Final Project

Eric Folsom, Amrit Singh

12/5/2021

For our project, we chose to forecast the future average earnings of drift gillnet vessels in Bristol Bay, Alaska. The data that we obtained was from the Commercial Fishing Entries Commission (CFEC) of Alaska. The dataset is a collection of data from each year of fishing in Bristol Bay Alaska from 1975-2020. The time period in which this data is collected is in a two month span from June through July each year. This forecasting is significant to our group because Eric commercial fishes in Bristol Bay. This data is also significant because we can use it to forecast the future health of the fishing industry.

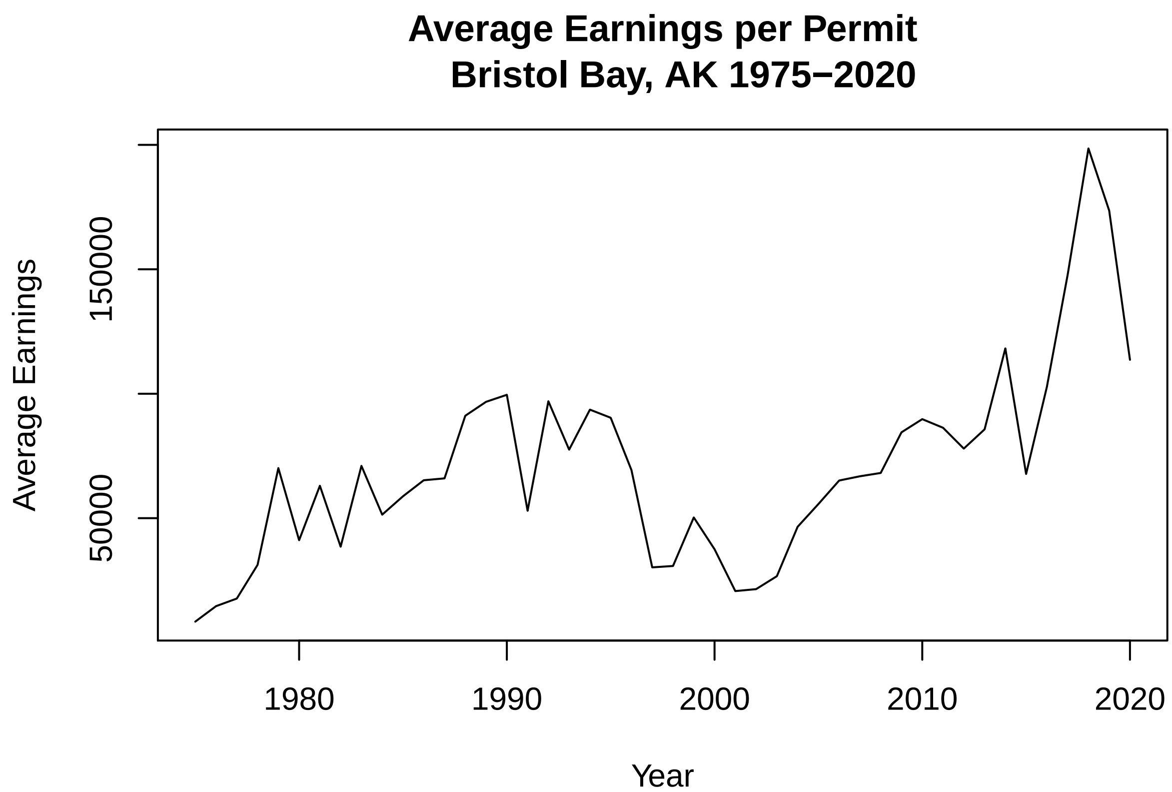
The dataset that we used is a collection of all the commercial fishing data from the state of Alaska. Because of this, we had to filter our dataset to only include data from the drift gillnet fishery in Bristol Bay. To begin forecasting, we had to use the astsa and the forecasting package in R.

We began our forecasting by taking a look at the original dataset.

```
library(forecast)
```

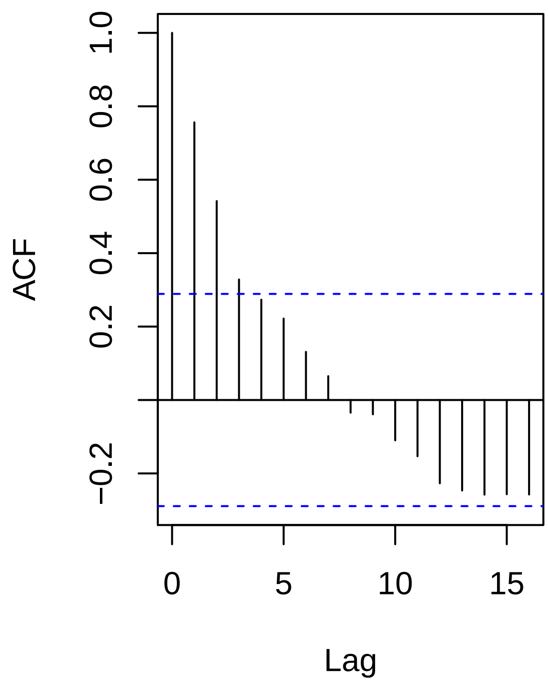
```
## Warning: package 'forecast' was built under R version 4.0.5

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

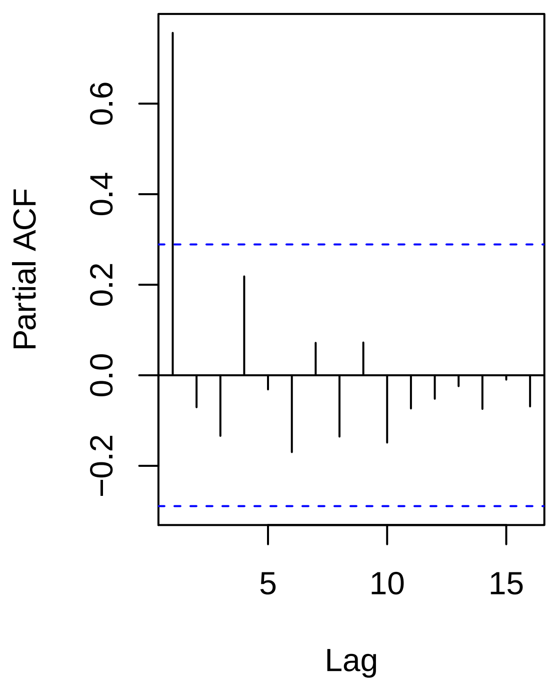
```
setwd("C:/Users/Eric Folsom/Desktop/School work/Stats 456")
data=read.csv("fishdat.csv")
earnings=rev(data$Average.Earnings)
salmonts=ts(earnings, start=1975)
plot(salmonts, xlab="Year", ylab="Average Earnings", main="Average Earnings per Permit
    Bristol Bay, AK 1975-2020")
```



We will also inspect the ACF and PACF plots of this series.

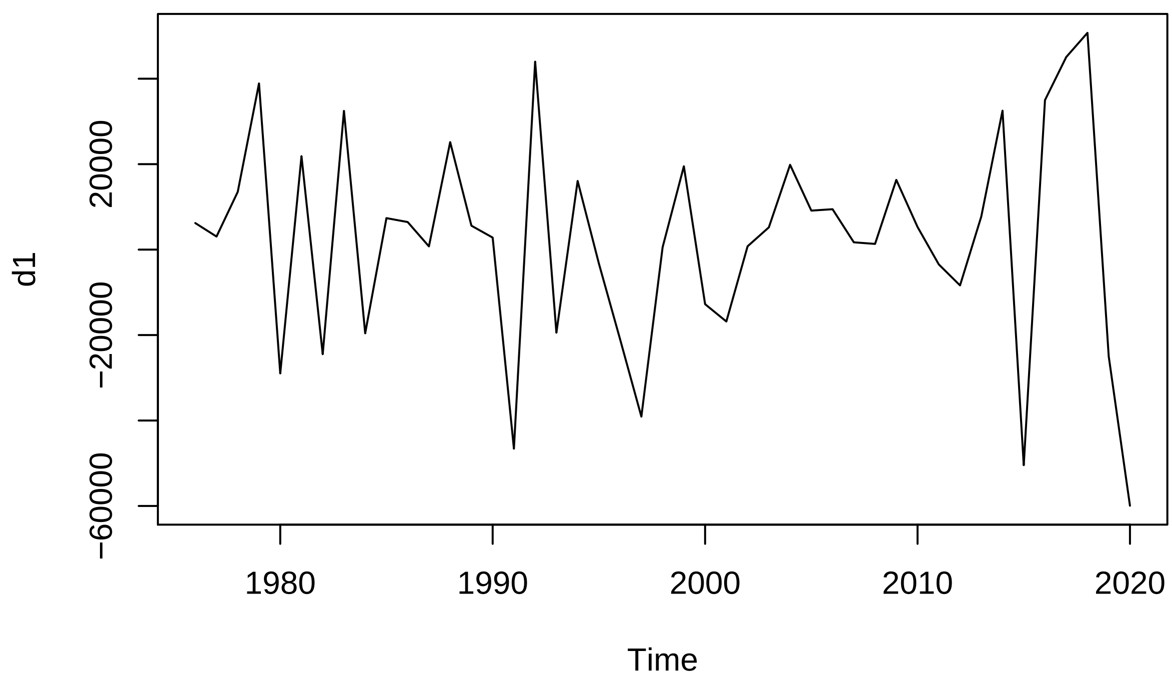
```
par(mfrow=c(1,2))
acf(salmonts, main="")
pacf(salmonts, main="")
```





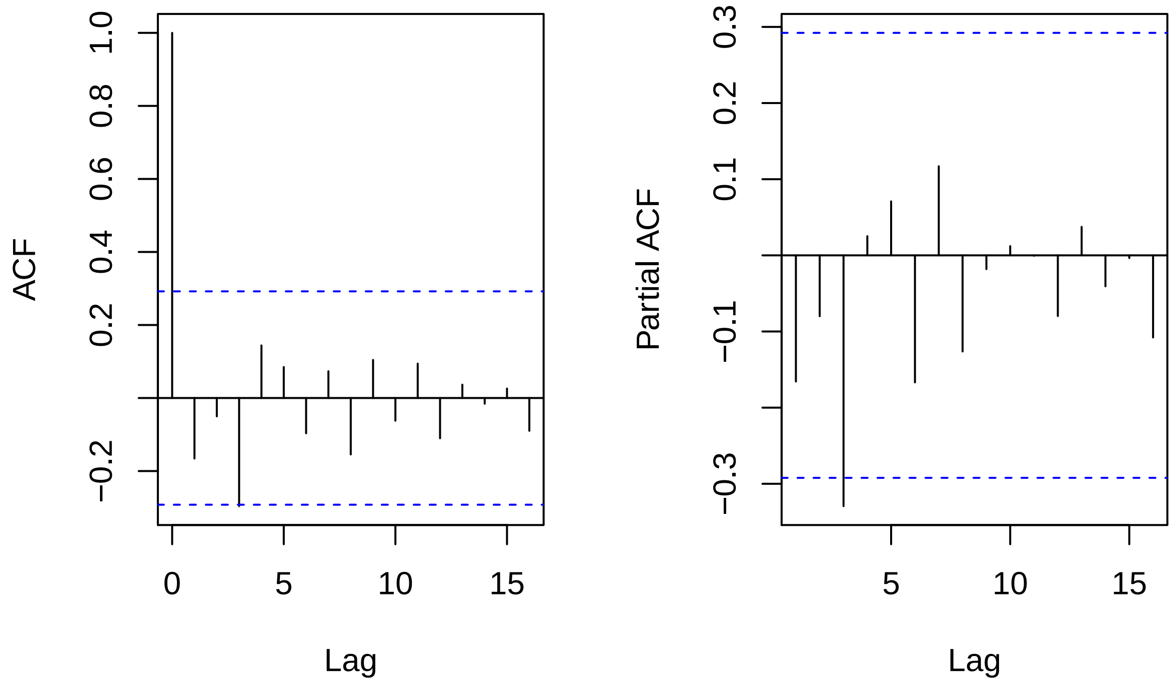
From the initial plot, can see that there is a trend in this time series. Additionally, the ACF shows a slow decay to zero which indicates that a possible remedy will be to difference the time series at lag 1.

```
d1=diff(salmonts)
plot(d1)
```



After differencing we can see that the series is now stationary. We then inspected the ACF and PACF to determine which model we will use for forecasting.

```
par(mfrow=c(1,2))
acf(d1, main="")
pacf(d1, main="")
```



From the ACF and PACF plots we can see that this series could be modeled by either an ARIMA(3,1,0), or an ARIMA(3,1,3). We will test both to see which model will be a better forecasting tool.

```
fit1=arima(salmonts,order=c(3,1,0))
fit2=arima(salmonts,order=c(3,1,3))
fit1
```

```
##
## Call:
## arima(x = salmonts, order = c(3, 1, 0))
##
## Coefficients:
##           ar1      ar2      ar3
##       -0.1774  -0.1486  -0.3617
## s.e.   0.1505   0.1525   0.1539
##
## sigma^2 estimated as 5.35e+08:  log likelihood = -516.28,  aic = 1040.56
```

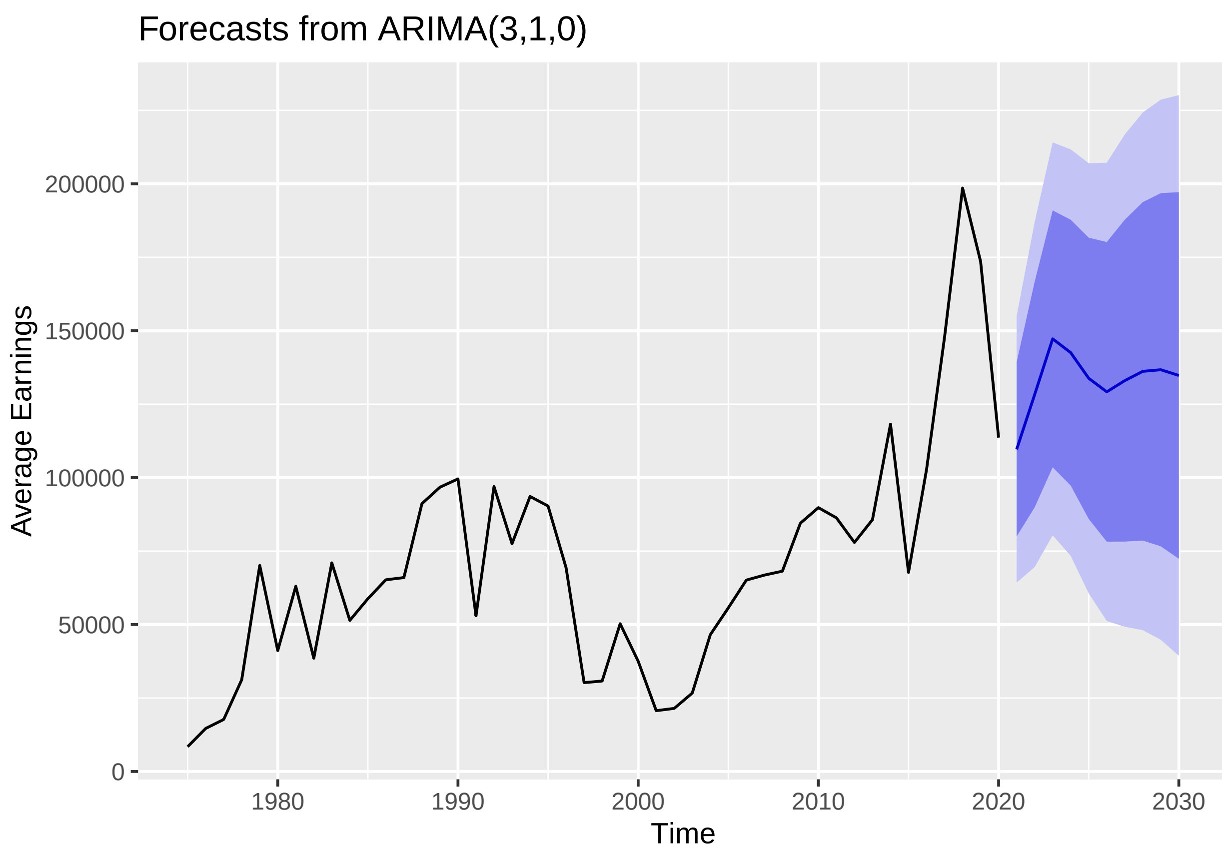
```
fit2
```

```
##
## Call:
## arima(x = salmonts, order = c(3, 1, 3))
##
## Coefficients:
##           ar1      ar2     ar3     ma1    ma2      ma3
##       -0.7171  -0.3545  0.1375  0.6137  0.267  -0.6015
## s.e.   0.3568   0.4203  0.3055  0.3390  0.402   0.3380
##
```

```
## sigma^2 estimated as 472347488:  log likelihood = -515.18,  aic = 1044.36
```

After looking at the AIC's of our model, we can see that the ARIMA(3,1,0) model fits our data the best. This is because the AIC is lower than the ARIMA(3,1,3) and it also includes less parameters than the ARIMA(3,1,3). Because of this we will choose an ARIMA(3,1,0) model for forecasting.

```r
fit1_fore=forecast(fit1)
autoplot(fit1_fore, ylab="Average Earnings")
```

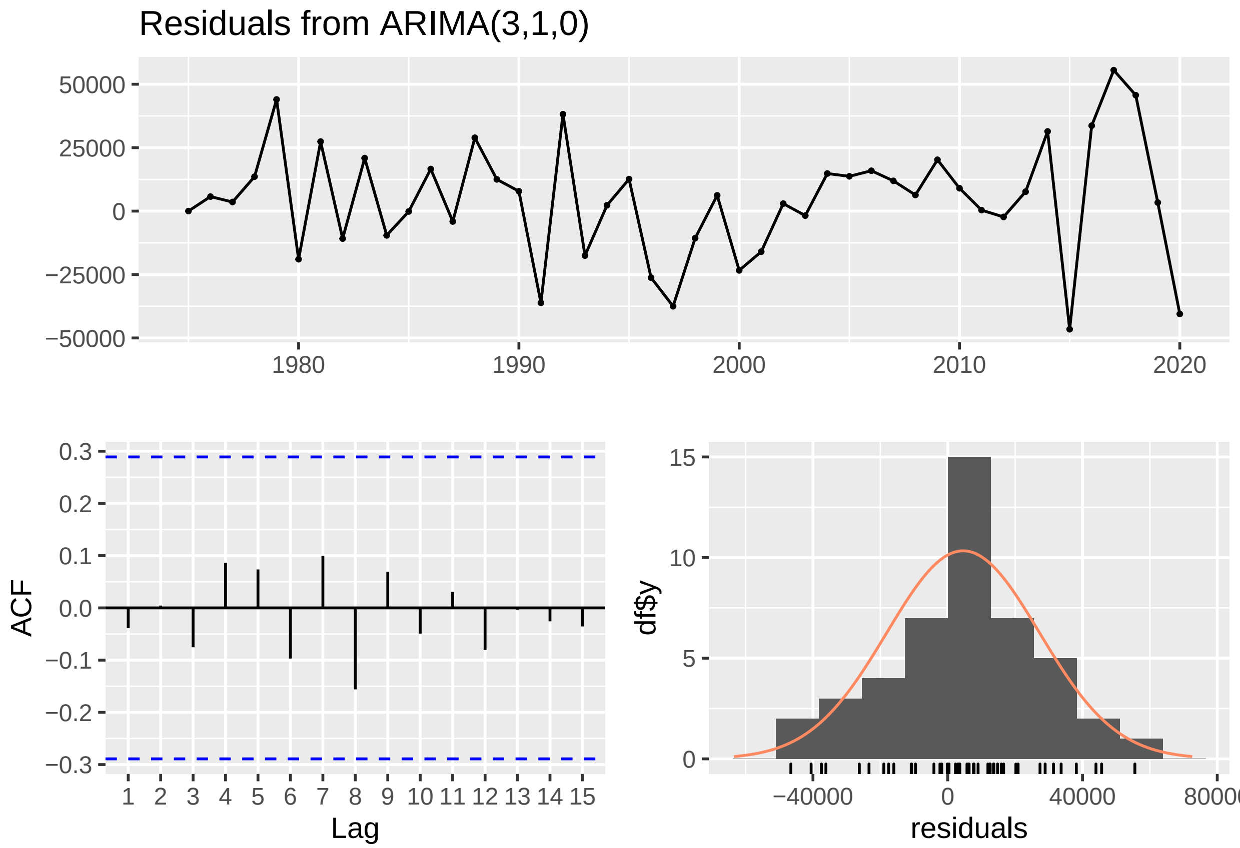



```r
fit1_fore
```

```
##      Point Forecast      Lo 80    Hi 80    Lo 95    Hi 95
## 2021       109621.3  79978.28 139264.4 64286.20 154956.5
## 2022       128273.2  89889.44 166656.9 69570.31 186976.1
## 2023       147240.8 103529.27 190952.3 80389.80 214091.7
## 2024       142553.2  97331.52 187774.9 73392.59 211713.9
## 2025       133820.9  85964.23 181677.7 60630.42 207011.5
## 2026       129206.1  78222.87 180189.3 51233.98 207178.2
## 2027       133017.4  78249.27 187785.5 49256.77 216778.0
## 2028       136185.2  78591.46 193778.9 48103.19 224267.2
## 2029       136726.2  76632.67 196819.6 44821.10 228631.2
## 2030       134781.1  72405.67 197156.6 39386.10 230176.1
```

After forecasting we then checked the residuals of the data to make sure that the residuals were IID and from a normal distribution.

```r
checkresiduals(fit1)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,1,0)
## Q* = 3.8308, df = 6, p-value = 0.6996
##
## Model df: 3.   Total lags used: 9
```

After checking the residuals we can see that the residuals are IID and come from a normal distribution.

This data comes from an aggregrate of data from the end of each commercial fishing season. The data, however, is limited by being broken down by region. Unfortunately, there is no data for the specific districts within each region. This causes issues because there are certain districts that perform better than others because the amount of fish that returns to each district varies based on the size and location of each district's river system. This flaw in the data is somewhat alleviated by the ability of drift gillnet vessels to fish in whichever district is performing the best. The data would also be improved if it included data from the 2021 fishing season. Unfortunately, the dataset ends after the 2020 season; this is an issue because the prices for the fish market were suppressed in 2020 due to COVID-19 effects on the global agriculture market. From personal experience, the prices rebounded in 2021 in a way that the model most likely cannot account for due to externalities. Because the data is collected at the end of the year, we are unable to forecast how much fishermen are able to earn during different parts of the fishing season. This would be useful for people who only fish for a portion of the season.

References:

https://www.cfec.state.ak.us/bit/X_S03T.htm

https://www.cfec.state.ak.us/PUBLIC/BIT.csv

Personal experience