Forecasting using ARIMA model

Srikar R

2022-04-22

Introduction

ARIMA, short for 'Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. An ARIMA model is characterized by 3 terms: p, d, q

where is the order of the AR term is the order of the MA term and d is the number of differencing required to make the time series stationary.

Data Description

The dataset describes Quarterly U.K. imports: goods and services (Pound millions) from 1960 – 1970.

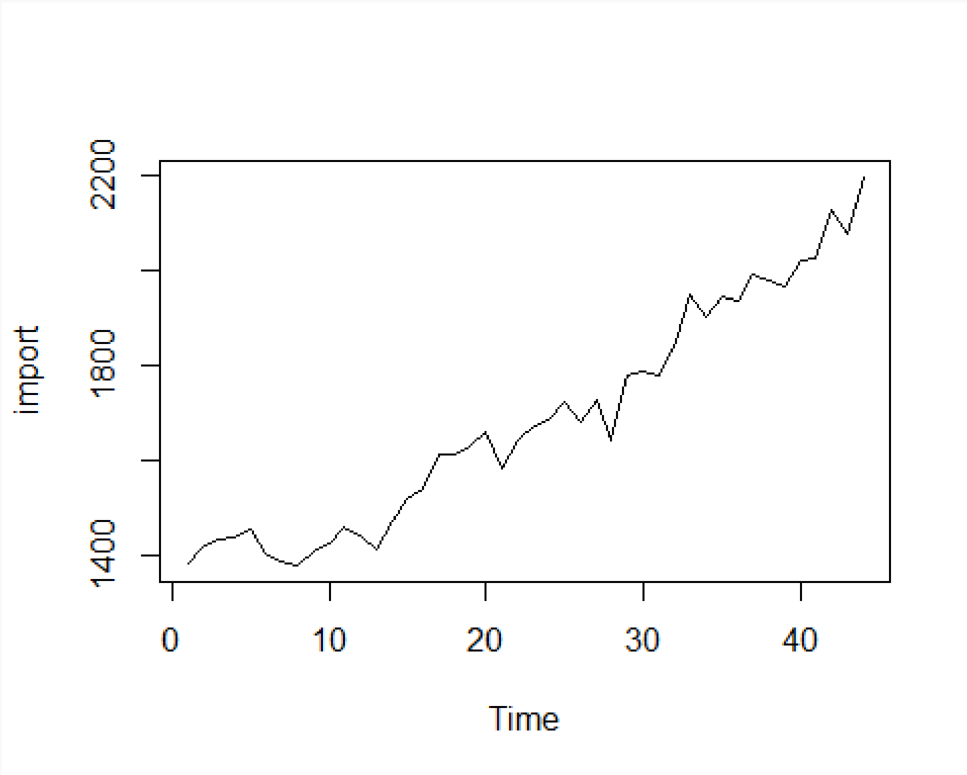
```
library(readxl)
import <- read_excel("C:/Users/Srikar/Desktop/SS/R/Sem 6/Practical
7/UKimport.xlsx")
head(UK)
```

```
##   Value
##   <dbl>
## 1  1382
## 2  1417
## 3  1432
## 4  1438
## 5  1457
## 6  1403
```

Analysis

1) Visualizing the data

```
ts.plot(import)
```



2) Checking for stationarity

Using ADF trst, we check for stationarity

```
library(tseries)

adf.test(import)

##
##  Augmented Dickey-Fuller Test
##
## data:  import
## Dickey-Fuller = -2.5343, Lag order = 3, p-value = 0.3623
## alternative hypothesis: stationary
```

We observe that it is a stationary dataset as the p-value is above the significance value.

2)Fitting an ARIMA model

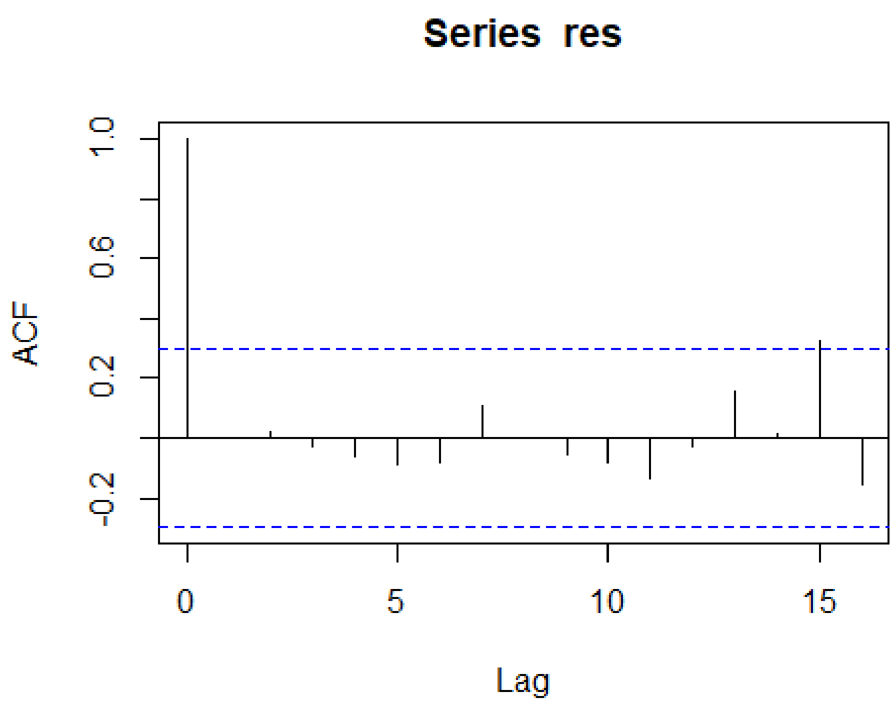
```
library(forecast)
fit=auto.arima(import,seasonal="FALSE")
fit
```

```
## Series: import
## ARIMA(1,1,0) with drift
##
## Coefficients:
##           ar1    drift
##       -0.4602  18.0900
## s.e.   0.1416   4.5349
##
## sigma^2 = 1945:  log likelihood = -222.93
## AIC=451.86   AICc=452.48   BIC=457.15
```

We observe that the model is an ar1 model with coefficient -0.4602

3)Residual Analysis

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



i)Testing normality of residuals assumption

```
shapiro.test(res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.9823, p-value = 0.7257
```

Since the p-value is above significance level (0.05), we accept the null hypothesis and say that it is normally distributed.

ii)Testing no-autocorrelation of errors

```
Box.test(res,lag=10,fitdf=1)
```

```
Box-Pierce test

data:  res
X-squared = 1.6596, df = 9, p-value = 0.9958
```

Since the p-value is greater than significance level (0.05), we accept the null hypothesis and say that there are no autocorrelation of errors.

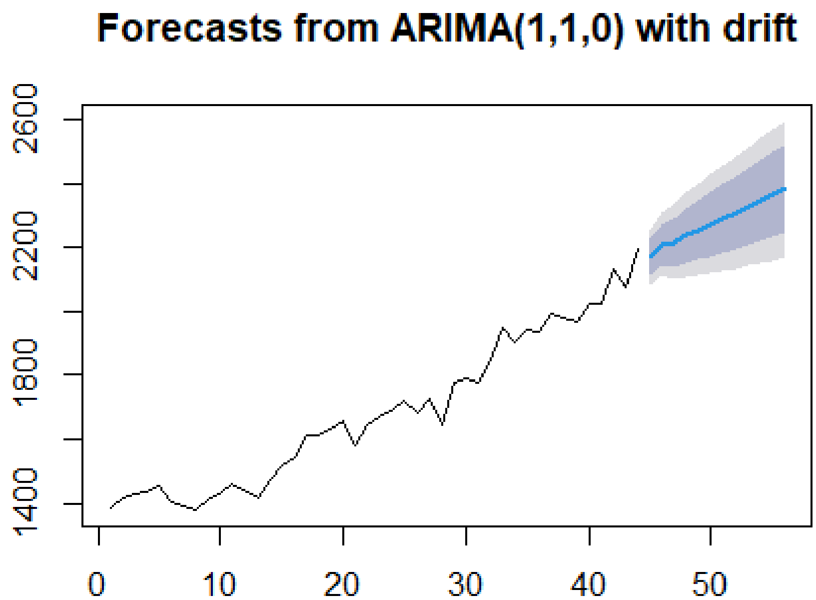
4) **Forecasting**

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

```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 45       2168.651 2112.128 2225.175 2082.206 2255.097
## 46       2208.112 2143.879 2272.345 2109.877 2306.348
## 47       2216.367 2139.357 2293.378 2098.590 2334.145
## 48       2238.983 2153.557 2324.409 2108.335 2369.631
## 49       2254.990 2160.871 2349.110 2111.047 2398.934
```

We see that the next 5 values are forecasted.

```
plot(forecast)
```



Conclusion

We observe that the data is already stationary and obtain an ARIMA model of (1,0,0) which is essentially an AR(1) model itself.