RESIDUAL ANALYSIS AND DIAGNOSTIC CHECKING OF ARMA (1, 1) PROCESS

Ananya Kaushal

2022-04-06

INTRODUCTION:

An ARMA model, or Autoregressive Moving Average model, is used to describe weakly stationary stochastic time series in terms of two polynomials. The first of these polynomials is for AutoRegression, the second for the moving average.

The "residuals" in a time series model are what is left over after fitting a model. For many (but not all) time series models, the residuals are equal to the difference between the observations and the corresponding fitted values: $e_t = y_t - \hat{y}_t$.

AIM:

The aim of this experiment is to fit the given data to an ARMA (1,1) model and then perform residual analysis on the same.

ABOUT THE DATASET:

The dataset gives information about the Annual Stock price in the US from the years 1871 to 1970.

PROCEDURE:

Calling the 'tseries', 'forecast' and 'astsa' 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

Loading the dataset and converting it to time series

```r
data <- read.csv("p8.csv")
datao = ts(data$stock,start = 1871,frequency = 1)
ts.plot(datao,main="Stock")
```
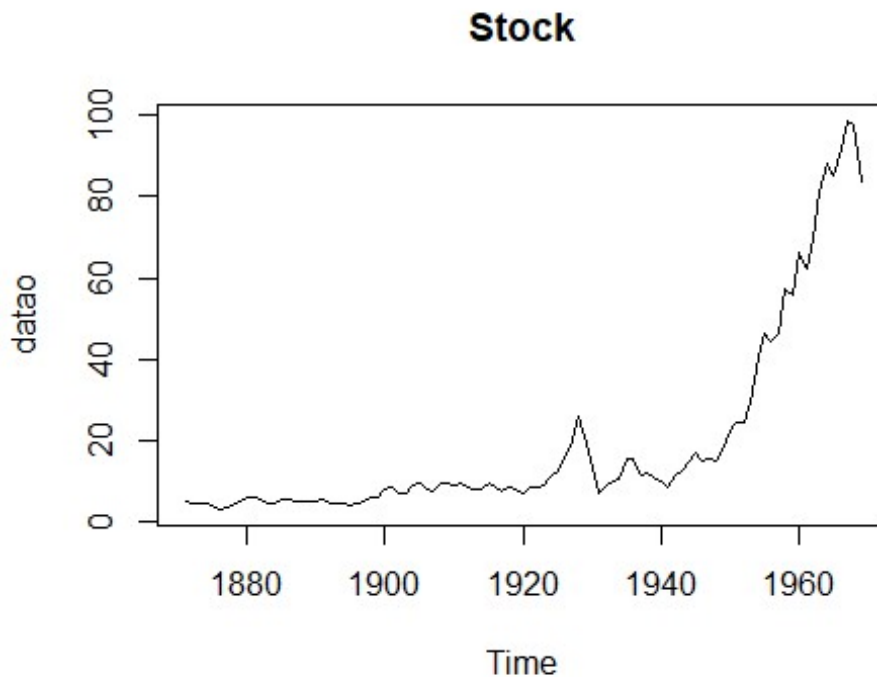
Stock

Figure 1: Graph for the Annual Stock Prices

Transforming the data into a stationary sequence

```r
diffdata = diff(datao)
```

Performing adf test to check the non-stationarity for the transformed data

```r
adf.test(datao)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  datao
## Dickey-Fuller = -1.1455, Lag order = 4, p-value = 0.9115
## alternative hypothesis: stationary
```

Since the p value is greater than 0.05, we will differentiate the data again

```r
newdata = diff(diffdata)
adf.test(newdata)
```

Warning in adf.test(newdata): p-value smaller than printed p-value

```
##
##  Augmented Dickey-Fuller Test
```

```
##
## data:  newdata
## Dickey-Fuller = -5.5647, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

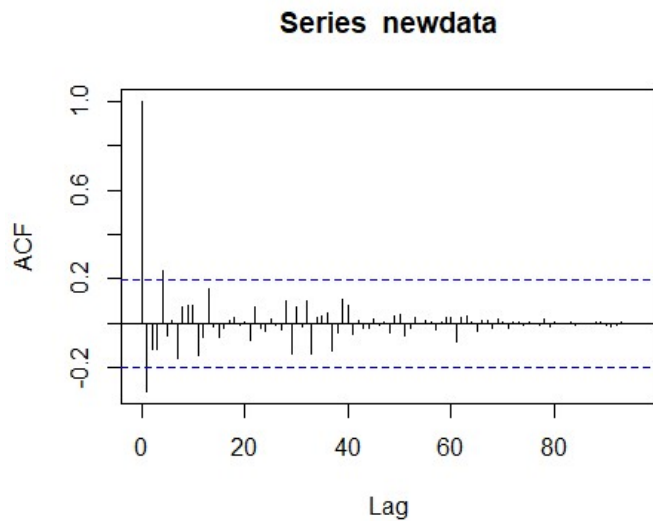Plotting the ACF and PACF plots

acf(newdata,lag=100)

Figure 2: ACF plot for the data that has been differentiated twice
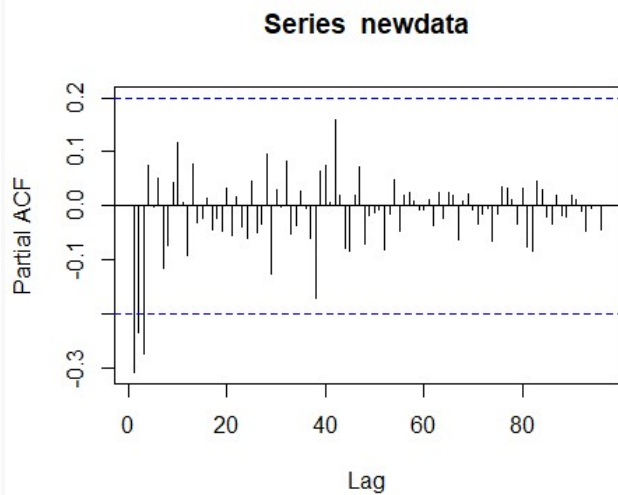
pacf(newdata,lag=100)

Figure 3: PACF plot for the data that has been differentiated twice

Fitting the dataset into an ARMA (1,1) model

```
arma_model = arma(newdata,order = c(1,1))
```

Warning in arma(newdata, order = c(1, 1)): Hessian negative-semidefinite

```
arma_model
```

```
##
## Call:
## arma(x = newdata, order = c(1, 1))
##
## Coefficient(s):
##      ar1      ma1  intercept
##  0.36132  -1.14394   0.02147
```

Extracting the residuals

```
res = resid(arma_model)
res1<- res[-1]
acf(res1)
```
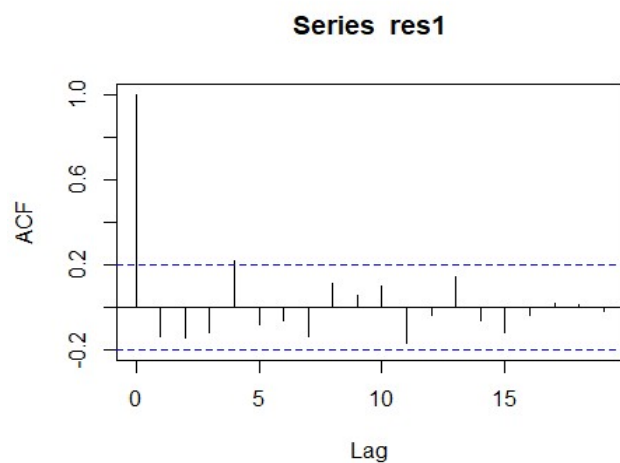

Figure 4: ACF plot for the residuals

Performing Portmanteu test to check if errors are uncorrelated

```
Box.test(res1)
```

```
##
##  Box-Pierce test
##
## data:  res1
## X-squared = 1.8815, df = 1, p-value = 0.1702
```

Plotting to check if the residuals are normally distributed
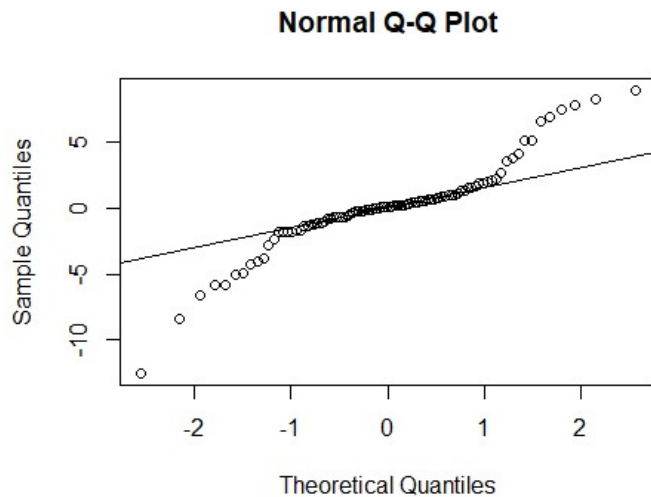
```
qqnorm(res1)
qqline(res1)
```


Figure 5: Normal Q-Q plot

Shapiro Wilk test to check for normality

```
shapiro.test(res1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res1
## W = 0.90217, p-value = 2.701e-06
```

CONCLUSION

The fitting of the dataset into an ARMA (1,1) model gave us the AR and MA coefficients as 0.36132 and -1.14394 respectively. After performing the residual analysis on the given dataset, we can conclude that the dataset satisfies only one assumption and that is the errors are uncorrelated. It does not satisfy normality.