chose was a Box-Cox transformation instead of a log transformation and a square root transformation.

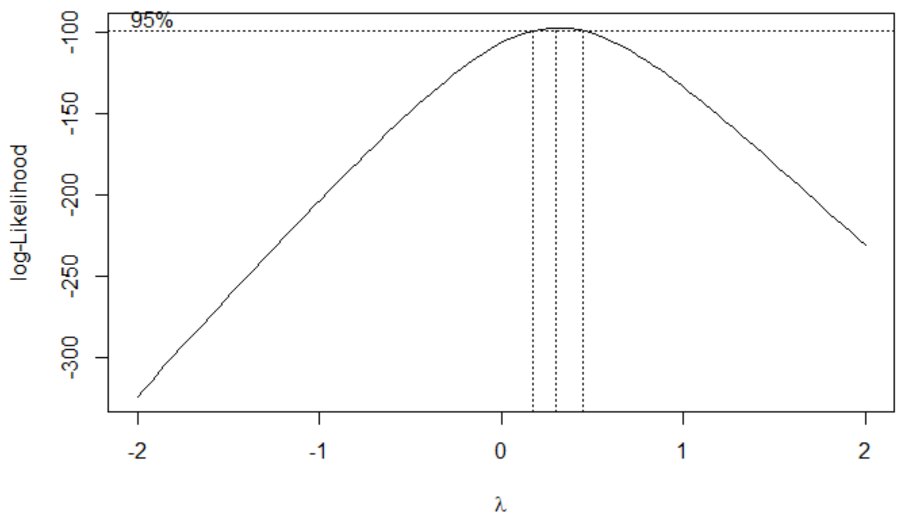


Figure 3

From figure 3, it justifies that Box-Cox should be used for this particular transformation because since zero is not within the confidence interval, the transformation is simply not a log transformation. Also, I did not choose a square root transformation because since Box-Cox yields the maximum likelihood for lambda, which was equivalent to .303030, it is justified that a Box-Cox transformation is a good transformation to implement instead of a square root which would have lambda at .50.

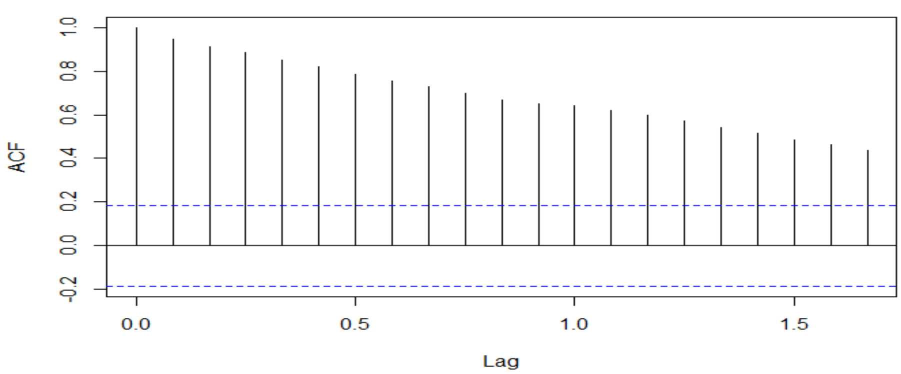


Figure 4

Now, after implementing this specific transformation, I analyzed the acf and pacf of the transformed Box-Cox data because I saw that differencing was necessary in order to remove the trend. From figure 4, one can see that there is a significant lag present at 1 in the acf and has a decreasing trend aftewards. In order to get rid of the trend, I differenced at lag 1 which should decrease the variance overall.

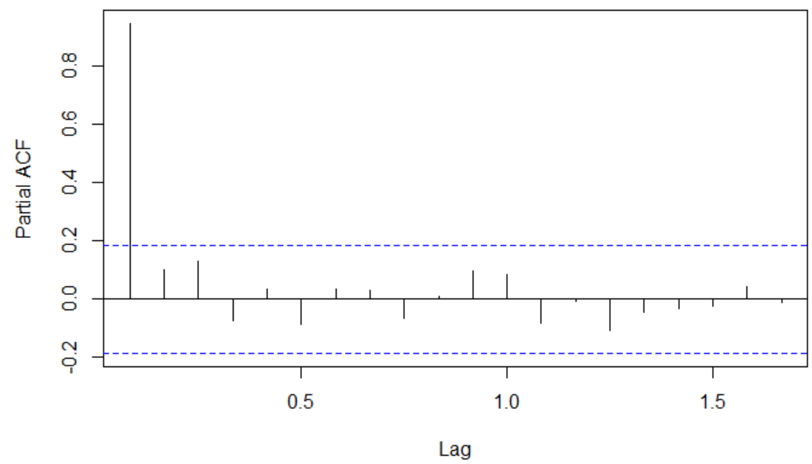


Figure 5

From figure 5, the pacf does not seem to distinguish any significant lags. After differencing, the new acf plot improved compared to the original acf plot of the undifferenced transformed data which is presented in figure 6. Also the variance decreaed from the box cox variance of 11.84745 to 0.8944784 for the box coxed data differenced at lag 1. Therefore, I analyzed the acf and pacf of the differenced box coxed data at lag 1. Both figure 6 and 7 exhibit the acf and pacf cutting off which suggests and AR model or an MA model.

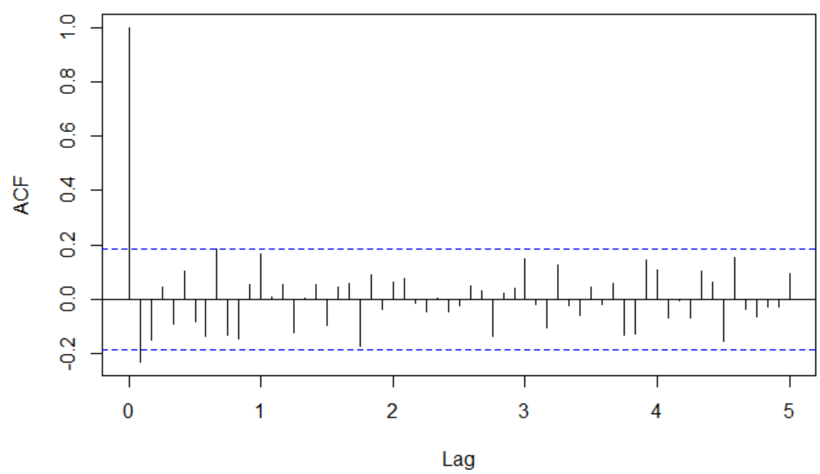


Figure 6

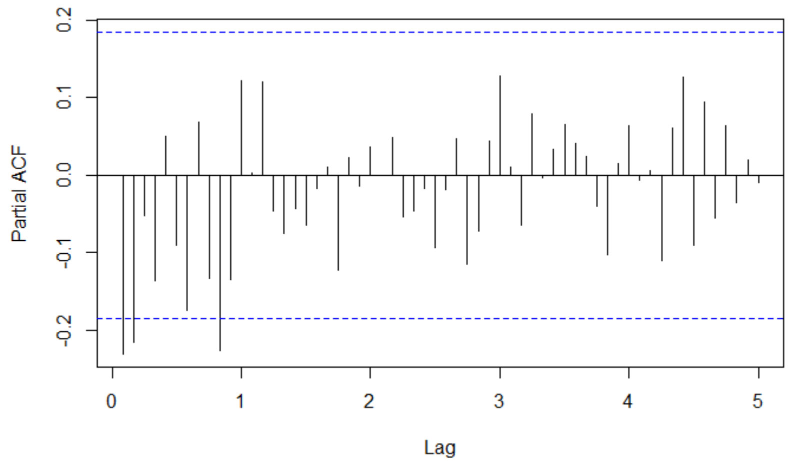


Figure 7

After differencing the transformed data at lag 1, I used the Augmented Dickey Fuller test to check if my differenced data was stationary. The null hypothesis is that the data is not stationary. At a significance level of .05, since my p-value was .01, I rejected the null hypothesis that the differenced box-coxed data at lag 1 is not stationary. The test can be viewed in figure 8 and therefore, the differenced data is stationary.

```
p-value smaller than printed p-value
	Augmented Dickey-Fuller Test

data:  diffmonthlyrobberiesboxcox1
Dickey-Fuller = -5.2908, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Figure 8

Fitting the model

From the acf and pacf of the differenced time series in figure 6 and 7 respectively, we can see that the lag cuts off after 1 for the acf and for the pacf cuts off after 10 because these two are where the last significant lags are present. Thus, to estimate p and q, I look at these particular lags and determine that p=10 and q=1. I then estimate model based on p and q and predict that it is an AR(10) or MA(1) model. Furthermore, I ran the ar function by using method of Yule-Walker to determine my coefficients for my proposed AR(10) function. As a result of the function output, the result was that of an AR(2) model shown in figure 9 on the next page. As we can see, the coefficients for this model were given as -.2805 and -.2154. Then I tried using the ar function again with the exception that this time I used the method of maximum likelihood. In figure 10 on the next page as well, one can see that the model came out to be AR(11). Based on the rule of parsimony, we choose to use less coefficients so for this particular model an AR(2) model via method of Yule-Walker estimate is the best since the function ar already chooses the model with lowest AIC value.

```
Call:
```

```
ar(x = diffmonthlyrobberiesboxcox1, aic = TRUE, order.max = NULL,     method
= "yule-walker")

Coefficients:
      1        2
-0.2805  -0.2154

Order selected 2  sigma^2 estimated as  0.8224
```

Figure 9

```
Call:
ar(x = diffmonthlyrobberiesboxcox1, aic = TRUE, order.max = NULL,     method
= "mle")

Coefficients:
      1        2        3        4        5        6        7        8
9       10       11
-0.3881  -0.3422  -0.2467  -0.2905  -0.1367  -0.2479  -0.3243  -0.1181  -
0.2925  -0.3140  -0.1432

Order selected 11  sigma^2 estimated as  0.6445
```

Figure 10

Also, besides using the ar function to estimate AR models via yule-walker and maximum likelihood method, I used the innovations algorithm as well to calculate the coefficient for an MA(1) model. The result is shown in Figure 11. We can see also that by checking the zeros of the MA(1) function via innovation algorithm, the model is invertible since the root lies outside of the unit circle.

```
source("innovations.r")
> acvf<-acf(diffmonthlyrobberiesboxcox1,plot=FALSE,lag.max=length(diffmonthly
robberiesboxcox1))$acf[ ,1,1]*var(diffmonthlyrobberiesboxcox1)
> k<-length(acvf)
> k1<-innovations.algorithm(k+1,acvf)
> k1$thetas[1,1:1]
[1] -0.230757
> polyroot(c(1,-0.2292035))
[1] 4.362935+0i
```

Figure 11

Besides using the ar function and innovation algorithm to choose the coefficients, I used the arima function to help estimate the coefficients for the MA(1), AR(2), and AR(11) models from the Yule-Walker method and maximum likelihood method since it estimated an AR(2) and an AR(11). So, from my previous prediction of the model being either AR(10) and MA(1), my prediction was not too far off due to the fact that the yule walker method and the maximum likelihood method for ar suggested an AR(10) model or an AR(2). Afterwards, I used the arima function with respect to the transformed data by analyzing an AR(11), MA(1) and an AR(2). I then calculated the AICc values and obtained the results found in figure 12.

```
library(qpcR)
```

```
> i1<-arima(monthlyrobberiesboxcox,order=c(11,1,0),method="ML",xreg=1:length(
monthlyrobberiesboxcox))
> i2<-arima(monthlyrobberiesboxcox,order=c(2,1,0),method="ML",xreg=1:length(m
onthlyrobberiesboxcox))
> i3<-arima(monthlyrobberiesboxcox,order=c(0,1,1),method="ML",xreg=1:length(m
onthlyrobberiesboxcox))
> AICc(i1)
[1] 299.6589
> AICc(i2)
[1] 300.9642
> AICc(i3)
[1] 300.2168
>
```

Figure 12

From the results in figure 12, it shows that the ARIMA(11,1,0) has the lowest AICc value, the ARIMA(2,1,0) has the third lowest value, and the MA(0,1,1) has the second lowest value. Even though the ARIMA(11,1,0) model has a lower AICc value compared to both the ARIMA(2,1,0) and ARIMA(0,1,1), based on the principle of parsimony, and since the AICc values are not too different, I chose the ARIMA(0,1,1). The value for the ARIMA(11,1,0) has an AICc value of 299.6589 and the value for ARIMA(0,0,1) has 300.2168. Therefore, since there are less coefficients (parsimony) and the AICc value is not too far apart, choose the ARIMA(0,1,1) model. As a result, the model obtained by using AICc yielded the same result as what the differenced acf suggest of an ARIMA(0,1,1). The fitted model is shown in figure 13.

```
Call:
arima(x = monthlyrobberiesboxcox, order = c(0, 1, 1), xreg = 1:length(monthly
robberiesboxcox),
    method = "ML")

Coefficients:
          ma1  1:length(monthlyrobberiesboxcox)
      -0.3854                            0.0811
s.e.   0.1143                            0.0525

sigma^2 estimated as 0.8079:  log likelihood = -147.05,  aic = 300.11
```

$$\nabla X_t = -0.3854 Z_{t-1} + Z_t + .0811 \;\; where\; Z_t \sim WN(0, \sigma^2)$$

Figure 13

Analyze Residuals

After now finding the best model described in the highlighted part of figure 13, I then tested the residuals of the model. I analyzed the residuals of the ARIMA(0,1,1) function and plotted it. This plot can be seen in figure 14. For the most part, the residuals do not have a surprising pattern. Afterwards, I plotted the histogram and QQ Normal Plot to see if the residuals were approximately normally distributed. Based on the histogram, the plot looks approximately normal and based on the QQ plot, there does not seem to be a presence of heavy tailed data or many outliers present. This can be seen in figures 15 and 16.

Plot of Residuals

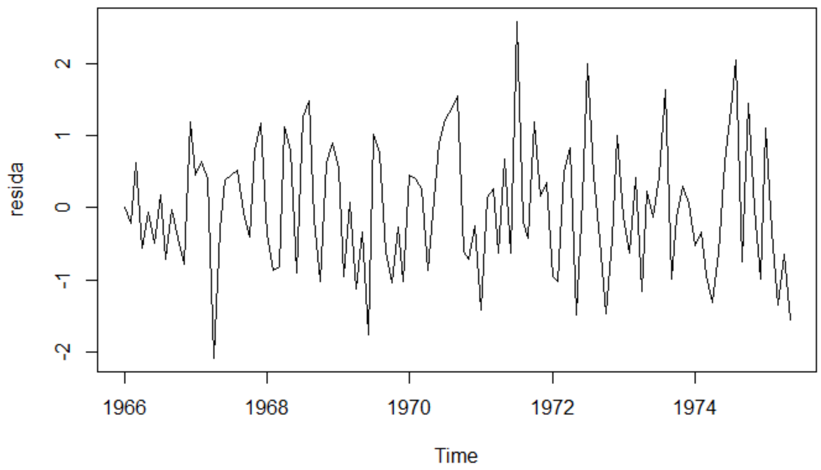


Figure 14

Histogram of Residuals

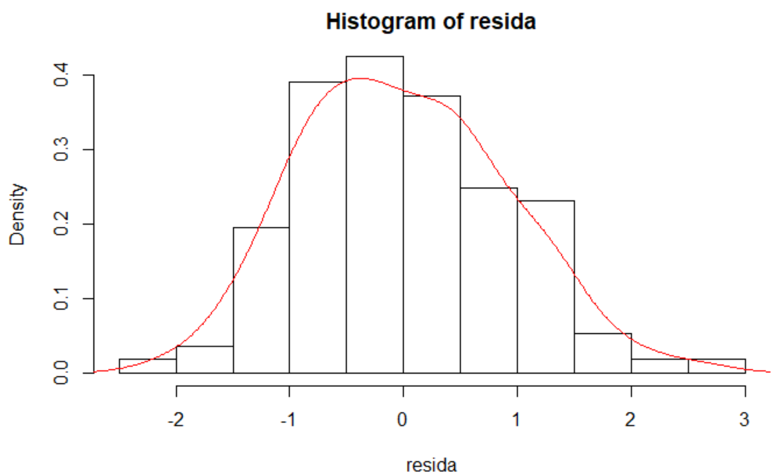


Figure 15

QQ Normal Plot of the Residuals

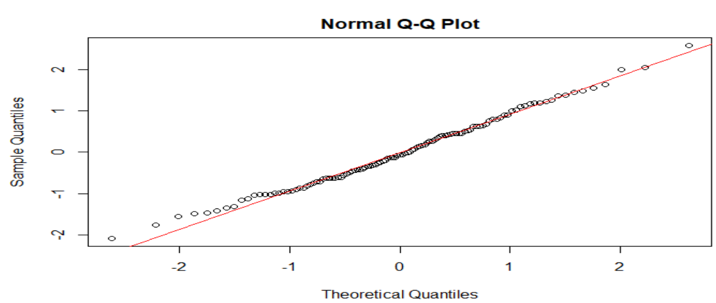


Figure 16

By looking at these three plots, the data appears to be normal and calculating the mean and variance, I obtained -.0001879711 for the mean and .807873 for the variance, so it is approximately normal from this analysis. However, there were more tests that I conducted in order to check the residuals for normality. I conducted an Ljung Box test, a Mcleod Li test, and a Shapiro Walk test. The results can be viewed in figures 17, 18, and 19. As a result of all three tests, they all fail to reject the null hypothesis at the 5% significance level. For the Shapiro Wilk test, I obtained the value of .6804>.05, obtained .08056>.05 obtained from the Ljung Box test where $h = \sqrt{n}$ and because it is ARIMA(0,1,1) the fit is just 1. Finally the McLeod-Li test passed since the .9253>.05 where again I failed to reject the Null. The null for the Shapiro Wilk states that the data is normally distributed, the Ljung Box test states that the autocorrelations are zero up to the lag we computed which was 11, and the McLeod-Li test tests for independence, which it passed. As a result, the model ARIMA(0,1,1) is a good forecasting model.

```
Shapiro-Wilk normality test

data:  resida
W = 0.99114, p-value = 0.6804
```

Figure 17

```
Box-Ljung test

data:  resida
X-squared = 16.73, df = 10, p-value = 0.08056
```

Figure 18

```
Box-Ljung test

data:  resida^2
X-squared = 5.1186, df = 11, p-value = 0.9253
```

Figure 19

Forecasting

As a result of the ARIMA(0,1,1) passing all the tests for its residuals, I then moved to forecasting five steps ahead for the months of June, July, August, September, and October. In the forecasting step, I used my ARIMA(0,1,1) model to predict the next five months for my Box-Coxed transformed data and calculated the lower and upper bounds for the confidence intervals. These bounds can be viewed in figure 20.

```
ltrans1
          Jun      Jul      Aug      Sep      Oct
1975 261.8882 251.1914 242.2332 234.4709 227.5964
> utrans1
          Jun      Jul      Aug      Sep      Oct
1975 414.5640 430.5156 444.7287 457.7458 469.8750
```

Figure 20

After this step, I then proceeded to plot the forecasted values within the time series and obtained this plot in figure 21.

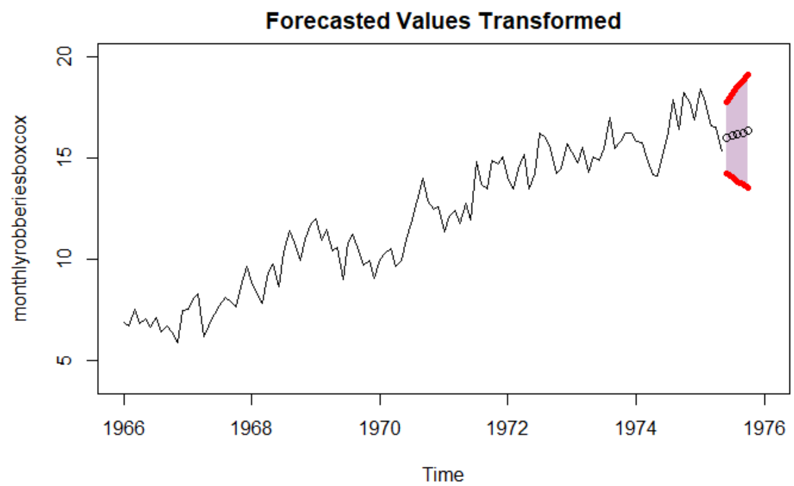


Figure 21

In figure 21, we can see that the predicted values are within the confidence interval which is represented by the red lines in the plot. Now, let us revert the data back to the original data set and invert the Box-Cox data to the original data. We can then see the given confidence intervals given in figure 22.

```
Jun       Jul       Aug       Sep       Oct
1975 261.8882 251.1914 242.2332 234.4709 227.5964
> utrans1
          Jun       Jul       Aug       Sep       Oct
1975 414.5640 430.5156 444.7287 457.7458 469.8750
```
Figure 22

Now, let us plot the time series of the original data.

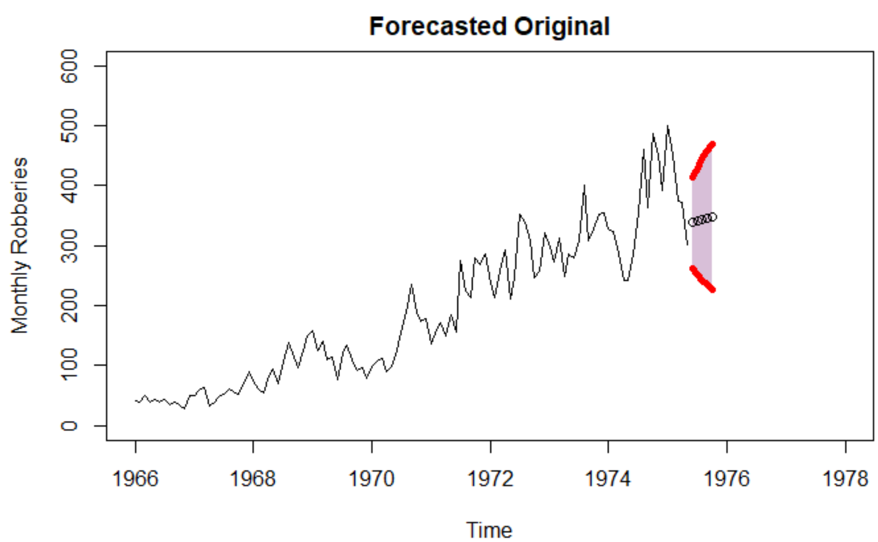


Figure 23

We can see here that after inverting the Box-Cox transformation, we obtain the next five values that are contained within the confidence interval. Viewing the actual data given by Data Market, the values for the June, July, August, September and October are given respectively by 316, 398, 394, 431 and 431, which are all within the confidence interval at the 95% confidence level given in figure 22. Thus, it was a success!

Conclusion

As a result of this whole project, the main goal was to answer the two questions given in the abstract of the report. The two questions were:

1) Can I create a time series model that could accurately model this behavior?
2) Can my model forecast a close estimate of the amount of robberies that would occur in the next five months?

These two questions were certainly answered as a result. The time series, ARIMA(0,1,1) passed all of the residual tests and was exactly one of the models I had guessed from the differenced data set. But most importantly, it answered the first question of coming up with this accurate model and secondly I was able to show that my data set was within the confidence interval at the 95% confidence level. I would like to thank Dr. Raya Feldman for teaching me the approach and methods for coming up with a valid time series forecasting model and my teacher assistant Nhan Huynh for being very helpful throughout the quarter. The recorded math ARIMA(0,1,1) formula that I chose to conclude this report was:

$$\nabla X_t = -0.3854Z_t + Z_t + .0811 \; where \; Z_t \sim WN(0, \sigma^2)$$

References

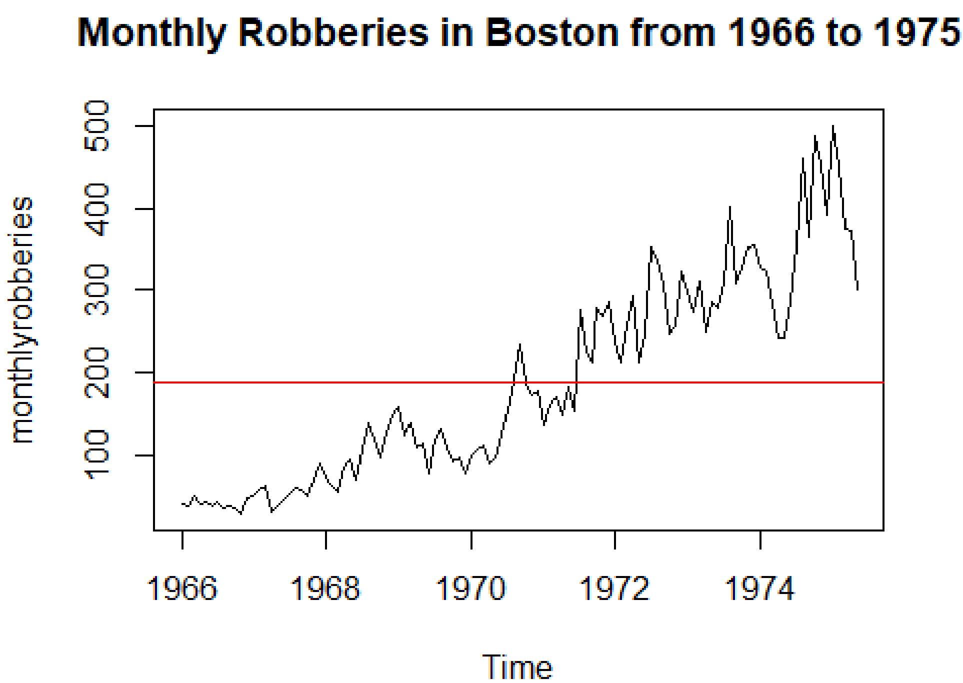
R Studio Statistical Software

Data Market

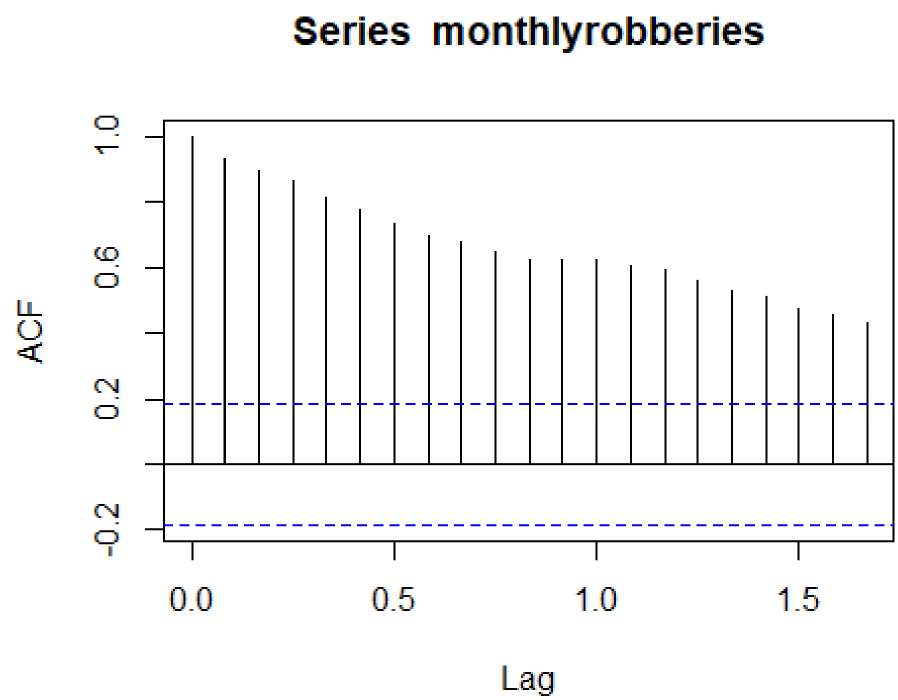
https://datamarket.com/data/set/22ob/monthly-boston-armed-robberies-jan1966-oct1975-deutsch-and-alt-1977#!ds=22ob&display=line

Appendix

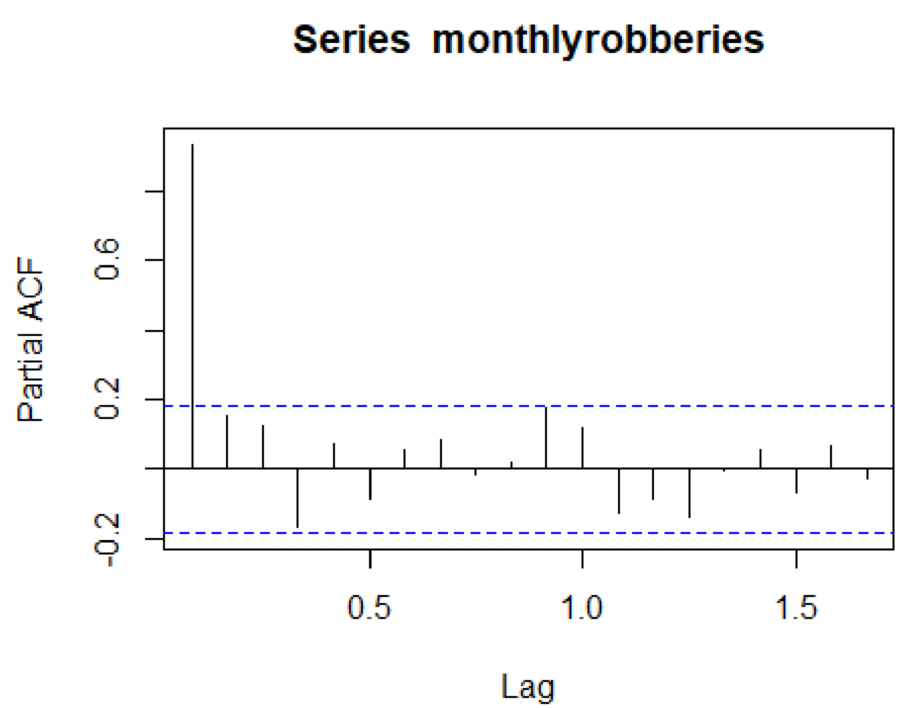
```
#Step 1: Read in the data file into R and convert to time series
monthlyrobberies.csv<-read.table("monthly-boston-armed-robberies-PSTAT174.csv
",sep=",",header=FALSE,skip=1,nrows=113)
head(monthlyrobberies.csv)

##         V1 V2
## 1 1966-01 41
## 2 1966-02 39
## 3 1966-03 50
## 4 1966-04 40
## 5 1966-05 43
## 6 1966-06 38

monthlyrobberies<-ts(monthlyrobberies.csv[,2],start=c(1966,1),frequency=12)


#Plot and analyze the time series. We see that the data has an upward
#trend and that the variance is not constant.
#Variance is very high at 15,225.8.
ts.plot(monthlyrobberies,main="Monthly Robberies in Boston from 1966 to 1975"
)
abline(h=0,lty=2)
abline(h=mean(monthlyrobberies),col="red")
```




```
acf(monthlyrobberies,lag.max=NULL)
```

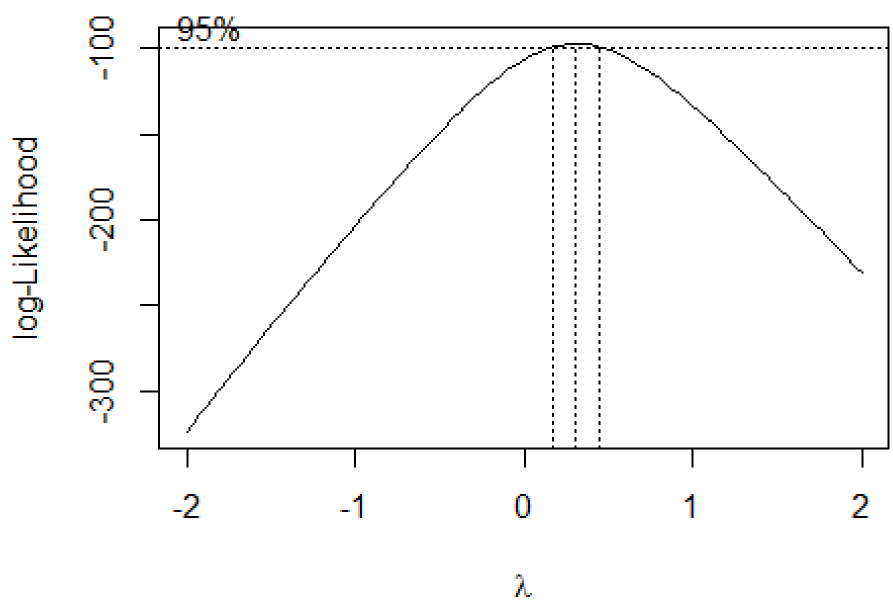



```
pacf(monthlyrobberies,lag.max=NULL)
```

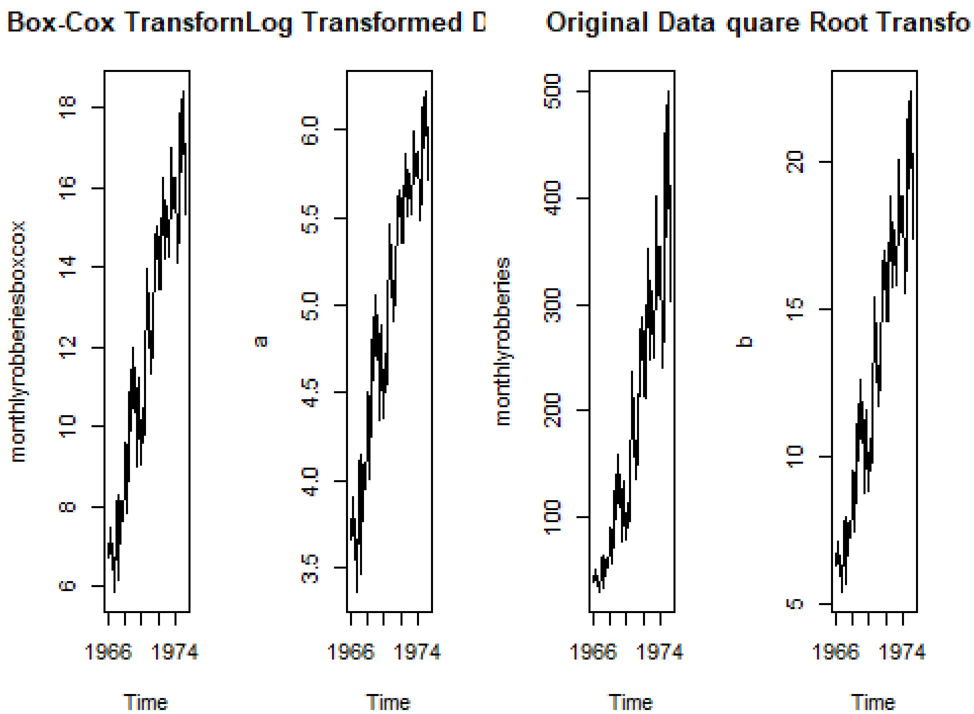



```
var(monthlyrobberies)
```

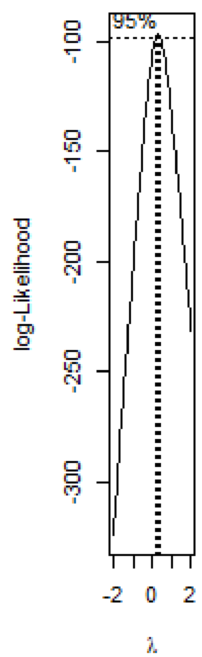
```
## [1] 15225.88

#Step 2: Based on data, data has upward trend and non-constant variance.
#Let's see three different types of transformation with original data
#All three transformations appear to have the same appearance. Therefore,
#Let's apply a Box-Cox transformation for this particular data set.
library(MASS)
time<-1:length(monthlyrobberies)
fit<-lm(monthlyrobberies~time)
boxcoxtransform<-boxcox(monthlyrobberies~time,plotit=TRUE)
```



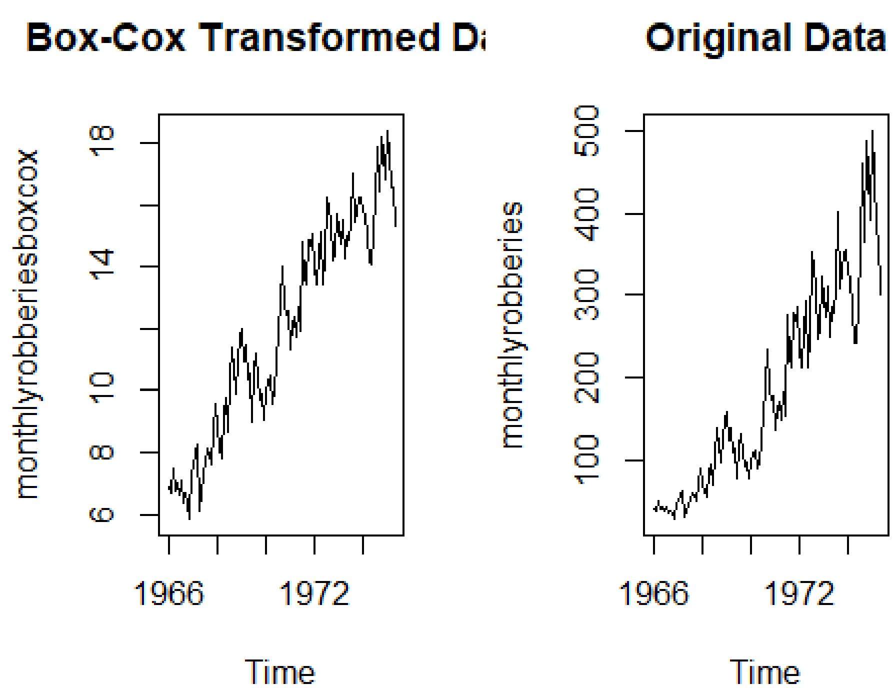

```
lamb<-boxcoxtransform$x[which(boxcoxtransform$y==max(boxcoxtransform$y))]
monthlyrobberiesboxcox<-(1/lamb)*(monthlyrobberies^lamb-1)
op4<-par(mfrow=c(1,4))
ts.plot(monthlyrobberiesboxcox,main="Box-Cox Transformed")
a<-log(monthlyrobberies)
ts.plot(a,main="Log Transformed Data")
ts.plot(monthlyrobberies,main="Original Data")
b<-sqrt(monthlyrobberies)
ts.plot(b,main="Square Root Transformed")
```



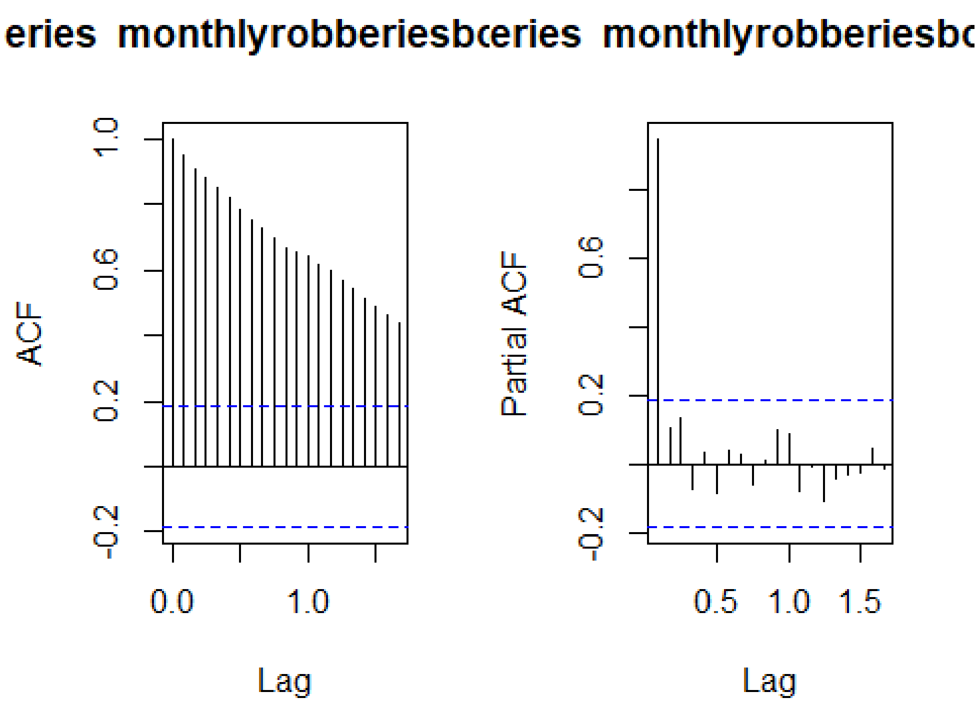

```
op4

## $mfrow
## [1] 1 1

#Step 3:#Let the Box-Cox transformation of the data set be denoted as
#monthlyrobberiesboxcox. The variance of the data is more stabilized.
time<-1:length(monthlyrobberies)
fit<-lm(monthlyrobberies~time)
boxcoxtransform<-boxcox(monthlyrobberies~time,plotit=TRUE)
lamb<-boxcoxtransform$x[which(boxcoxtransform$y==max(boxcoxtransform$y))]
monthlyrobberiesboxcox<-(1/lamb)*(monthlyrobberies^lamb-1)
par1<-par(mfrow=c(1,2))
```




```
ts.plot(monthlyrobberiesboxcox,main="Box-Cox Transformed Data")
ts.plot(monthlyrobberies,main="Original Data")
```

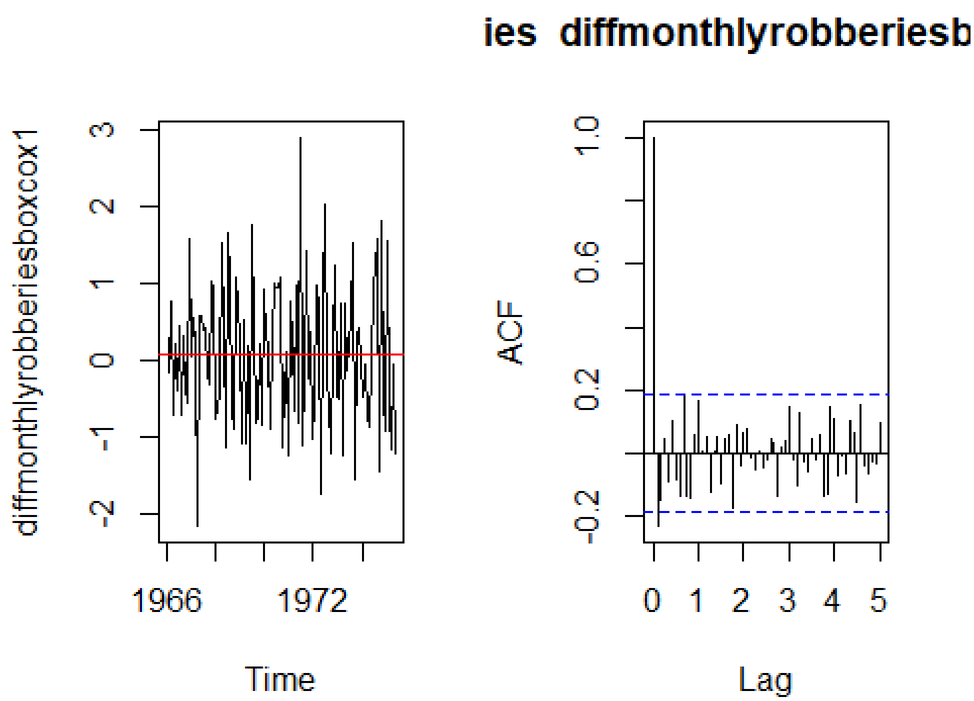


```r
par1
## $mfrow
## [1] 1 4
var(monthlyrobberiesboxcox)
## [1] 11.84745
acf(monthlyrobberiesboxcox)
pacf(monthlyrobberiesboxcox)


#Check if Box-Cox transformed data is stationary
#Box-Cox transformed data is not stationary because
#we fail to reject the null hypothesis of non-
#stationarity at the 5% significance level.
library(tseries)
```




```r
adf.test(monthlyrobberiesboxcox)
##
##  Augmented Dickey-Fuller Test
##
## data:  monthlyrobberiesboxcox
## Dickey-Fuller = -3.4072, Lag order = 4, p-value = 0.05686
## alternative hypothesis: stationary
```

```r
#Data not stationary, thus from
#analyzing the acf of a, we difference at lag 1
#in order to remove the trend.
#We can see that the variance greatly decreased.
#Also acf and pacf look better. Denote diffa as
#a differenced at lag 1
diffmonthlyrobberiesboxcox1<-diff(monthlyrobberiesboxcox,1)
ts.plot(diffmonthlyrobberiesboxcox1)
abline(h=mean(diffmonthlyrobberiesboxcox1),col="red")
acf(diffmonthlyrobberiesboxcox1,lag.max=60)
```

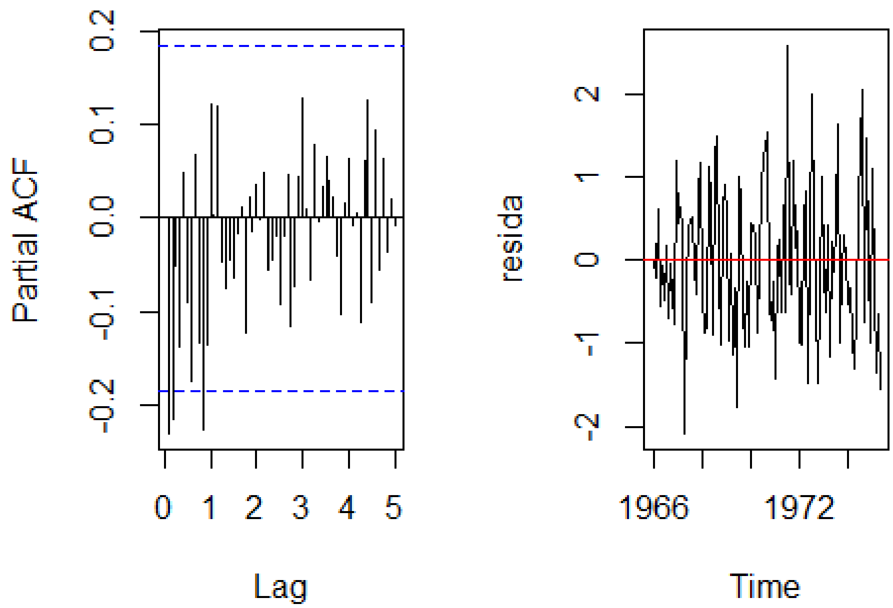



```r
pacf(diffmonthlyrobberiesboxcox1,lag.max=60)
var(diffmonthlyrobberiesboxcox1)

## [1] 0.8944784

#Check if data is stationary time series. As a result
#p-value is .01 which is less than .05 and therefore
#reject the null that the data is not stationary
library(tseries)
adf.test(diffmonthlyrobberiesboxcox1)

## Warning in adf.test(diffmonthlyrobberiesboxcox1): p-value smaller than
## printed p-value

##
##  Augmented Dickey-Fuller Test
##
```

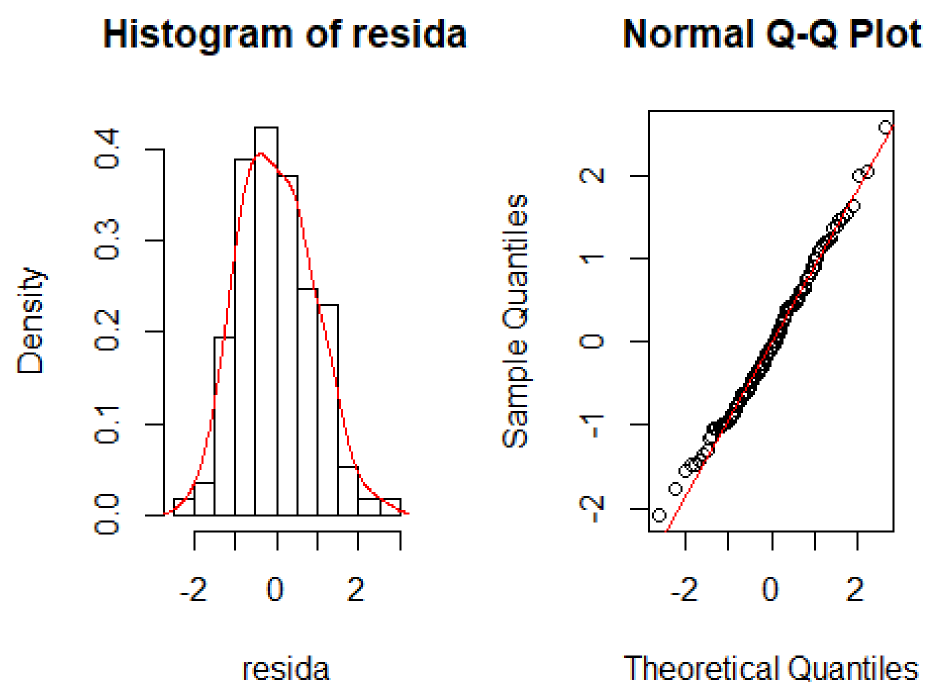
```
## data:  diffmonthlyrobberiesboxcox1
## Dickey-Fuller = -5.2908, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary

#Suggested models based on the stationary differenced
#time series acf is MA(1) since it cuts off after lag 1
# and for pacf AR(10) since pacf cuts off after lag 10
#Estimate best AR model using Yule-Walker method of ar and use
#ML method of ar. Denote ay as ar estimated by yule-walker
#and aml as ar estimated by maximum likelihood method
ay<-ar(diffmonthlyrobberiesboxcox1,aic=TRUE,order.max=NULL,method="yule-walke
r")
ay

##
## Call:
## ar(x = diffmonthlyrobberiesboxcox1, aic = TRUE, order.max = NULL,     meth
od = "yule-walker")
##
## Coefficients:
##       1        2
## -0.2805  -0.2154
##
## Order selected 2  sigma^2 estimated as  0.8224

aml<-ar(diffmonthlyrobberiesboxcox1,aic=TRUE,order.max=NULL,method="mle")
aml

##
## Call:
## ar(x = diffmonthlyrobberiesboxcox1, aic = TRUE, order.max = NULL,     meth
od = "mle")
##
## Coefficients:
##       1        2        3        4        5        6        7        8
## -0.3881  -0.3422  -0.2467  -0.2905  -0.1367  -0.2479  -0.3243  -0.1181
##       9       10       11
## -0.2925  -0.3140  -0.1432
##
## Order selected 11  sigma^2 estimated as  0.6445

#Based on the ar function method using Yule-Walker method, the model
#is presented as an AR(2) but with the mle method, it is presented as
#an AR(11).Estimate coefficients using innovations algorithm for MA(1)
#Then see if MA(1) fits because only significant acf is at lag 1.
source("innovations.r")
acvf<-acf(diffmonthlyrobberiesboxcox1,plot=FALSE,lag.max=length(diffmonthlyro
bberiesboxcox1))$acf[ ,1,1]*var(diffmonthlyrobberiesboxcox1)
k<-length(acvf)
k1<-innovations.algorithm(k+1,acvf)
k1$thetas[1,1:1]
```

```
## [1] -0.230757

polyroot(c(1,-0.2292035))

## [1] 4.362935+0i

#Compare the three models below.
#The AR(11) and MA(1) model only have a slight difference.
#Therefore, based on parsimony, we will choose the
#MA(1) model instead of the AR(11) with the ar function.
#Also, AR(2) has higher AICc value than MA(1) and AR(11)
#therefore, not the best model.
library(qpcR)

## Loading required package: minpack.lm

## Loading required package: rgl

## Loading required package: robustbase

## Loading required package: Matrix

i1<-arima(monthlyrobberiesboxcox,order=c(11,1,0),method="ML",xreg=1:length(mo
nthlyrobberiesboxcox))
i2<-arima(monthlyrobberiesboxcox,order=c(2,1,0),method="ML",xreg=1:length(mon
thlyrobberiesboxcox))
i3<-arima(monthlyrobberiesboxcox,order=c(0,1,1),method="ML",xreg=1:length(mon
thlyrobberiesboxcox))
AICc(i1)

## [1] 299.6589

AICc(i2)

## [1] 300.9642

AICc(i3)

## [1] 300.2168

#Step 5 analyze the resid
#forecast
resida<-resid(i3)
plot(resida)
abline(h=mean(resida),col="red")
```

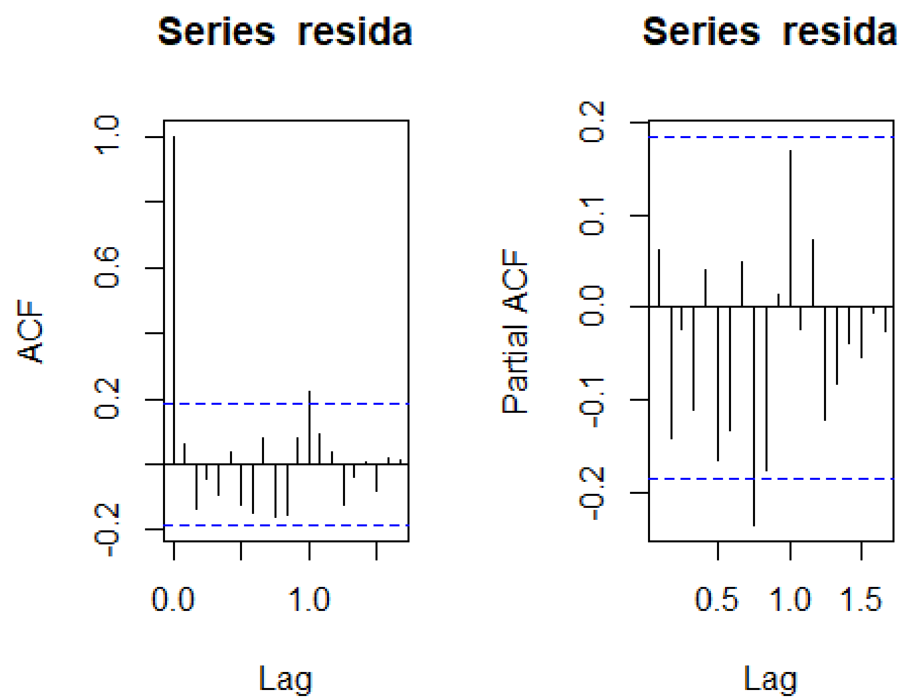
ies diffmonthlyrobberiesb




```
hist(resida,breaks=10,probability=TRUE)
lines(density(resida),col="red")
mean(resida)

## [1] -0.0001879711

var(resida)

## [1] 0.807873

qqnorm(resida)
qqline(resida,col="red")
```

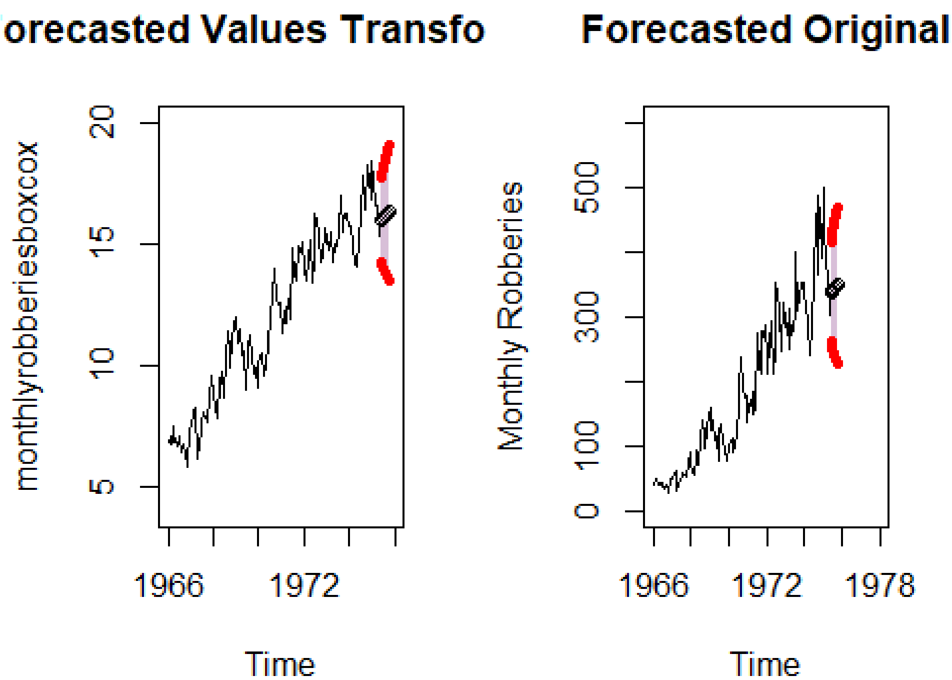



```
shapiro.test(resida)

##
##  Shapiro-Wilk normality test
##
## data:  resida
## W = 0.99114, p-value = 0.6804

Box.test(resida,lag=11,type="Ljung",fit=1)

##
##  Box-Ljung test
##
## data:  resida
## X-squared = 16.73, df = 10, p-value = 0.08056

Box.test(resida^2,lag=11,type="Ljung",fit=0)

##
##  Box-Ljung test
##
## data:  resida^2
## X-squared = 5.1186, df = 11, p-value = 0.9253

acf(resida)
pacf(resida)
```




```
#Step 6
#First best model is the AR(2) because of the lower AICc value
#Forecast next five months of June, July, August, September, October
library(forecast)
fitma1<-arima(monthlyrobberiesboxcox,order=c(0,1,1),method="ML",xreg=1:length
(monthlyrobberiesboxcox))
predtrans<-predict(fitma1,n.ahead=5,newxreg=(length(monthlyrobberiesboxcox)+1
):(length(monthlyrobberiesboxcox)+5))
ltrans<-predtrans$pred-1.96*predtrans$se
ltrans

##           Jun      Jul      Aug      Sep      Oct
## 1975 14.23740 14.01235 13.82712 13.66930 13.53186

utrans<-predtrans$pred+1.96*predtrans$se
utrans

##           Jun      Jul      Aug      Sep      Oct
## 1975 17.76076 18.14797 18.49537 18.81536 19.11497

#Forecast the next five months June, July, August, October, September, and Oc
tober of the transformed data
ts.plot(monthlyrobberiesboxcox,xlim=c(1966,1976),ylim=c(4,20),main="Forecaste
d Values Transformed")
# Add the predicted values:
space=1/12
k=0:11*space
```

```
indextoadd=1975+k[6:10]
polygon(c(indextoadd,rev(indextoadd)),c(ltrans,rev(utrans)),col="thistle",bor
der=NA)
points(indextoadd,predtrans$pred,pch=1)
lines(indextoadd,ltrans,col="red",lwd=5)
lines(indextoadd,utrans,col="red",lwd=5)


#Transform the transformed data back to the original data
unboxcoxed<-((monthlyrobberiesboxcox*lamb)+1)^(1/lamb)
fitma1unboxcoxed<-arima(unboxcoxed,order=c(0,1,1),method="ML",xreg=1:length(u
nboxcoxed))
predtrans1<-predict(fitma1unboxcoxed,n.ahead=5,newxreg=(length(unboxcoxed)+1)
:(length(unboxcoxed)+5))
ltrans1<-predtrans1$pred-1.96*predtrans1$se
ltrans1

##           Jun      Jul      Aug      Sep      Oct
## 1975 261.8882 251.1914 242.2332 234.4709 227.5964

utrans1<-predtrans1$pred+1.96*predtrans1$se
utrans1

##           Jun      Jul      Aug      Sep      Oct
## 1975 414.5640 430.5156 444.7287 457.7458 469.8750

ts.plot(unboxcoxed,xlim=c(1966,1978),ylim=c(0,600),main="Forecasted Original"
,ylab="Monthly Robberies")
space1<-1/12
k1<-0:11*space1
indextoadd1=1975+k1[6:10]
polygon(c(indextoadd1,rev(indextoadd1)),c(ltrans1,rev(utrans1)),col="thistle"
,border=NA)
points(indextoadd1,predtrans1$pred,pch=1)
lines(indextoadd1,ltrans1,col="red",lwd=5)
lines(indextoadd1,utrans1,col="red",lwd=5)
```


orecasted Values Transfo
monthlyrobberiesboxcox
20
15
10
5
1966
1972
Time
Forecasted Original
Monthly Robberies
500
300
100
0
1966
1972
1978
Time

```
```