Time Series Project

Dr Noha Youssef

Maya Elshweikhy - 900204233

Alaa Aoun - 900204512

Karim AbouDaoud - 900212779

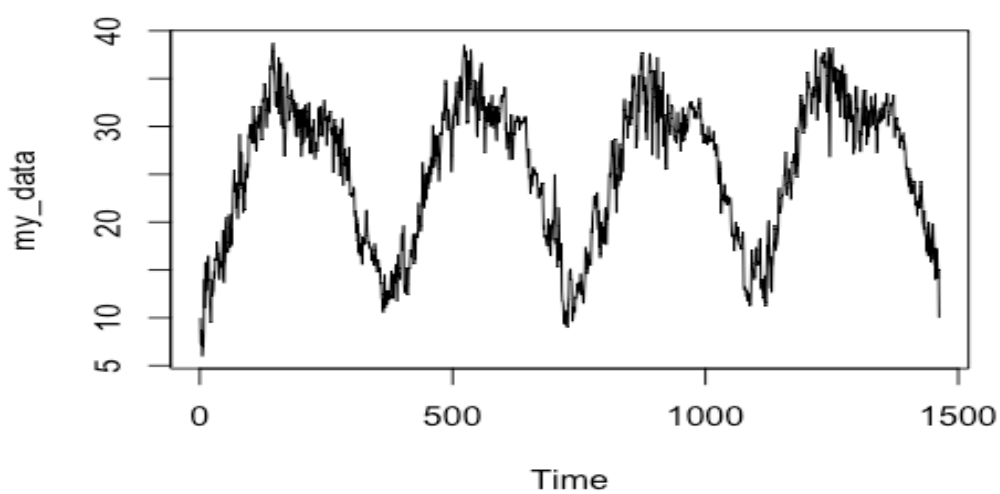
23 May 2024

Daily Climate Time Series Data Analysis

- The data was obtained from kaggle and we used the mean temprature column for our analysis.
- Data Source: https://www.kaggle.com/datasets/sumanthvrao/daily-climate-time-series-data/data

1. Reading the dataset

```
setwd("~/Desktop/Spring 2024/Time Series/Project")
data<-read.csv("Final Data.csv")
my_data<-data[,2]
head(my_data)

## [1] 10.000000  7.400000  7.166667  8.666667  6.000000  7.000000

plot.ts(my_data)
```



- This is the graph of our data. There doesn't seem a trend. We will now check for stationarity.

2. Testing for stationarity

```
library(urca)
df1<-ur.df(my_data,type="none",lags=1)
summary(df1)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6439 -0.8393  0.1195  1.0747  6.5548
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -0.001632   0.001626  -1.003    0.316
## z.diff.lag -0.160051   0.025914  -6.176   8.5e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.649 on 1458 degrees of freedom
## Multiple R-squared:  0.02644,    Adjusted R-squared:  0.0251
## F-statistic:  19.8 on 2 and 1458 DF,  p-value: 3.287e-09
##
##
## Value of test-statistic is: -1.0032
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

Tests the null hypothesis H0 : $\phi 1 = 1$ against H1 : $\phi 1 < 1$.

H_0: series is not stationary , H_1: series is stationary , alpha=0.05

- By looking at z.lag.1 we notice that the pr is 0.316 > 0.05, we fail to reject H_0. This means we will try to take difference and then check for stationarity again.

3. Taking first difference

```
dif_data1<-diff(my_data, diff=1)
```

After taking the first difference, we will test for stationarity again.

4. Checking for stationarity again

```
df1new<-ur.df(dif_data1,type="none",lags=1)
summary(df1new)

##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6674 -0.8636  0.0764  1.0176  6.5795
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -1.27974    0.03980 -32.156  < 2e-16 ***
## z.diff.lag  0.10209    0.02612   3.909 9.71e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.641 on 1457 degrees of freedom
## Multiple R-squared:  0.5835, Adjusted R-squared:  0.5829
## F-statistic:  1020 on 2 and 1457 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -32.1559
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

Tests the null hypothesis H0 : ϕ1 = 1 against H1 : ϕ1 < 1.

H_0: series is not stationary , H_1: series is stationary , alpha=0.05

- When we did the test for stationarity again, the z.lag.1 gave pr of 2e-16 which is much less than 0.05. This means that we reject H_0. This implies that now our series is stationary.

5. Checking for stationarity using a different method

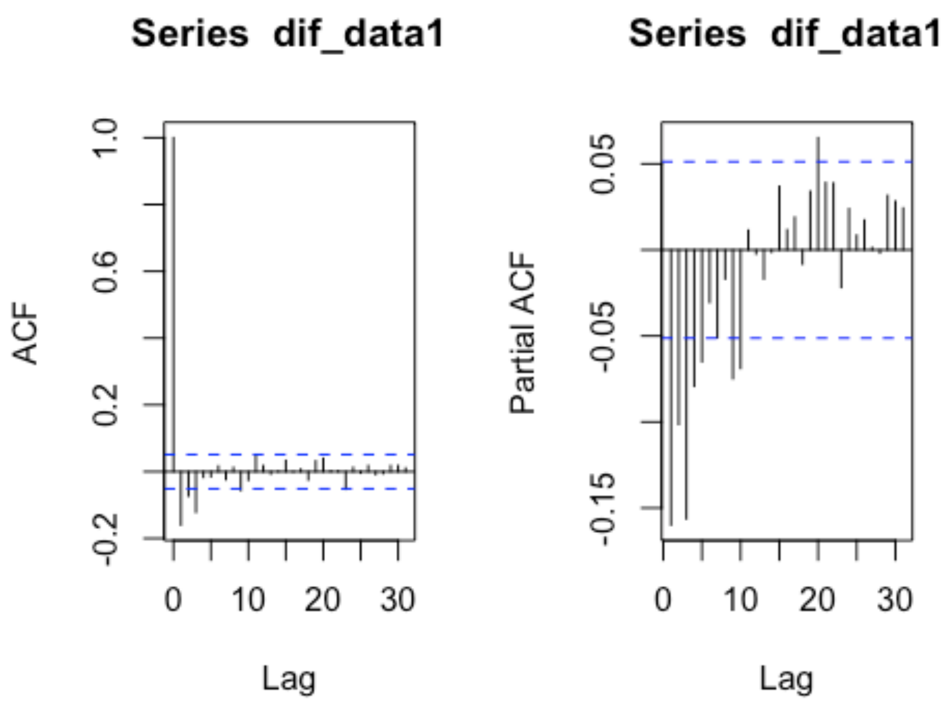
```
library(tseries)

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

df2<-adf.test(dif_data1,k=1)

## Warning in adf.test(dif_data1, k = 1): p-value smaller than printed p-
value

df2

##
##  Augmented Dickey-Fuller Test
##
## data:  dif_data1
## Dickey-Fuller = -32.169, Lag order = 1, p-value = 0.01
## alternative hypothesis: stationary
```

H_0: series is not stationary , H_1: series is stationary , alpha=0.05

- This output shows that p-value is 0.01<0.05. This implies that we reject H_0 and thus we can confirm that our series is stationary and we can proceed with our analysis.

6. Plotting ACF and PACF

```
par(mfrow=c(1,2))
acf(dif_data1)
pacf(dif_data1)
```



- It is noticeable that our ACF cuts off at lag 3 and PACF decays. We can say that it is MA(3).

7. Modeling the data (MA(3))

```
ma3<-arima(my_data, order=c(0,1,3))
ma3

##
## Call:
## arima(x = my_data, order = c(0, 1, 3))
##
```

```
## Coefficients:
##           ma1      ma2      ma3
##       -0.2261  -0.1242  -0.1522
## s.e.   0.0260   0.0274   0.0259
##
## sigma^2 estimated as 2.583:  log likelihood = -2766.25,  aic = 5540.51
```

Model Equation: X_t= -0.2261 e_t-1 -0.1242 e_t-2 -0.1522 e_t-3 + e_t

8. Checking a different model

We would like to check ARMA(1,1) for us to be able to compare AICs.

```
arma11<-arima(my_data,order=c(1,1,1))
arma11
```

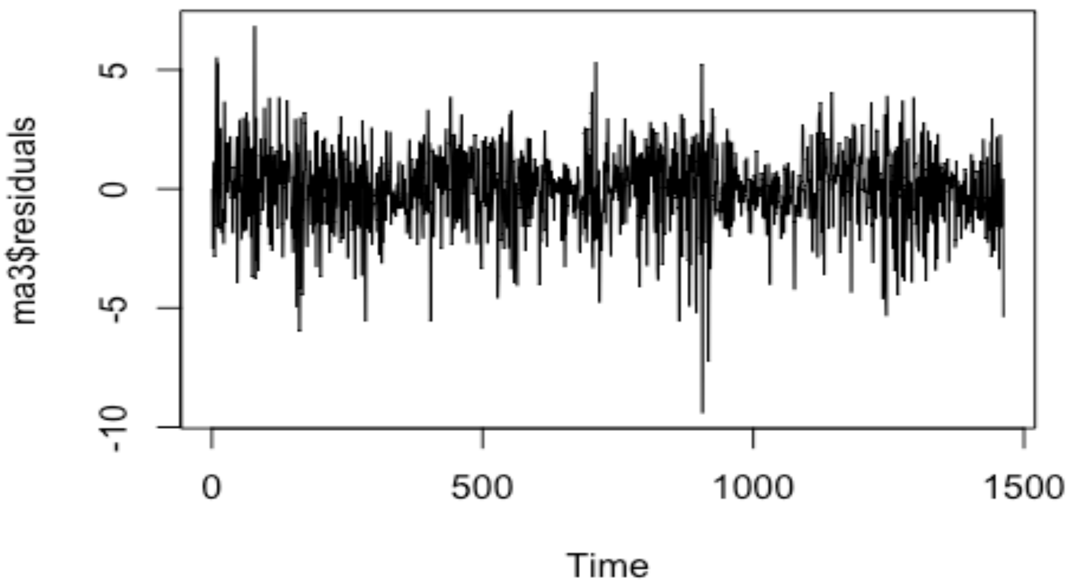
```
##
## Call:
## arima(x = my_data, order = c(1, 1, 1))
##
## Coefficients:
##          ar1      ma1
##       0.5727  -0.8027
## s.e.  0.0419   0.0289
##
## sigma^2 estimated as 2.59:  log likelihood = -2768.24,  aic = 5542.48
```

Model Equation: X_t= 0.5727 X_t-1 -0.8027 e_t-1 + e_t

- The AIC of the first model is smaller than the AIC of the second model. Therefore, we will use the first model.

9. Checking the model assumptions, NICE assumptions

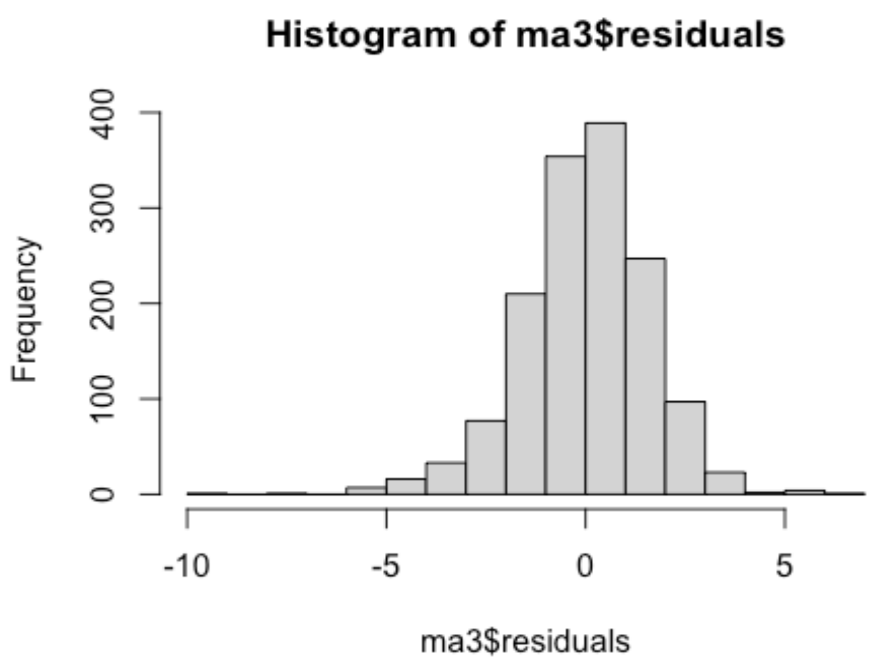
```
plot.ts(ma3$residuals)
```



- The residuals has no pattern in time.

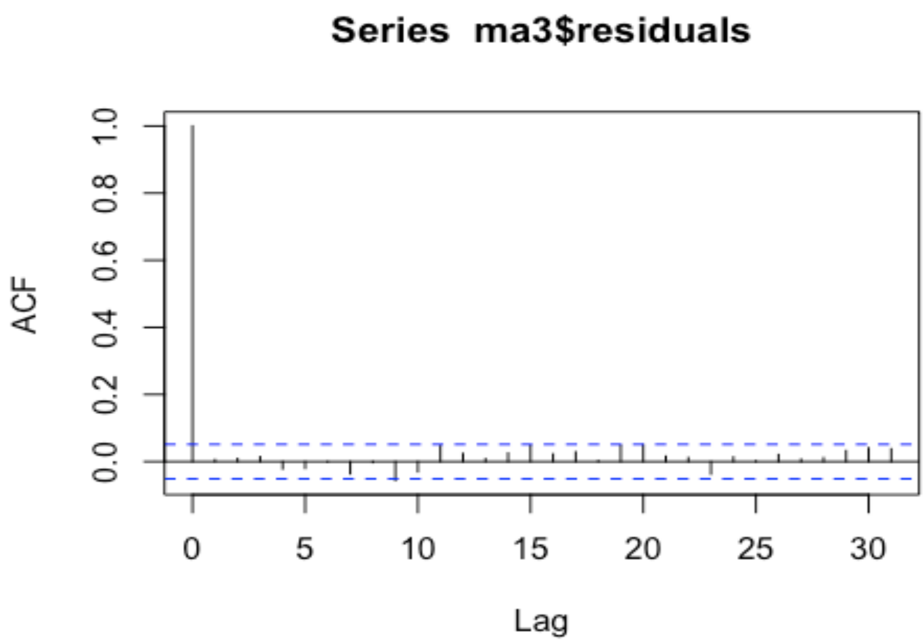
Test for normality

```
hist(ma3$residuals)
```



- The residuals are normally distributed.

```
acf(ma3$residuals)
```



- There are no spikes which means that the residuals are independent.

Testing for Indepenndence of Residuals

```
Box.test(ma3$residuals,lag=20,fitdf = 3)

##
##  Box-Pierce test
##
## data:  ma3$residuals
## X-squared = 27.009, df = 17, p-value = 0.05793
```

H_0: residuals are independent , H_1: residuals are not independent , alpha=0.05

- Since the p value is 0.05793 which is greater than alpha (0.05), therefore we fail to reject H_0. This means residuals are independent which agrees with the ACF plot.

10. Forecasting

```
data<-read.csv("Final Data.csv")
my_data2<-data[1:1460,2]
head(my_data2)

## [1] 10.000000  7.400000  7.166667  8.666667  6.000000  7.000000
```

We took the first 1460 observations as our data and left 2 observations for forecasting.

```
library(forecast)
fit <- Arima(my_data2, order=c(0,1,3))
```

```
forecasts <- forecast(fit, h=2)
forecasts

##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 1461       14.61767 12.56266 16.67268 11.47480 17.76053
## 1462       15.05462 12.45519 17.65406 11.07913 19.03012
```

- The forecasted values for observations 1461 and 1462 are 14.61767 and 15.05462 respectively. The actual values for observations 1461 and 1462 are 15.05263158 and 10 respectively.

Calculating Forecast Error

```
forecasted_values <- c(14.61767 , 15.05462)
actual_values <-c(15.05263158, 10)
errors = abs(actual_values - forecasted_values)
errors

## [1] 0.4349616 5.0546200
```

- The forecast error of the first observation is less than that of the second. This could indicate that this forecast is good for short-term forecast, because it seems that the error would increase for later observations.