UNIVERSITY OF LEEDS

You must sign this (digitally signing with your name is acceptable) and include it with each piece of work you submit.

I am aware that the University defines plagiarism as presenting someone else's work, in whole or in part, as your own. Work means any intellectual output, and typically includes text, data, images, sound or performance.

I promise that in the attached submission I have not presented anyone else's work, in whole or in part, as my own and I have not colluded with others in the preparation of this work. Where I have taken advantage of the work of others, I have given full acknowledgement. I have not resubmitted my own work or part thereof without specific written permission to do so from the University staff concerned when any of this work has been or is being submitted for marks or credits even if in a different module or for a different qualification or completed prior to entry to the University. I have read and understood the University's published rules on plagiarism and also any more detailed rules specified at School or module level. I know that if I commit plagiarism I can be expelled from the University and that it is my responsibility to be aware of the University's regulations on plagiarism and their importance.

I re-confirm my consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the UK) to monitor breaches of regulations, to verify whether my work contains plagiarised material, and for quality assurance purposes.

I confirm that I have declared all mitigating circumstances that may be relevant to the assessment of this piece of work and that I wish to have taken into account. I am aware of the University's policy on mitigation and the School's procedures for the submission of statements and evidence of mitigation. I am aware of the penalties imposed for the late submission of coursework.

Student Signature		Date	24/11/23
Student Name	Omkar Pawar	Student Number	201694894

1. Introduction:

The data we will use in this report is the data we obtained from the slavery travel data. The comprehensive compilation of slave voyages known as the Trans-Atlantic Slave Trade Database reveals key details about over 35,000 voyages involving more than 10 million enslaved Africans from the 16th to 19th centuries[1]. Ship information, voyage details, slave demographics, crew details and mortality rates during the transatlantic voyage are carefully documented[1]. When analyzing the data, it was found that there was an occasional increase in the number of slaves embarked under the British flag during this period.There was no particular time of year when there was a decrease in the estimated number of slaves traveling on these ships, but we did see that there were occasional decreases and increases in the number of estimated slaves. When reviewing the periodograms and analyzing the data, it was concluded that the order one autoregressive model was used, which meant that the number of slaves depended on the number of the previous year. After verifying that the chosen model was correct, the model was fitted to the data after correctly removing the linear trend in the data. The model predicted a decline in the slave trade trend from 1808 to 1812, with the trade ending in Britain in 1807. However, the model after the predictions showed a decreasing trend in values, despite not considering the end of the slave trade in Great Britain.

2. Data:

After loading the data into a variable called xf, exploration of the data was done. The data consisted of two data frame columns; one consisted of years from 1654 to 1807, and the other column was the number of estimated slaves in that year. The following plot shows the number of slaves that were embarked in the voyages in accordance to that year.

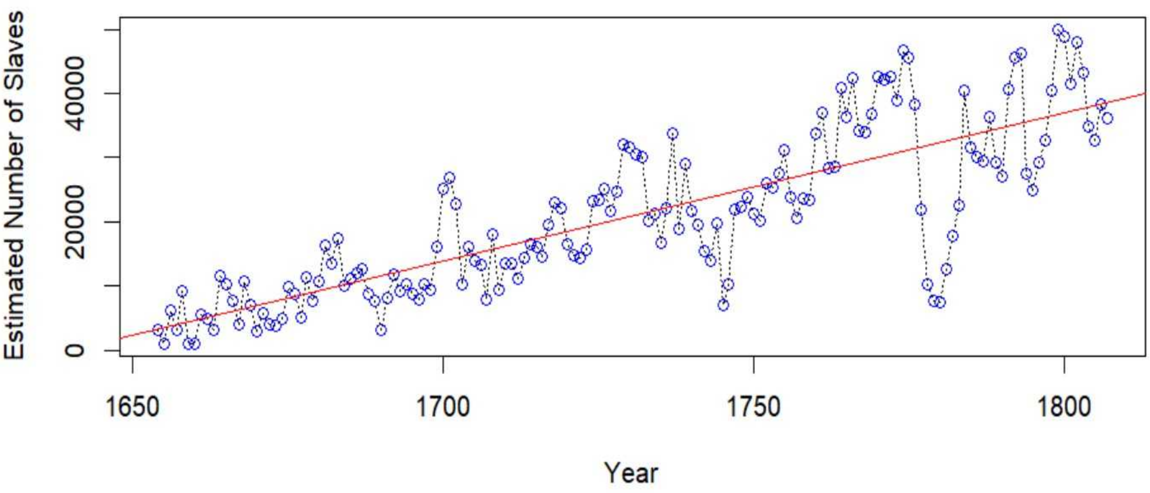

Figure 2.1 Estimated number of Slaves

On reading the graph we can notice that there is no visible seasonality in the plot but there is a trend in the data that increases across the years. We can conclude that there is a linear trend in the data that needs to be removed in order to make the series a bit stationary. Since the data shows a linear trend, we can remove it by fitting it through a linear model. So after the fit we obtain the residuals as follows.

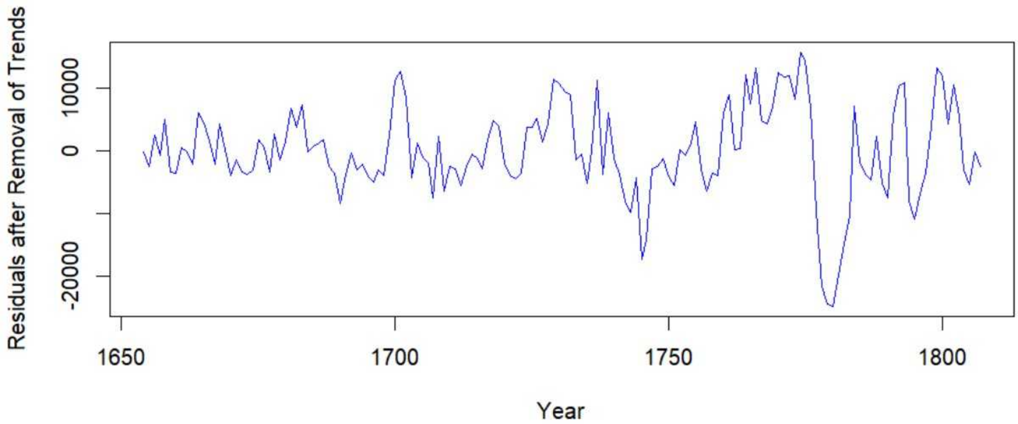

Figure 2.2 Residuals After Removing the Trend

After the removal of the trend we can see that there are no specific patterns. Thus, no check for seasonality was done. These residuals were saved into a variable called Y.

3. **Processes:**

To check which model to go with we need to check the autocorrelation plots (ACF) and partial autocorrelation plots (PACF) for the residual values we got after removal of trends. The plots are as follows:

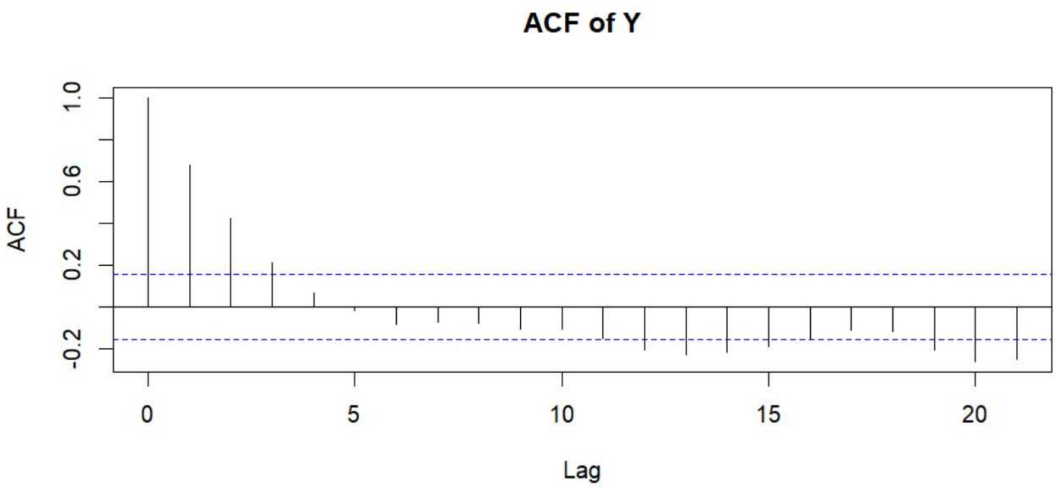

Figure 3.1 ACF plot for residuals Y

The ACF plot shows a gradual decay in the lags. The lags cut off after the third lag. It means that the current y value is correlated to its past values and the correlation decreases as the lag period increases. In the plot the lag does not cut off after lag 2, thus we can say that a MA(Moving Averages) model is not applicable here.

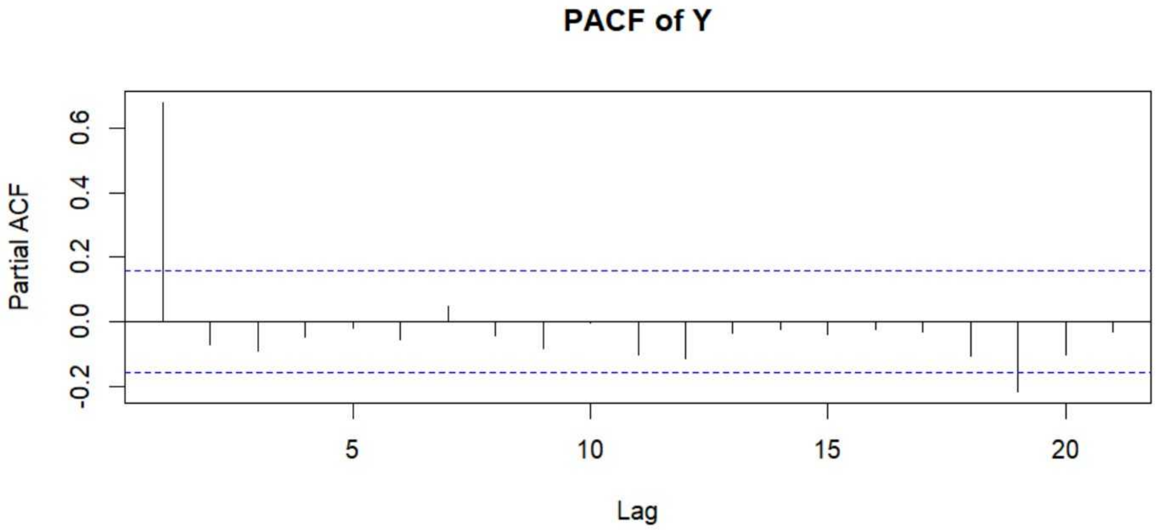

Figure 3.2 PACF for residuals of Y

The current value of y is linearly dependent on its value one lag period ago, after accounting for the linear dependence of y on its more distant lags. The PACF suggests that the series y has an AR(1)(Autoregressive) component.The fact that the PACF decays after lag 1 suggests that the series y does not have any higher-order AR components. Hence, we can assume that AR(1) fits the data best.

To cross verify the value of p, we will consider the values for p as 1,2,3 for the AR(p) model. Here are the outputs after calling the models in R:

Call: AR(1)
```
ar(x = Y, aic = FALSE, order.max = 1, method = "yule-walker")

Coefficients:
     1
0.6777

Order selected 1  sigma^2 estimated as  28766227 . For p as 1
```

Call: AR(2)
```
ar(x = Y, aic = FALSE, order.max = 2, method = "yule-walker")

Coefficients:
      1        2
 0.7241  -0.0685

Order selected 2  sigma^2 estimated as  28820780. for p as 2
And
```

Call: AR(3)

```
ar(x = Y, aic = FALSE, order.max = 3, method = "yule-walker")

Coefficients:
      1        2        3
 0.7180  -0.0045  -0.0884

Order selected 3  sigma^2 estimated as  28786148. for p as 3
```

The coefficients in an autoregressive (AR) model indicate the strength and direction of the relationship between the current observation and its past values.

AR(1) Model:

- Coefficient: 0.6777
- Interpretation: This suggests that each observation is positively correlated with its previous observation. Specifically, for every unit increase in the previous observation, the current observation is expected to increase by 0.6777 units.

AR(2) Model:

- Coefficients: 0.7241, -0.0685
- Interpretation: The first coefficient (0.7241) represents the influence of the immediate past observation, similar to the AR(1) model. The second coefficient (-0.0685) suggests a negative influence from the observation two time periods ago. So, the current observation is influenced positively by the recent past and negatively by the observation two time periods ago.

AR(3) Model:

- Coefficients: 0.7180, -0.0045, -0.0884
- Interpretation: Similar to the AR(2) model, the first coefficient (0.7180) represents the influence of the immediate past. The second coefficient (-0.0045) suggests a negligible negative influence from the observation two time periods ago. The third coefficient (-0.0884) introduces a negative influence from the observation three time periods ago.

In the AR(1) model, you have a positive coefficient for lag one. In the AR(2) and AR(3) models, you have negative coefficients for lags beyond one. In this case, an AR(1) model seems to capture the primary autocorrelation structure of the time series effectively. Therefore, the AR(1) model is likely sufficient to explain the temporal dependencies in your time series data, as it captures the essential autocorrelation pattern observed in the ACF and PACF plots.

4. **Plots:**

Now we will plot ACFs for all the residuals of the AR(p) models with p ranging from 1 to 3. Also the plots for the squared residuals for these models. Following is the correlogram for the stated:

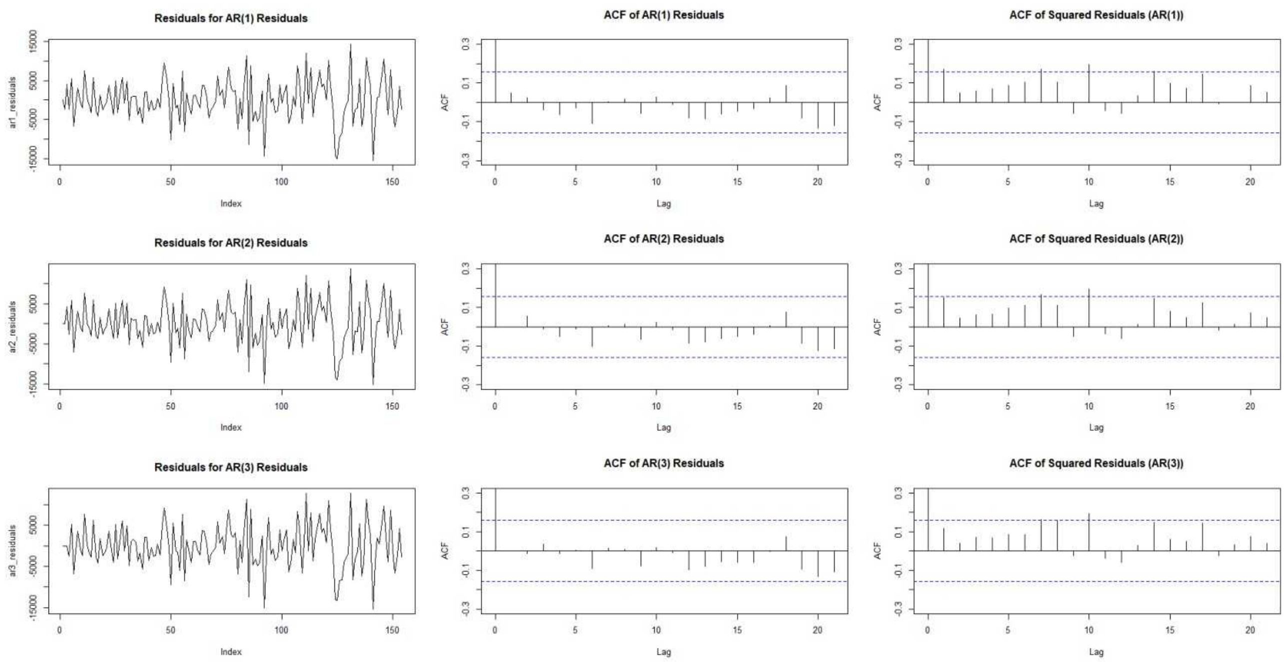

Figure 4.1 Correlogram for AR(1), AR(2), AR(3)

Looking at the residuals ACF plot for AR(1) , AR(2) and AR(3) we can see that no lags after lag 1 are crossing the significance threshold (the blue line), which means that there is no significant correlation and the lags are basically a white noise. The ACFs for the residuals do not tell us any significant difference among the three. Now checking the ACFs for the squared residuals. We can see that the AR(1) shows a significant correlation at lag 1 and cuts to zero. But for it we can see that there are some possible autocorrelations at previous lags which can affect the model. For AR(2) and AR(3) we can see that their correlation decays to zero just after the first lag, but there are still some significant correlations in the residuals. There should be less autocorrelations in the ACF plots, that is, there must be few lags that cross the significance threshold(the blue line). The ACF plot for AR(3) shows that there are fewer lags that are above the blue line than the AR(2) plot and hence we can say that AR(1) models capture the autocorrelation of the model effectively. However , we must be careful that it can overfit the model.

The following are the periodograms for the variables X (original number of estimated slaves) , Y(residuals after removal of trend) and Z(residuals of AR(1))

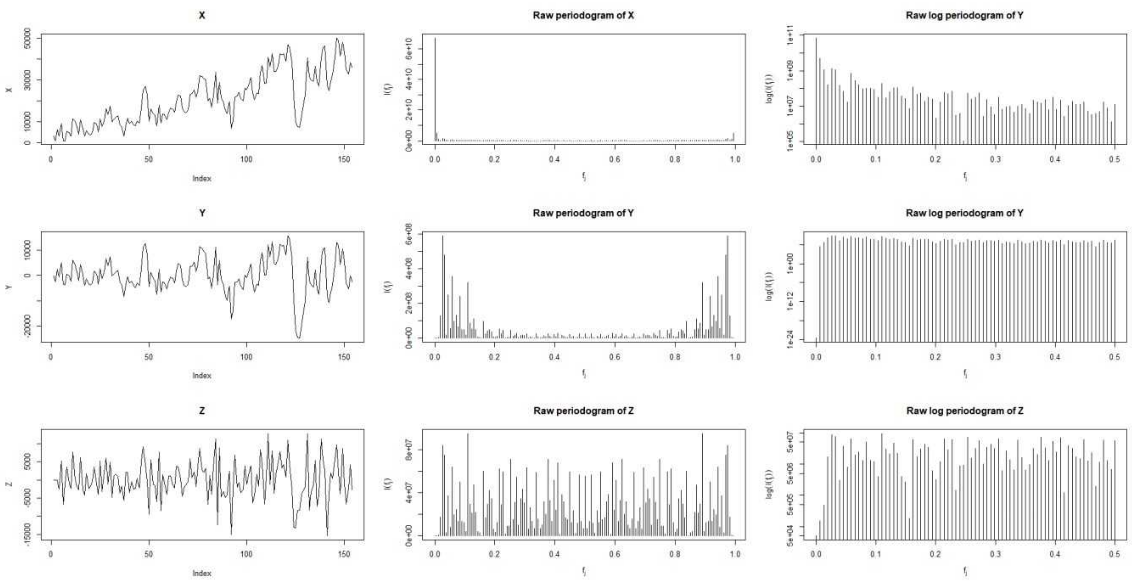

Figure 4.2 Raw periodograms and raw log periodograms for X,Y and Z

In the above diagram, X shows a clear upward trend. This suggests that there is a non-stationary component to the series. Y shows a more stationary pattern than the original series. However, there is still some autocorrelation and volatility in the residuals. This suggests that the linear model may not be adequately capturing all of the dynamics in the data. The residuals of the AR(1) model; Z show a more white noise-like pattern than the residuals after removal of trend. This suggests that the AR(1) model is capturing most of the dynamics in the data, but we can still move to models with higher order as it will help capture more complex autocorrelation patterns. However it it preferred to keep the model less complex as to reduce overfitting.

5. Fitting the model:

Before we continue, we must keep in mind that the values for log likelihood must be higher and AIC(akaike information criterion) must be low for the model that fits the data the best.

For arima(1,0,0)

```
arima(x = arr_fir, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.9471  20849.082
s.e.  0.0263   5369.395
sigma^2 estimated as 15308007:  log likelihood = -1493.53,  aic = 2991.06
```

For arima(2,0,0)

```
arima(x = ar_fit2, order = c(2, 0, 0))

Coefficients:
         ar1     ar2  intercept
      0.7994  0.1494  20844.457
```

```
s.e.   0.0796  0.0806   5817.808
sigma^2 estimated as 17489329:  log likelihood = -1503.75,  aic = 3013.5
```

For arima(3,0,0)

```
arima(x = ar_fit3, order = c(3, 0, 0))
Coefficients:
         ar1      ar2     ar3  intercept
      0.8899  -0.1054  0.1682  20698.213
s.e.  0.0795   0.1075  0.0803   6146.358
sigma^2 estimated as 17236140:  log likelihood = -1502.65,  aic = 3013.31
```

The AIC values the lowest is for ARIMA(1,0,0) and the log likelihood highest is for the same. Hence we can say that the ARIMA(1,0,0) model is the best fit among them. Also ARIMA(1,0,0) generally has smaller standard errors for its coefficients compared to the other models. Smaller standard errors often indicate more precise parameter estimates.Here we fitted the estimated no of slaves by removing the AR(1) residuals into the ARIMA(1,0,0) model.

6. Predictions:

Since we finalized our model as ARIMA(1,0,0) we can now predict the model to forecast future values on the given series. As we can see that after the removal of residuals and fitting the data the plot gives us a rough idea as to how the fitted values are.

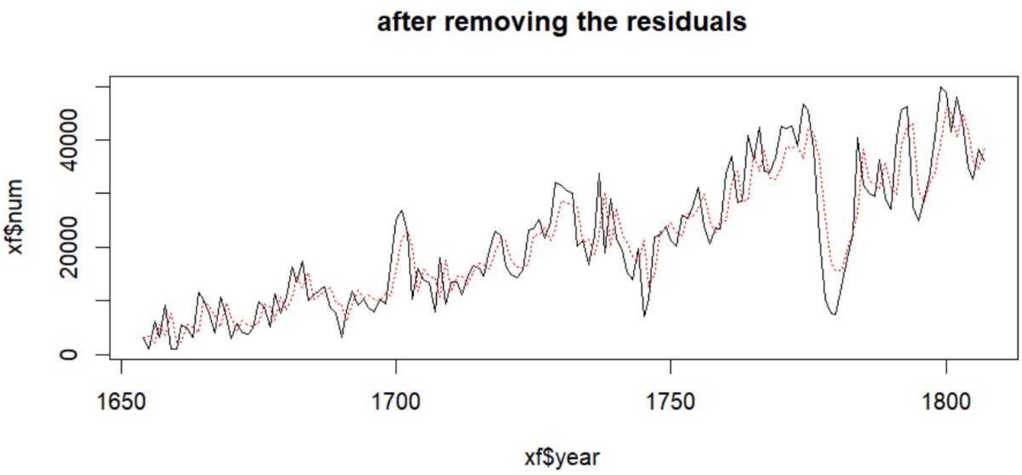

Figure 6.1 Plot after removal of residuals

We apply simple code called predict on the model and it gives the output as:

```
Time Series:
Start = 155
End = 159
Frequency = 1
[1] 37552.29 36668.89 35832.22 35039.79 34289.28

$se
Time Series:
```

```
Start = 155
End = 159
Frequency = 1
[1] 3912.545 5388.842 6430.954 7239.219 7894.149
```

Here we have predicted for the additional 5 years. Here we can see that there is a decrease in trend of the predicted values and an increase in the standard errors. This is due to the uncertainty that increases as we predict more in the future.

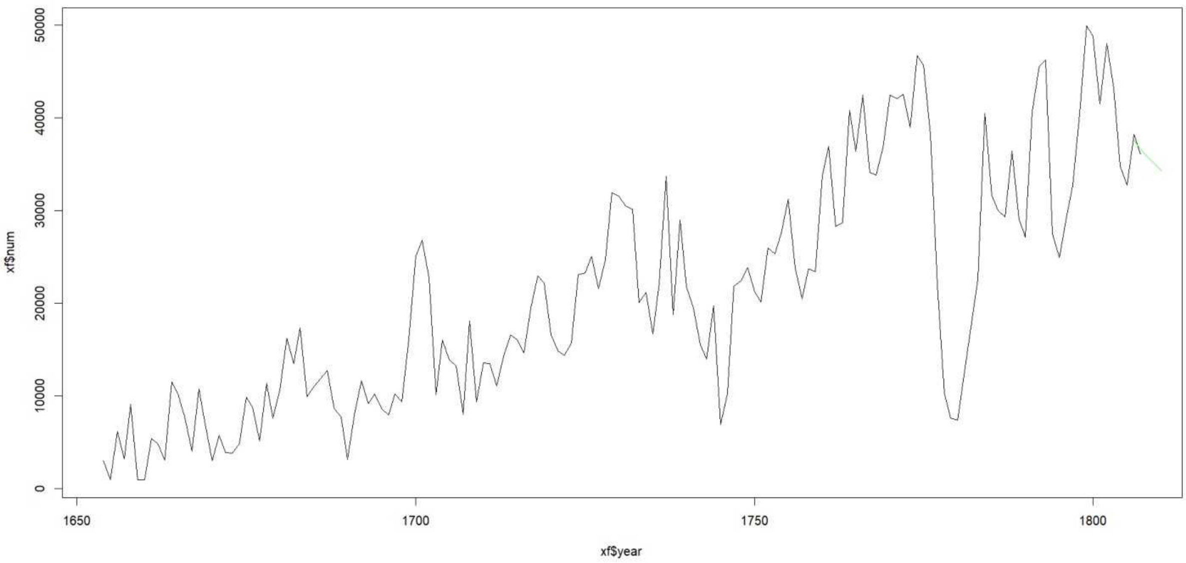

Figure 6.2 Plot for predicted values (In Green)

7. Conclusion

Therefore, we can conclude by saying this that our model showed characteristics of an AR(1) process and thus we fitted the model by removing the trend and then the residuals. After that once we verified that the best model is the AR(1) we predicted in the consecutive years from 1807 to 1812. In the end we say a decreasing trend in the values for the predicted values.

8. References:

1. https://www.slavevoyages.org/