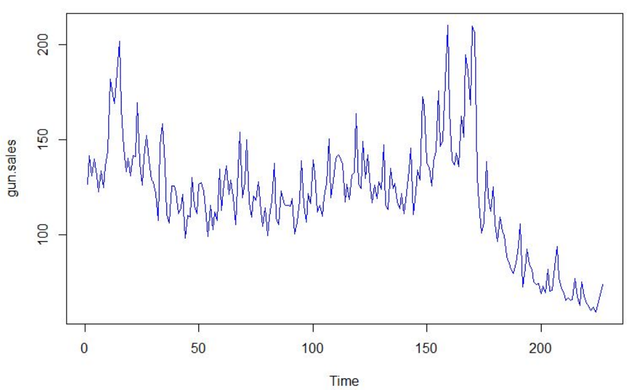
Jordan Brennan

STA 137 - Final Project

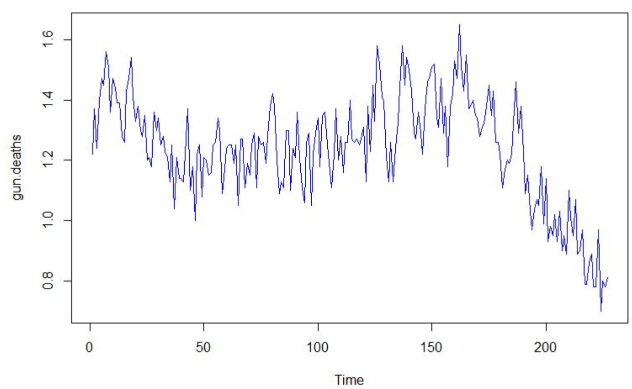
Introduction:

The purpose of this study is to analyze the dataset - GD.dat, that contains the monthly handgun sales and firearms related deaths in California from (1980-1998).The reason that the data is a time series data because each column can be plotted over time (18 years), and the data arrives in time order. It is important to analyze because we are able to see trends, variances, and be able to make important predictions for the future based on the past. To better understand these time series, let's first start by looking at the time series plots, acf, and the distribution of the residuals.

Gun sales

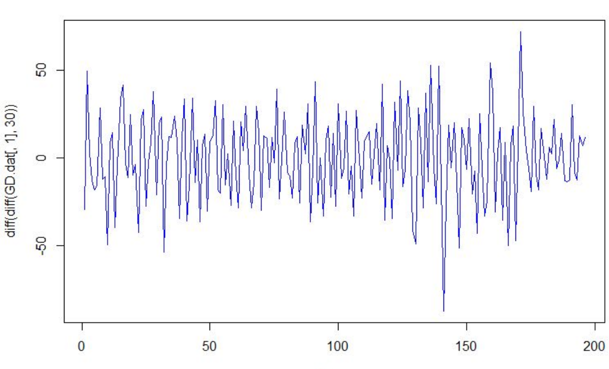


Deaths

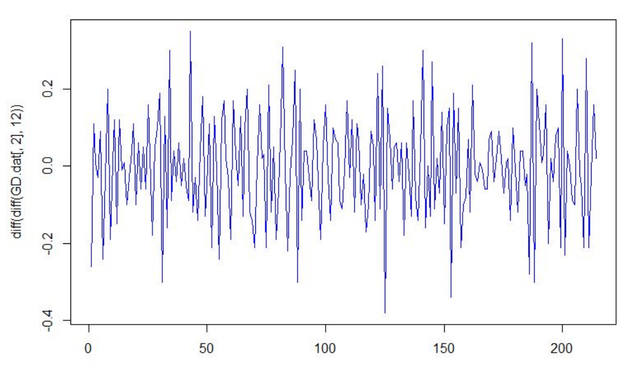


From the plots above we can see that the mean and variance of both gun sales and deaths are not consistent over time, this implies that the series is not stationary. In order to make these time series stationary, we need to apply some kind of transformations. I first decided to use differencing on both time series to eliminate trend. After differencing the time series only once, it is time to assess the seasonal component. The time series both respond differently to seasonal differencing, which is why I used a seasonal operator of 30 for gun sales, but 12 for deaths. The reason why I chose these values is because when looking at ACF Gun sales, the values start becoming negative after lag 30. This isn't the same as deaths because the periodic trend of deaths can be broken down by 12 months of each year. The reason I went with 30 for Gun sales is because it yields a better fitting ARIMA model, which is discussed later.

Gun sales



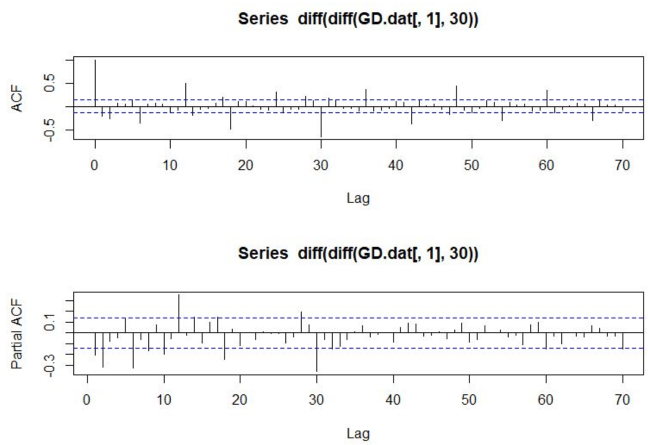
Deaths



Methodology:

Next, we can analyze the ACF and PACF of the Gun sales time series in order to help suggest a possible model to fit the data - ARIMA(p,d,q) x (P,D,Q)s. Below are the plots of the acf and pacf of the time series after making the transformations.

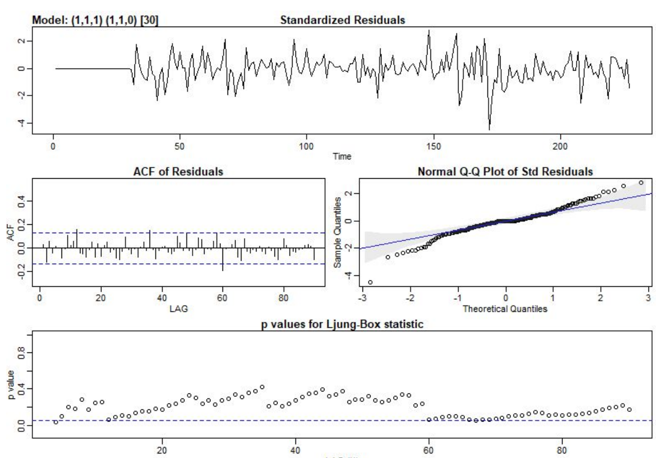
Gun sales



When looking at the ACF and PACF to decide a model, the first step is to identify the seasonal component. For gun sales, the appropriate seasonal component would be an AR(1) model since ACF seems to tail off and PACF cuts off. Now to identify the ordinary component, it can be seen in between the seasonal time lags in the ACF and PACF, and in this case looks like an ARMA (1,1). Now the complete suggested ARIMA model: ARIMA(1,1,1) x (1,1,0)s =30.

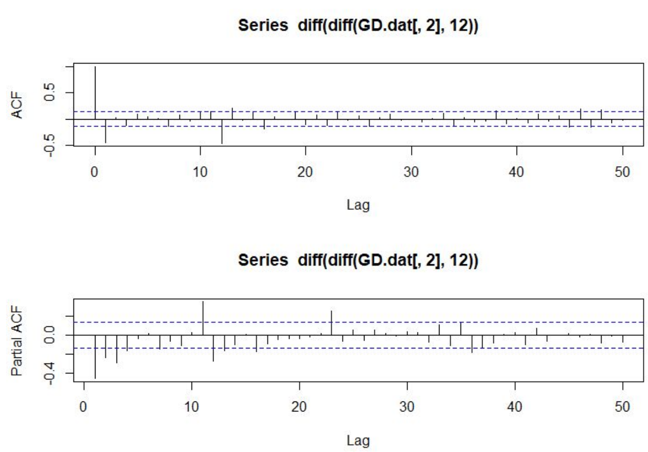
Results:

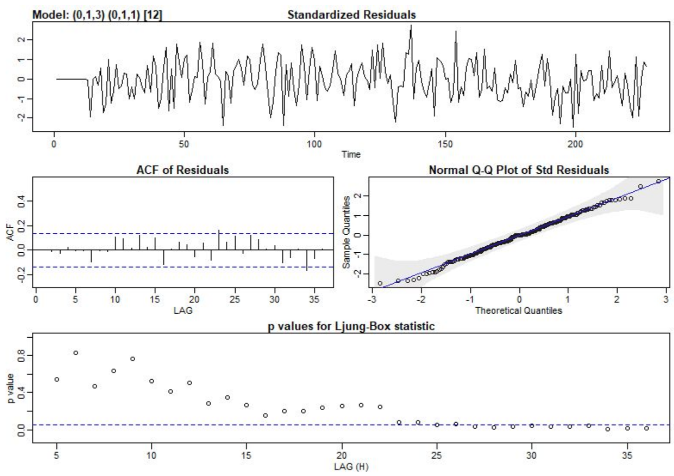
To see if this model is a good fit for the data, we need to look at the ACF of residuals, Normal Q-Q plot, and the p values for Ljung-Box statistic.



According to the plots, the ARIMA(1,1,1) x (1,1,0)s =30 model seems to be a good fit, not perfect, but will do. To be certain, I used other models and compared model selection criterion. This model gave the lowest values of AIC, AICc, and BIC respectively: 7.428133 7.428616 7.486411. We choose AIC as the lowest.

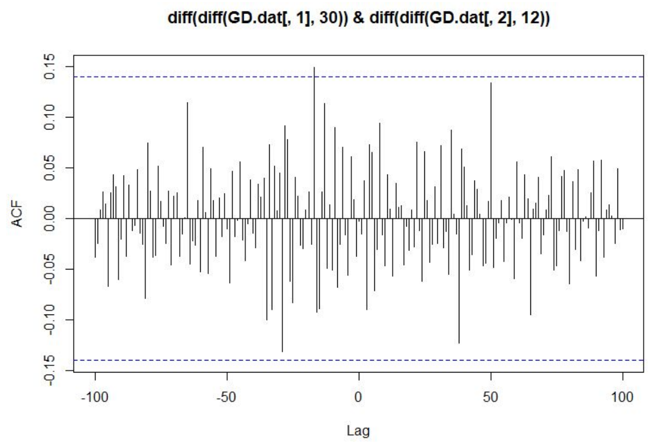
Next, let's analyze the ACF and PACF of the deaths time series in order to find a model that fits the data. For deaths, the appropriate seasonal component would be an MA(1) model since ACF seems to tail off and PACF cuts off. Now to identify the ordinary component, it can be seen in between the seasonal time lags in the ACF and PACF, and in this case looks like a MA (0,3). Now the complete suggested ARIMA model: ARIMA(0,1,3) x (0,1,1)s =12.





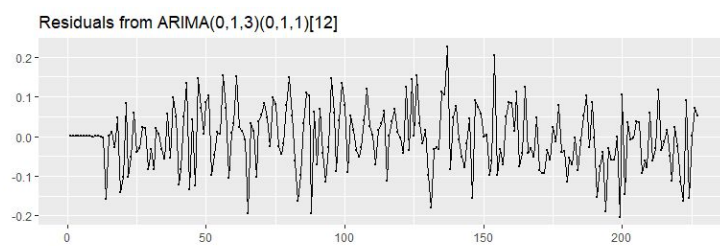
According to the plots, the ARIMA(0,1,3) x (0,1,1)s =12 seems to be a fairly decent fit. To be certain, I used other models and compared model selection criterion. This model gave the lowest values of AIC, AICc, and BIC respectively: -1.853644, -1.852836, and -1.778845. Choose BIC as the lowest.

Next, the CCF can be computed since both series are stationary and gives the following plot:



From the CCF, we can establish that not only are each time series stationary, but they also have joint stationarity since the CCF plot shows that only one time lag is beyond the boundary. This means that the CCF is a function of time difference only.

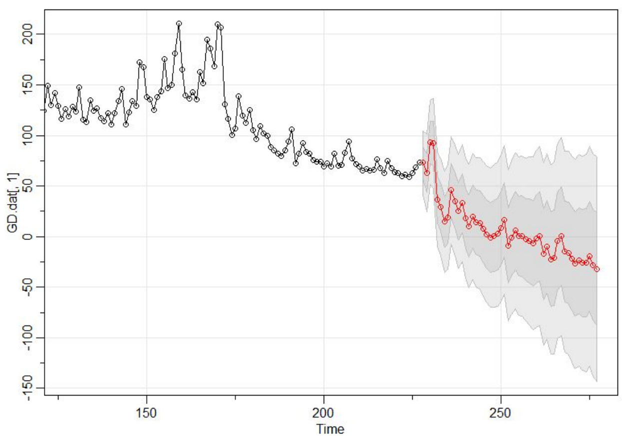
Now, the regression with autocorrelated errors can be used to explore the relationship of firearms deaths to handgun sales. The following plot is of the residuals of the linear model, showing the strength of the linear relationship.



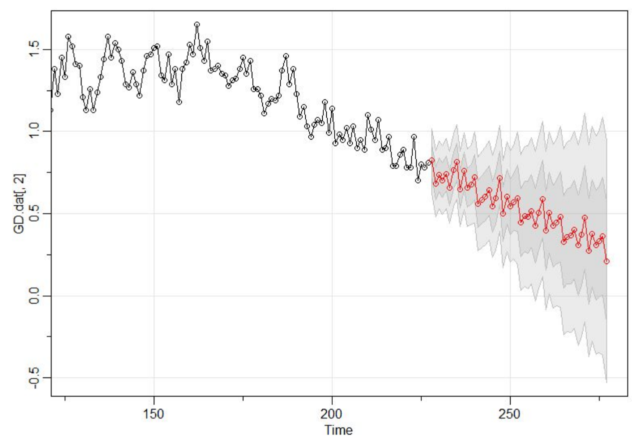
Conclusion and Discussion:

The gun sales and firearms deaths dataset initially was a non-stationary time series, but after transforming the data, it became stationary and able to analyze. The ACF and PACF of the transformed data helped aid in selecting an appropriate ARIMA model to fit the data. The ARIMA(1,1,1) x (1,1,0)s=30 ended up being the best model for the gun sales data and ARIMA(0,1,3) x (0,1,1)s=12 is the best model for firearms deaths. We also established that the two time series are jointly stationary by calculating the CCF after transformation. Lastly, we constructed a linear model with regression with autocorrelated errors to explore the linear relationship between gun sales and firearms deaths. After the time series analysis that has been done on this dataset, we can show a relationship between gun sales and firearms deaths and can make predictions, as shown in the graphs below.

Gun sales



Deaths



Appendix:

```
attach(GD.dat)
dim(GD.dat)

library(astsa)
library(ggplot2)

gun.sales = GD.dat[,1]
deaths = GD.dat[,2]
ts.plot(gun.sales, col =4)
ts.plot(deaths, col =4)

par(mfrow=c(3,1))
ts.plot(GD.dat[,1], col =4)
ts.plot(diff(GD.dat[,1]), col=4)
ts.plot(diff(diff(GD.dat[,1],30)), col=4)

par(mfrow=c(2,1))
acf(diff(diff(GD.dat[,1],30)), lag.max = 70)
pacf(diff(diff(GD.dat[,1],30)), lag.max = 70)

m1=sarima(GD.dat[,1], p=1,d=1,q=1, P=1,D=1,Q=0,30)
m2=sarima(GD.dat[,1], p=1,d=1,q=1, P=1,D=1,Q=0,12)
m3=sarima(GD.dat[,1], p=3,d=1,q=1, P=1,D=1,Q=0,30)
c(m1$AIC,m1$AICc,m1$BIC)
c(m2$AIC,m2$AICc,m2$BIC)
c(m3$AIC,m3$AICc,m3$BIC)
```

```
par(mfrow=c(2,1))
ts.plot(GD.dat[,1], col =4)
sarima.for(GD.dat[,1], 50, 1,1,1, 1,1,0,30)

par(mfrow=c(3,1))
ts.plot(GD.dat[,2], col=4)
ts.plot(diff(GD.dat[,2]), col=4)
ts.plot(diff(diff(GD.dat[,2],12)), col=4)

par(mfrow=c(2,1))
acf(diff(diff(GD.dat[,2],12)), lag.max =50)
pacf(diff(diff(GD.dat[,2],12)), lag.max =50)

M1 =sarima(GD.dat[,2], p=0,d=1,q=1, P=1,D=1,Q=0,12)
M2 =sarima(GD.dat[,2], p=1,d=1,q=1, P=0,D=1,Q=1,12)
M3 =sarima(GD.dat[,2], p=0,d=1,q=3, P=0,D=1,Q=1,12)
c(M1$AIC,M1$AICc,M1$BIC)
c(M2$AIC,M2$AICc,M2$BIC)
c(M3$AIC,M3$AICc,M3$BIC)
c(M4$AIC,M4$AICc,M4$BIC)

par(mfrow=c(2,1))
ts.plot(GD.dat[,2], col =4)
sarima.for(GD.dat[,2], 50, 1,1,1, 1,1,0,12)

ccf(diff(diff(GD.dat[,1],30)),diff(diff(GD.dat[,2],12)), lag.max = 100)


library(forecast)
trend = time(GD.dat[,2])
fit = lm(deaths~ trend+sales, na.action = NULL)
acf2(resid(fit), 60)
checkresiduals(fit)

fit2 = sarima(GD.dat[,2], p=0,d=1,q=3, P=0,D=1,Q=1,12)
acf2(resid(fit2$fit), 60)
checkresiduals(fit2$fit)
```