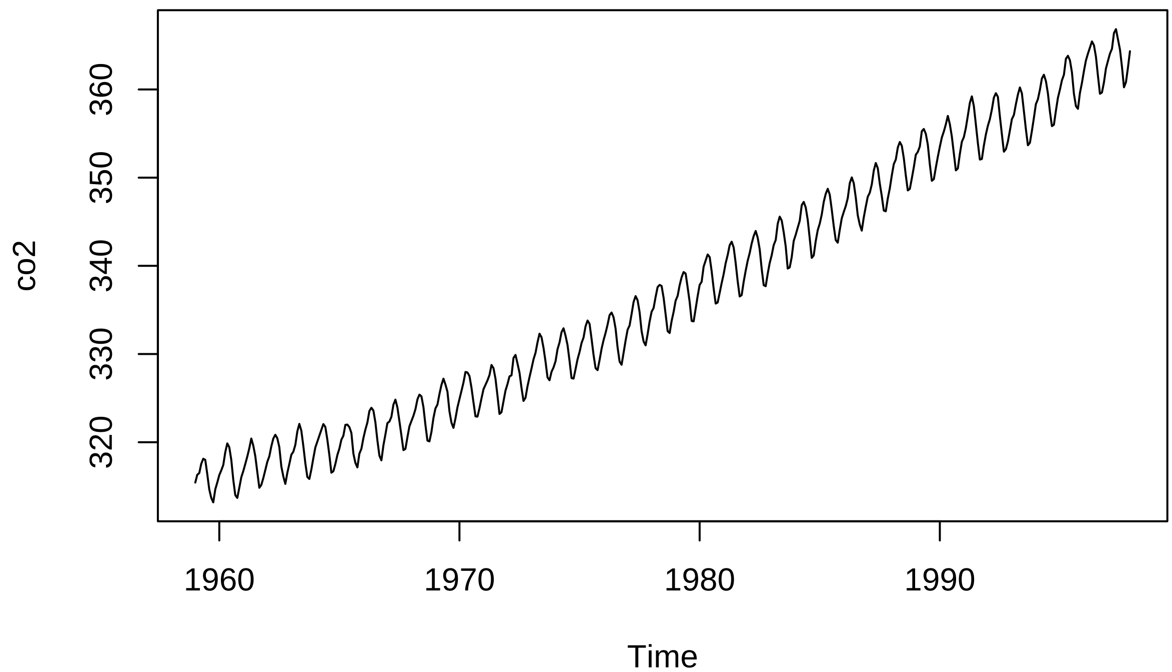
Assignment-1

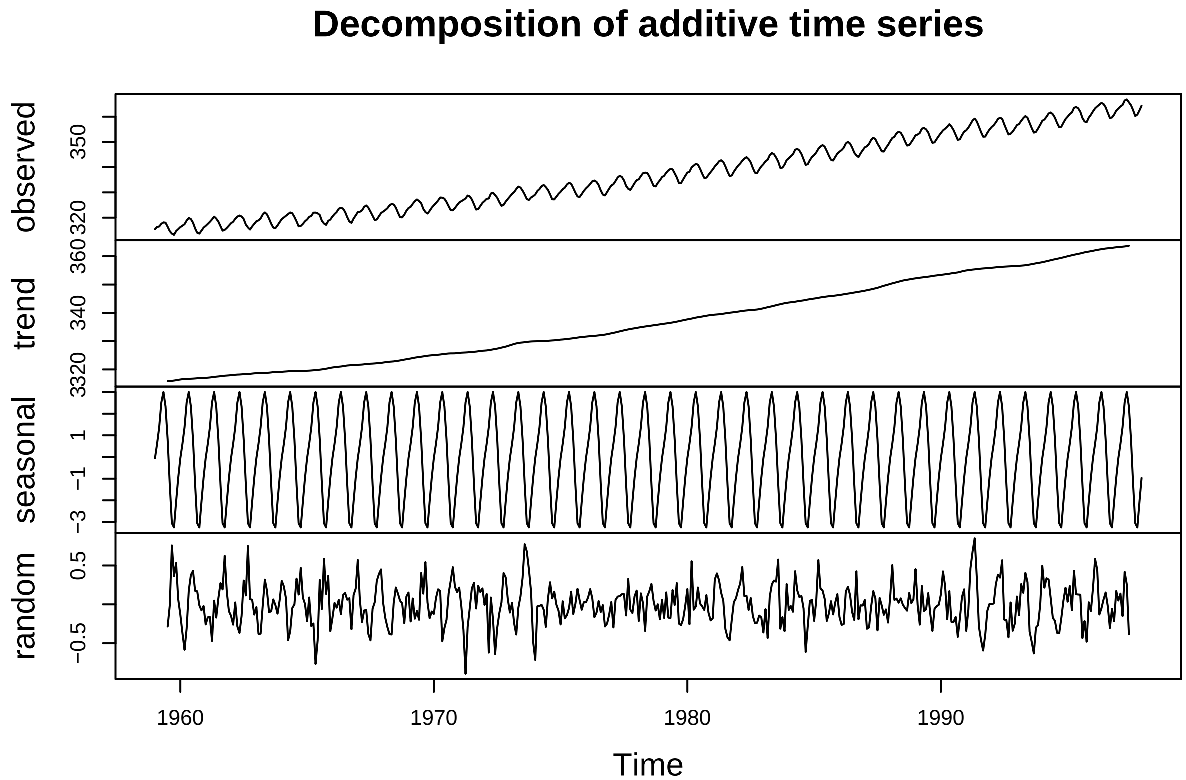
2023-03-01

Excercise 1

Time-Series [1:468] from 1959 to 1998: 315 316 316 318 318 ...

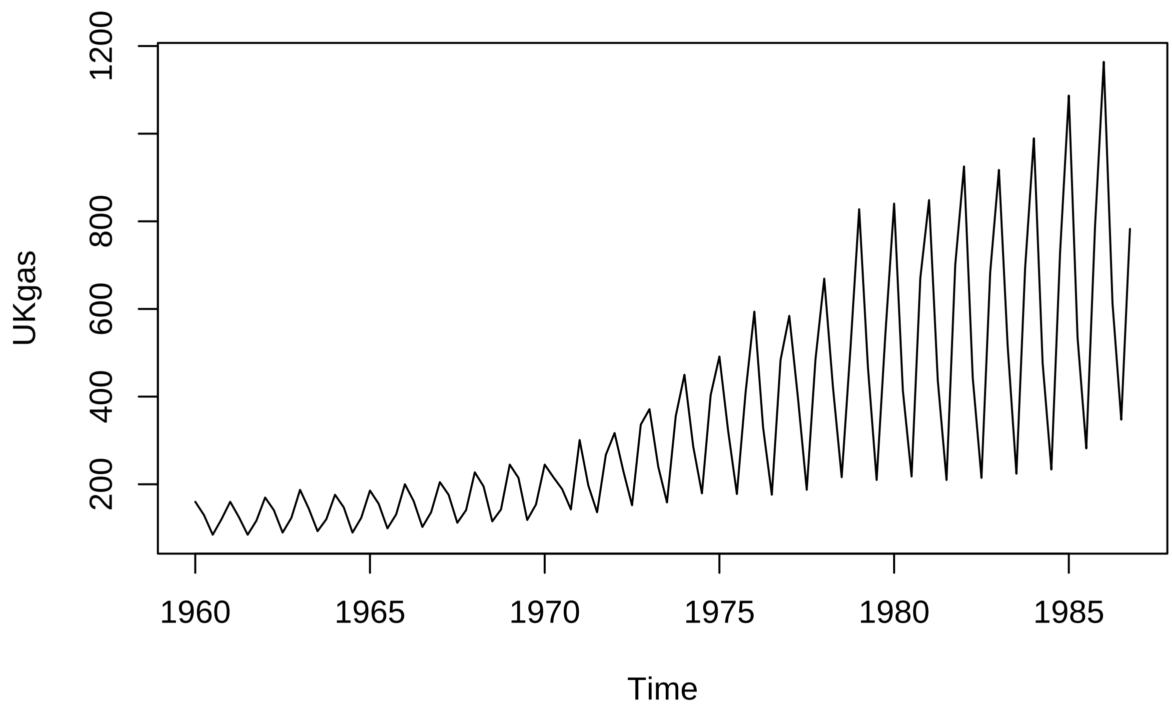


By looking at the decomposition we can clearly see that there is a linear increasing trend, while seasonality is regular and its magnitude doesn't change over time.

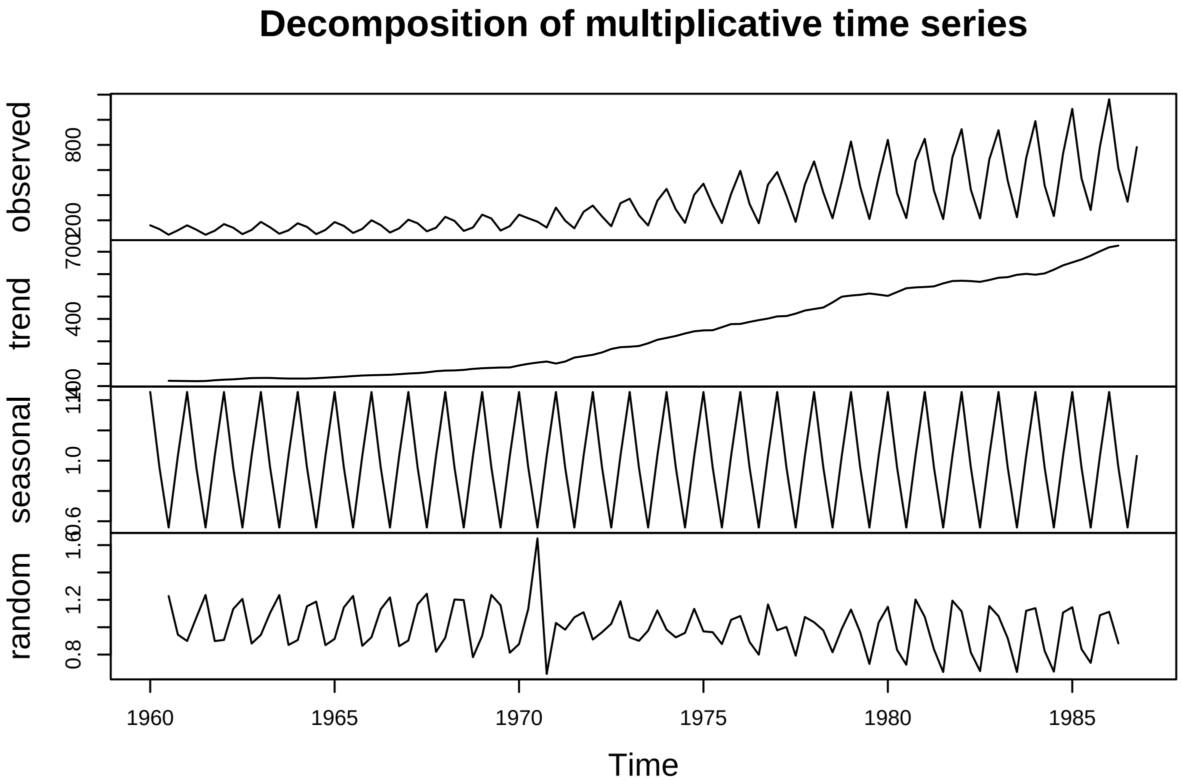


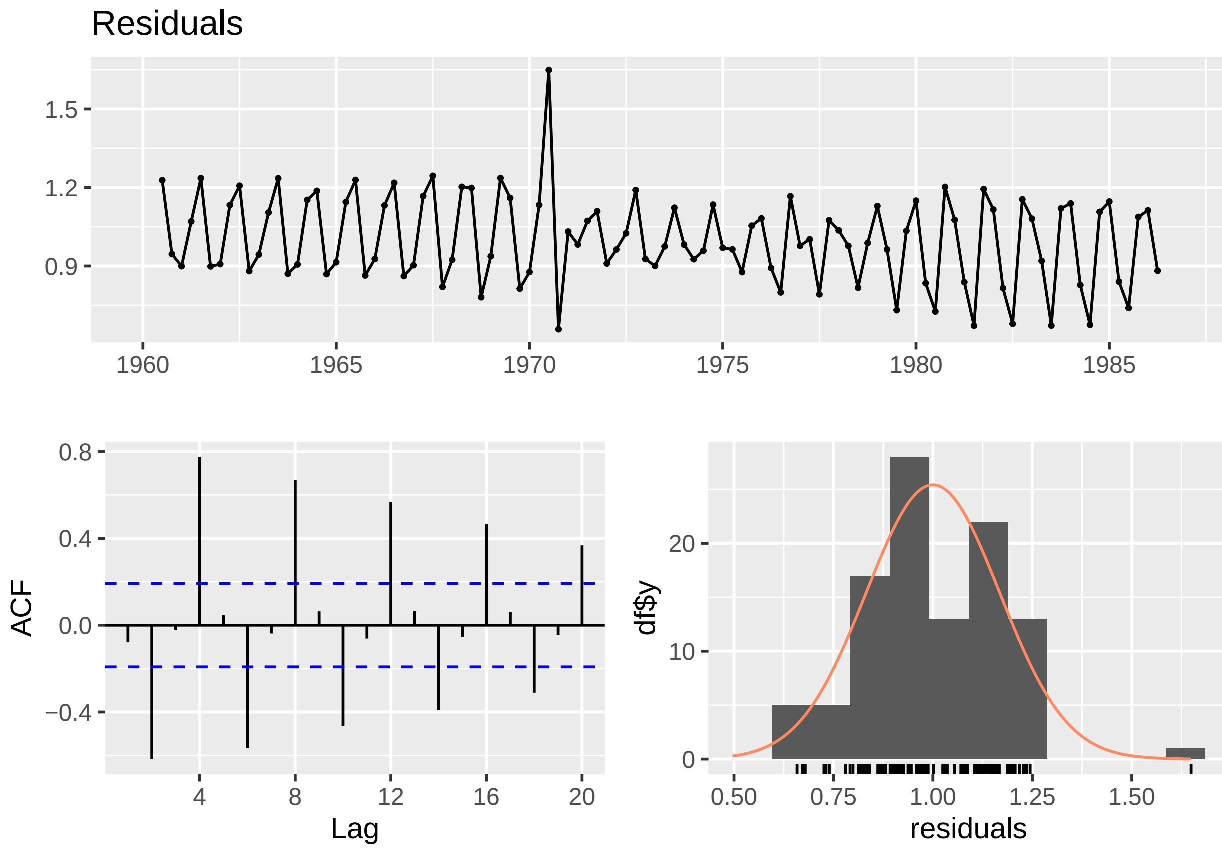
Excercise 2

Time-Series [1:108] from 1960 to 1987: 160.1 129.7 84.8 120.1 160.1 . . .



As we can see from the plotted data, there is a change in the seasonal component from the begin of the 70's. At the same time there is also a change in the trend, which becomes steeper around the same year. In this case, a multiplicative decomposition is the more appropriate approach, since we do not have a constant seasonal component. That can be shown by analyzing the residuals of the two decompositions: we can see that in the case of the multiplicative time series the residuals are smaller, suggesting a better fit of the raw data


Decomposition of multiplicative time series
observed
trend
seasonal
random
800
200
700
400
100
1.4
1.0
0.6
1.6
1.2
0.8
1960
1965
1970
1975
1980
1985
Time

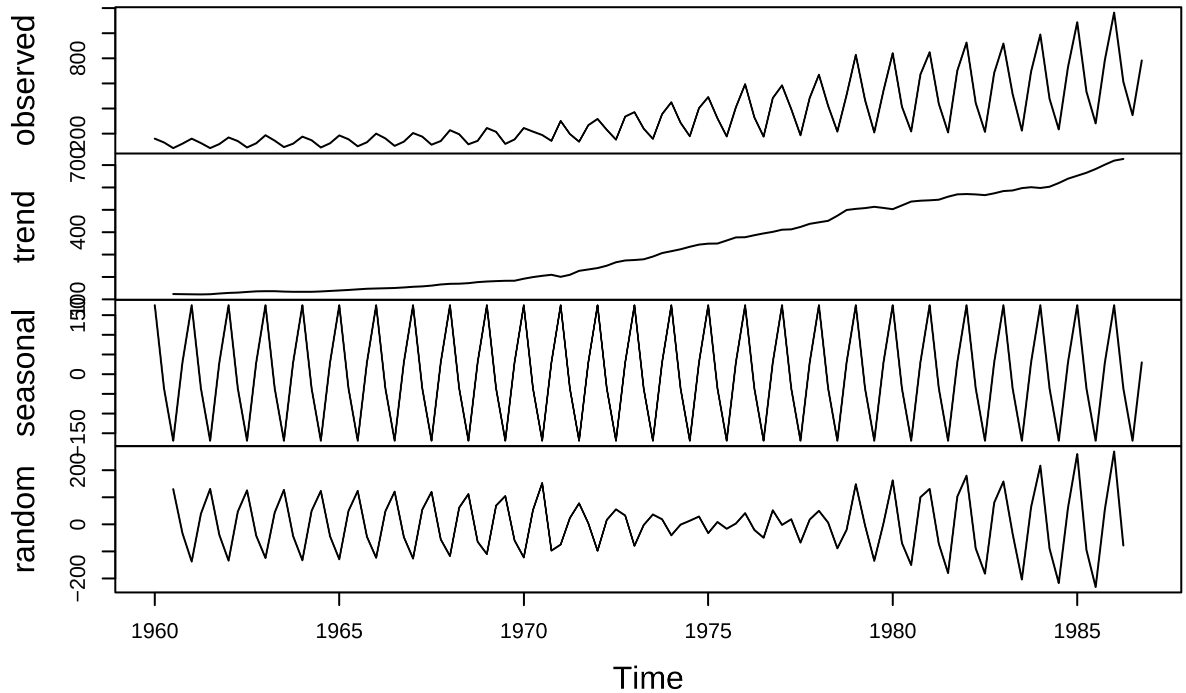


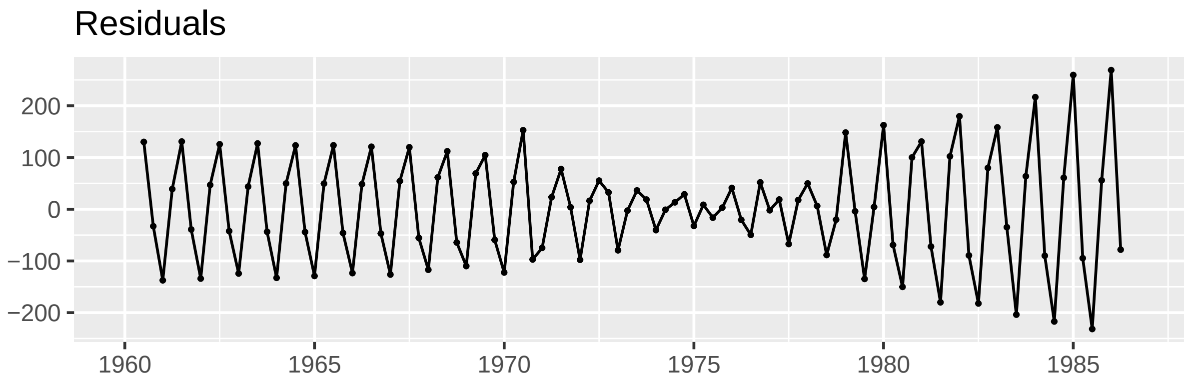
Ljung-Box test

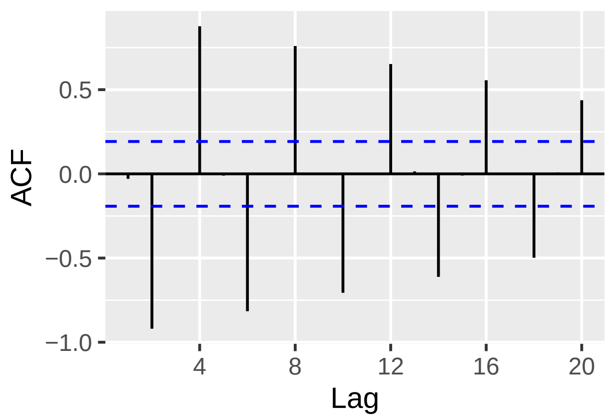
data: Residuals Q* = 195.83, df = 8, p-value < 2.2e-16

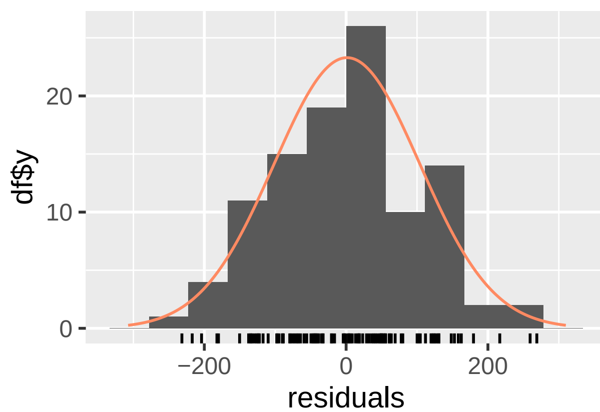
Model df: 0. Total lags used: 8

Decomposition of additive time series







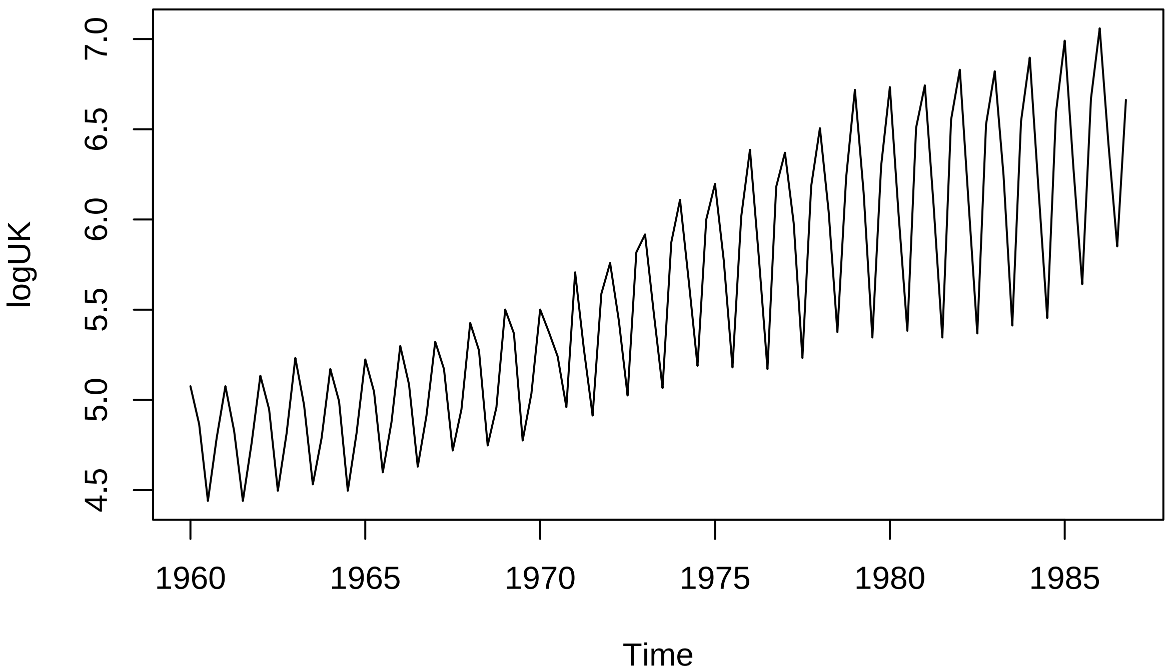


Ljung-Box test

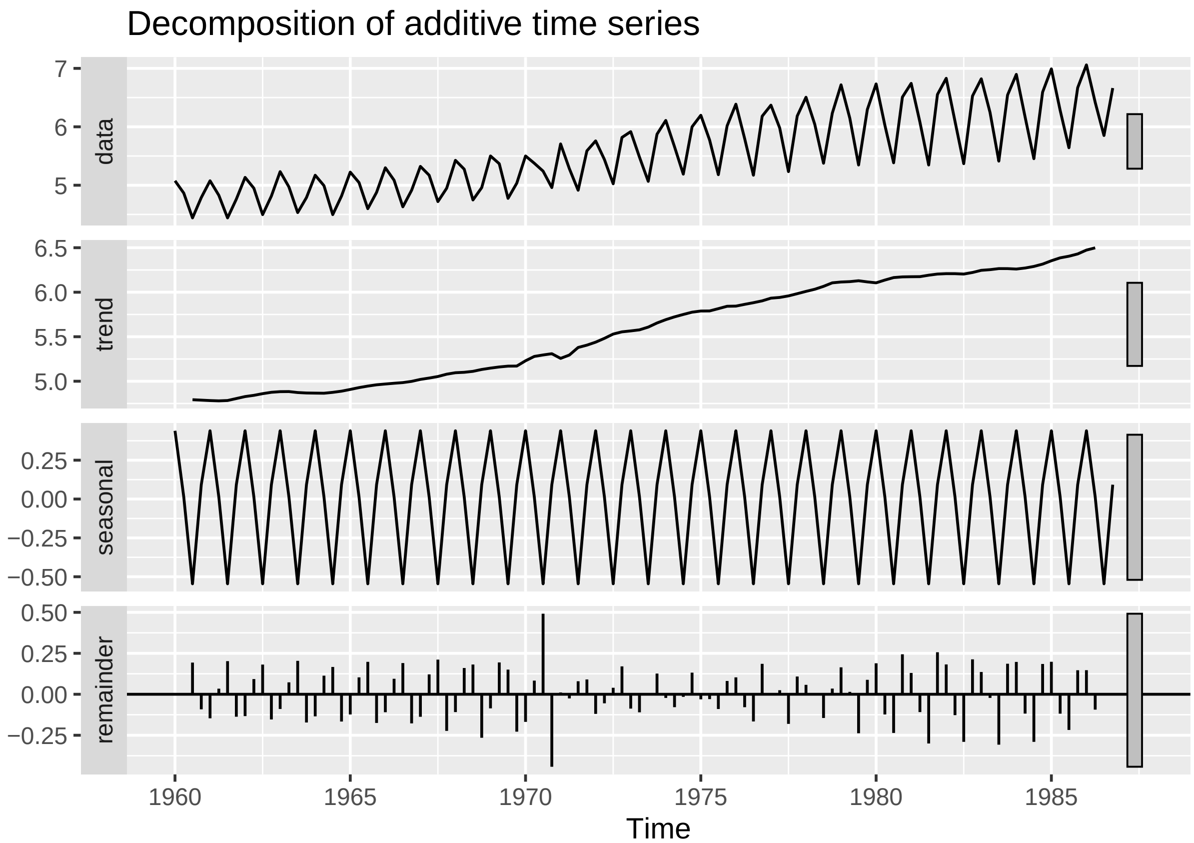
data: Residuals Q* = 317.25, df = 8, p-value < 2.2e-16

Model df: 0. Total lags used: 8

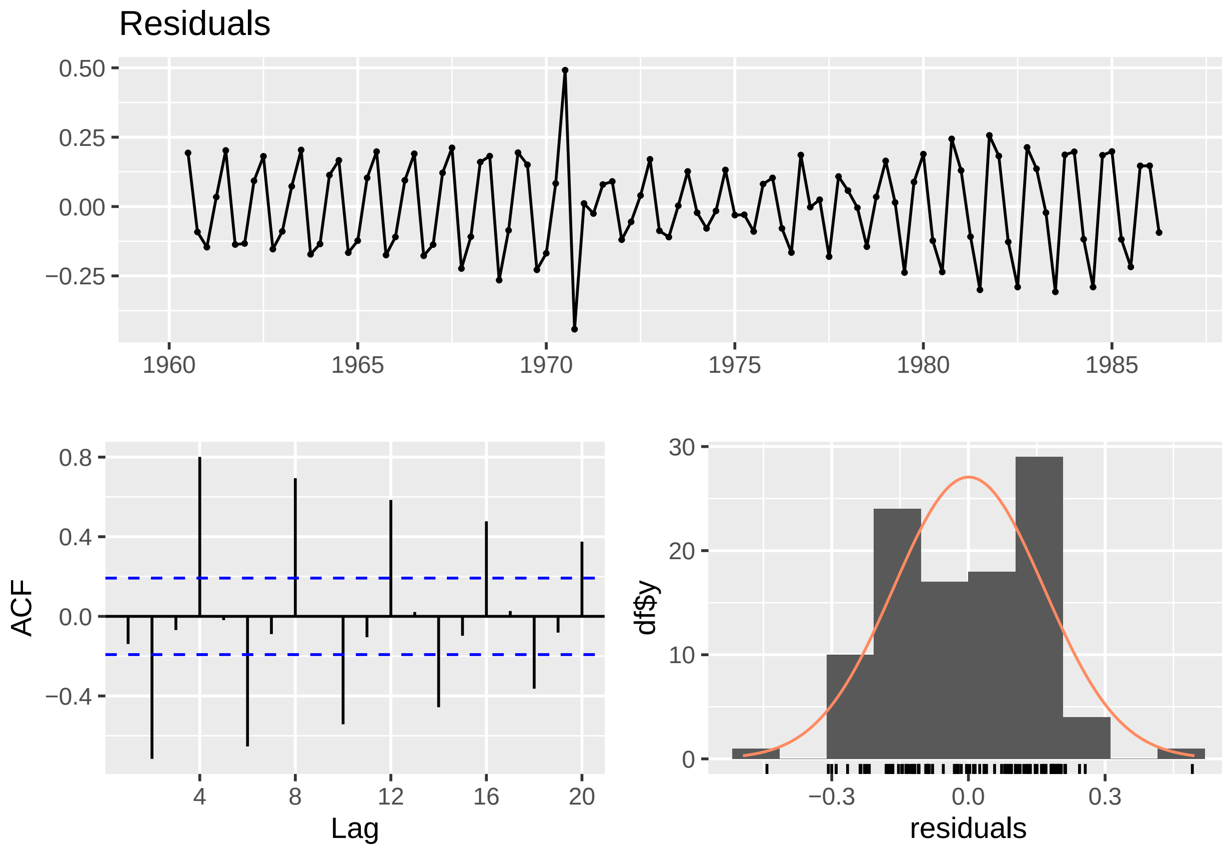
If we take the log of the time series we can see that the heteroskdasticity of the data diminishes. By using log(UKgas) we can use an additive time series decomposition and still get a good representation of the trend and seasonal components. At the same time, the decomposition remains a good fit of the raw data, as we can see from the residuals.



logUK
7.0
6.5
6.0
5.5
5.0
4.5
1960
1965
1970
1975
1980
1985
Time



Decomposition of additive time series
data
trend
seasonal
remainder
7
6
5
6.5
6.0
5.5
5.0
0.25
0.00
−0.25
−0.50
0.50
0.25
0.00
−0.25
1960
1965
1970
1975
1980
1985
Time

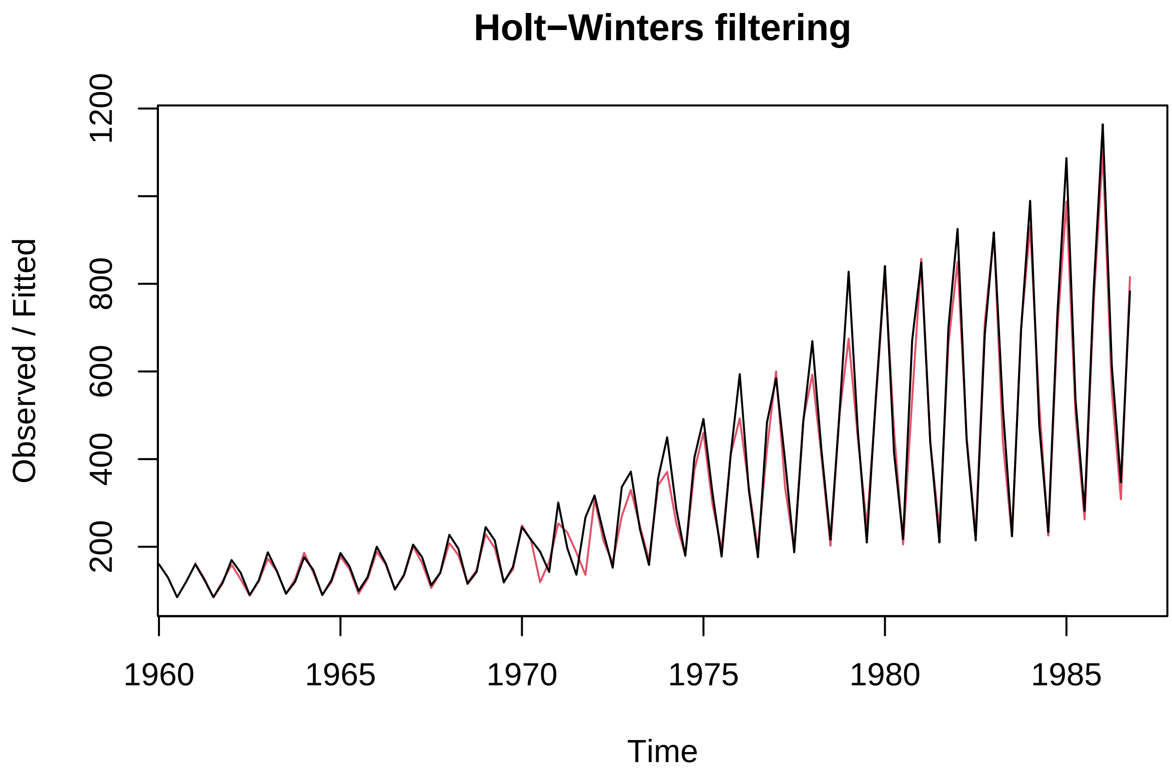


Ljung-Box test

data: Residuals Q* = 232.99, df = 8, p-value < 2.2e-16

Model df: 0. Total lags used: 8

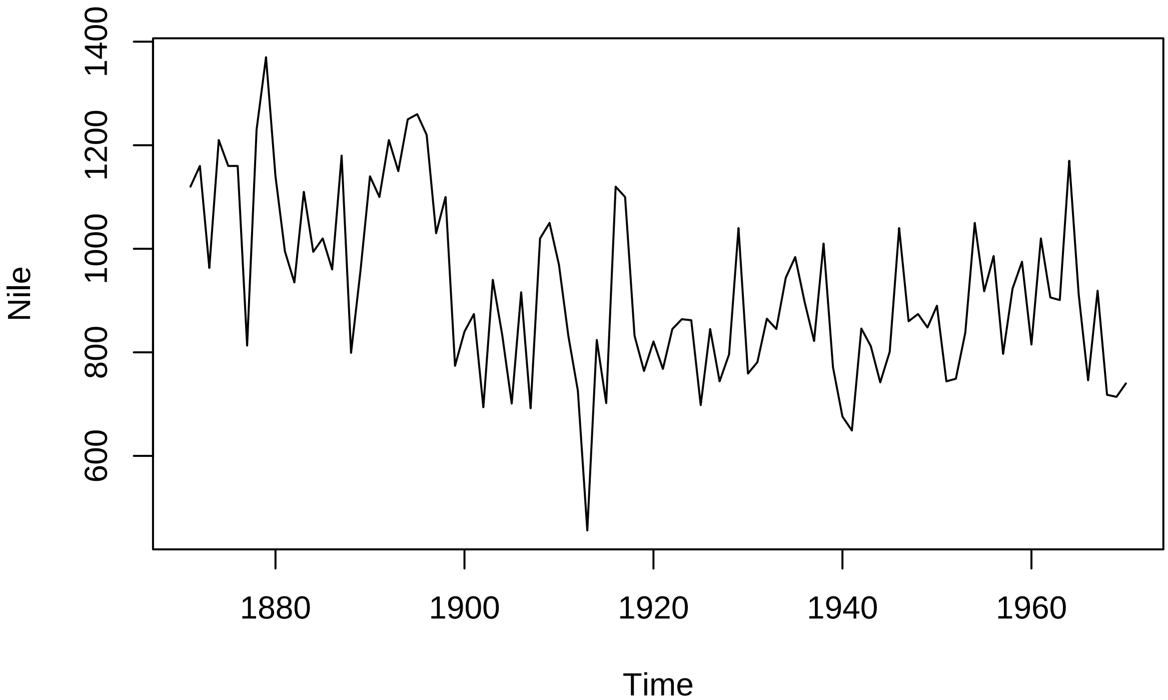
We can try to create a forecast of our time series and compare it with the original dataset.



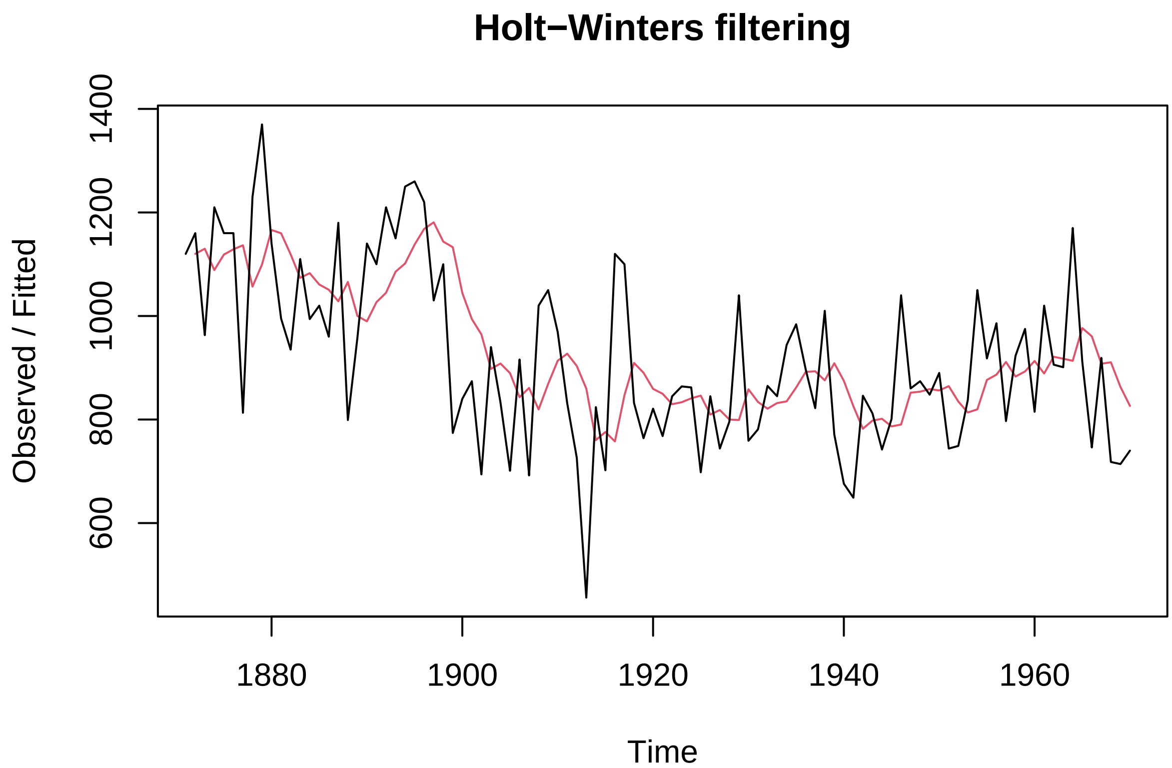
From the plot comparison, it seems that H&W is a good predictor of the raw data. However, we can get a more accurate value of the predictive performance by using the MAPE. As we can see, the predictive performance of the H&W estimates is poor: on average the predictions are 42.45% away from the target.

```
[1] 0.4245779
```

Exercise 3



The value of alpha chose is 0.2465579



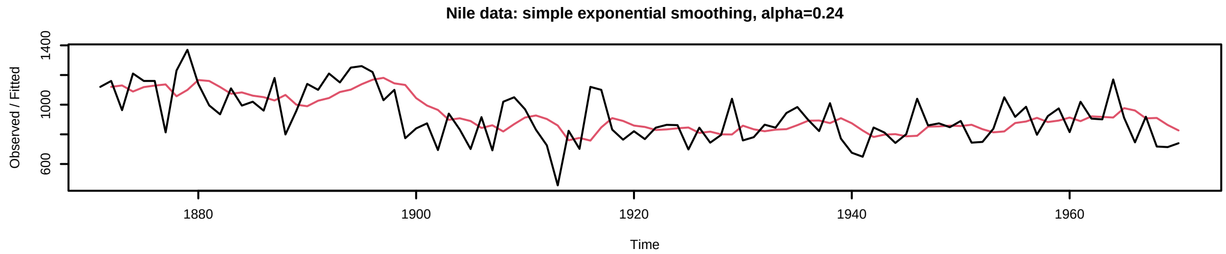
Holt-Winters exponential smoothing without trend and without seasonal component.

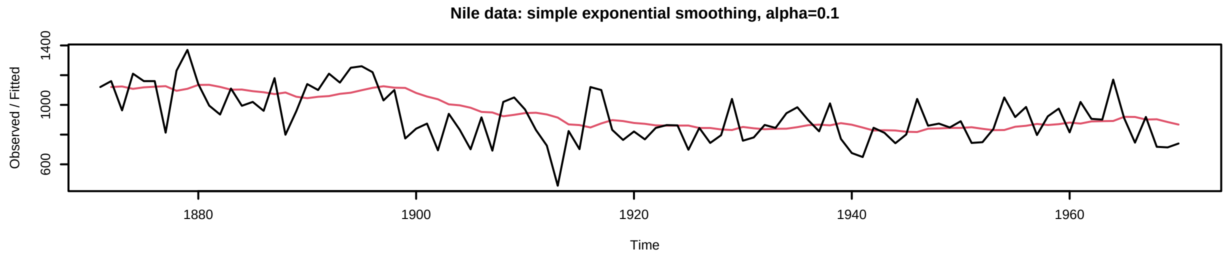
Call: HoltWinters(x = Nile, beta = F, gamma = F)

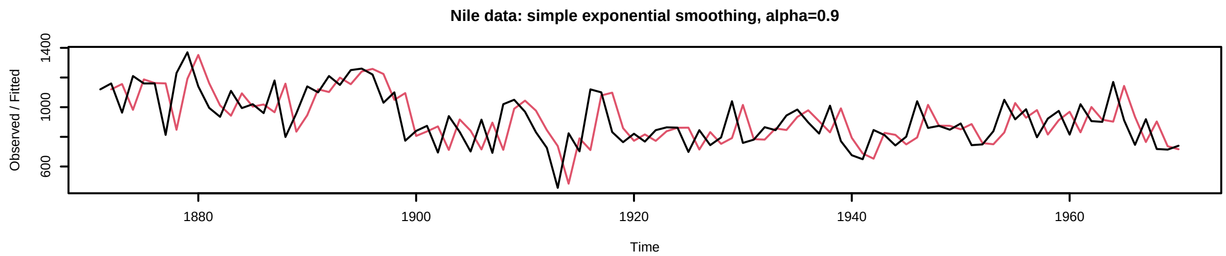
Smoothing parameters: alpha: 0.2465579 beta : FALSE gamma: FALSE

Coefficients: [,1] a 805.0389

We can see from the graphs that as α approaches 1, the estimated value converges to the actual data. If instead α approaches 0, the fitted values converge to a constant.





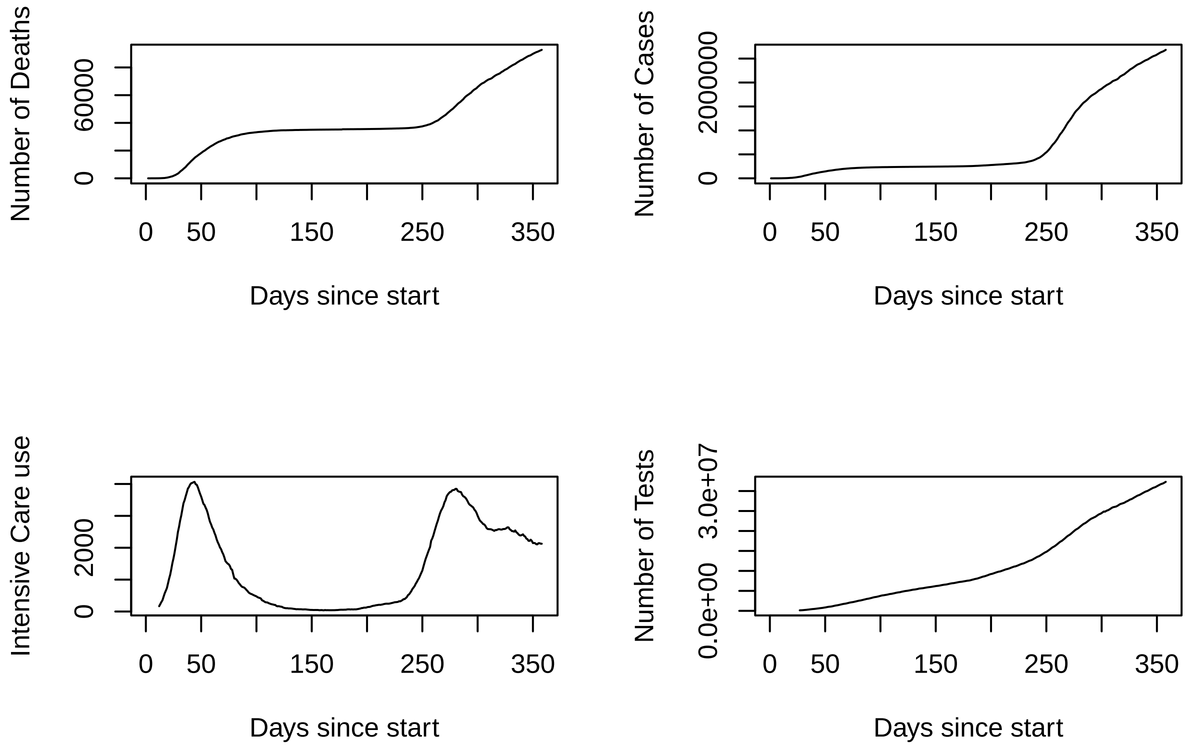


[1] 0.1307089 [1] 0.1348183 [1] 0.144959

By looking at the 3 MAPE we can see that the HW with α=0.24 is the one with the highest predicting performance, with predictions that are on average 13.07% away from the actual values.

Exercise 4

'data.frame': 358 obs. of 5 variables: $ Day : chr "20-feb" "21-feb" "22-feb" "23-feb" ... $ Contagion : int 3 16 79 157 229 322 400 650 888 1128 ... $ Deaths : int NA 1 2 3 7 11 12 17 21 29 ... $ IntensiveCare: int NA NA NA NA NA NA NA NA NA NA ... $ TestsTamponi : int NA NA NA NA NA NA NA NA NA NA ...


Number of Deaths
0
60000
0
50
150
250
350
Days since start
Number of Cases
0
2000000
0
50
150
250
350
Days since start
Intensive Care use
0
2000
0
50
150
250
350
Days since start
Number of Tests
0.0e+00
3.0e+07
0
50
150
250
350
Days since start