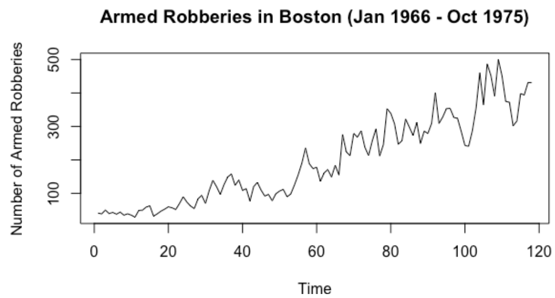
Boston Armed Robberies Report

Anna Lee, 999231359
Isabel Kraus-Liang, 999350609
STA 137, WQ 2017
Final Project
D. Izyumin

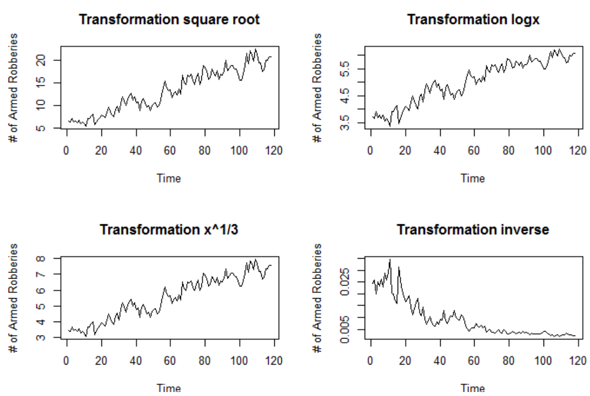
Step 1: Inspection of Data

Over time, the variance increases; this can be seen in the fact that the peaks and troughs have larger space in between them. The mean also changes as an upward trend is evident. Therefore, the data is not stationary.



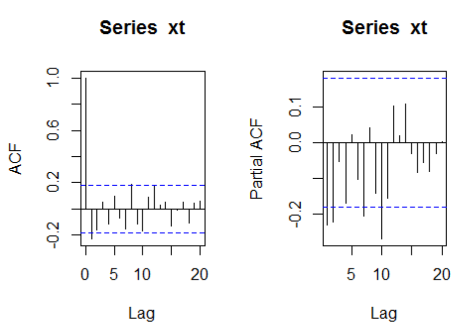
Step 2: Transformations

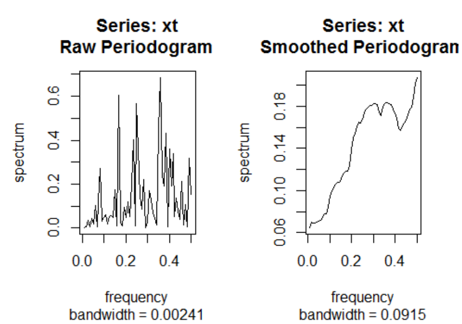
From inspecting the data, we notice that the data is monthly. There is no clear seasonality present; thus, we decided to transform the data and to address the problem with the variance first. As you can see from below, the $x^{1/3}$ transformation provides the best solution. In theory, if we were to draw bands around the data, they would be equidistant throughout. From this point on, we shall be using the $x^{1/3}$ transformation.



Step 3: ACF, PACF, and Periodogram

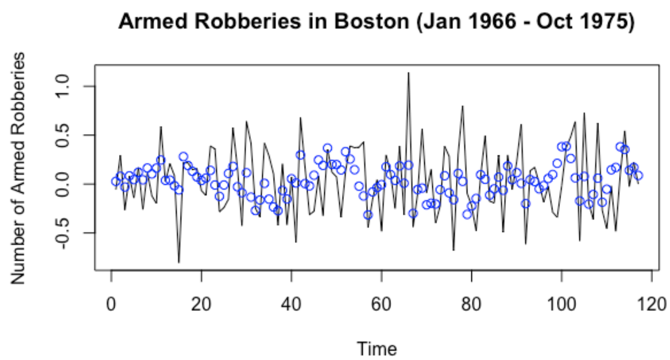
In evaluating the ACF and PACF, we see that the process is i.i.d. which means we can fit a model. Based on the ACF, we choose q=2; through the PACF, p=2. Because we are differencing, we decide upon an ARIMA(2, 0, 2) model. From the periodograms, there are no significant peaks, and therefore, no evidence of seasonality.



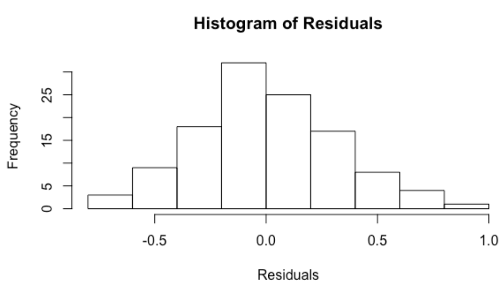


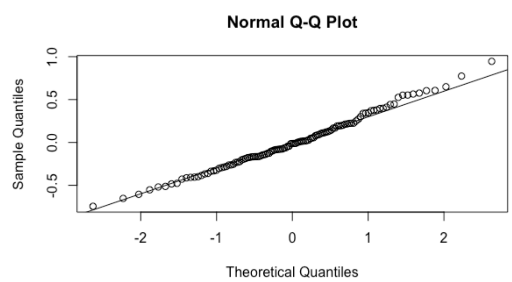
Step 4: Fitting a model
The model we fit is an ARIMA(2,0,2) with a non zero-mean. The coefficients are given below along with the variance. The blue points are the fitted values.

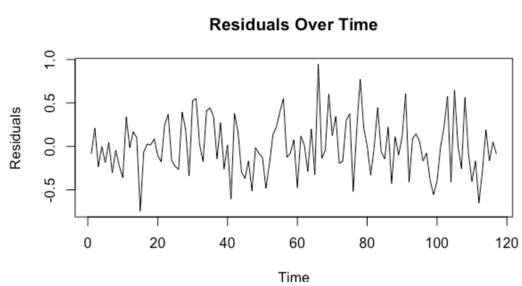
```
ARIMA(2,0,2) with non-zero mean

Coefficients:
          ar1     ar2      ma1      ma2   mean
      -0.4023  0.5752  -0.0203  -0.9797  0.038
s.e.   0.1016  0.0834   0.0775   0.0773  0.002

sigma^2 estimated as 0.1079
```



Step 5: Residuals
From evaluating the residuals in the histogram, they seem to be distributed similarly to Gaussian white noise. Most points on the qqplot lie on the qqline. The plot of residuals over time shows that all structures are accounted for and there is no discernable trend or seasonal component. The Box Ljung test produced a large p-value, supporting the idea that there is not enough evidence to say the residuals are dependent.



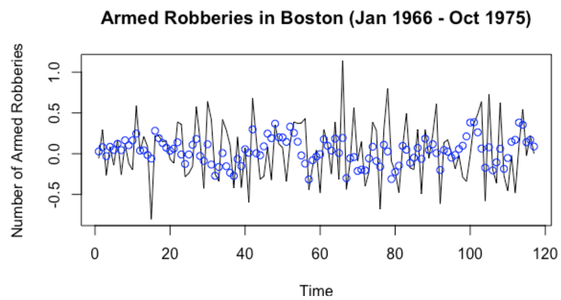




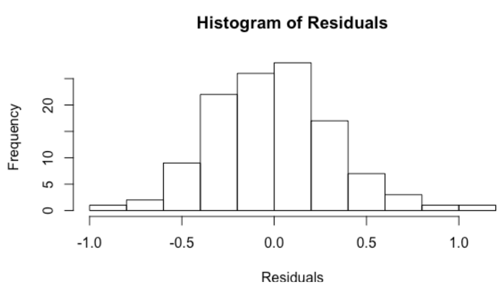

```
        Box-Ljung test

data:  mod1$residuals
X-squared = 22.838, df = 20, p-value = 0.2968
```

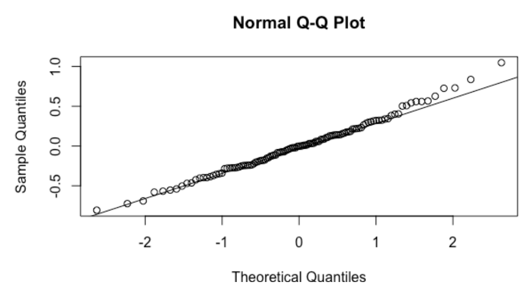
Step 6: Fit Model with auto.arima
The model we fit is an ARIMA(0,0,2) with a zero-mean. The coefficients are given below along with the variance. The blue points are the fitted values.

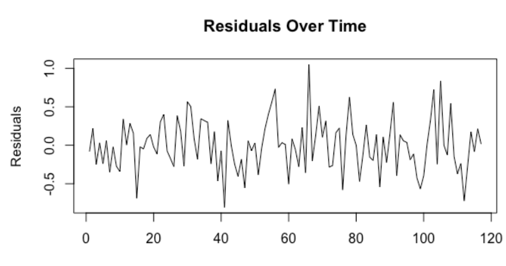
```
ARIMA(0,0,2) with non-zero mean

Coefficients:
          ma1      ma2    mean
      -0.3659  -0.3222  0.0352
s.e.   0.0963   0.1238  0.0100

sigma^2 estimated as 0.116; AICc=85.9
```



Step 7: Residuals
Through the histogram, the residuals seem to be distributed as Gaussian white noise. The qqplot shows some points being off the qqline near the tails; however, this is negligible as the residuals in the middle are close or on the line. Plotting the residuals over time shows there are no dependent structures that have yet to be accounted for. This is further supplemented by the results of running the Box-Ljung test, in which our p-value is greater than our chosen alpha of 0.05; from this test, we can conclude there is not enough evidence to say it is dependent. Therefore, the fit of the model is good.

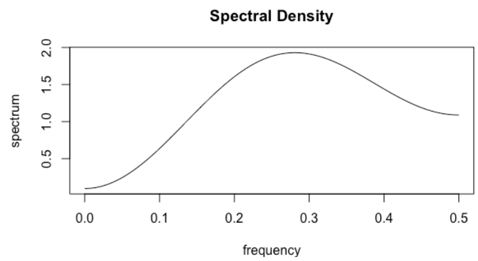







```
Box-Ljung test

data:  mod$residuals
X-squared = 26.001, df = 20, p-value = 0.1658
```

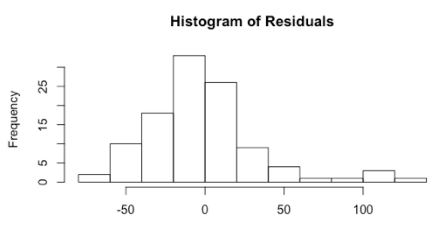
Step 8: Theoretical Spectral Density
We chose L through creating a vector of candidates of odd numbers from 3 to 117. We then used the modified Daniell kernel to choose an L of 39 for our smoothed periodogram. The plot given below is the theoretical spectral density, which is similar to our smoothed version from step 3.

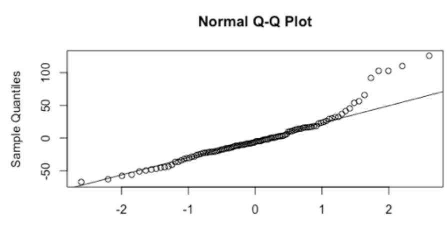


Step 9: Forecasting
In using auto.arima, R chose an ARIMA(1,1,1) model with the following coefficients. The AICc is 1074. We kept the original data rather than transform it because the resulting ARIMA model is not efficient; it overfit the data and had a variance of 0.

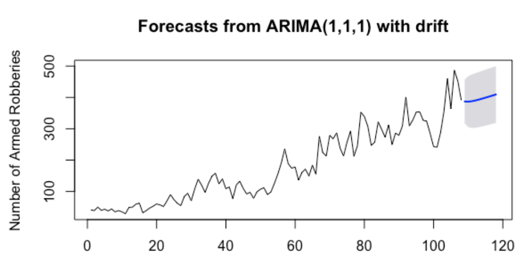
```
ARIMA(1,1,1) with drift

Coefficients:
         ar1      ma1   drift
      0.5309  -0.9439  3.4134
s.e.  0.1132   0.0633  0.4846

sigma^2 estimated as 1270
```

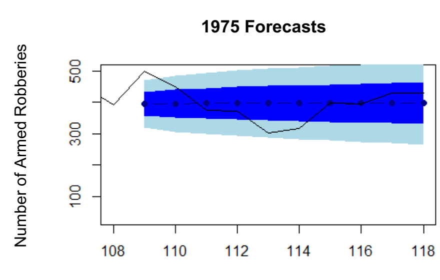
The residuals are an approximate Gaussian distribution.





The performance of the forecast shows an upward trend, consistent with the trend in the observed data. Unlike the observed data, the forecast does not have peaks and troughs but rather, is closer to a straight line. Most points are close to the observed. This is a sufficient model given that crime is dependent on economic environment, which is hard to include in our model as it is not a trend nor a seasonal thing.





APPENDIX

```
install.packages("astsa")
library(astsa)
library(forecast)

data=read.table("bostonArmedrobberies.txt"
)
x = data[,2]

plot.ts(x,ylab="Number of Armed
Robberies")
n=length(x)
m=floor(n/2)

# Tranform the data
par(mfrow=c(1,2))
xt = sqrt(x)
ts.plot(xt,ylab="# of Armed
Robberies",main="Transformation square
root")
xt = log(x)
ts.plot(xt,ylab="# of Armed
Robberies",main="Transformation logx")
xt = x^(1/3)
ts.plot(xt,ylab="# of Armed
Robberies",main="Transformation x^1/3")
xt = 1/x
ts.plot(xt,ylab="# of Armed
Robberies",main="Transformation inverse")

#chosen transformation
x1=x
xt = x^(1/3)
xt = diff(xt)

# ACF and PACF
acf(xt)
pacf(xt)

# get the raw periodogram values at the
Fourier frequencies
pgrm.raw = spec.pgram(xt,
plot=F,log='no')$spec

# vector of candidate L values for smoothing
spans = (1:(m-1))*2+1

# vector to store criterion values for each L
Q = numeric(length(spans))

# go through the L values and compute Q
for each
for(j in 1:length(spans)){
  L = spans[j]
  pgrm.smooth = spec.pgram(xt,
spans=L,log='no', plot=F)$spec
  Q[j] = sum((pgrm.smooth - pgrm.raw) ^ 2)
+ sum((pgrm.raw)^2)/(L-1)
}

# plot the values
plot(x=spans, y=Q, type='b')

# figure out which L is best
L = spans[which.min(Q)]; L

par(mfrow=c(1,2))
spec.pgram(xt, log='no')
spec.pgram(xt, spans=L, log='no')
par(mfrow=c(1,1))

##################################
#step 4
mod1=Arima(xt, order=c(2,0,2))
mod1$coef
summary(mod1)
fitted.values=fitted(mod1)
plot.ts(xt,ylab="Number of Armed
Robberies")
points(fitted.values, type = "p", col="blue")

#step 5
hist(mod1$residuals)
res1=mod1$residuals
qqnorm(res1)
plot.ts(res1)
Box.test(mod1$residuals, lag= 20, type=
"Ljung-Box")

#step 6
```

```
# fit arima model
mod=auto.arima(xt)
mod$coef
summary(mod)
fitted.values=fitted(mod)
plot.ts(xt,ylab="Number of Armed
Robberies")
points(fitted.values, type = "p", col="blue")

#step 7
hist(mod$residuals)
res1=mod$residuals
qqnorm(res1)
plot.ts(res1)
Box.test(mod$residuals, lag= 20, type=
"Ljung-Box")

#step 8
arma.spec(ma=c(-0.365, -0322.), log='no',
main='Spectral Density')

#step 9
x2=x[1:108]

mod2=auto.arima(x2)
summary(mod2)


res = mod2$residuals
ts.plot(res)
acf(res)
pacf(res)
Box.test(res,lag=20,type="Ljung")
hist(res)
qqnorm(res)
qqline(res)

plot(forecast(mod2))
points(x, col="black", type="l")

h = 10; n=108
model = Arima(x2, order=c(1,1,1))
summary(model)
fcasts = predict(model,n.ahead=10)
f.vals=cbind(394.4054, 395.8202, 396.4080,
396.6522, 396.7536,
        396.7958, 396.8133, 396.8206,
396.8236, 396.8248)
f.se = cbind(37.24433, 44.86115, 49.08646,
52.21815, 54.89536,
        57.34153, 59.64566, 61.84736,
63.96665, 66.01526)

upper1 = f.vals + f.se
lower1 = f.vals - f.se
upper2 = f.vals + 2 * f.se
lower2 = f.vals - 2 * f.se

plot.ts(x,xlim=c(108,n+h))
polygon(x=c(n+(1:h),n+(h:1)),
y=c(upper2,lower2[h:1]), col='lightblue',
border=NA)
polygon(x=c(n+(1:h),n+(h:1)),
y=c(upper1,lower1[h:1]), col='blue',
border=NA)
points(x=n+1:h, y=f.vals,
col='darkblue',type='b',pch=19)
points(x, type="l")
```