Fall 2019 STA 137 Final Project

Instructor Name: Jamshid Namdari
Student Name: Xuecheng Zhang
Student ID:915942842
2019/12/05

Introduction: Our data is about the monthly handgun sales and firearms related deaths in California (1980-1998) and prepare a report. Time series analysis is use in order to understand the underlying structure and function that produce the observations. In our data, a time series gives the relationship between two variables, one of them being time.
A time series carries profound importance in business and policy planning. It's uses are :

- It is used to study the past behavior of the monthly handgun sales and firearms-related deaths under consideration.
- It can also be used to predict monthly handgun sales and firearms-related deaths after 1998.

Nature of the Variation:

Figure 1.1 Time Series of GUN SALES

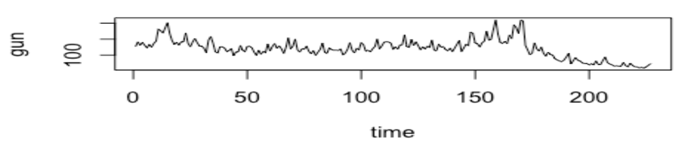
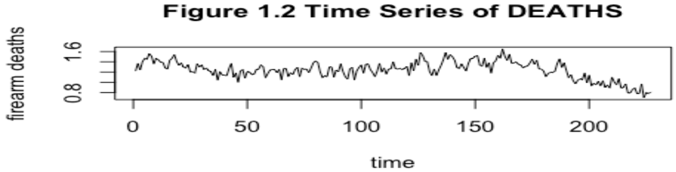


Figure 1.2 Time Series of DEATHS



From the time series data of gun sales, the sales volume of guns was stable after 1980. However, after 1994, the sales volume of guns gradually decreased as shown in figure 1.1.As same as the firearms related deaths, The deaths have been stable since 1980, but the sales of guns have been declining in the following years, but it's not as obvious as gun sales as shown in Figure 1.2.

Figure 2.1 LOG transformed GUN SALES

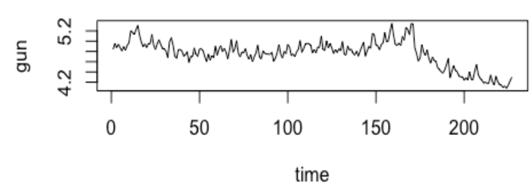
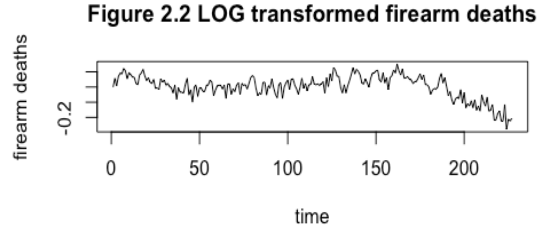


Figure 2.2 LOG transformed firearm deaths



Stationary & Transformed Series

Stationary:

A stationary series is one in which the properties – mean, variance and covariance, do not vary with time. From the figure 1.1 and figure 1.2, the mean of the handgun sales and firearms-related deaths changed with the time, so they are not stationary. I will want to transform data.
Method 1: I try the log transformation, because the log difference are helpful for making non-stationary data stationary.

After I did the log transformation, the time series of gun sales is still nonstationary, the mean still change by time as shown in figure 2.1. As same as the firearm deaths, after the of transformation, the time series of deaths is still not stationary, because the mean changed by time.

Method 2: We can take differences of a time series as well.
As we can see, taking the difference is an effective way to remove a trend and make a time series stationary. The mean of monthly handgun sales doesn’t change by time. Also the firearms related deaths in California(1980-1998) doesn’t change by time. So both of them are stationary now.

Figure 3.1 take differences transformed GUN

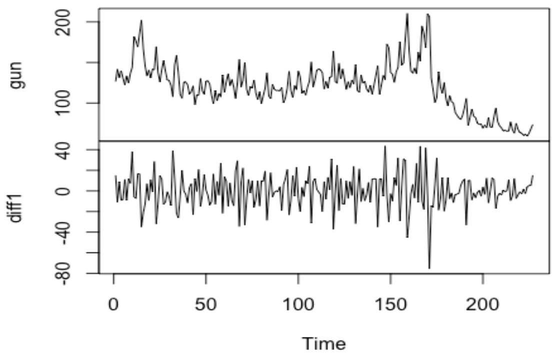
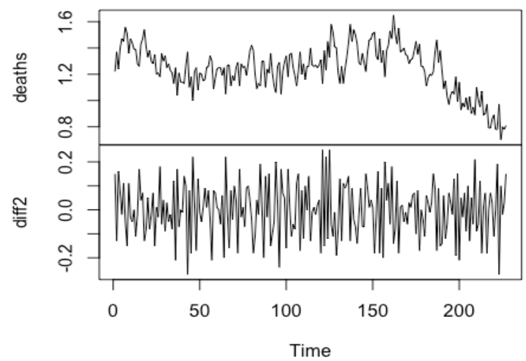


Figure 3.2 take differences transformed deaths



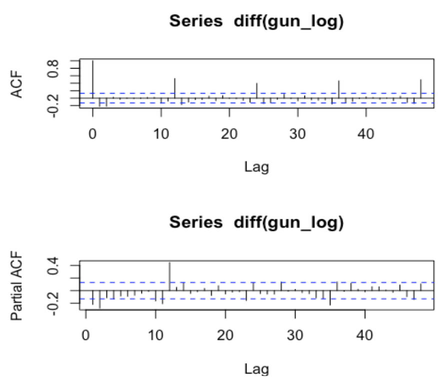
Sample Autocorrelation and Partial Autocorrelation:

Method of autocorrelation: Autocorrelation is the linear dependence of a variable with itself at two points in time. Define: $Cov(y_{t+h}, y_t) = \gamma_h$ and the lag-h autocorrelation is given by

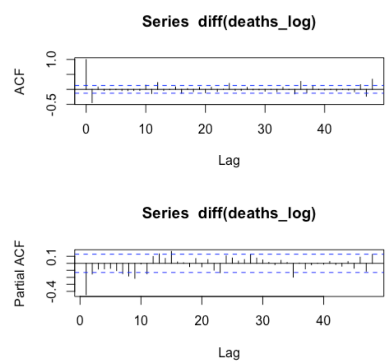
$$\rho_h = Corr(y_{t+h}, y_t) = \frac{\gamma_h}{\gamma_0}.$$

Method of partial autocorrelation: Partial autocorrelation is autocorrelation between y_t and y_{t-h} after removing any linear dependence on $y_1, y_2 \dots y_{t-h+1}$. PACF is defined as

$$\varphi_{hh} = Corr(x_{t+h} - \widehat{x_{t+h}}, x_t - \widehat{x_t})$$



For monthly handgun sales:
From the ACF of the gun sale ,we can see that ACF circulates periodically every 12 months, and the ACF tails off and PACF tails off, we can model given ARIMA process ARIMA(1,1,2).



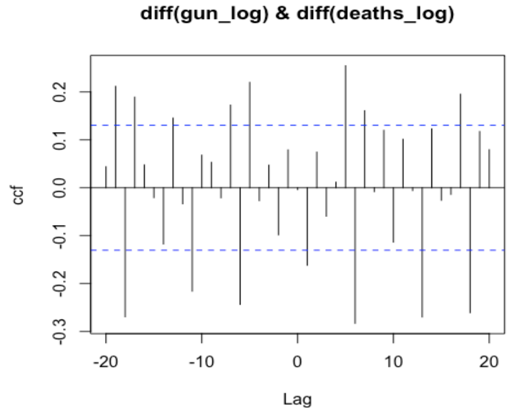
For Firearms related deaths in California:
From the ACF of the firearms death ,we can see that ACF cut off at lag 2, Since ACF cuts off at lag 2 and the PACF tails off, we can model given process using linear combination of first 2 lag ARIMA(0,1,2).

Cross Correlation:

Method of Cross Correlation: Cross correlation is a standard method of estimating the degree to which two series are correlated. The cross correlation at delay d is defined as

$$\rho_{y,x}(\mathrm{t,s}) = \frac{\gamma_{yx}(t,s)}{\sqrt{\gamma_y(t,t) * \gamma_x(s,s)}}$$

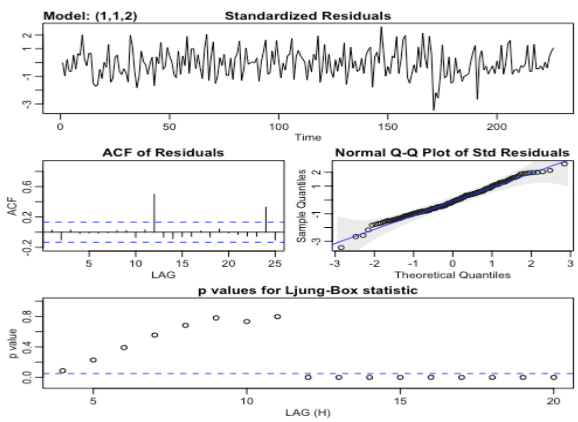
The correlation plot (ccf between gun sales and death), indicating that the monthly handgun sales and firearms related deaths in California (1980-1998) are most strongly correlated. When |h|<5, the gun sales and death have negative correlation when the gun sale increase the firearms related deaths decrease. When |h|>5 the gun sales and death have positive correlation when the gun sale increase the firearms related deaths increase as shown in diff(gun_log) & diff(deaths_log) figure.

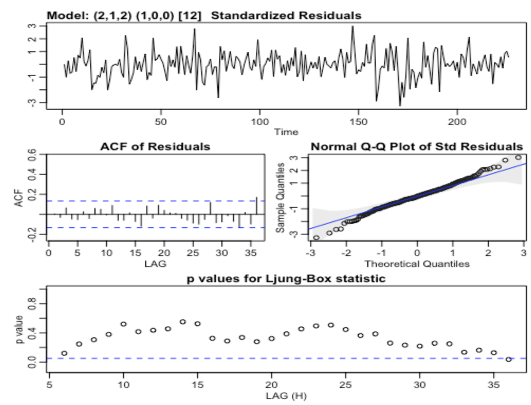


ARIMA Model:

Method of ARIMA Model: we need to see the PACF and ACF plot to choose p, d, q, P, D, Q, S. In the previous analysis, we decided that handgun follows ARIMA(1,1,2) model, and firearms deaths follows ARIMA(0,1,2) model.

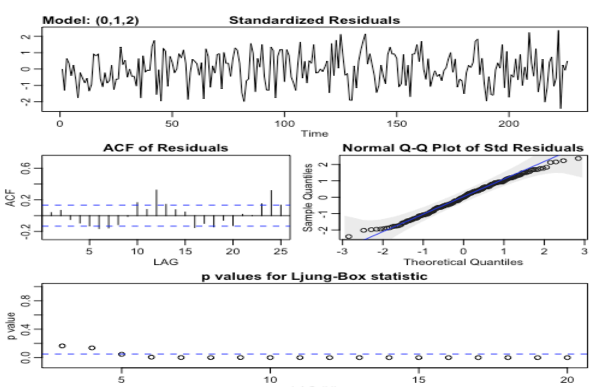
For monthly handgun sales: I also try p=2,d=1,q=2,P=1,D=0,Q=0,S=12 as model 2

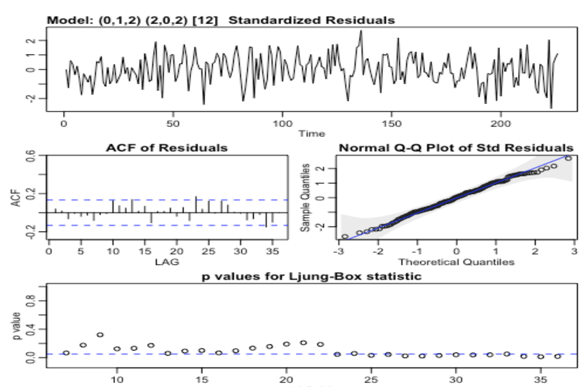




A few outliers in the standardized residuals and their normal Q-Q plot. I have 277 sample size, I think this data set have a large sample size, so I use AIC to choose my model, the AIC of ARIMA(1,1,2)= -1.410255, and the AIC of SARIMA(2,1,2)*(1,0,0)[12] = -1.697271, science AIC of SARIMA(2,1,2)*(1,0,0)[12] is smaller, I chose it.

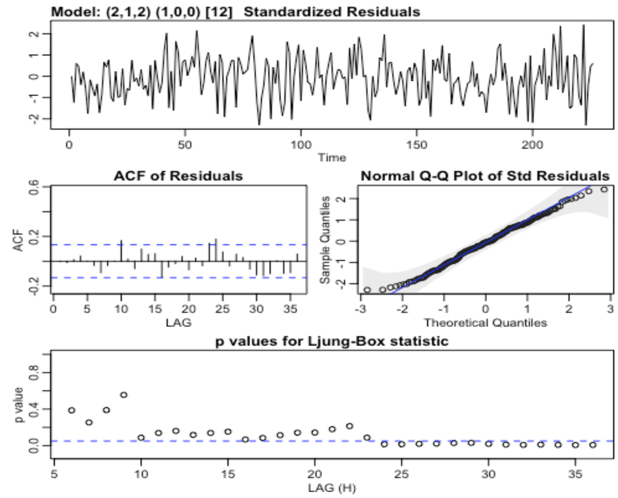
For Firearms related deaths: I also try p=0,d=1,q=2,P=2,D=0,Q=2,S=12 as model 2

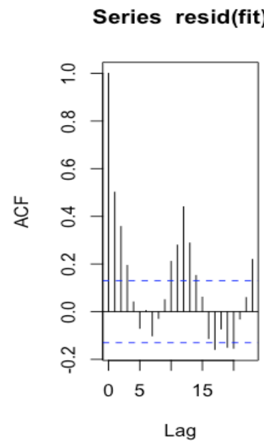


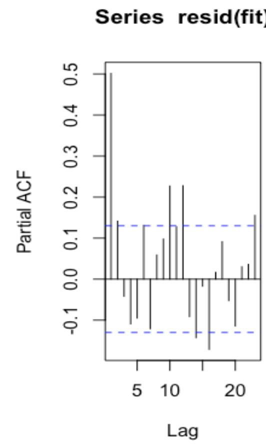


In the p-value plot, I think model 2 is better since we fail to reject the null hypothesis. I use AIC to choose my model, the AIC of ARIMA(0,1,2)= -2.064370, and the AIC of SARIMA(0,1,2)*(2,0,2)[12] = -2.316904, science AIC of SARIMA(0,1,2)*(2,0,2)[12] is smaller, I chose it.

Explore regression:







Linear regression modelling is a powerful tool to explore causal relationships between events and predict future outcomes. The fitted model is y(firearms related deaths) = -1.73 + 0.4064*x(monthly handgun sales), and the ACF and PACF residuals of fitted model is tails off.

Conclusion & Discussion:

Through my analysis, first of all, the monthly handgun sales and firearms related deaths in California follows a time series, and their changes are related to time. From 1980 to 1998, monthly handgun sales and firearms related deaths were stable at first, and then declined. So they're not from stationary, and then after I transform the data by taking differences of a time series. Finally I found the best model of the monthly handgun sales is $\boldsymbol{\hat{x}_t = -0.3126w_{t-1} - 0.2623w_{t-2} + 0.5116x_{t-12} + w_t}$ and the firearms related deaths best model is $\boldsymbol{\hat{x}_t = -1.6573w_{t-1} + 0.6573w_{t-1} + 0.1986x_{t-12} + 0.8002x_{t-24} - 0.1966w_{t-12} - 0.756w_{t-24} + w_t}$. Moreover, When time difference is smaller than 5, the gun sales and death have negative correlation when the gun sale increase the firearms related deaths decrease. When time difference is larger than 5 the gun sales and death have positive correlation when the gun sale increase the firearms related deaths increase.

```
# explain the data
my_data <- read.table("~/Desktop/GD.dat.txt", quote="\"", comment.char="")
dim(my_data)
summary(my_data)
boxplot(my_data)
# understand the nature of the variations
par(mfrow=c(1,2))
gun<-my_data$V1
deaths<-my_data$V2
ts.plot(gun,gpars=list(xlab="time",
  ylab="gun", main="Figure 1.1 Time Series of GUN SALES", lty=c(1:3)))
ts.plot(deaths,gpars=list(xlab="time",
   ylab="firearm deaths", main="Figure 1.2 Time Series of DEATHS", lty=c(1:3)))
#stationary #transformation(s)
# A stationary series is one in which the properties –
#mean, variance and covariance, do not vary with time.
##transform data
##log
# I use log transform, but it is still non
gun_log<-log(gun)
plot.ts(gun_log,xlab="time",
       ylab="gun", main="Figure 2.1 LOG transformed GUN SALES")

deaths_log<-log(deaths)
plot.ts(deaths_log,xlab="time",
       ylab="firearm deaths", main="Figure 2.2 LOG transformed firearm deaths")

#take diff
#We can take differences of a time series as well.
#This is equivalent to taking the difference between
#each value and its previous value:
diff1<-diff(gun,lag = 1)
tm1<-cbind(gun,diff1)
head(tm1)
plot.ts(tm1,
       main="Figure 3.1 take differences transformed GUN")
diff2<-diff(deaths,lag = 1)
tm2<-cbind(deaths,diff2)
head(tm2)
plot.ts(tm2,main="Figure 3.2 take differences transformed deaths")
#Often in time series analysis and modeling, we will want to transform data. There are a number of different functions that can be used to
#transform time series data
#such as the difference, log, moving average, percent change, lag, or cumulative sum.
```

```
#These type of function are useful for both visualizing time series data and for modeling time series.
#For example, the moving average function can be used to more easily visualize a high-variance time series and is also a critical part the ARIMA family of models.
#Functions such as the difference, percent change, and log difference are helpful for
#making non-stationary data stationary.
#As you can see, taking a difference is an effective way to
#remove a trend and make a time series stationary.
#Fit an ARIMA model for each one of the series.
#Use a model selection criterion to select the best models.

##acf and pacf of gun
par(mfrow=c(2,1))
acf(diff(gun_log),48)
pacf(diff(gun_log),48)

##acf and pacf of death
par(mfrow=c(2,1))
acf(diff(deaths_log),48)
pacf(diff(deaths_log),48)

###ccf
par(mfrow=c(1,1))
ccf(diff(gun_log), diff(deaths_log),laf.max=20, ylab="ccf")

## modle of gun seal
par(mfrow = c(1,2))
m1_gun<-sarima(diff(gun_log), p=0, d=0, q=2)
m2_gun<-sarima(diff(gun_log), p=0, d=0, q=2,P=1,D=0,Q=0,S=12)

##modle of death
m1_deaths<-sarima(diff(deaths_log), p=0, d=1, q=3)
m2_deaths<-sarima(diff(deaths_log), p=0, d=1, q=2,P=2,D=0,Q=2,S=12)


# modle selection
c(m1_gun$AIC,m1_gun$AICc,m1_gun$BIC)
c(m2_gun$AIC,m2_gun$AICc,m2_gun$BIC)


c(m1_deaths$AIC,m1_deaths$AICc,m1_deaths$BIC)
c(m2_deaths$AIC,m2_deaths$AICc,m2_deaths$BIC)
##？ questions : if we do the log transformation the data is still non stationary
## after we de the diff of log data ot orangional data??

##for the gun and death data , do we have the sensonal or not
```

```
# if we use arima modle, we neef to use data or diff data



sarima(diff(deaths_log),2,1,2,1,0,0,12,xreg=rbind(diff(gun_log)))

trend=time(deaths_log)
fit = lm(deaths_log~gun_log)
summary(fit)

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