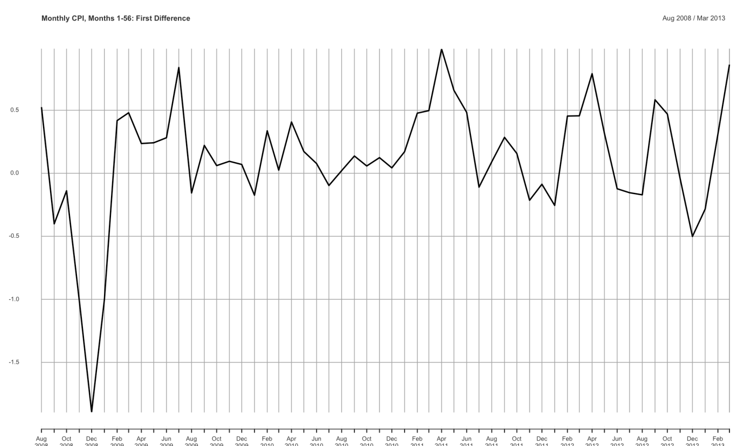
Olamide OLADEJI

Problem 6.1. BPP Data Analysis
Q6.1a: Order Selection.

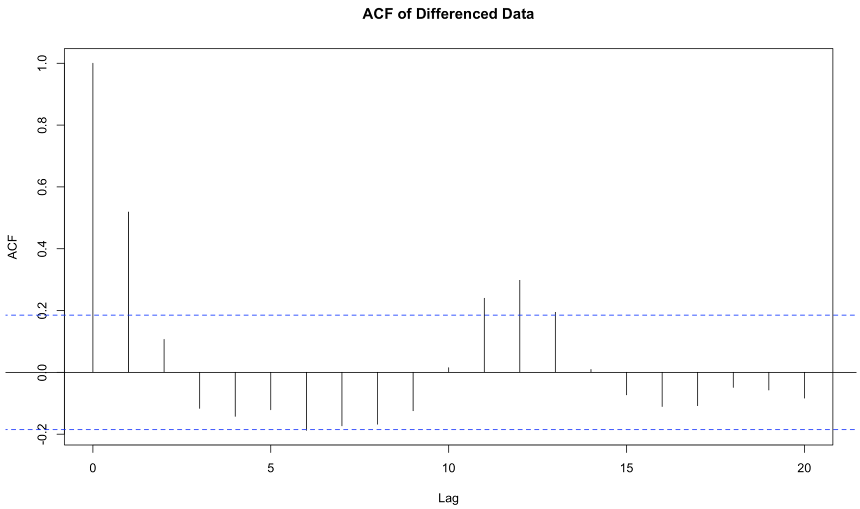
To determine the best order to use for model fitting, the first step undertaken was a visualization of the data shown below.

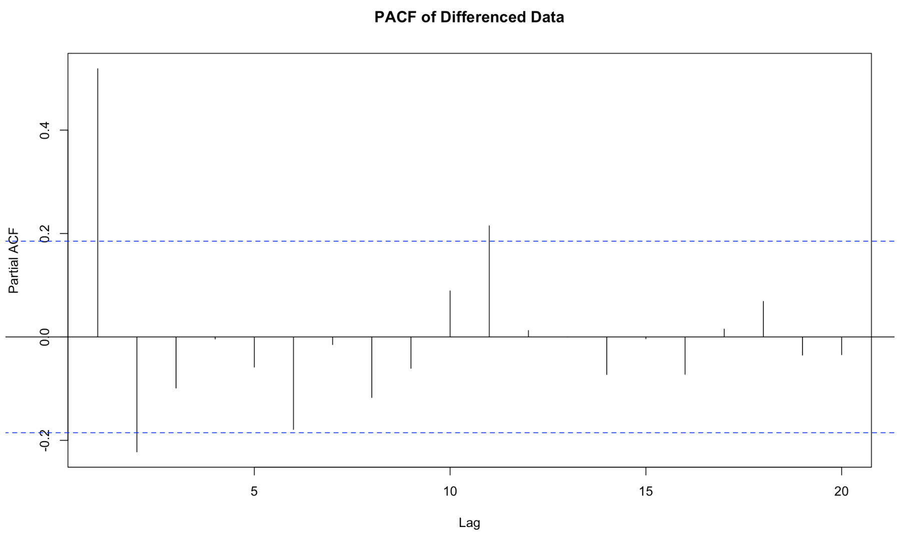
. The data was seen to be non-stationary and so I differenced the data to stationarize the data. The differenced data is shown below:



I then plotted the AutoCorrelation Functions and Partial AutoCorrelation Functions of the differenced monthly CPI data.

Olamide OLADEJI

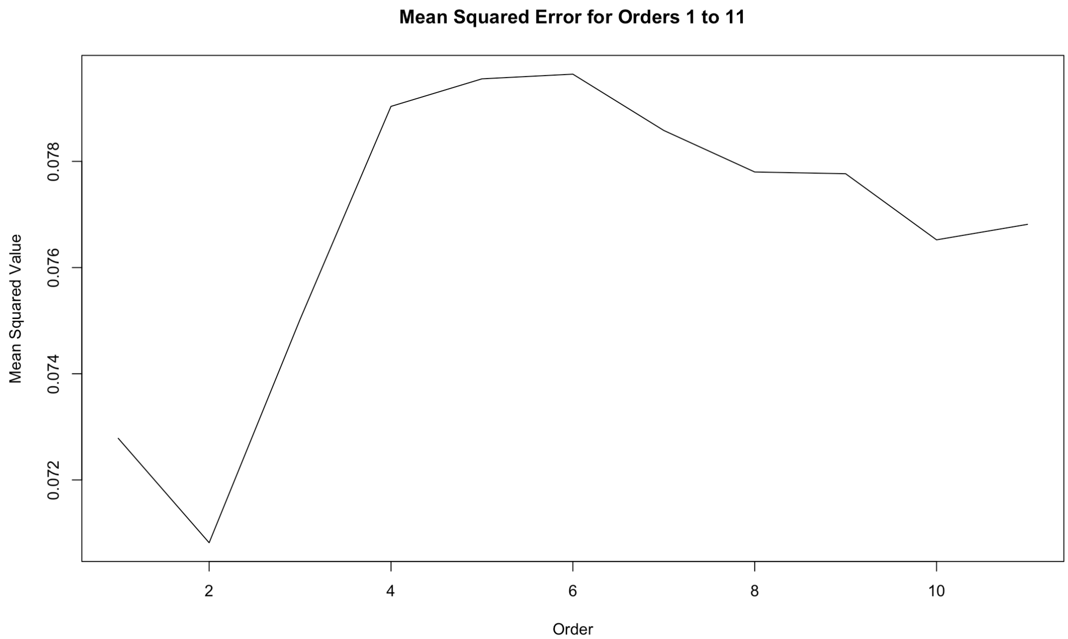




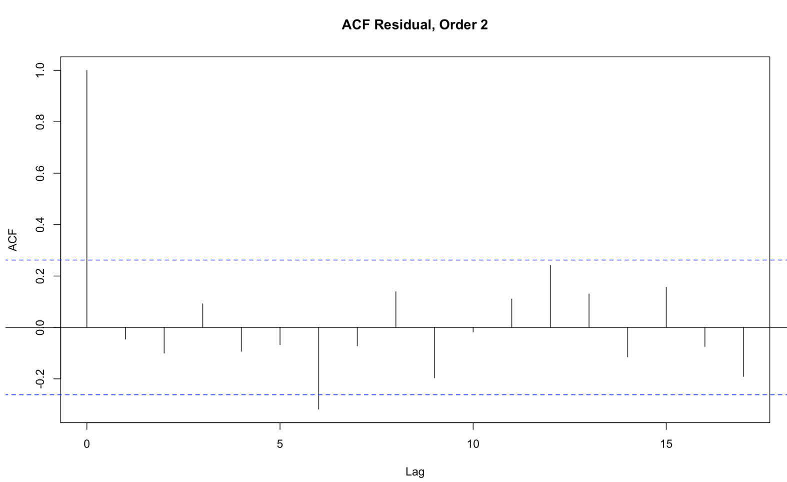
The ACF suggests an order of 1 should fit the model. However, we try to explore other orders as instructed to see if there is any other order that can lead to lower prediction metric (Mean Squared Error – MSE) than that of order 1.

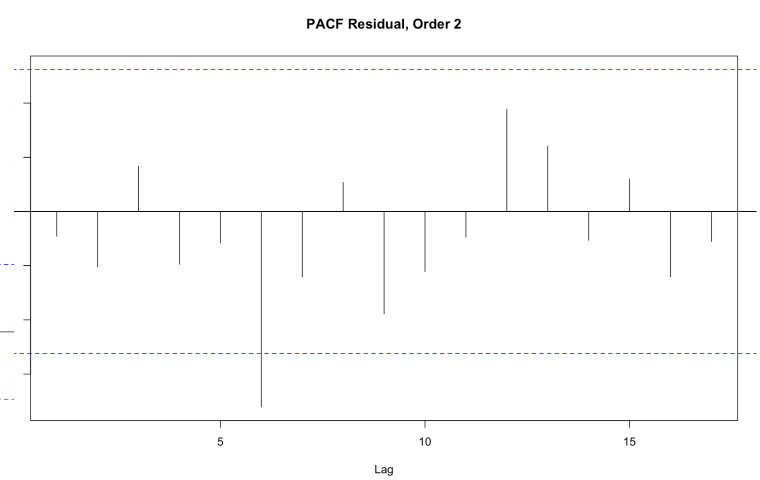
I examined fits of AR models of orders from p= 1 to 11 on the differenced data, i.e. *"order = c(p,1,0)" in an arima R function,* using the first 56 months. As instructed in the question, for the remaining months for evaluation, we then continuously make 1 month ahead forecasts that applying the originally fitted model and its coefficients on all the previous months data before the month to be forecasted, going all the way to the last month.

The comparisons of prediction metric for different orders resulted in the following plots of MSE versus order.



The prediction metric (MSE) plot comparisons suggest that order 2 might be a better fit to the data. This is confirmed by the plot of the auto-correlation and partial auto-correlation functions of the residuals after prediction with AR(2). These, as shown below, show no correlation in the residuals as the function values are all within the 'noise' band.

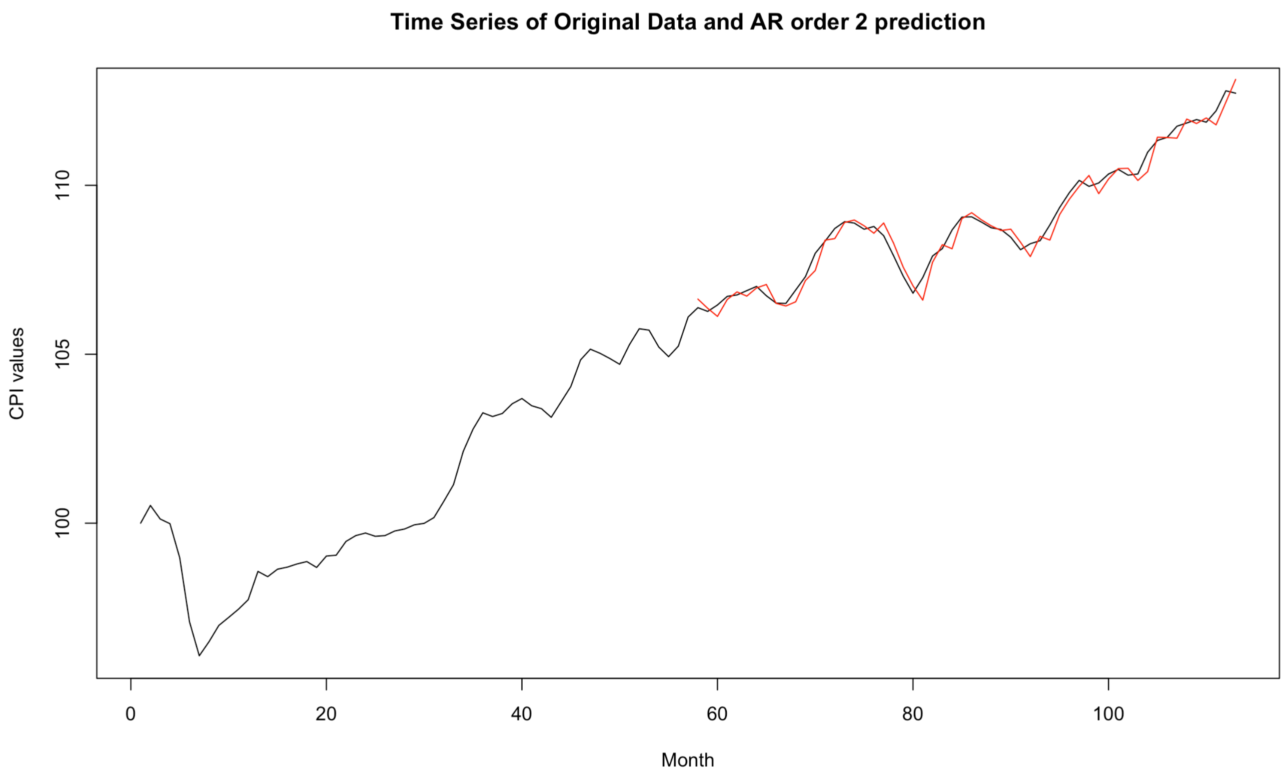




Thus our order 2 result makes sense and we can fit an order 2 AR model to the data and achieve **the lowest prediction metric, the Mean Squared Error(MSE) = 0.070817**.

A plot of the predicted data is shown below. The original data is in black while the predicted CPI is indicated with the red line plot.

Olamide OLADEJI

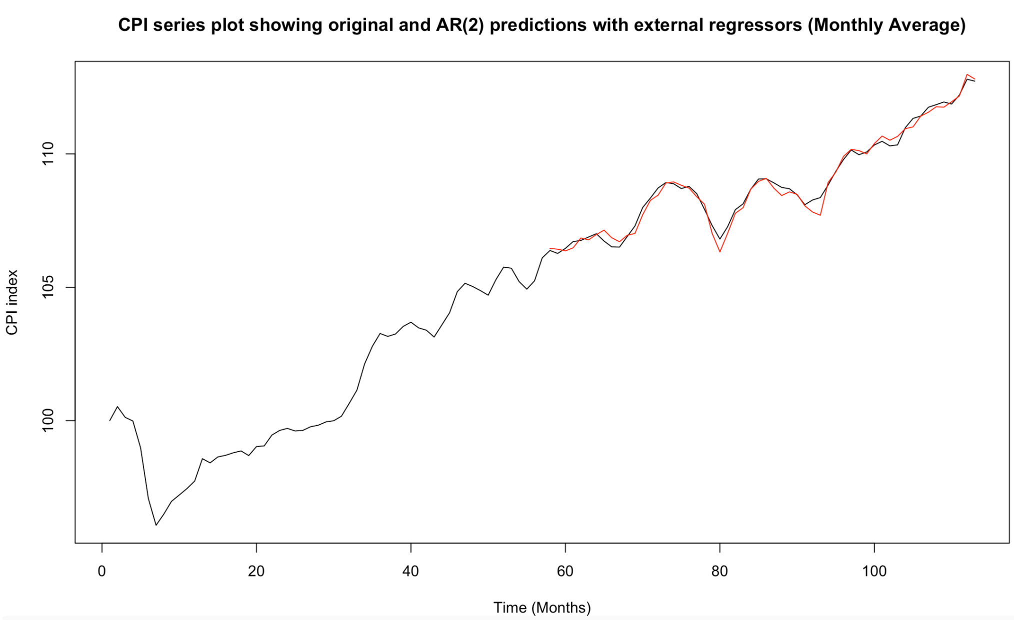


Q6.1b: External Regressors.

i. **Monthly Average of BPP and BER**

I computed a monthly average of billion prices data (BPP) and the break-even rate (BER) data, using the 'xreg' argument in R.

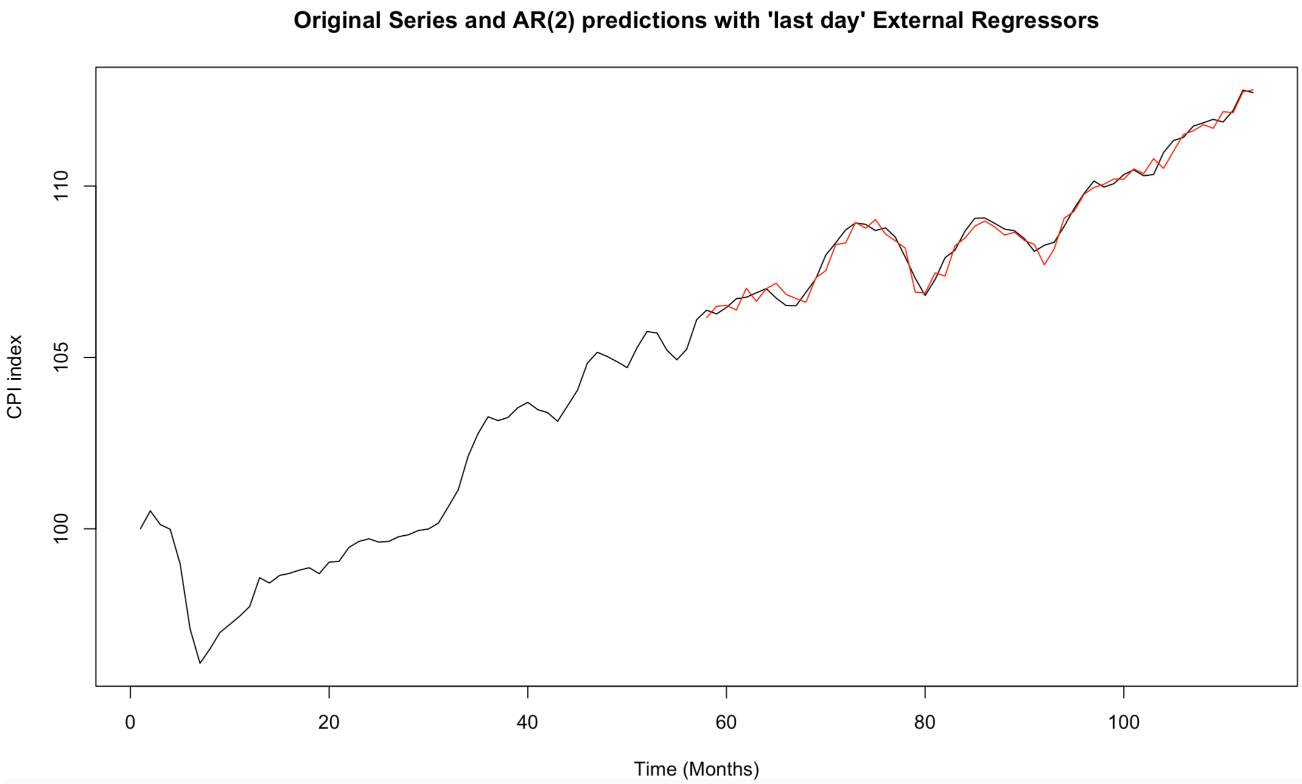
This provided the following plot of predicted CPI.



Again, note that the predicted values are the red line while the original data is the black line.

The Mean Squared Error for this prediction is: **0.04351768.** This is lower than just using only CPI data. Thus, the external regressor data of average monthly has improved our prediction, reducing the error by 38.5%[1].

ii. **Last Day of the Month External Regressor**
I similarly extracted the last day of the month data values from the BPP and BER data and used this as an external regressor to fit the model on the first 56 months as required. This model was then used in evaluation of the last 56 months, taking care to incorporate all the previous months values in the prediction as has been done.
The prediction metric, MSE, for the resulting prediction obtained was**: 0.0586081**



Black line = Original CPI data
Red line = Predicted values

[1] Note that we used previous month's data as external regressor value for current month in the fitting and prediction of the data. There was no explicit instruction on whether to use current month bpp + Ber value as external regressor for the month or previous month like we did. Using previous month seemed more intuitive, and provided lower MSE error so that's what is presented above. We seemingly interpreted 'last day of month' as "last day of PREVIOUS month".

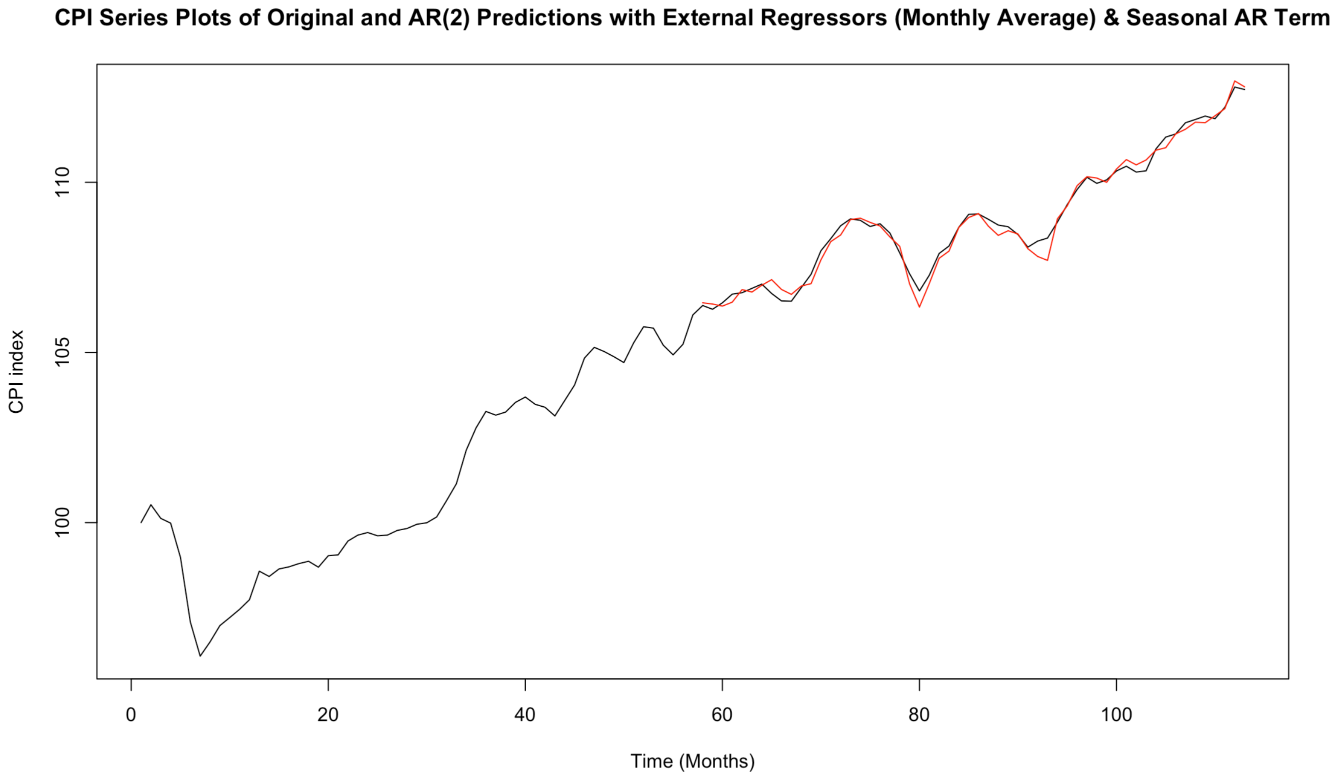
Olamide OLADEJI

Which performs better?: We can see that the MSE value of the monthly averaged BPP and BER data as external regressor is lower than that using last day values. Thus we choose the monthly average as our better model.

Q6.1c: Improving the better model of b.

To improve the monthly average model from part b, I played around with the effect of adding MA terms as well as AR seasonal terms as suggested. Adding MA terms did not really lead to a decrease in MSE. I however obtained a decrease in MSE when I finally added an AR seasonal term of 2 with no MA terms.

The resulting fitted model gave a prediction plotted below:



The MSE with the addition of the seasonal AR 2 term to the model is**: 0.04315607.**

This is lower than without the seasonal term, so we have a final improved model which as an order 2 AR model, on the first differenced data, and seasonal AR of 2 -- i.e. "`order=c(2,1,0), seasonal=c(2,0,0))`" -- with the monthly averaged BPP and BER data as external regressors.[2]

[2] I also tried to improve the worse model from 6.1b: the one incorporating last day of month model, by experimenting with adding seasonal and MA terms but the MSE never got lower so they're not reported here as I decided to report the overall final best improved model.

Olamide OLADEJI

Q6.1d: Autocovariance function Of MA(1) model.

d)

$$\gamma(h) = \mathbb{E}(X_t X_{t+h}) = \mathbb{E}[(w_t + \theta w_{t-1})(w_{t-h} + \theta w_{t-h-1})]$$

$$= \mathbb{E}[w_t w_{t-h} + \theta w_{t-1} w_{t-h} + \theta w_t w_{t-h-1} + \theta^2 w_{t-1} w_{t-h-1}]$$

$$= \mathbb{E}[w_t w_{t-h}] + \theta \mathbb{E}[w_{t-1} w_{t-h}] + \theta \mathbb{E}(w_t w_{t-h-1}) + \theta^2 \mathbb{E}(w_{t-1} w_{t-h-1})$$

when $h = 0$

$$\gamma(0) = \mathbb{E}(w_t w_t) + \theta \mathbb{E}[w_{t-1} w_t] + \theta \mathbb{E}[w_t w_{t-1}] + \theta^2 \mathbb{E}[w_{t-1} w_{t-1}]$$

$$= \sigma_w^2 + \quad + \theta^2 \sigma_w^2$$

$$= \sigma_w^2 (1 + \theta^2)$$

when $h = 1$

$$\gamma(1) = \mathbb{E}(w_t w_{t-1}] + \theta \mathbb{E}(w_{t-1} w_{t-1}) + \theta \mathbb{E}(w_t w_{t-2}) + \theta^2 \mathbb{E}(w_{t-1} w_{t-2}]$$

$$= \theta \sigma_w^2$$

when $h > 1$

$$\gamma(h) = 0$$

so
$$\gamma(h) = \begin{cases} \sigma_w^2 (1 + \theta^2 & \text{when } h = 0 \\ \theta \sigma_w^2 & \text{when } h = 1 \\ 0 & \text{when } h > 1 \end{cases}$$

where $\sigma_w^2 = \sigma^2$ ~~of w~~

Q6.1e: Autocovariance function Of AR(1) model.

Olamide OLADEJI

Collaborator for this question: Wenjia Wang

c)

$$\gamma(h) = \mathbb{E}[X_t X_{t-h}] = \mathbb{E}[(\phi X_{t-1} + W_t)(\phi X_{t-h-1} + W_{t-h})]$$

$$= \mathbb{E}[\phi^2 X_{t-1} X_{t-h-1} + \phi X_{t-h-1} W_t + \cancel{\phi X_{t-h-1} W_t} + \phi X_{t-1} W_{t-h} + \cancel{\phi X} + W_t W_{t-h}]$$

$$= \phi^2 \mathbb{E}[X_{t-1} X_{t-h-1}] + \phi \mathbb{E}[\cancel{X_{t-h-1} W_t}] + \phi \mathbb{E}[\cancel{X_{t-1} W_{t-h}}] + \mathbb{E}[W_t W_{t-h}]$$

$$= \phi^2 \mathbb{E}[X_{t-1} X_{t-h-1}] + \mathbb{E}[W_t W_{t-h}]$$

when $h = 0$

$$\gamma(0) = \phi^2 \mathbb{E}[X_{t-1}^2] + \mathbb{E}[W_t^2]$$

$$\Rightarrow \mathbb{E}[X_t^2] = \phi^2 \mathbb{E}[X_{t-1}^2] + \mathbb{E}[W_t^2]$$

or $\gamma(0) = \phi^2 \gamma(0) + \sigma_w^2$

$$\Rightarrow \gamma(0) = \frac{\sigma_w^2}{1-\phi^2} = \frac{\sigma_w^2}{1-\phi^2}$$

when $h \neq 0$

$$\gamma(h) = \mathbb{E}[X_t X_{t-h}] = \mathbb{E}[X_t(\phi X_{t-h-1} + W_{t-h})]$$

$$= \phi \mathbb{E}[X_t X_{t-h-1}] + \mathbb{E}[\cancel{X_t W_{t-h}}]$$

$$\gamma(h) = \phi\, \gamma(h-1)$$

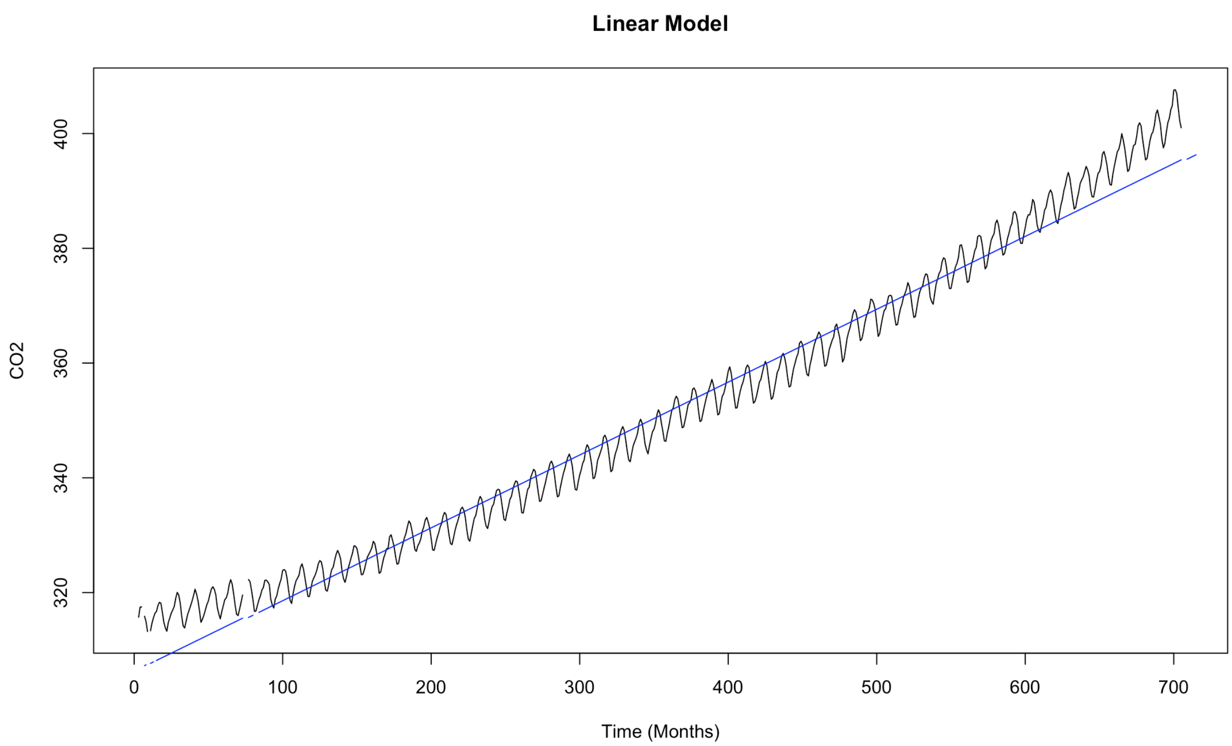
$$= \phi^h \gamma(0)$$

$$= \phi^h \frac{\sigma_w^2}{1-\phi^2}$$

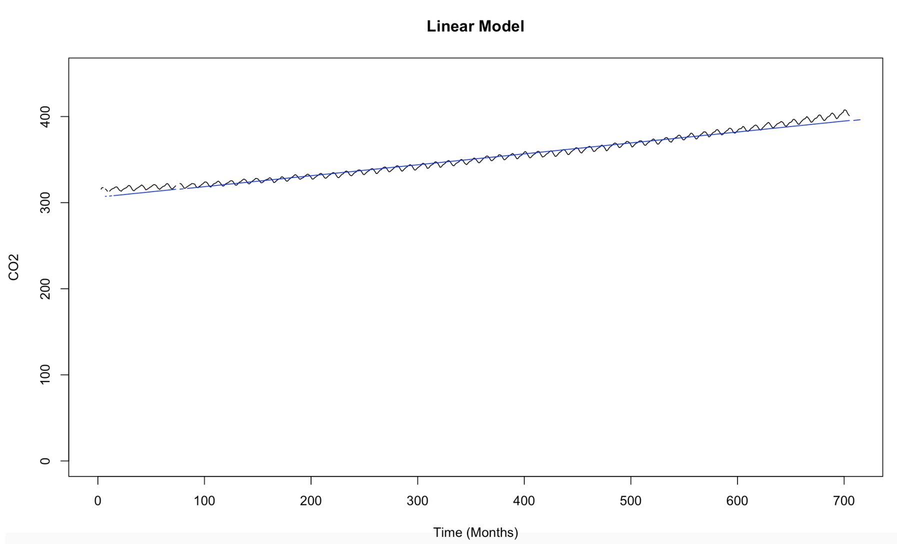
Problem 6.2: The Mauna Loa CO2 concentration

Olamide OLADEJI

Q6.2a. Linear Model
I fitted the data to a linear model in the Form $F_1(t) \sim \alpha 1 + \alpha 2 * t$ as required.
The plot of the data(black) with the linear fit (blue line) superimposed is shown below:



Note that the y axes of the above plot is compressed (for visualization) and the plot does not actually have a positive x intercept. An uncompressed plot is seen below:



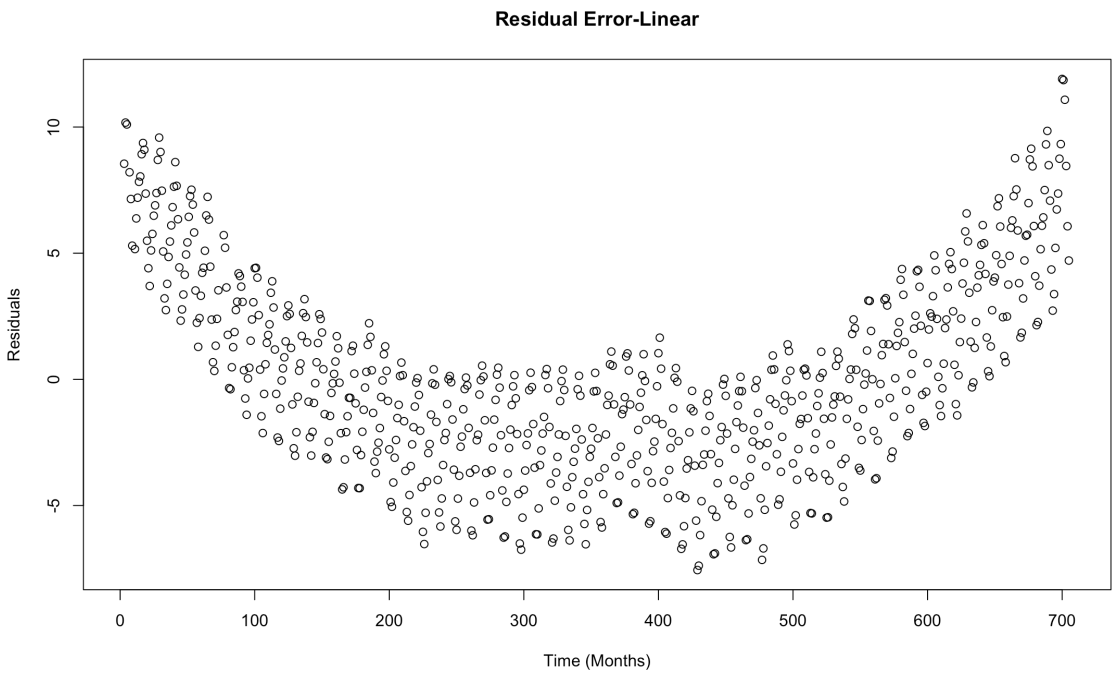
The values for alpha1 and alpha2 obtained from fitting this in R are below.

```
(Intercept) df_2$Date.1
```

```
306.834686     1.523946
```

i.e. $\boldsymbol{\alpha 1}$ **= 306.8347 and** $\boldsymbol{\alpha 2 = 1.524}$

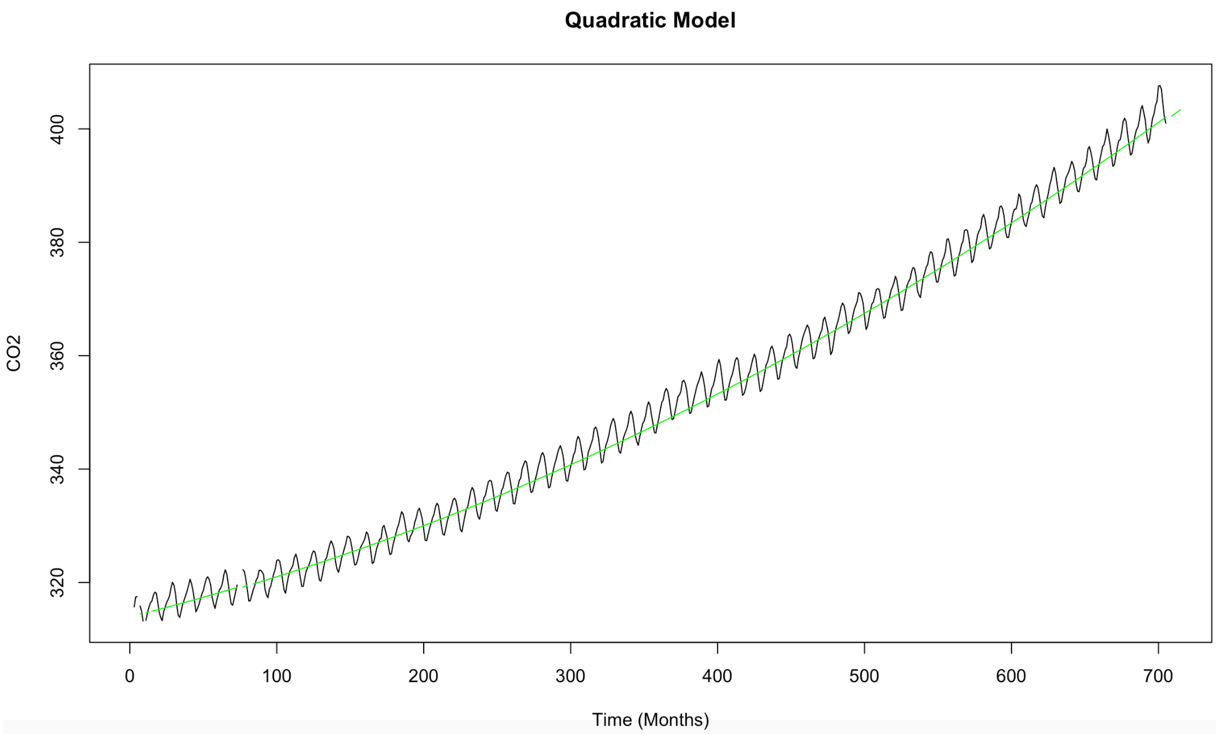
The residual error plot is shown below.



Comment: We can see that the residuals are not randomly distributed and have a distinct shape/trend. We can see the residual-error values go from 10 to -5 and back up to 10 over the months.They are also not centered around 0.

Q6.2b: Quadratic Model

Fitting to a quadratic model F2, using the R function ‘tslm’ yielded the following plot of the data and fit. The quadratic model is the green line.

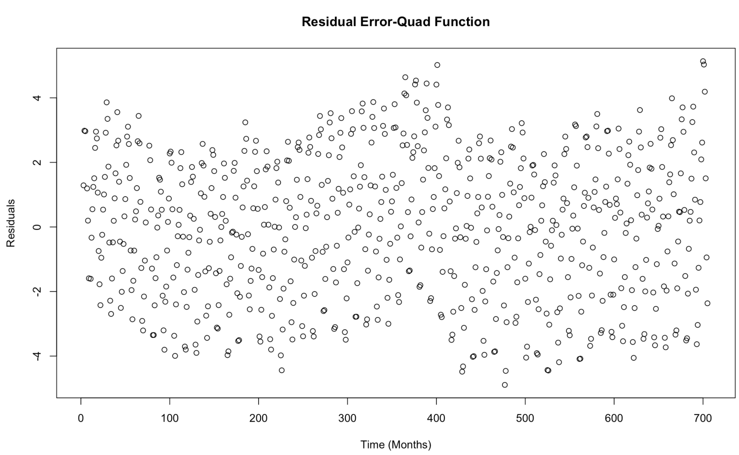


The coefficients are as follows:

```
(Intercept)      df_2$Date.1 I(df_2$Date.1^2)
    314.24134495        0.78341100       0.01251704
```

$$Thus\ \beta1 = 314.2413, \quad \beta2 = 0.7834\ and\ \beta3 = 0.0125$$

The residuals of this fitting are as follows:



The residuals of the quadratic fitting as seen in the plot are centered around 0.

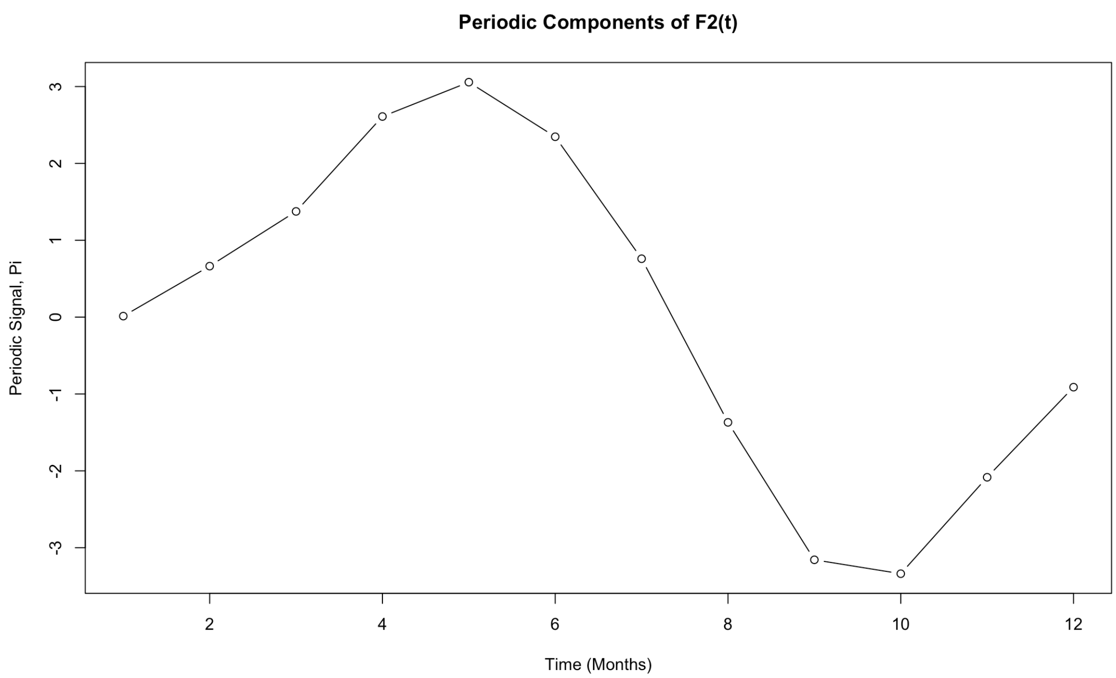
Olamide OLADEJI

6.2c: Better Fit

The quadratic fit F2 is definitely better at capturing the trend in the data. The differences in the residuals show this. With the linear fit, the residual errors still had some obvious long-term trend. However, the residual errors of the quadratic do not seem to have any (significant) trend of upwards or downwards, so I'd say that it is better in capturing the trend in the data.

6.2d: Periodic Signal

We averaged all the data for each month as was required.
This was plotted as shown below



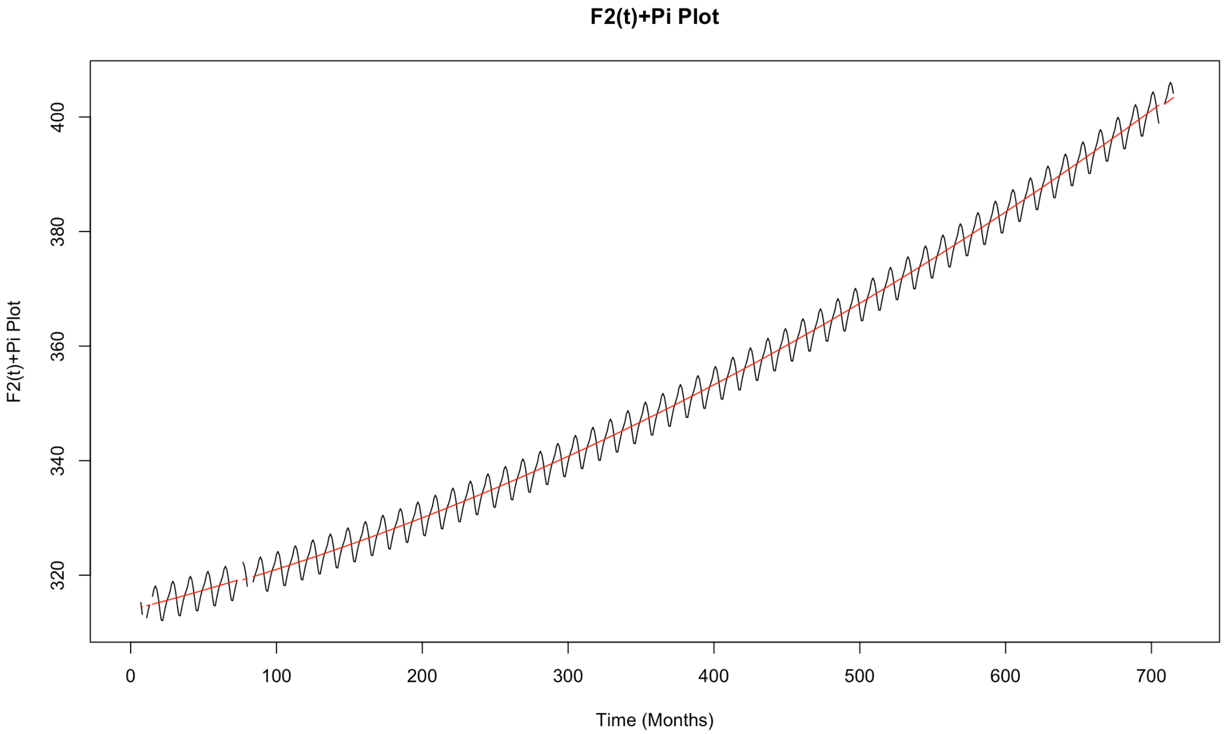
The plot actually shows a seemingly periodic signal as we assumed, with a period of 1 year.

6.2e: Trend + Periodic & Variation

The fit F2(ti) and the Periodic signal were combined and plotted as shown below.

NB: In the plot, I also superimposed the quadratic fit plot for reference.

Olamide OLADEJI



We can see that the mean of the C02 concentration since 1958 has increased quadratically over time. There are also seasonal – periodic – variations in the concentration levels whose amplitude seem constant so perhaps they're due to some natural cyclical processes such as plants uptake or local weather. The increasing mean trend however could be sign of underlying long term issues such as climate change.