STA 137 Final Project

By John Tran

916232852

Introduction: Statement of the problem.

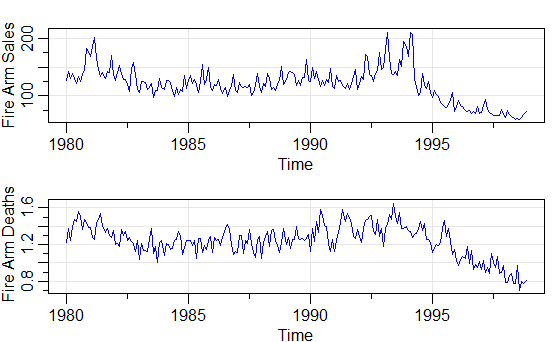
In this report, I analyze two time series on gun sales and gun deaths. The data given to me was presented in two columns, each column presenting a time series. This is all monthly data so each row represented the amount of gun sales and firearm deaths per month. This data has been recorded from 1980 -1998 in California. With these two time series, I analyze any correlation between the two time series and try to prove a relationship if possible using Rstudio, to present my results. I look at the ACF, PACF, CCF, stationarity, and fit the best ARIMA model

Material and Methods: Description of your data and the methods you have used for the analysis.

In the data set given, it is a data frame with two columns V1 and V2. V1 representing gun sales and V2 representing fire-arm related deaths. I converted the columns into time series using “astsa” package in R. This package is primarily used to analyze time series. Using the ts function on each column, I aligned the columns to the dates years starting from January 1980 to November 1998. Gun sales was named ts\_sales and gun deaths was named ts\_deaths.

**Plotting the time series**

I used to tsplot function to plot both time series. Here are the plots. I plot them to check stationarity.

Figure 1

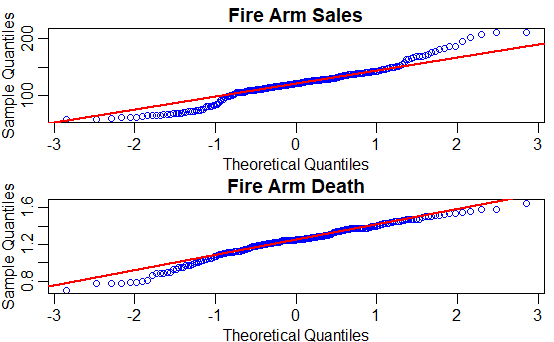
**Checking if the data is normal** 

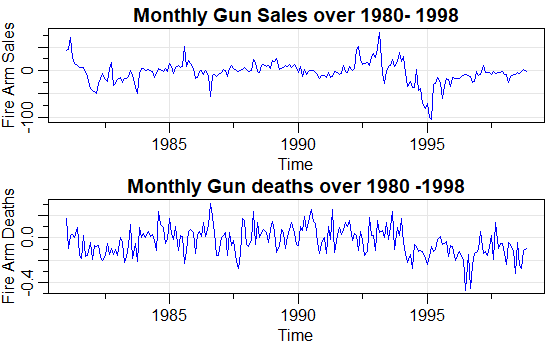
Figure 2

I used the qqnorm function to check if the time series are normally distributed.

Based on these plots, the gun sales time series shows a heavy tail meaning there is more extreme values in the end so the qqplot. While on the other hand, the fire-arm deaths is less heavy tail and is more linear with the line means that the data is not as extreme as the other one.

**Differencing the Time Series**

I used the diff function to difference the time series because the time series was not stationary due based on the ts plot I plotted above. The expected mean did not look constant. Plus there was a seasonality and a trend. So I decided to diff(ts\_sales, lag =12) and diff(ts\_deaths , lag =12) to add seasonal differences to make the time series stationary, and to move the trend I difference it one more time and it resulted in these plots in the results section.



**Calculating PACF and ACF,**

The reasoning behind to compute the PACF and ACF is to show us the autocorrelation values over the time lag to confirm what the method we did prior. To show if there is still trend or seasonality. I computed the ACF and PACF of the original time series and the difference time series that I created previously and compared the ACF and PACF to each other for seasonality or trend.

**Calculating CCF**

I computed the CCF to determine if a certain time series leads another such as sales time series leads death time series. To show whether or not it is future independent or not. The results and output would be listed below in the results section.

**Fitting the correct ARIMA model for each time series using model selection.**

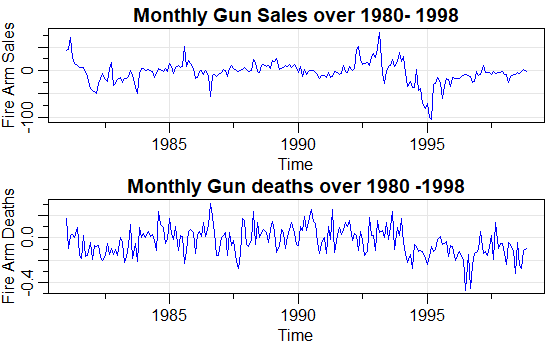
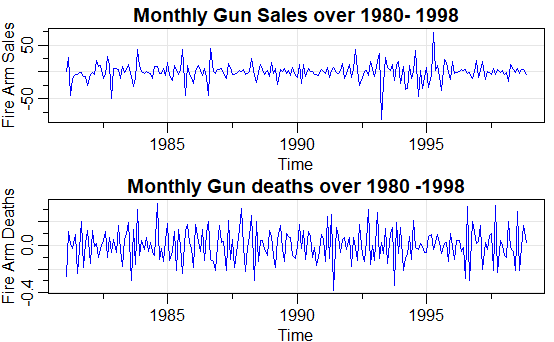
In this portion, I fitted each time series based on the ACF and PACF. I estimated the which order it was based on the ACF and PACF plots previously. Using model selection, I determine the best AIC and compare it to relative models I created before. I created about 3 models each for gun sales and gun deaths. After finding the lowest AIC, I determine that this was the correct sARIMA model for these time series.

**Create a linear regression model for autocovariance errors between gun deaths and gun sales**

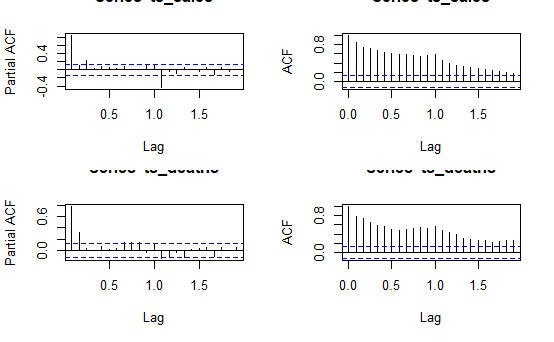
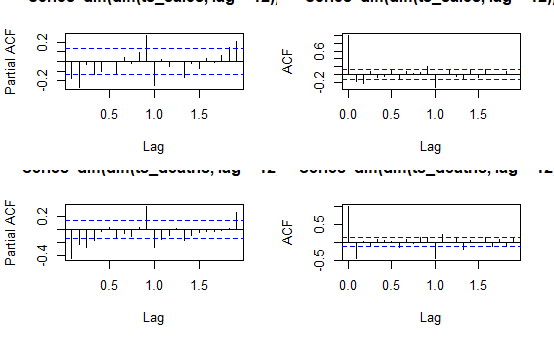
I want to know the behavior of the noise prior to gun sales time series. If the residuals shows any association with whiteness. I also created a trend and temp variables (trend = time(ts\_deaths); temp = ts\_sales - mean(ts\_sales)). So I created a linear model fit = lm(ts\_deaths ~trend + temp, na.action=NULL). And applied the acf function to the residuals to the fit. Based on that plot, I fitted an arima model to it to find the candidate model.

Results: Explanation of the results of your analyses. You can cut and paste some of your computer outputs and refer to them in explaining your results.

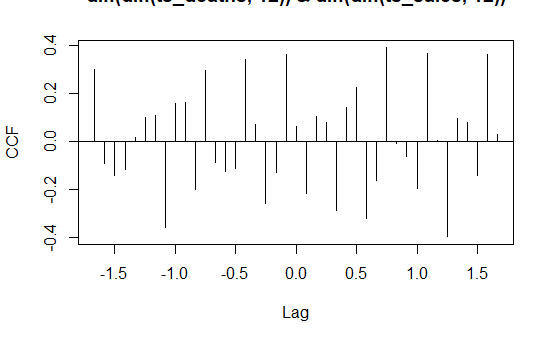
**Plots of the first difference and without seasonality**

Plot without seasonal component, but not detrended Plot without seasonal component, and detrended (differencing)

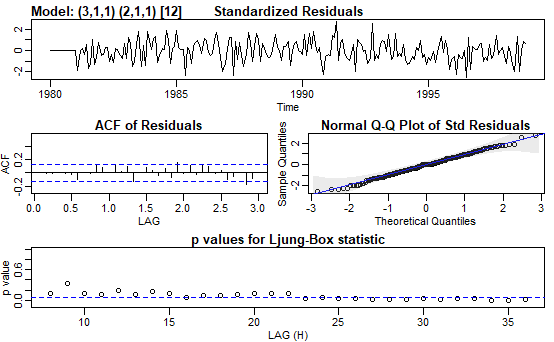
I can now compare the time series between these two and the original time series to show stationarity.

Nonstationary Time Series Stationary Time Series

Here I compared the results of the original time series and the transformed time series. I observe that there is seasonality based on the ACF it trends downward. The transformed ACF cuts off for both death and sales. Based on these ACF and PACFs the transform time series is most likely to be a sARIMA model because the seasonal lag of 12 and they both tail off.

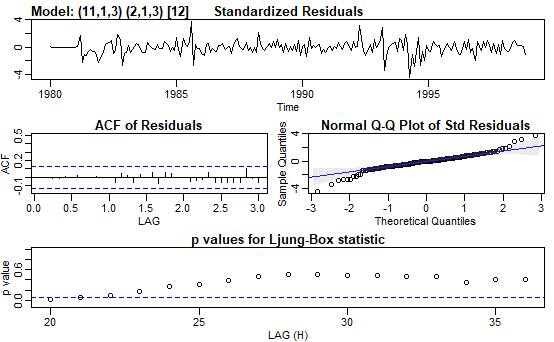
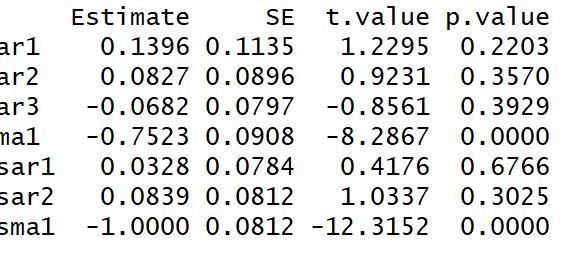
**CCF**

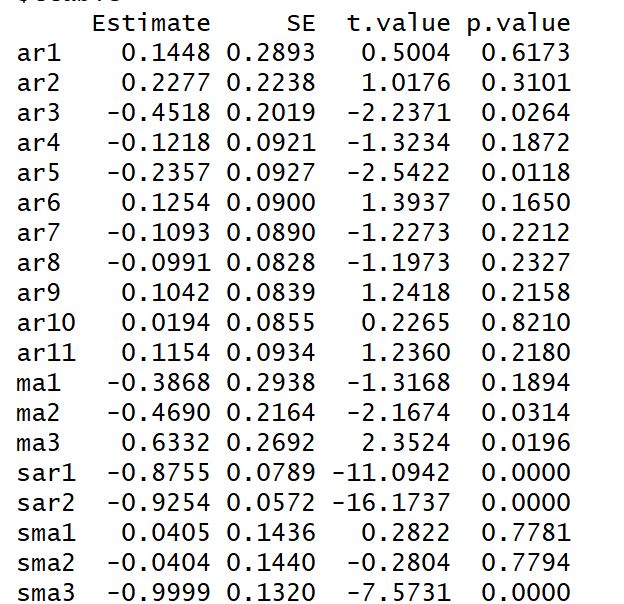
This is the CCF for the transformed time series. The peak is at time lag 9 positive side. That means the sales time series leads the death time series. This helps us predict the deaths.

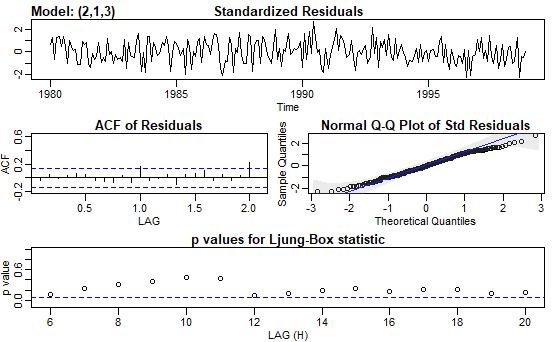
**Best sARIMA model**

The AIC is -1.8322 which was the lowest compared to most of the models for the death sarima model and the BIC is -1.712. Based on the results, the p- values are slightly above the line in the test statistics. This model is (3,1,1)(2,1,1)[12]. I believe this is the best model because it had the lowest AIC.

The residuals look random like noise.

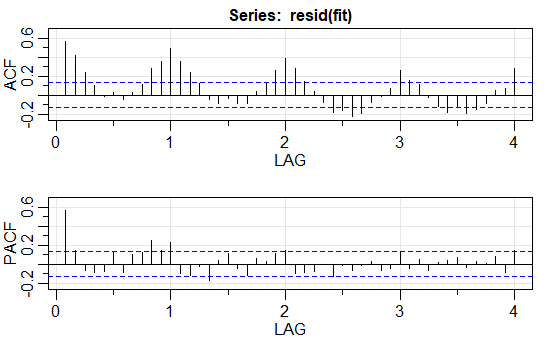
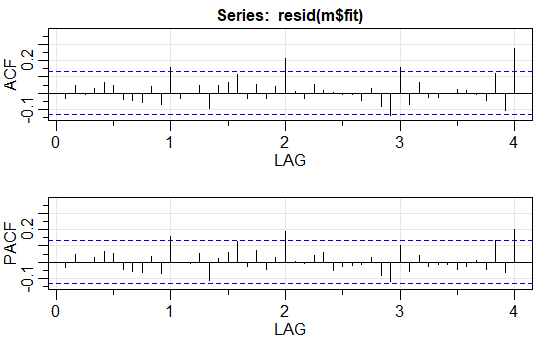


The AIC is 7.55 which was the lowest compared to most of the models for the gun sales time series and the BIC is 7.85. I believe the model l(11,1,3,)(2,1,3)[12] is the best model.

Autocorrelated Regression

I tested out the residuals for the linear model. However, the residuals were correlated which violates an assumption of linear regression. So I created a new sArima and found a model with its AR not correlated and it looks like white noise.

The model is sARIMA with (2,1,3).



Conclusion and Discussion: Highlighting the main points and discussing.

In conclusion, there is a correlation between gun deaths and firearm sales. At first, I detrended and remove both seasonal components through differencing. In this report, I created a fitted model based on the estimates I chose through model selection. Then I utilized regression to check any effects of gun sales on deaths.I analyzed the satisfactory model and found the residuals to be similar to white noise. Based on my results this shows a correlation between the two time series..

Appendix: Include all the code you used to analyze the data in the appendix.