NonSeasonal

Suyog

2023-04-03

Analyzing and Forecasting Globle CO2 Emissions using Time Series Analysis

Introduction:

Global carbon emissions refer to the amount of carbon dioxide (CO_2) and other greenhouse gases (GHGs) released into the atmosphere from various human activities, such as burning fossil fuels, deforestation, and agriculture. Carbon emissions are the primary driver of climate change, and reducing these emissions is critical to mitigating the impacts of climate change.

Step 1: Collecting Data

```
library(TSA)
```

```
##
## Attaching package: 'TSA'
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
library(tseries)
```

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

```
library(forecast)
```

```
## Registered S3 methods overwritten by 'forecast':
##   method       from
##   fitted.Arima TSA
##   plot.Arima   TSA
```

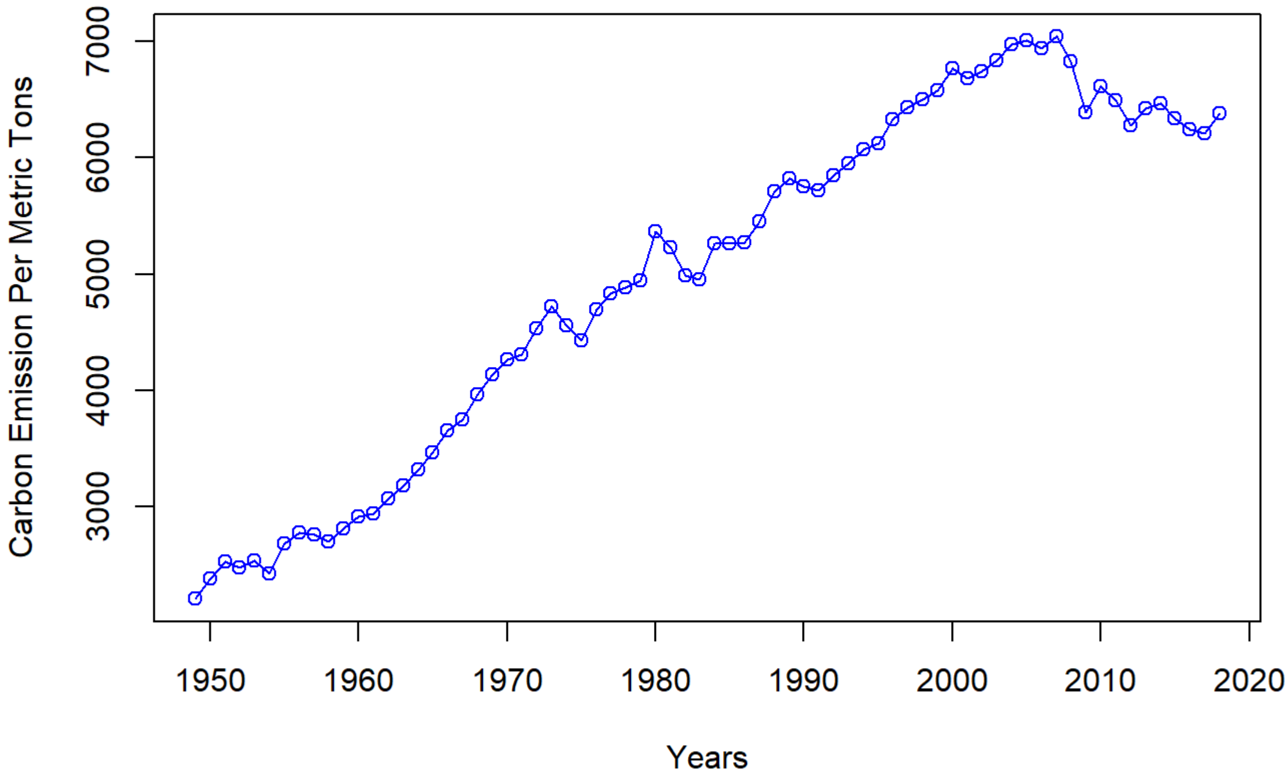
```
dataCo2 <- datasetCO2Emission <- read.csv("C:/Users/win/Desktop/datasetCO2Emission.txt", header=
FALSE)
names(dataCo2)[2] <- "Carbon_Emission"
names(dataCo2)[1] <- "Year"

tsdata <- ts(dataCo2[,2], start = c(1949, 1), frequency = 1)
```

```
#NO NA values in the dataset
sum(is.na(tsdata))
```

```
## [1] 0
```

```
plot(tsdata,ylab='Carbon Emission Per Metric Tons',xlab='Years',type='o' , col="blue")
```



The graph of global carbon emissions over time shows a steadily increasing trend, indicating a growing concern for climate change.

```
adf.test(tsdata)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  tsdata
## Dickey-Fuller = -0.4857, Lag order = 4, p-value = 0.9798
## alternative hypothesis: stationary
```

It is not stationary. Let us consider the First Difference.

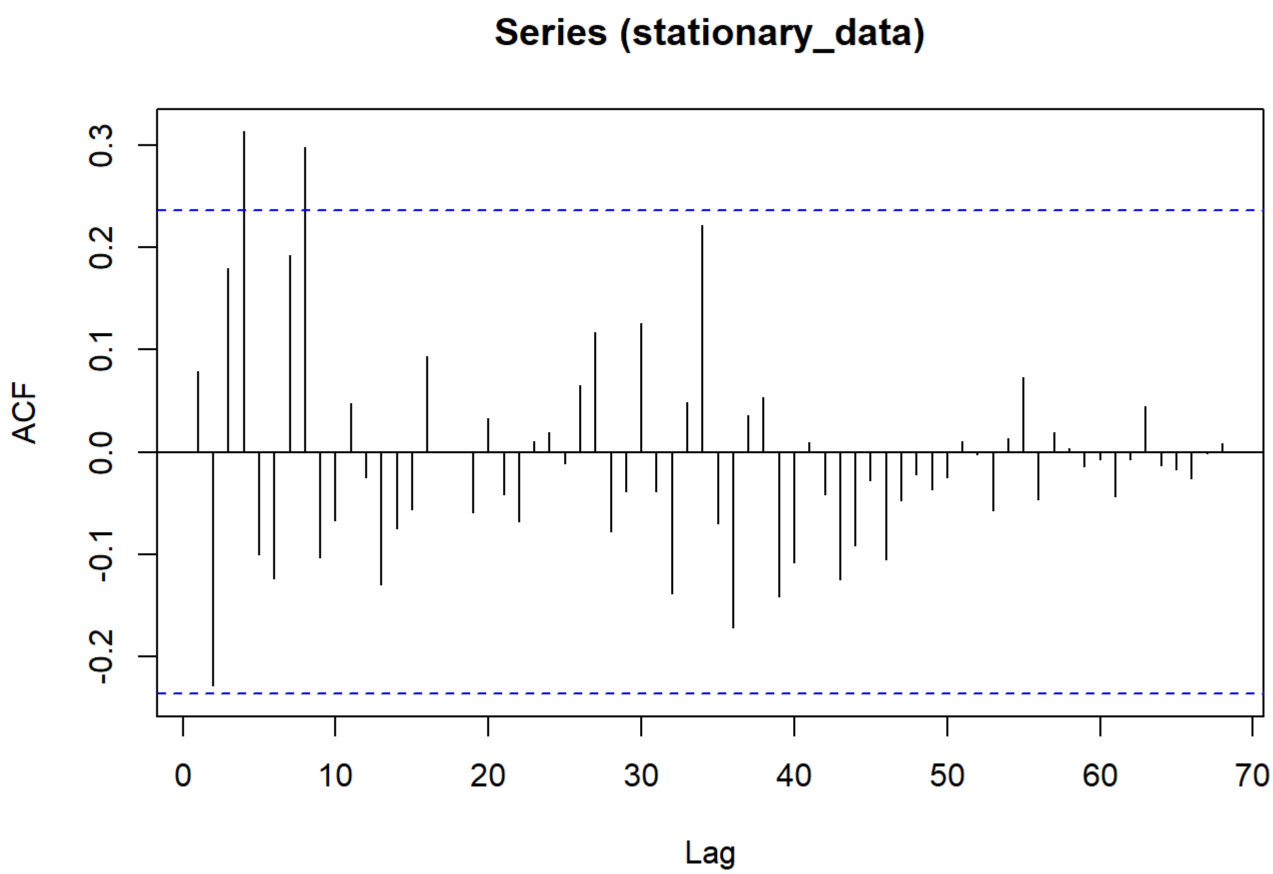
```
stationary_data=(diff(tsdata))
adf.test(stationary_data)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  stationary_data
## Dickey-Fuller = -3.5324, Lag order = 4, p-value = 0.0457
## alternative hypothesis: stationary
```

As the test Suggest data is Stationary after first difference.

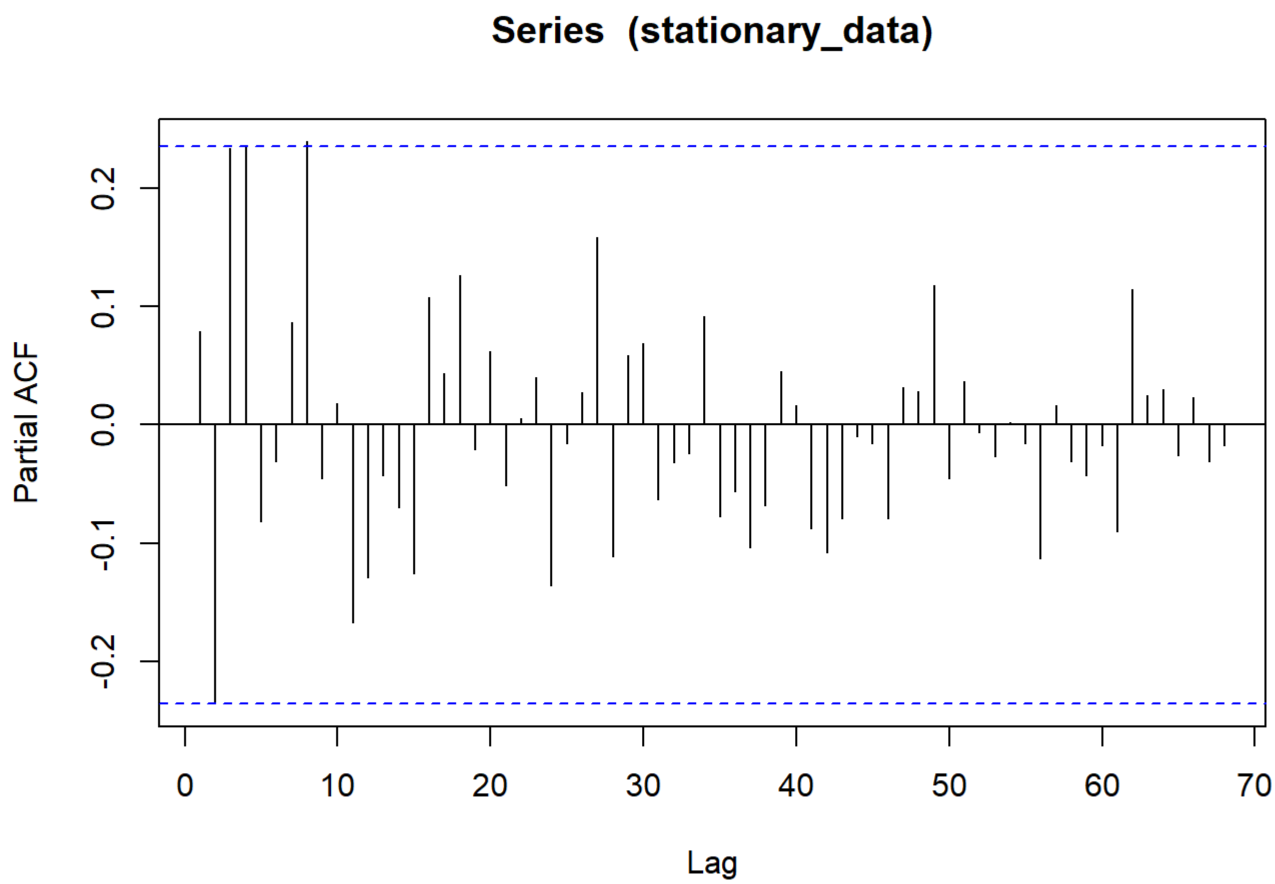
Step 2: Finding Models

```
acf((stationary_data),lag.max = 70)
```



We got Tailing off ACF. Now let us see the PACF

```
pacf((stationary_data),lag.max = 70)
```



PACF Tails off as well suggests that there might be an ARMA() process.

ACF and PACF suggest it is type of ARIMA(). Now our task is to get which ARIMA() it is?

```
eacf((stationary_data))
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 o o o x o o o x o o o  o  o  o
## 1 x o o x o o o x o o o  o  o  o
## 2 x o o o o o o x o o o  o  o  o
## 3 x x x o o o o x o o o  o  o  o
## 4 x o o o o o o x o o o  o  o  o
## 5 x o o o o o o x o o o  o  o  o
## 6 x o o x o o o x o o o  o  o  o
## 7 o o o x o o o x o o o  o  o  o
```

We have two models to consider ARIMA(2,1,3) and ARIMA(2,1,1)

Step 3/4: Parameter Reduency | Parameter Estimation

```
auto_arima_co2 <- auto.arima((stationary_data))
summary(auto_arima_co2)
```

```
## Series: (stationary_data)
## ARIMA(2,1,1)
##
## Coefficients:
##           ar1      ar2      ma1
##       -0.0950  -0.4092  -0.7860
## s.e.   0.1372   0.1278   0.1213
##
## sigma^2 = 19810:  log likelihood = -432.3
## AIC=872.6   AICc=873.23   BIC=881.47
##
## Training set error measures:
##                     ME    RMSE      MAE     MPE     MAPE     MASE         ACF1
## Training set -11.59272 136.608 98.00923 63.1395 123.1049 0.676612 -0.001885468
```

```
best=99999999999
best_Index=0
for (p in c(0,1,2,3,4,5)){
  for (q in c(0,1,2,3,4,5)){
    arima_model <- arima((stationary_data), order = c(p,1,q))
    #print(c(AIC(arima_model),p,1,q))
    if(AIC(arima_model)<best){
      best=AIC(arima_model)
      best_Index=c(p,q)
    }
  }
}
```

```
## Warning in log(s2): NaNs produced
```

```
cat("Best AIC so far", best_Index)
```

```
## Best AIC so far 2 3
```

By Brute-force we got ARIMA(2,1,3) model with lowest AIC score. We should consider this model for final fit.

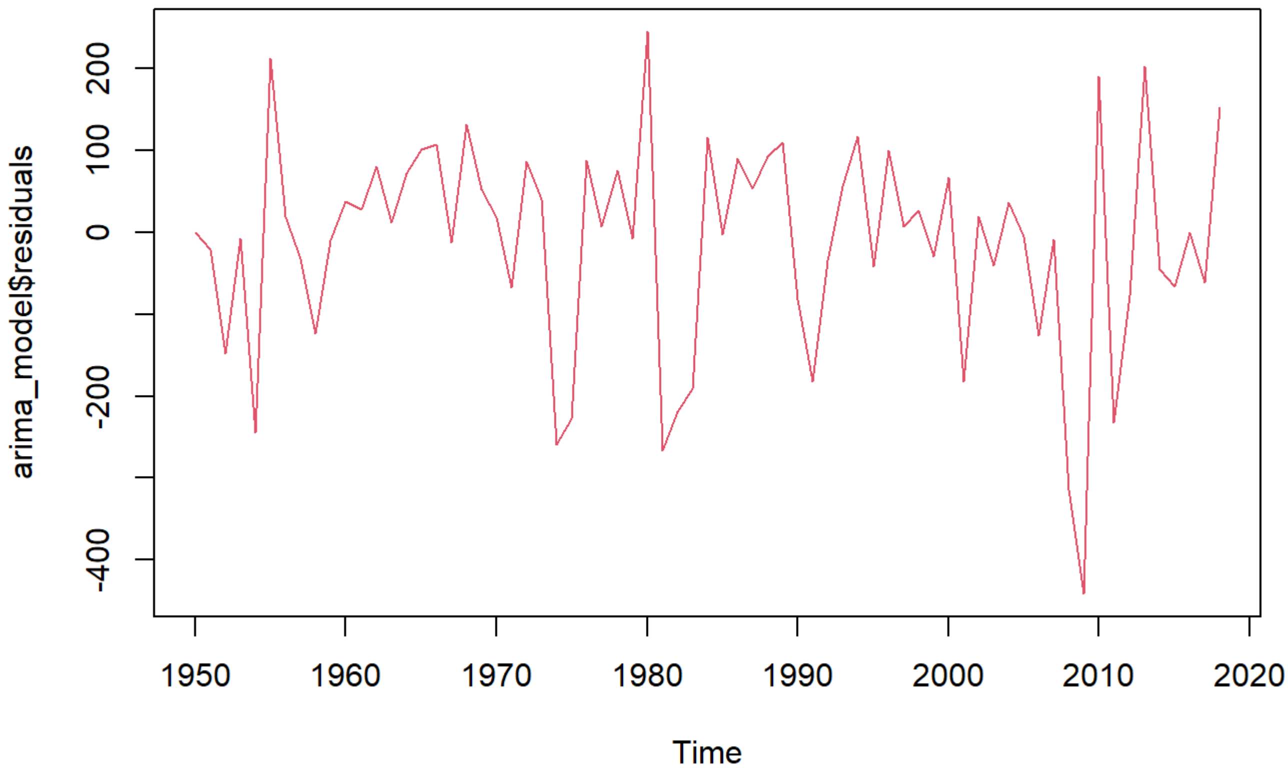
```
arima_model <- arima((stationary_data), order = c(0,1,0))
arima_model$aic
```

```
## [1] 912.0986
```

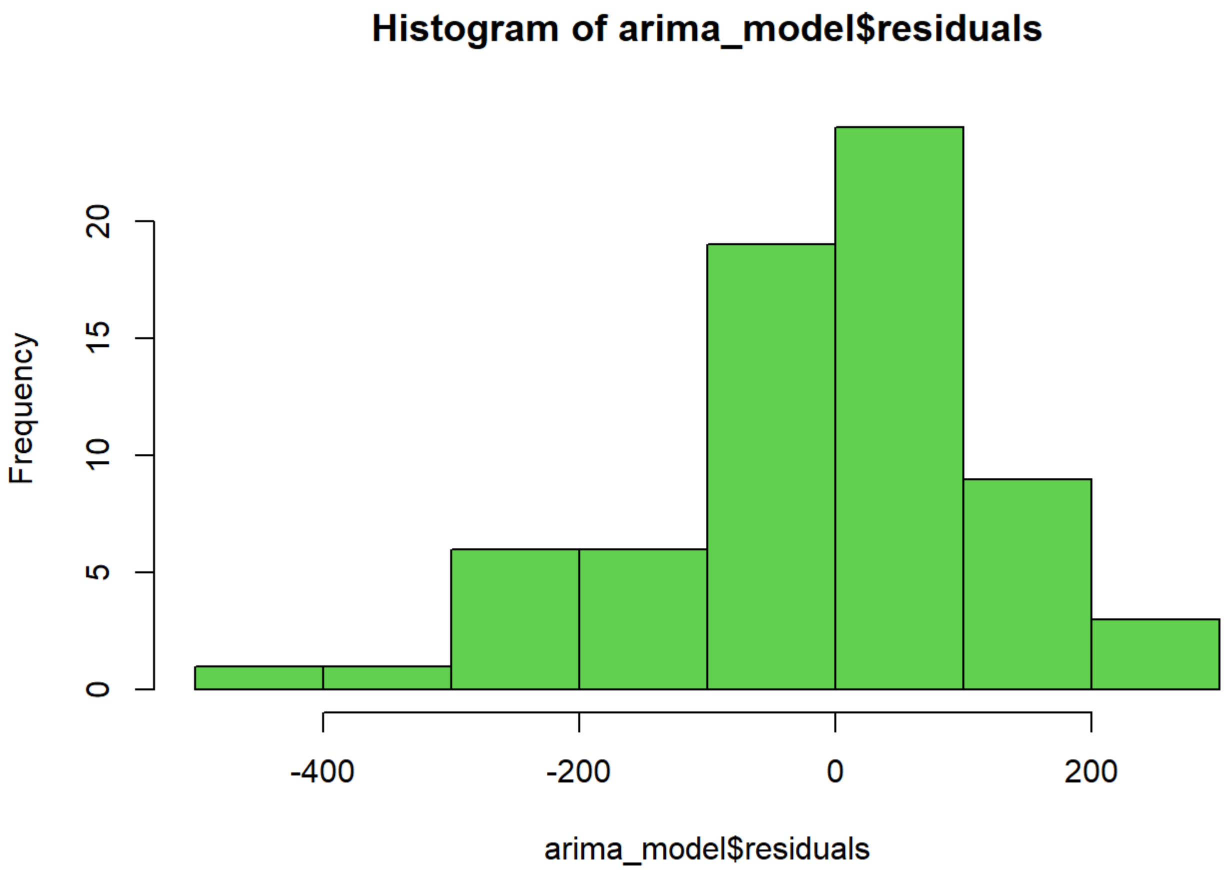
Now we are finally fitting ARIMA(2,1,3) and move ahead with residual analysis.

Step 5: Residule Analysis

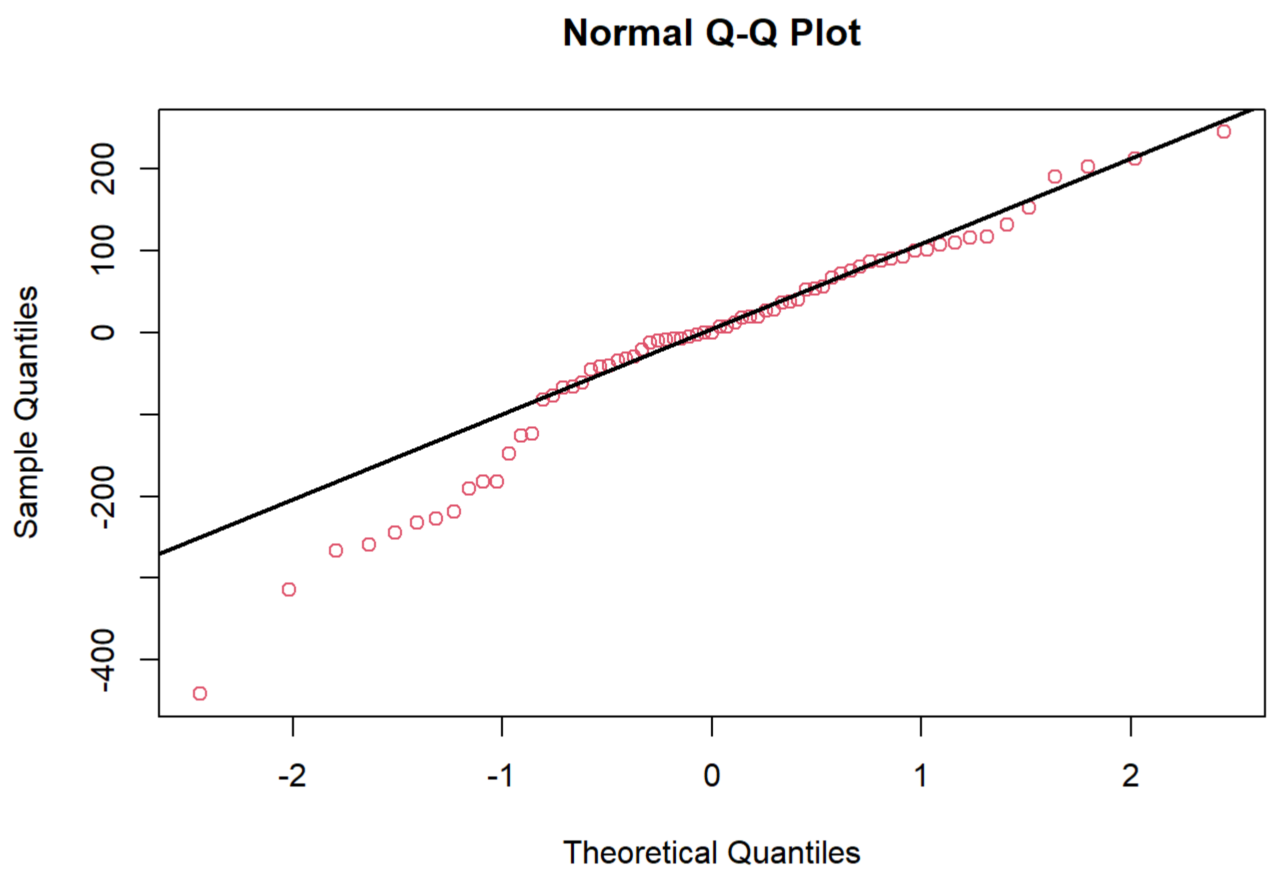
```
arima_model <- arima((stationary_data), order = c(2,1,3))
plot(arima_model$residuals,col=2)
```




```
hist(arima_model$residuals,col=3)
```




```
qqnorm(arima_model$residuals,col=2)
qqline(arima_model$residuals,lwd=2)
```




```
shapiro.test(arima_model$residuals)
```

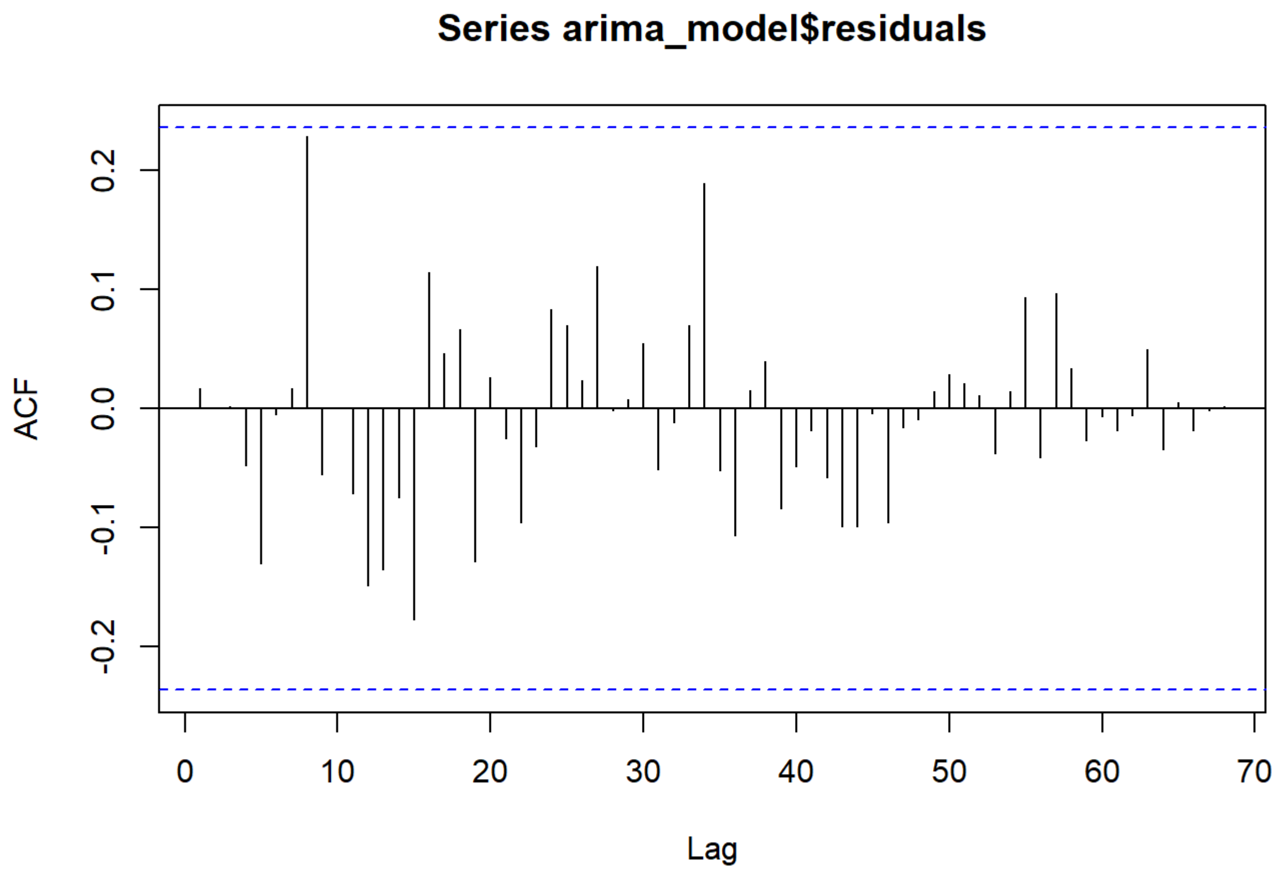
```
##
##  Shapiro-Wilk normality test
##
## data:  arima_model$residuals
## W = 0.95038, p-value = 0.008184
```

The residuals are not normally distributed. overall performance of model seems good. we will move ahead with LB test.

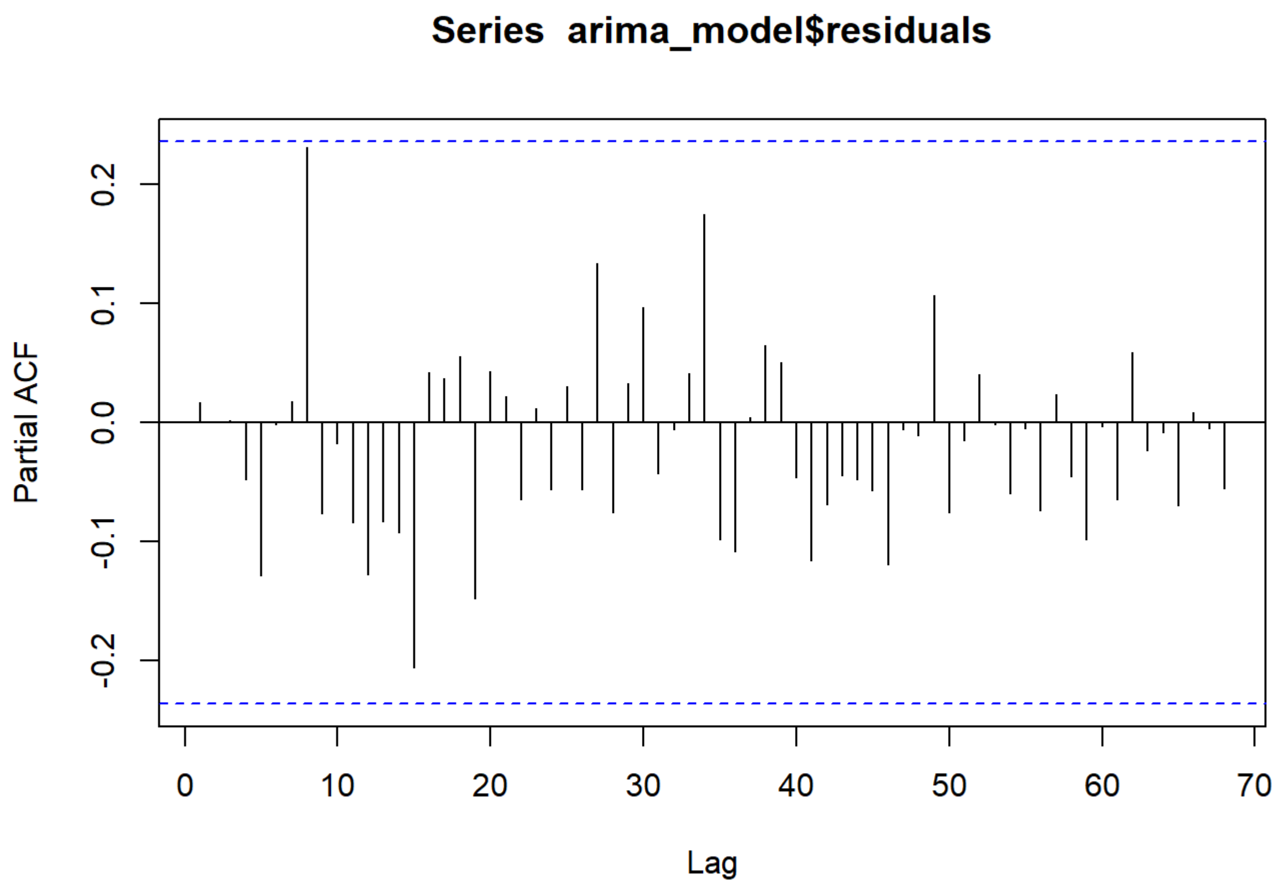
```
#Changes this to professors LB plot
Lb=Box.test(arima_model$residuals, lag = 10, type = "Ljung-Box")
Lb
```

```
##
##  Box-Ljung test
##
## data:  arima_model$residuals
## X-squared = 5.9643, df = 10, p-value = 0.8183
```

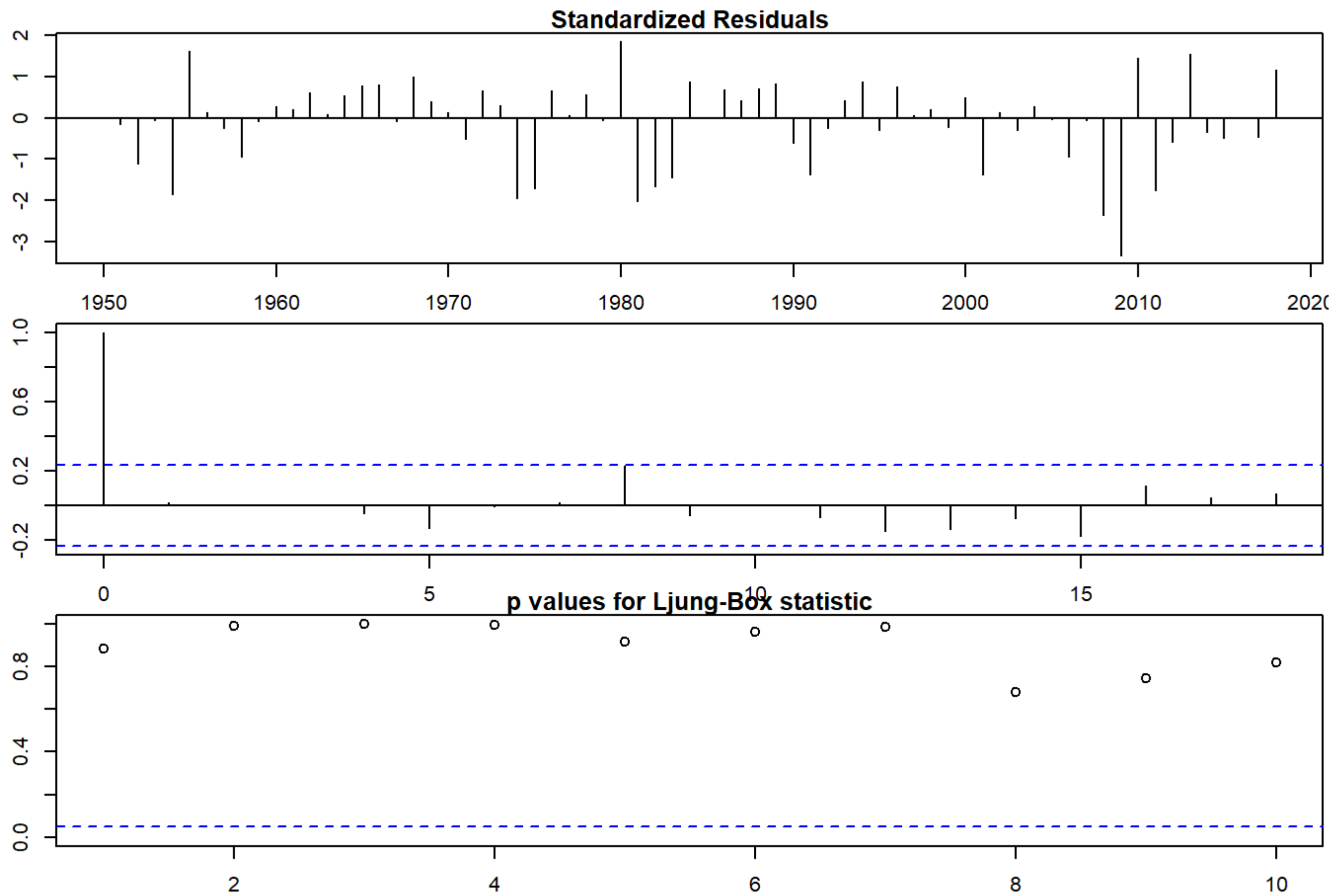
```
acf(arima_model$residuals, lag.max = length(tsdata))
```




```
pacf(arima_model$residuals, lag.max = length(tsdata))
```




```
par(mfrow = c(1, 1), mar = c(1, 0, 1, 0) + 0.2, oma = c(1, 2, 2, 0))
tsdiag(arima_model, main = "Residuals of Ljung-Box Test")
```

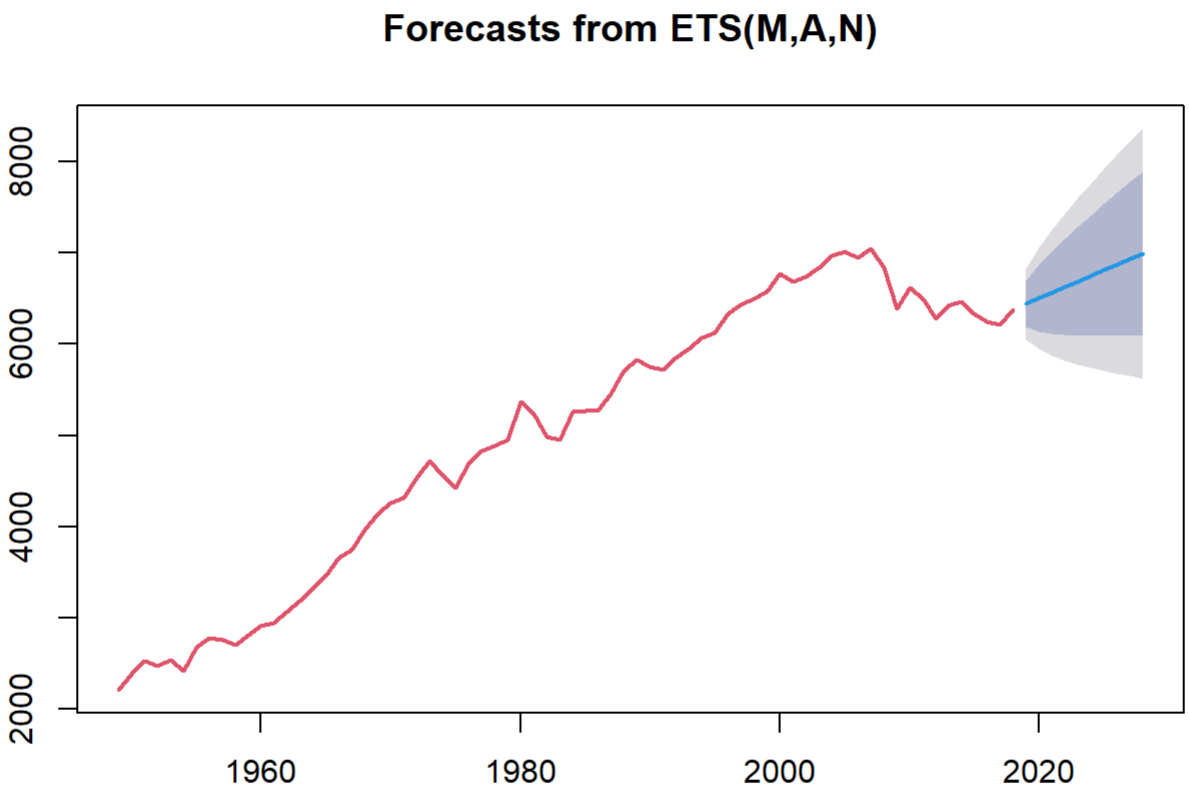


Step 6: Forecasting

```
library(forecast)


forecast_data <- forecast(tsdata, h = 10)

plot(forecast_data,col = 2,lwd=2)
```



Conclusion: In this project, time series data was analyzed using various models, including ARIMA(2,1,1), ARIMA(2,1,3), and ARIMA(2,1,3) was finilized. The ARIMA(2,1,3) model was found to be the best fit with lowest AIC value. The model was further evaluated using residual analysis techniques, including ACF plot, histogram, qq plot, Shapiro Wiki test, and Ljung-Box test. The results of these tests showed that the model was a good fit for the data. Finally, the model was used to forecast future values based on the original data. Overall, the results of this analysis suggest that the ARIMA(2,1,3) model is a suitable approach for modeling and forecasting the given time series data.