FITTING AN ARIMA MODEL AND FORECASTING THE NEXT FIVE VALUES

Ananya Kaushal

2022-04-26

INTRODUCTION

An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A statistical model is autoregressive if it predicts future values based on past values.

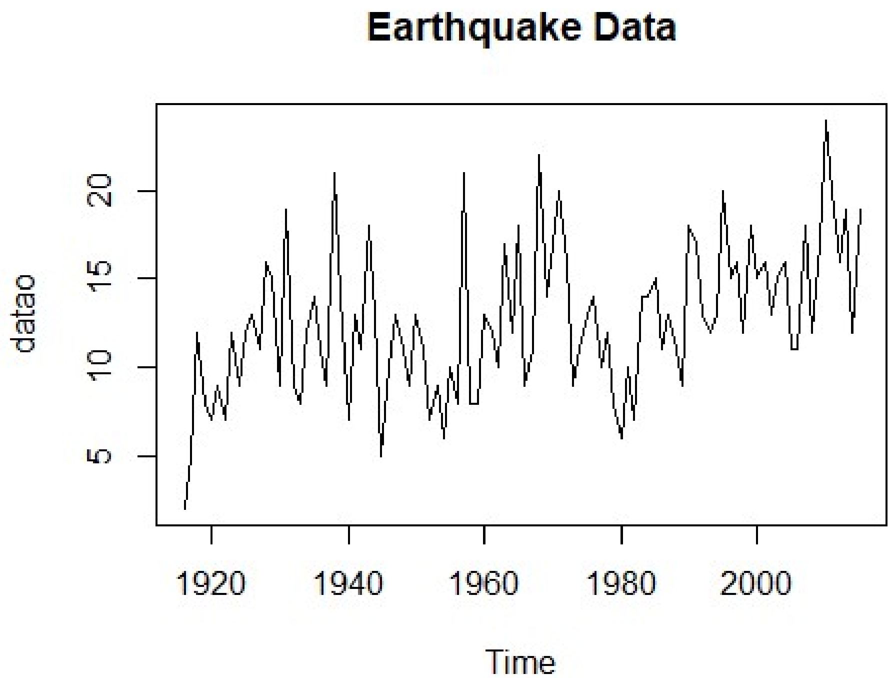
PROCEDURE

Calling the ‘astsa’, ‘forecast’ and ‘tseries’ libraries

```
library(astsa)

## Warning: package 'astsa' was built under R version 4.0.5

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

##
## Attaching package: 'forecast'

## The following object is masked from 'package:astsa':
##
##     gas

library(tseries)

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

Importing the dataset and plotting it as time series

```
data <- read.csv("p9.csv")

datao = ts(data$Quakes,start = 1916,frequency = 1)
ts.plot(datao,main="Earthquake Data")
```



Performing the adf test

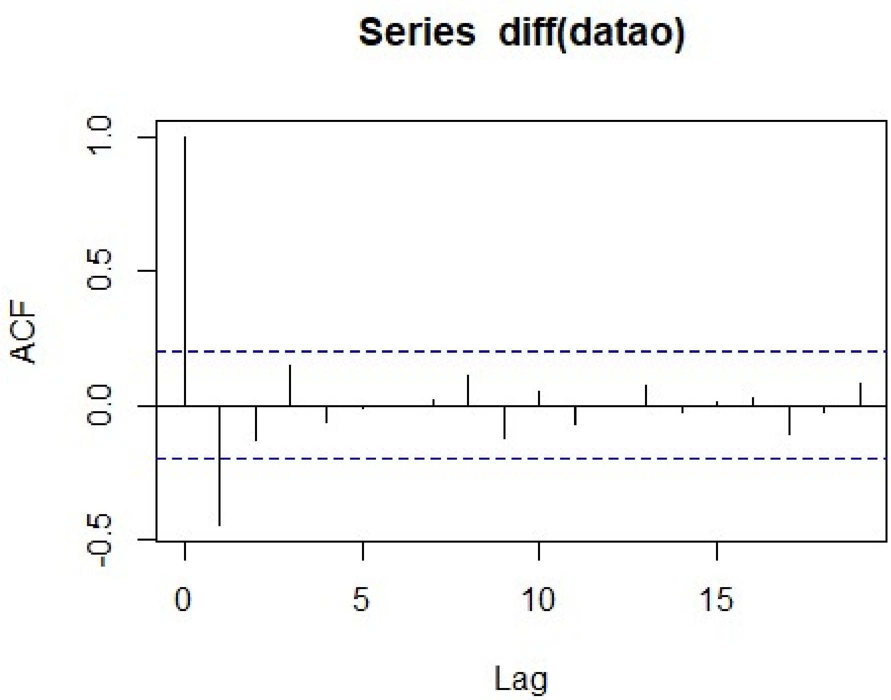
```
adf.test(datao)

##
##  Augmented Dickey-Fuller Test
##
## data:  datao
## Dickey-Fuller = -3.452, Lag order = 4, p-value = 0.04991
## alternative hypothesis: stationary
```

Fitting the arima model onto the dataset

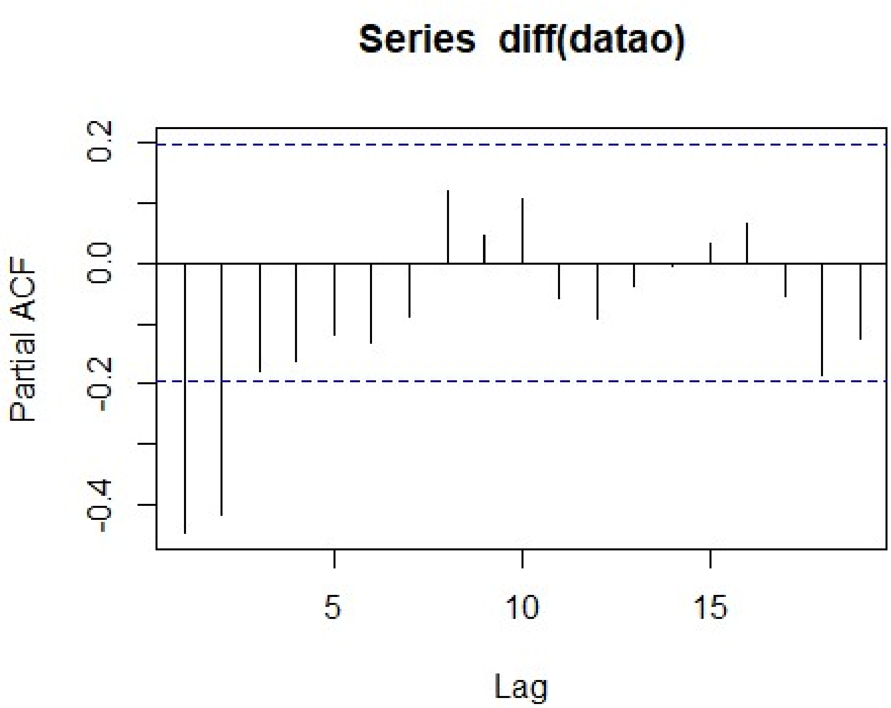
```
fit = auto.arima(datao, seasonal="F")
fit

## Series: datao
## ARIMA(0,1,1)
##
## Coefficients:
##           ma1
##       -0.8092
## s.e.   0.0710
##
## sigma^2 = 15.53:  log likelihood = -276.26
## AIC=556.53   AICc=556.65   BIC=561.72
```

Plotting the ACF and PACF plots of the differentiated data to verify if the arima model is the best fit or not

```
acf(diff(datao))
```

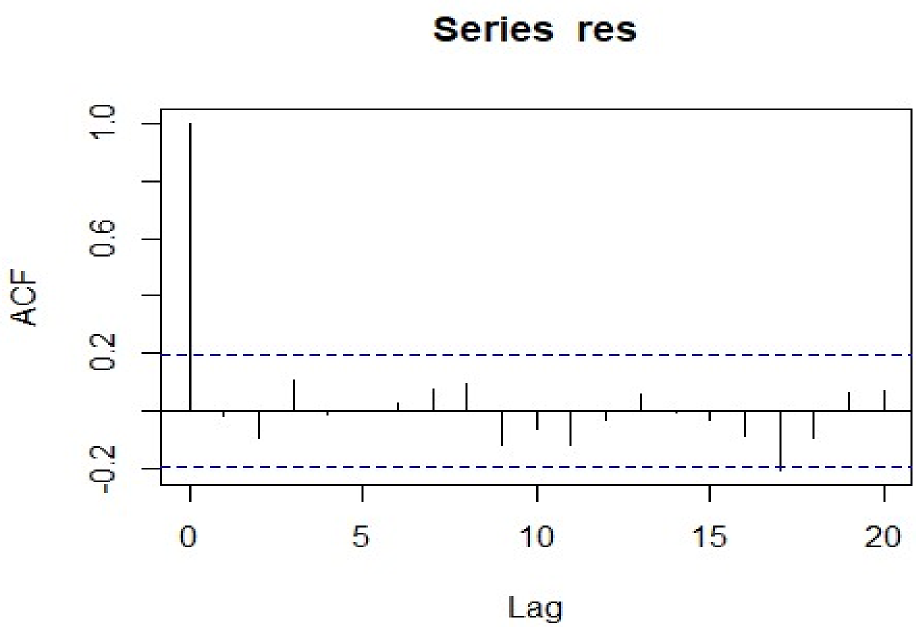



```
pacf(diff(datao))
```



Performing residual analysis

```
res = resid(fit)
acf(res)
```



Performing Box Test to check if the residuals are uncorrelated

```
Box.test(res)

##
##  Box-Pierce test
##
## data:  res
## X-squared = 0.037841, df = 1, p-value = 0.8458
```

Performing Shapiro test to check for normality

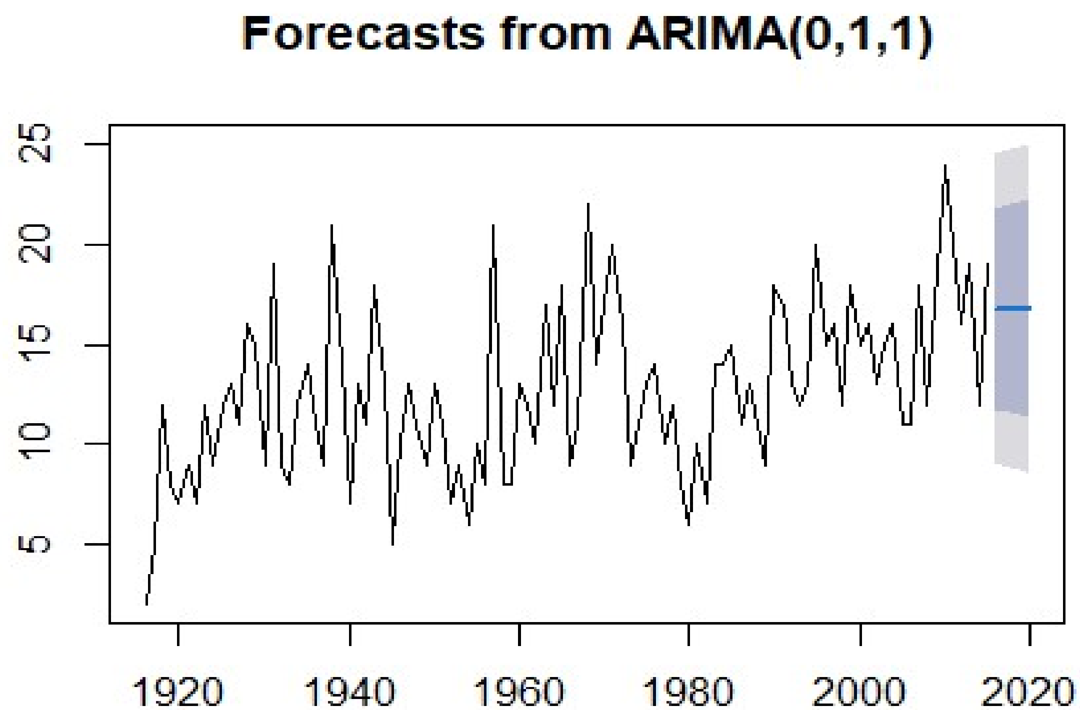
```
shapiro.test(res)

##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.98273, p-value = 0.2156
```

Forecasting future values and plotting them

```
forecast = forecast(fit, h=5)
forecast

##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2016        16.7859 11.73592 21.83589 9.062616 24.50919
## 2017        16.7859 11.64481 21.92700 8.923275 24.64853
## 2018        16.7859 11.55528 22.01652 8.786360 24.78545
```

```
## 2019        16.7859 11.46727 22.10454 8.651749 24.92006
## 2020        16.7859 11.38068 22.19112 8.519330 25.05247
```

```
plot(forecast)
```



CONCLUSION

From the above experiment we can conclude that the best fit ARIMA model for the given dataset is ARIMA(0,1,1). Differentiating the data once also gives us the same result. Forecasting the future five values shows us that they follow a stagnant trend.