Project 1

TMS088: Financial Time Series

João Fernandes
Markus Edlund

Chalmers University of Technology

2021/2022

Question 1

a)

Show that $\nabla^q m_t = q!\, a_q$. Proceeding by induction on q:

Base: $q = 1$, we must show that $\nabla m_t = a_1$

$$\nabla m = m_t - m_{t-1} = \sum_{j=0}^{1} a_j t^j - \sum_{j=0}^{1} a_j (t-1)^j = \sum_{j=0}^{1} a_j (t^j - (t-1)^j) = a_0 \underbrace{(t^0 - (t-1)^0)}_{=0} + a_1 (t - (t-1)) = a_1$$

Step: Firstly, for clarity, let's denote on an extra subscript to denote the value of q in the polynomial m_t. That is, let's denote

$$m_{t,\,q} = \sum_{j=0}^{q} a_j\, t^j$$

We assume that $\nabla^{q-1} m_{t,\,q-1} = (q-1)! a_{q-1}$ and we must prove that $\nabla^q m_{t,\,q} = q! a_q$. Let's start by noting that

$$\nabla^q m_{t,\,q} = \nabla^{q-1} \left(\nabla m_{t,\,q} \right)$$

If we prove that $\nabla m_{t,\,q}$ is a $(q-1)^{\text{th}}$ degree polynomial with leading coefficient $q\, a_q$, then we can use the hypothesis to conclude our result:

$$\nabla^q m_{t,\,q} = \nabla^{q-1} \left(\nabla m_{t,\,q} \right) = (q-1)!\, q\, a_q = q!\, a_q$$

So let's prove those properties about $\nabla m_{t,\,q}$. Note that:

$$\nabla m_{t,\,q} = m_{t,\,q} - m_{t-1,\,q} = \sum_{j=0}^{q} a_j \left(t^j - (t-1)^j \right) = \sum_{j=0}^{q} a_j \left(t^j - \sum_{i=0}^{j} t^i \, (-1)^{j-i} \binom{j}{i} \right) =$$

$$= \sum_{j=0}^{q} a_j \left(t^j - t^j + \sum_{i=0}^{j-1} t^i \, (-1)^{j-i+1} \binom{j}{i} \right) = \sum_{j=0}^{q} a_j \sum_{i=0}^{j-1} t^i \, (-1)^{j-i+1} \binom{j}{i}$$

It should be immediate that the degree of $\nabla m_{t,\,q}$ is $q-1$, achieved when $j = q$ and $i = j-1 = q-1$, and for those values we obtain that the leading coefficient is then

$$a_q \, (-1)^{q-(q-1)+1} \binom{q}{q-1} = q\, a_q$$

As stated before, since $\nabla m_{t,\,q}$ is a $(q-1)^{\text{th}}$ degree polynomial with leading coefficient $q\, a_q$, we can use the induction hypothesis on $\nabla m_{t,\,q}$ which immediately yields the result.

b)

Note that:

$$\nabla^q\, Y_t = (1 - B)^q\, Y_t = \sum_{i=t-q}^{t} b_i\, Y_i$$

What this means is that $\nabla^q\, Y_t$ can be written as some linear combination of $Y_t, Y_{t-1}, ..., Y_{t-q}$. More specifically, the coefficients b_i are given by Newton's binomial $\left(|b_i| = \binom{q}{i-t+q}, \text{ keeping in mind that } sgn(b_i) \text{ alternates} \right)$. This implies:

$$\mathbb{E}[\, \nabla^q\, Y_t \,] = \mathbb{E}\left[\sum_{i=t-q}^{t} b_i\, Y_i \right] = \sum_{i=t-q}^{t} b_i\; \mathbb{E}[Y_i]$$

Since $\mathbb{E}[Y_i] = 0, \forall i$, then we get $\mathbb{E}[\, \nabla^q\, Y_t \,] = 0$, meaning it is in fact constant in t.

c)

$\text{Var}[\,\nabla^q\, Y_t\,] = \mathbb{E}[(\,\nabla^q\, Y_t\, - \mathbb{E}[\,\nabla^q\, Y_t\,]\,)^2] = \mathbb{E}[(\,\nabla^q\, Y_t\,)^2]$, since $\mathbb{E}[\,\nabla^q\, Y_t\,] = 0$ by last question. Note that:

$$(\nabla^q\, Y_t)^2 = (1-B)^{2q}\, Y_t^2 = \sum_{j=t-2q}^{t} c_j\, Y_j^2$$

Again, $(\nabla^q\, Y_t)^2$ can be written as some linear combination of $Y_t^2, Y_{t-1}^2, ..., Y_{t-2q}^2$, where $|c_j| = \binom{2q}{j-t+2q}$

We know that Y is stationary and therefore $\text{Var}[\,Y_j\,] < +\infty, \forall j$. This means:

$$\text{Var}[\,Y_j\,] = \mathbb{E}[Y_j^2] - \mathbb{E}^2[Y_j] < +\infty \implies \mathbb{E}[Y_j^2] < +\infty, \forall j$$

Thus, we get:

$$\text{Var}[\,\nabla^q\, Y_t\,] = \mathbb{E}[(\,\nabla^q\, Y_t\,)^2] = \sum_{j=t-2q}^{t} c_j\, \mathbb{E}[Y_j^2]$$

Because it's a finite sum, all coefficients c_j are finite and $\mathbb{E}[Y_j^2] < +\infty, \forall j$, then $\text{Var}[\,\nabla^q\, Y_t\,] < +\infty$

d)

$$\gamma_{\nabla^q Y}(r,s) := \text{Cov}(\nabla^q\, Y_r, \nabla^q\, Y_s) = \mathbb{E}[\,(\nabla^q\, Y_r - \underbrace{\mathbb{E}[\,\nabla^q\, Y_r\,]}_{=0}\,)\,(\,\nabla^q\, Y_s - \underbrace{\mathbb{E}[\,\nabla^q\, Y_s\,]}_{=0}\,)\,] = \mathbb{E}[\,\nabla^q\, Y_r\, \nabla^q\, Y_s\,]$$

It suffices to show that $\mathbb{E}[\,\nabla^q\, Y_r\, \nabla^q\, Y_s\,] = \mathbb{E}[\,\nabla^q\, Y_{r+h}\, \nabla^q\, Y_{s+h}\,], \forall r, s, h$.

Note that since Y is stationary, $\text{Cov}(Y_s, Y_r) = \text{Cov}(Y_{s+h}, Y_{r+h}), \forall r, s, h$ and because $\mathbb{E}[Y_t] = 0$ this implies:

$$\text{Cov}(Y_s, Y_r) = \text{Cov}(Y_{s+h}, Y_{r+h}) \Leftrightarrow \mathbb{E}[\,Y_r\, Y_s\,] = \mathbb{E}[\,Y_{r+h}\, Y_{s+h}\,] \longrightarrow (*)$$

Using the same thought process of b) and c), we can write

$$\nabla^q\, Y_r\, \nabla^q\, Y_s = \sum_{i=r-q}^{r} b_i\, Y_i \sum_{j=s-q}^{s} c_j\, Y_j = \sum_i \sum_j b_i\, c_j\, Y_i\, Y_j$$

Therefore:

$$\mathbb{E}[\,\nabla^q\, Y_r\, \nabla^q\, Y_s\,] = \mathbb{E}\left[\sum_i \sum_j b_i\, c_j\, Y_i\, Y_j\right] = \sum_i \sum_j b_i\, c_j\, \mathbb{E}[Y_i\, Y_j] \overbrace{=}^{\text{by}(*)} \sum_i \sum_j b_i\, c_j\, \mathbb{E}[Y_{i+h}\, Y_{j+h}]$$

We can 'shift' the sum limits by h without changing the rearrangement of the coefficients! That is, we assume b_i '=' b_{i+h} and c_j '=' c_{j+h}. From this change we obtain that the last member of the equation above is equal to:

$$\sum_{i=r+h-q}^{r+h} \sum_{j=s+h-q}^{s+h} b_i\, c_j\, \mathbb{E}[Y_i\, Y_j] = \mathbb{E}\left[\sum_{i=r+h-q}^{r+h} \sum_{j=s+h-q}^{s+h} b_i\, c_j\, Y_i\, Y_j\right] = \mathbb{E}[\,\nabla^q\, Y_{r+h}\, \nabla^q\, Y_{s+h}\,]$$

e)

1) $\mathbb{E}[\nabla^q X_t] = \mathbb{E}[\nabla^q m_t] + \underbrace{\mathbb{E}[\nabla^q Y_t]}_{=0} = q!\, a_q + 0 = q!\, a_q \longrightarrow$ Mean of $\nabla^q X_t$ is $q!\, a_q, \forall t$

2) $\text{Var}[\nabla^q X_t] = \text{Var}[\,\underbrace{\nabla^q m_t}_{\text{constant}} + \nabla^q Y_t] = \text{Var}[\nabla^q Y_t] < +\infty, \forall t$, by question c).

3) $\text{Cov}(\nabla^q X_r, \nabla^q X_s) = \mathbb{E}[\,(\nabla^q X_r - \mathbb{E}[\nabla^q X_r]\,)\,(\nabla^q X_s - \mathbb{E}[\nabla^q X_s]\,)\,]$

Since $\mathbb{E}[\nabla^q X_t]$ is constant for all t, it suffices to show that $\mathbb{E}[\nabla^q X_r\, \nabla^q X_s] = \mathbb{E}[\nabla^q X_{r+h}\, \nabla^q X_{s+h}], \forall r, s, h$

Furthermore, because $\nabla^q X_t = \nabla^q m_t + \nabla^q Y_t$ and $\nabla^q m_t$ is constant for all t, we can further simplify this condition to only proving that

$$\mathbb{E}[\nabla^q Y_r\, \nabla^q Y_s] = \mathbb{E}[\nabla^q Y_{r+h}\, \nabla^q Y_{s+h}], \forall r, s, h$$

But this condition was already verified in question d), so the proof is done.

By 1), 2) and 3), we have proved that $\nabla^q X$ is stationary with mean $q!\, a_q$

Question 2

After loading the data set, a simple calculation shows that it contains $N = 302$ observations and mean of this time series is $\hat{\mu} = 5.444073$.

We start by presenting a plot of mean corrected data:

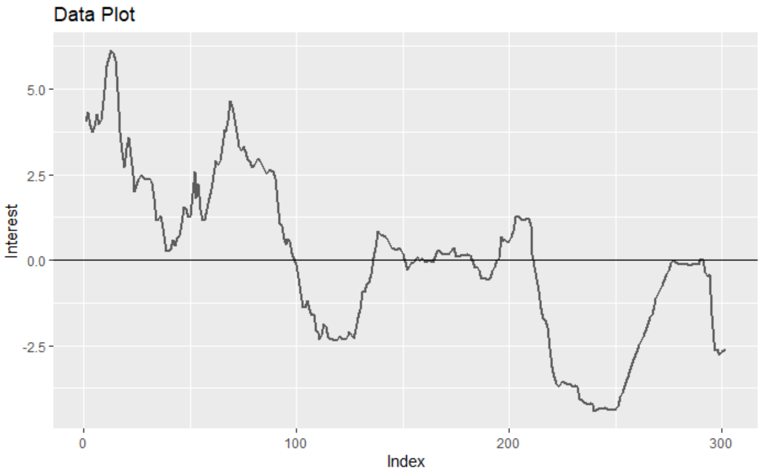
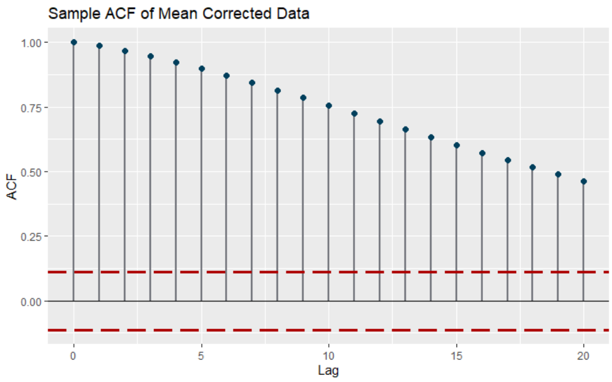
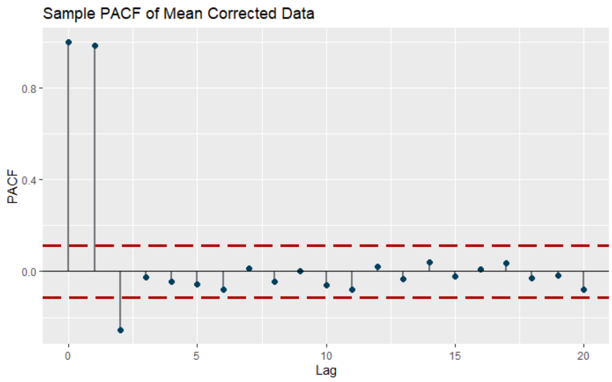


Figure 1: Mean corrected data

Afterwards, we can plot the sample auto correlation function (ACF) and the sample partial auto correlation function (PACF), shown in the plots below.



(a) Sample ACF



(b) Sample PACF

Figure 2: ACF and PACF for the mean corrected data

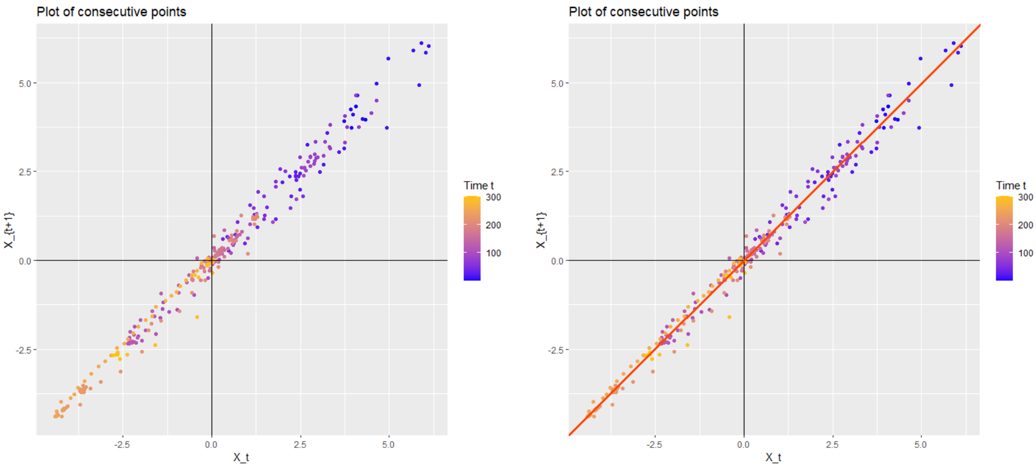
Note that both graphs start for a lag of $h = 0$. The values are calculated with manually programmed functions, and compared with the automatic `Acf` and `Pacf` functions, obtaining the same values. The red dashed lines in both graphs represent the bounds $\pm 1.96/\sqrt{N}$, used to informally test whether a time series can be classified as i.i.d. noise.

The values of the sample ACF corroborate the theory that the times series is an autoregressive process of order 1, since they are slowly decreasing. However, the PACF show a non-negligible value for $h = 2$, whereas a AR(1) process should have a negligible PACF for any lag $h > 1$. This could indicate the possibility of having an AR(2) process instead, since the PACF values are negligible for $h > 2$.

Question 3

a) and b)

The graphs below contain the plots for the data set consisting of consecutive points. The second includes (red line) the linear fit without intercept of both variables.



(a) Without linear fit (b) With linear fit

Figure 3: Plots of consecutive points

The data points seem to indeed fall into a straight line and it's fairly coincident with the linear fit shown in the graph 3(b), which could suggest an appropriate linear fit. However, with the color coded points, we see that the points "shift directions" through time. This means that a simple linear regression model isn't accurate to predict an increasing or decreasing tendency of the data. For a fitted model of the form

$$S_t = \varphi S_{t-1}$$

The linear regression gives us the value $\hat{\varphi} = 0.9962448$, which is equivalent to a nearly constant time series (apart from the noise). As for the residual standard error, we obtain the value $\hat{\sigma} = 0.2722802$. With those values we can simulate some random paths using that recurrence and simulating noise from a $\mathcal{N}(0, \sigma^2)$ distribution. The graph below displays an example with 3 random paths.

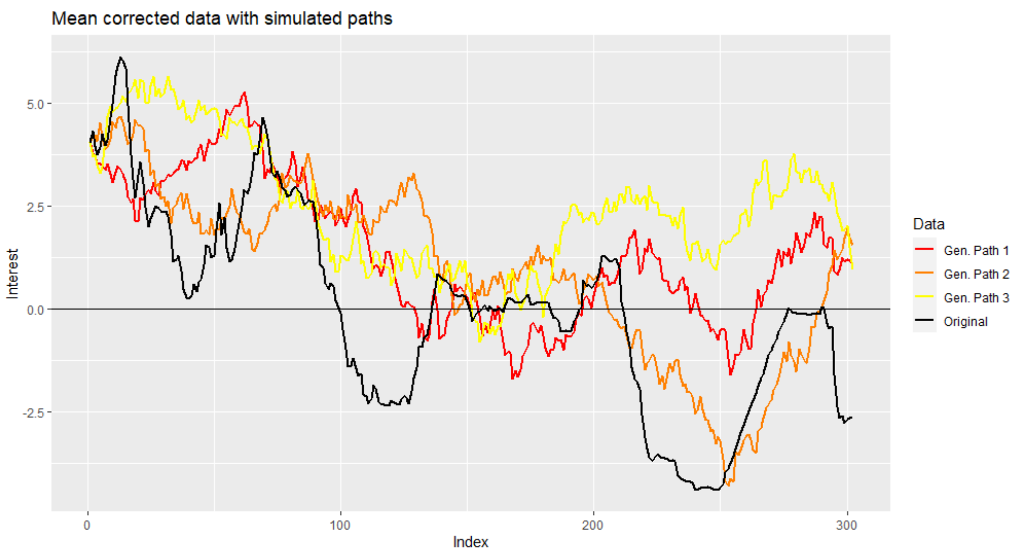


Figure 4: Simulated paths

c)

The graph below contains a scatter plot containing the residuals.

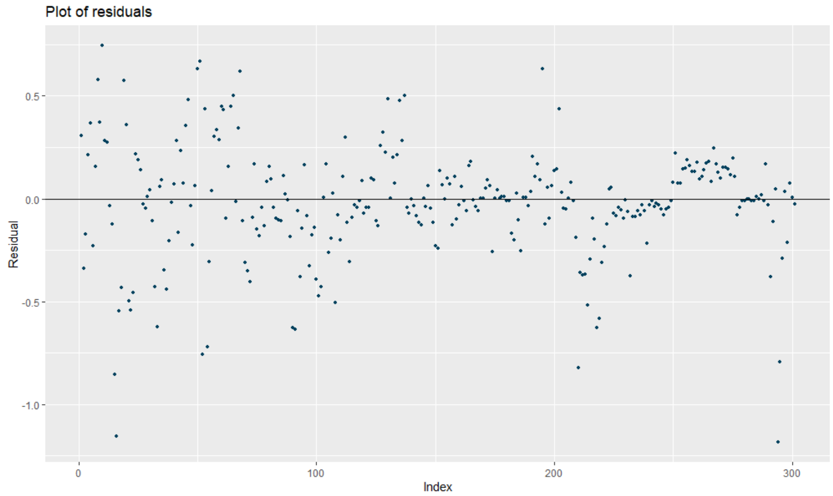


Figure 5: Plot of residuals

If the residuals were iid noise, it would be expected that they were evenly scattered along the y-axis. However, it appears that there's a "thinning" of the scatter, that is, later residuals seem to be more concentrated around 0, while earlier residuals appear to be more disperse, which could already indicate a lack of fit for an iid distribution.

To test whether or not the residuals behave like iid noise, we calculate the ACF values for every lag up to $N/4$ and compute a Ljung-Box test for the same lags. The Ljung-Box test results were both manually and automatically (using the `Box.test` function) and the results are shown below for some selected lags.

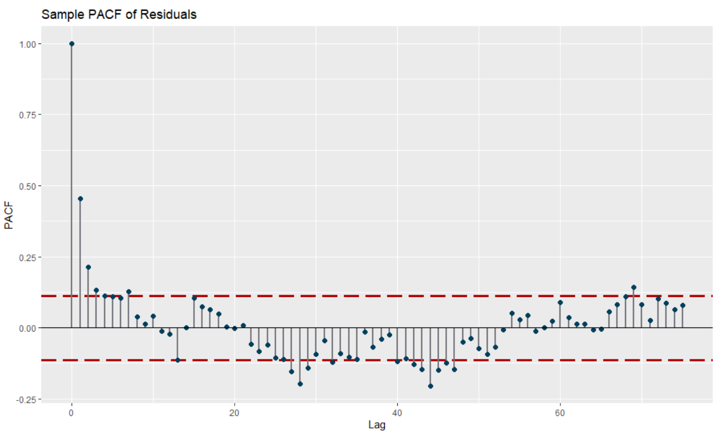


Figure 6: ACF of residuals

Lag	Test stat	Manual p-val	Auto p-val
1	62.546	2.55e-15	2.55e-15
2	62.580	2.57e-14	0
3	63.050	1.31e-13	0
6	64.683	5.01e-12	0
12	69.927	3.305e-10	8.88e-16

Table 1: Ljung-Box test results

It should be mentioned that results for the p-value obtained were all equal or incredibly close to zero, contributing for a very strong evidence that supports the alternative hypothesis of the Ljung-Box test, meaning the residuals do not behave like iid noise.

This is further proved by the ACF plot. If so, the values would have approximately a $\mathcal{N}(0, N^{-1})$ distribution, which would imply that, for a 5% significancy test, 95% of the ACF values should lie within the bounds marked by the red dashed lines, which is not the case. Most prominently, we notice that $\rho(1)$ is substantially high, indicating that neighbouring values are strongly correlated.

Question 4

First the data set is divided into two parts, the training set and the test test. Then we will use the 20 previous months to predict the next month. Here the 20 last points of the training set are between the vertical lines:

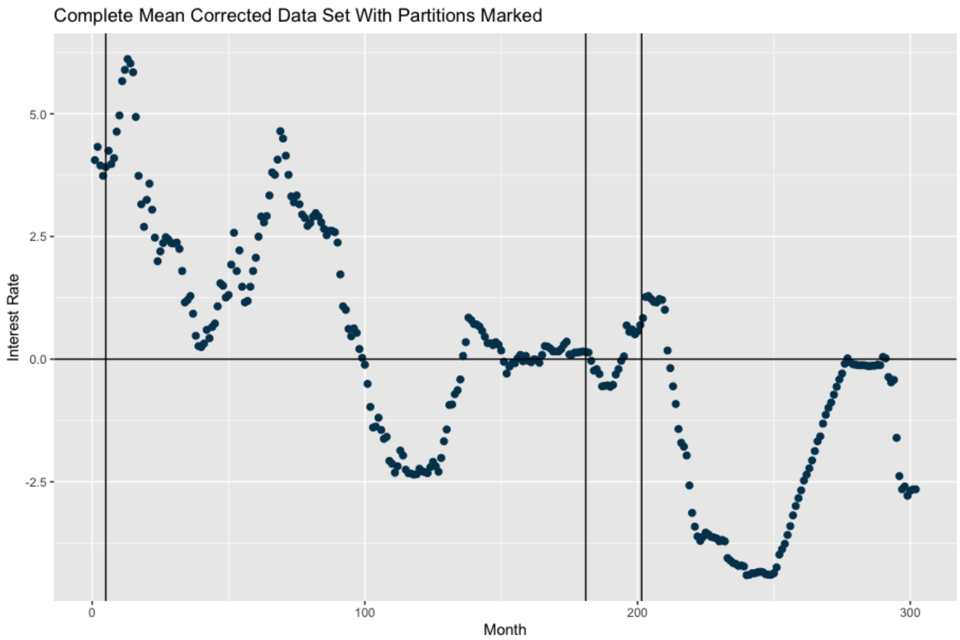


Figure 7: Mean corrected data with partitions

These data points will be used to predict the first data point of the test set. For the second data point in the test set the 19 last data points of the training set and the first of the test set will be used. This is repeated for all the data points in the test set.

To do this we solve the equations from Proposition 2.4.5 using the sample ACVF, with lag $h = 20$, to get the coefficients of the best linear predictor. This is what the predictions look like compared to the test data:

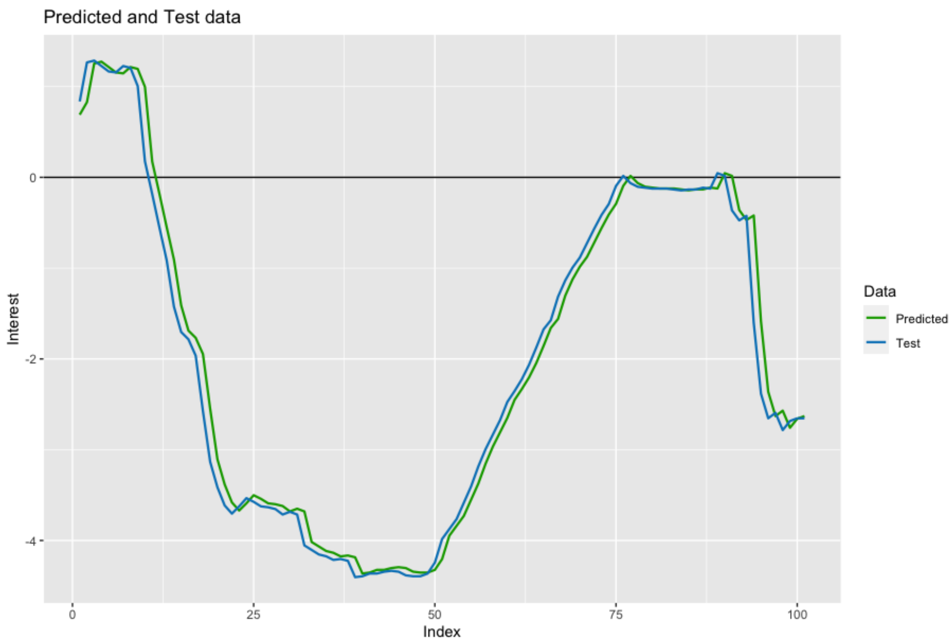


Figure 8: Prediction

This results in a mean square error for the predicted data of $MSE_{np} = 0.045$. For comparison, using the mean, which is zero for the mean corrected data, as a future prediction gives a mean square error of $MSE_{\mu} = 7.33$. In a mean square sense the prediction is significantly better than just guessing the mean.

Question 5

To begin we need the following result from Question 3: $\hat{\phi} = 0.996244$. Then we use the following to make our one step predictions:

$$X_{t+1} = \hat{\phi} X_t \tag{1}$$

Repeating this process for all the points in the test set makes our parametric prediction. The plot is the result combined with the previous predictions from Question 4, the non-parametric prediction, and the test data.

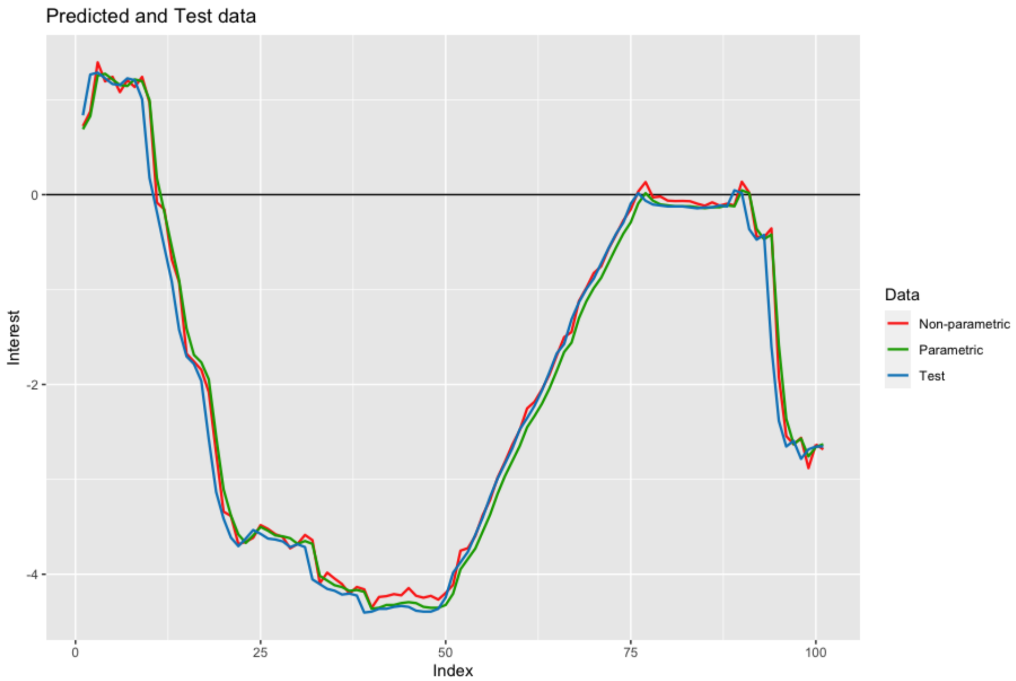


Figure 9: Predictions

Yet again we want to compare the mean square errors of the predictions. For the parametric prediction we have $MSE_{pp} = 0.058$ and as established in Question 4 $MSE_{np} = 0.045$ and $MSE_{\mu} = 7.33$.

The results show that the parametric prediction is worse than the non-parametric one. As stated in Question 3 the simple linear regression model will not be effective for this data. This indicates that for similar time series making the prediction using the parametric assumption is pointless.

To try and improve the prediction further one might consider a seasonal component. Just by inspection the data appears to have peaks and valleys at roughly equal distance in conjunction with a negative trend. This however might be considered over-fitting.