Time Series Analysis
Assignment 1

Authors

Ghassen Lassoued - s196609

March 5, 2020

Contents

1 Question 1.1

The last 25 observation in the data will be used for testing. The rest is used for training.

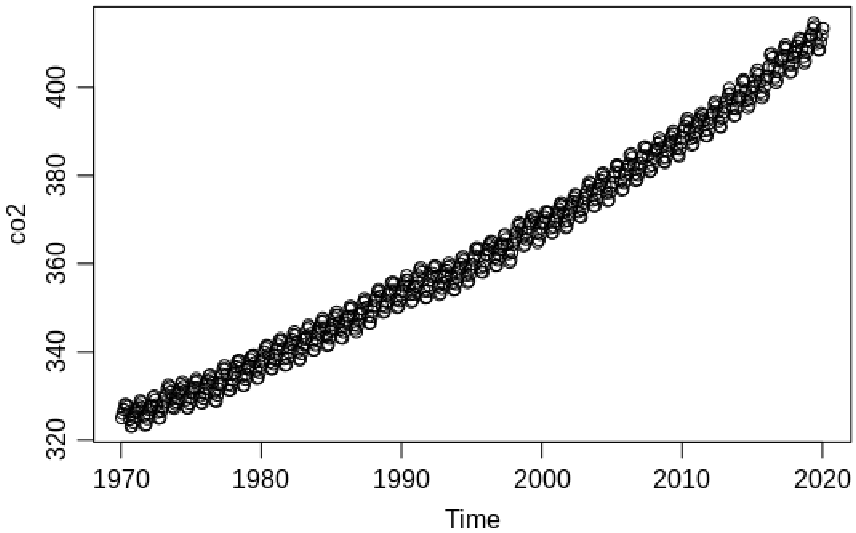

Figure 1: CO2 as a function of time

From this plot, it is relevant to search for a linear regression. We can see that the value of CO2 are increasing by the time.

2 Question 1.2

The anticipated model is of the form:

$$Y_t = \alpha + \beta_t + \beta_s sin((\frac{2\pi t}{p}) + \beta_c cos(\frac{2\pi t}{p}) + \epsilon_t \tag{1}$$

$$\hat{\theta} = \begin{bmatrix} \alpha \\ \beta_t \\ \beta_s \\ \beta_c \end{bmatrix} \tag{2}$$

The period p is chosen 12.
1)Using the OLS linear regression model we can estimate the parameters .
We use the theorem 3.1 for θ .

$$\hat{\theta} = \begin{bmatrix} -2999.454 \\ 1.685373e \\ 0.09939631 \\ 0.2875789 \end{bmatrix} \tag{3}$$

For $\hat{\sigma}$ we use the expression given in the book in Definition 3.2 : $\hat{\sigma}$ = 3,057715

2)The measure of uncertainty is given by the formula in page 37
$\hat{\theta}$ - θ =
$\begin{bmatrix} -2,723734e-09 \\ 1,365926e-12 \\ 4,431779e-13 \\ 1,260103e-14 \end{bmatrix}$
3)We plot the fitted values together with the data :

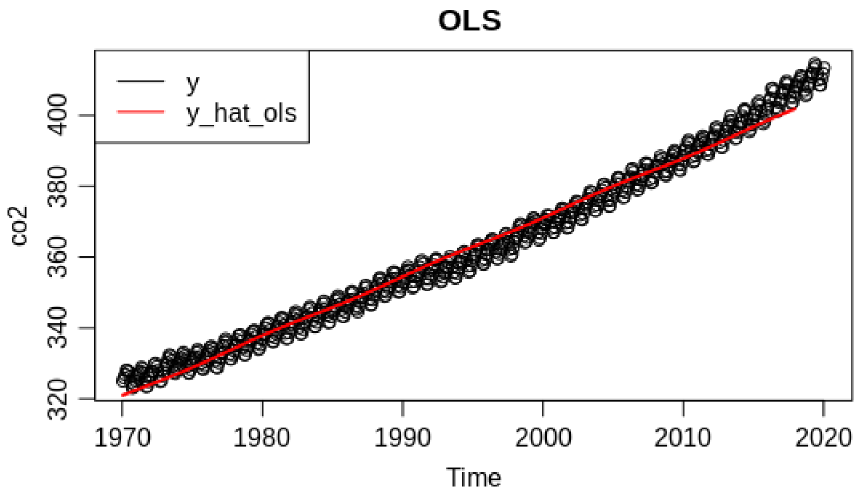

Figure 2: CO2 as a function of time and the OLS fitted model

4-5)In this part, we will be applying WLS model .
After applying the relaxation algorithm in the book page 41, we find ρ = 0,9161076
6)The WLS gives :

$$\hat{\theta} = \begin{bmatrix} -3004,2016984 \\ 1,6878318 \\ 0,3263573 \\ 0,3924329 \end{bmatrix} \tag{4}$$

A measure of uncertainty on θ can be obtained as before.
$\hat{\theta}$ - θ =

$$\begin{bmatrix} 4,747890573 \\ -0,002459173 \\ -0,226961026 \\ -0,104854022 \end{bmatrix}$$

An unbiased estimater for σ^2 is given by the formula (3.44) page 39. It gives : $\sigma =$ 10,0184

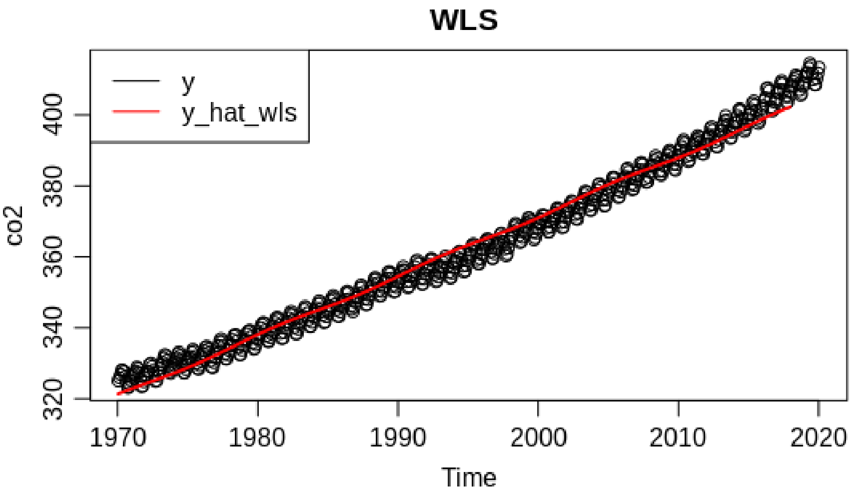

Figure 3: CO2 as a function of time and the WLS fitted model

8)We have obtained bigger σ in WLS model. To compare these 2 models, it is logical to check the σ distribution, which is assumed to be normally distributed. The sum of OLS residuals squared is equal to 5347,985.
The sum of WLS residuals squared is equal to 5380,496 which is higher.
More over, the uncertainty on θ for OLS is much smaller.
It seems that the OLS model fits the training data better than the WLS model.

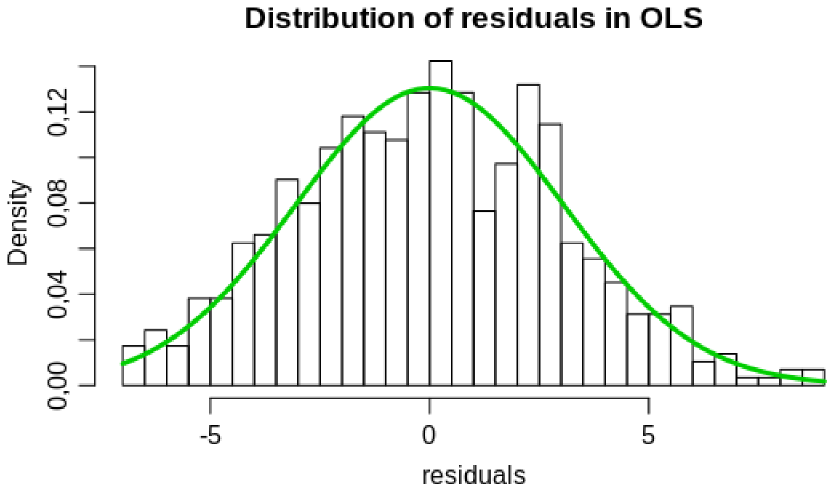

Figure 4: Distribution of residuals in OLS

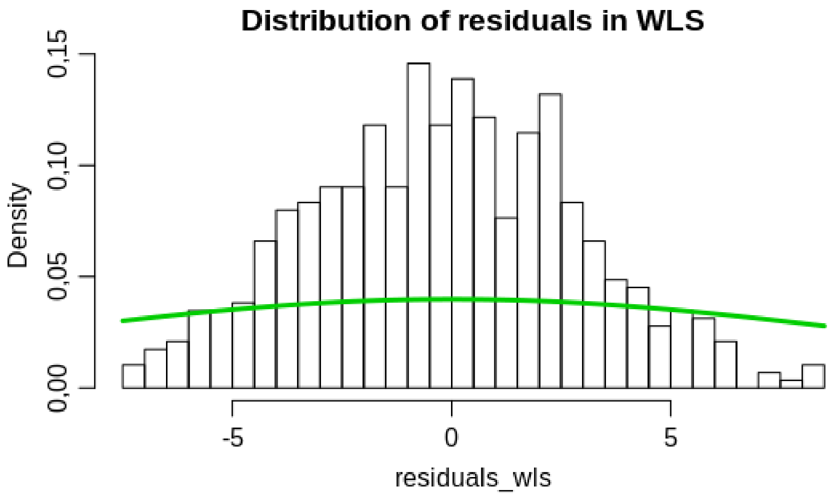

Figure 5: Distribution of residuals in OLS

3 Question 1.3

1)In this part, we will be using a local model trend. By applying the formulas for the harmonic model with the period p on page 54 of the book, we find

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & cos((\frac{2\pi t}{p}) & sin((\frac{2\pi t}{p}) \\ 0 & 0 & -sin((\frac{2\pi t}{p}) & cos((\frac{2\pi t}{p}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0,8660254 & 0,5000000 \\ 0 & 0 & -0,5000000 & 0,8660254 \end{bmatrix}$$

$$f(j) = \begin{bmatrix} j \\ sin((\frac{2\pi j}{p}) \\ cos((\frac{2\pi j}{p}) \end{bmatrix} \text{ which gives } f(0 = \begin{bmatrix} 0 \\ 0 \\ 1) \end{bmatrix}$$

2-3)For this model, we have to initialize F_N and h_N using the first 10 observations. For that, the formulas (3.100) that gives F_N and h_N are on page 56.
Then step by step we update F_N and h_N using the formulas (3.104) on page 57. $th\hat{e}ta$ is provided as well by updating using formula (3.103) on page 57. σ is estimated using the formulas given in the solution of the exercice 3.6 question 1.
The first 10 observations are skipped as transient. The one step prediction errors and the resulting estimates of σ for each observation are below.

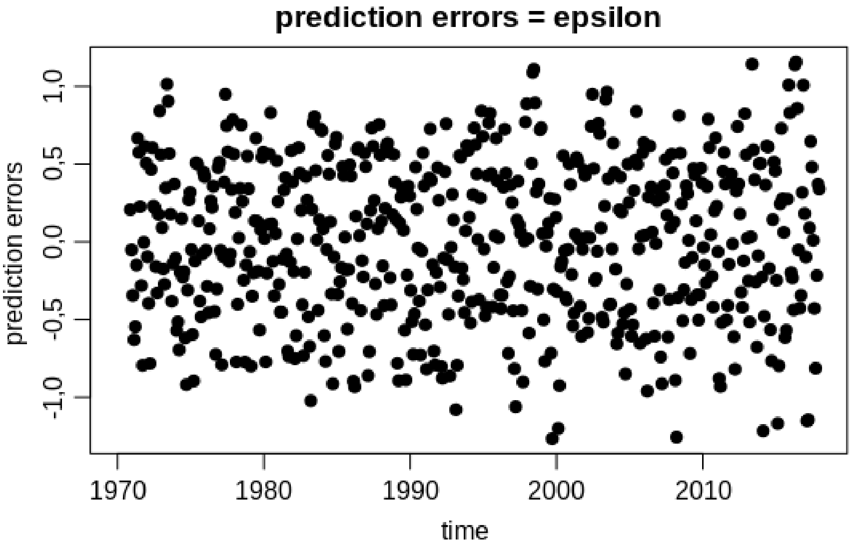

Figure 6: Prediction errors

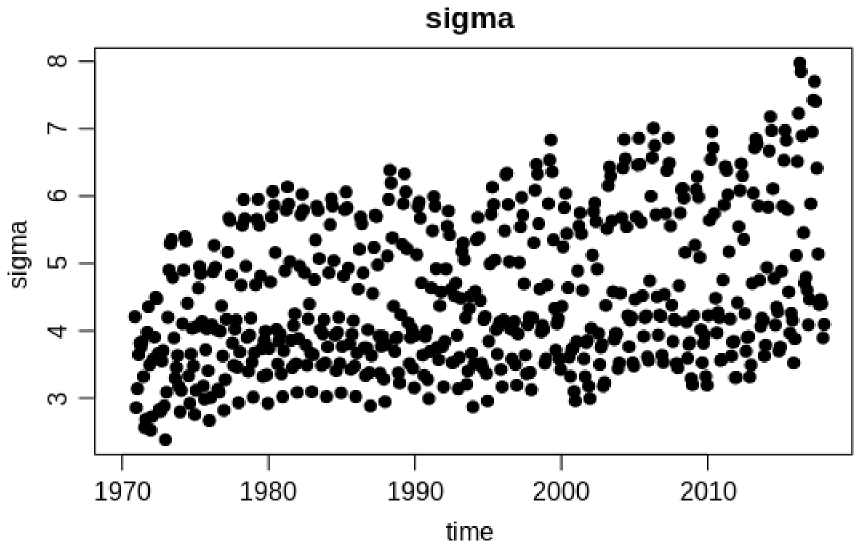

Figure 7: Sigma

4)For the one step prediction we use the formula (3.90) on page 55. For the prediction intervals (3.92) of the same page is used.

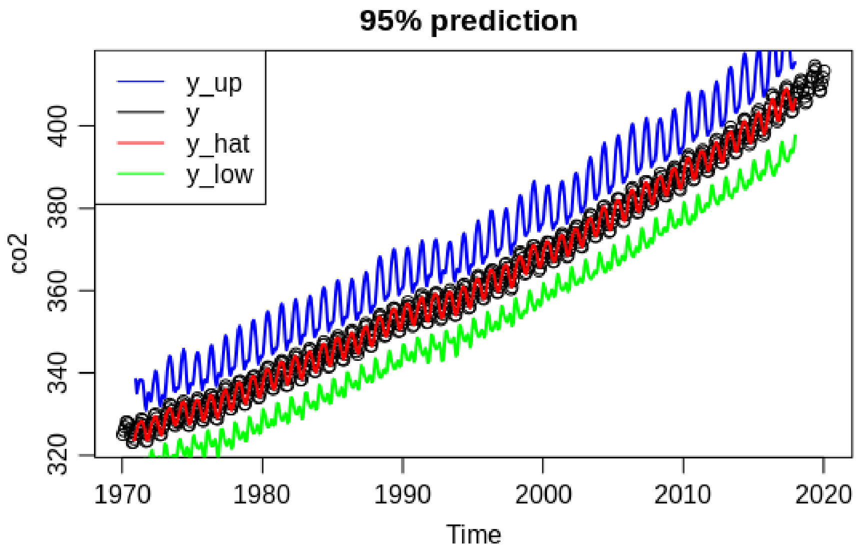

Figure 8: Predictions

The black graph correspond to the training data. The red line is the predicted values , the blue line is obtained by adding half of the prediction interval,the green line is obtained by adding half of the prediction interval. All the training values and the predicted values are in the prediction interval. However, the shape of the blue and green lines are different. This should not be the case likely.

5)The prediction of the test data is included here.

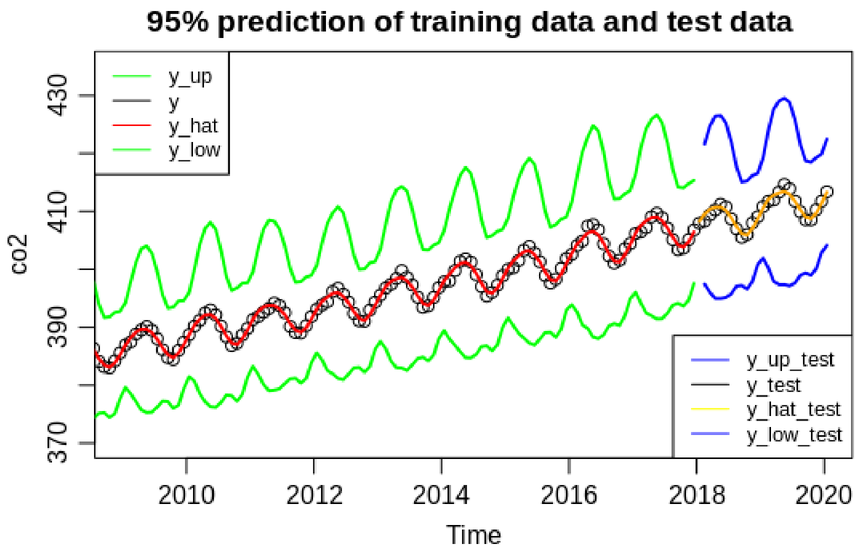

Figure 9: Predictions on training data and test data

The test data and the predicted test data are always in the prediction interval. However, there is a discontinuity in the prediction interval.

6-7) Starting from the last observation that correspond to [2017 12 2017.958 406.81] , which is the observation number 576 = 601 - 25 .

Time ahead in months	1	2	6	12	24
Prediction	408,2867	409,5451	410,2187	408,6895	411,3833
Test data	407,96	408,32	410,79	409,07	411,76
Half of prediction interval	NA	12,09666	15,07477	7,89025	8,503613

Due to the discontinuity in the prediction interval, a NA is generated in one month ahead. It should be a problem with indexing in the code.

The prediction is Prediction $\pm$ Half of prediction interval. E.g. the second prediction is 409,5451 $\pm$ 12,09666 with 95 % confidence. The predicted data and the test data are very close. This model generates well on the new test data . However, it might not be the best model since the λ used here 0.9 may be not the best choice. The prediction interval is relatively small when compared with data values.

8)The estimated mean is the first element in θ i.e it is α in the model.

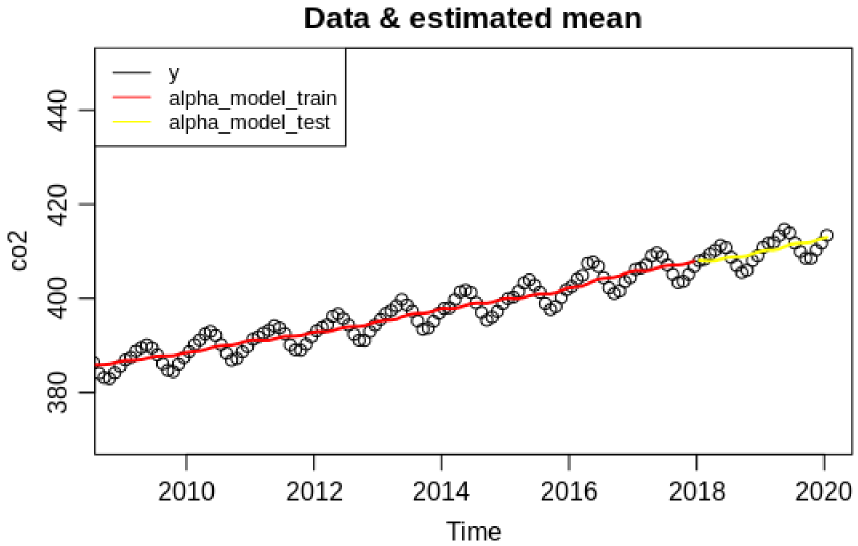

Figure 10: Data estimated mean

4 Question 1.4

1) λ is in $]0;1[$, so is α
The burning period is 100months i.e 100 steps. As it was demanded to skip first 10 observations, we will start the 100 steps from observation 11.

5 Code

List of Figures