**BIA 656**

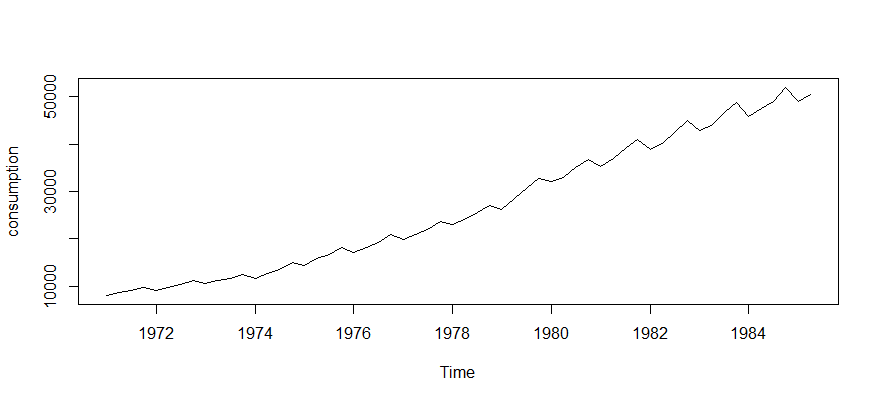
**Advance Data Analytics and Machine Learning**

Assignment : Lab 3

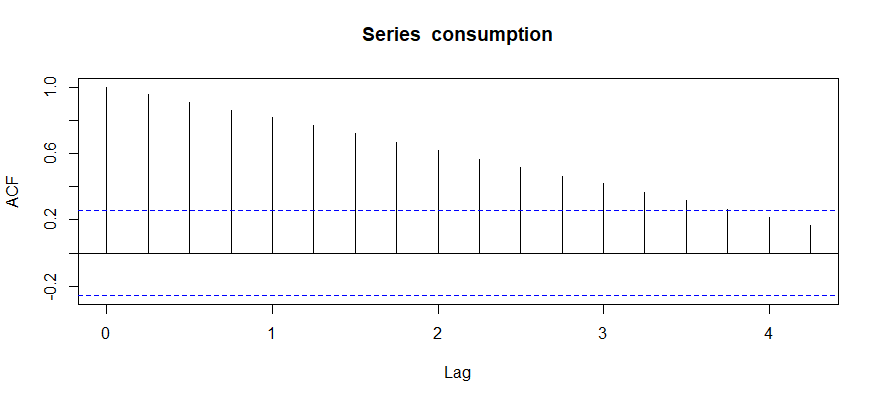
Shrey Kshatriya

**Q1. Describe the behaviour of consumption. What types of differencing, seasonal, nonseasonal, or both, would you recommend? Do you recommend fitting a seasonal ARIMA model to the data with or without a log transformation? Consider also using ACF plots to help answer these questions.**

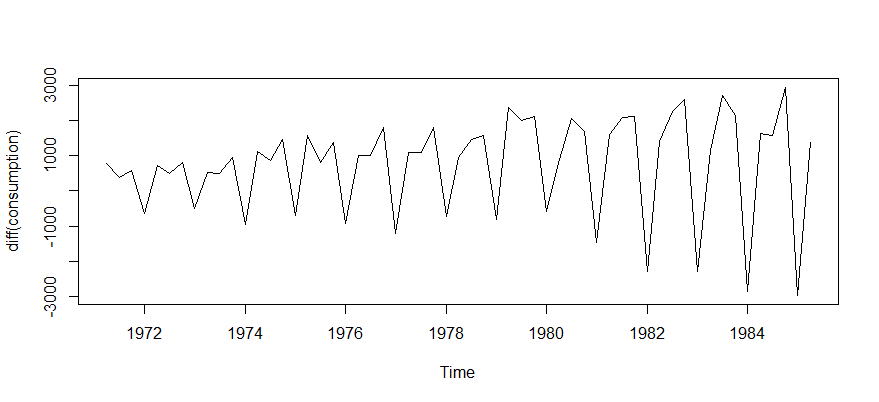
A1.



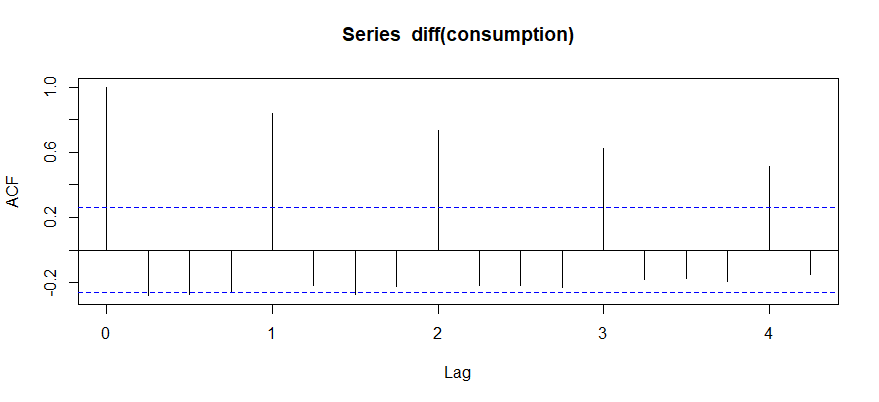
Here we can see that the plot has noticeable trends and changing levels over the period. Therefore, we can say that it is a non-stationary time series. However, to confirm it, we can plot the ACF graphs and confirm it.



The above ACF graphs shows us that since ACF is decreasing slowly, it confirms that the data is non-stationary.

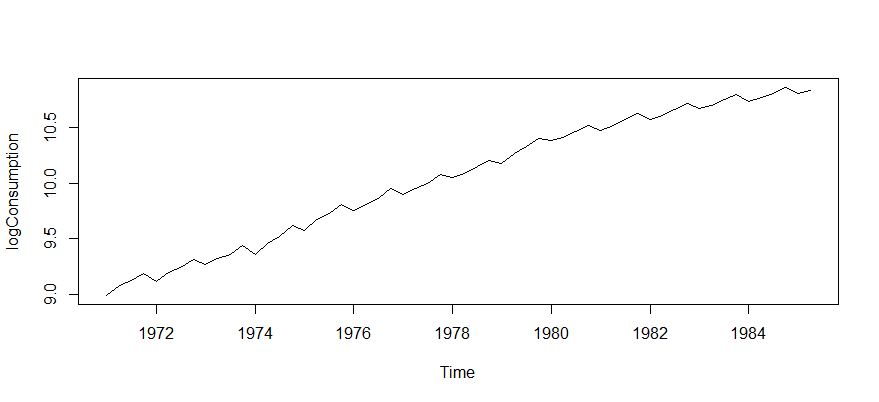


This time series seems to be stationary however, the trends can be seen later, and we can conclude that this is non-stationary and there might be some seasonality.



When we plot the ACF of the differentiation, we can see that that there is spike at regular intervals in the multiple of 4. So, we can conclude that there is quarterly seasonal effect in the time series. Thus, it can be said that we require both seasonal and non-seasonal differencing. We recommend both seasonal and non-seasonal differencing.

We perform log transform to convert a time series into stationary. However, even after performing log transform, we can see the results shown below are like results before log transform.



Plotting logconsumption



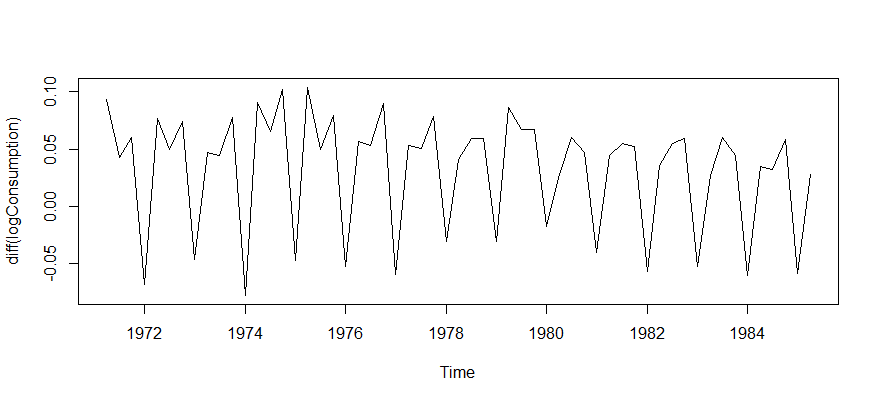
Plotting ACF of logconsumption

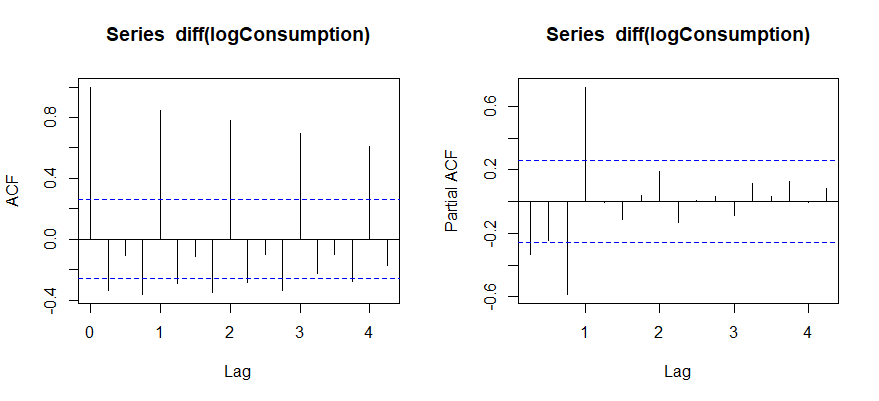
As seen above, the log transform did not make much difference. This is also seen when we plot the ACF, that the decreasing in ACF is slow, hence appears non-stationary.

It would be recommended to fit the ARIMA model without log transform.

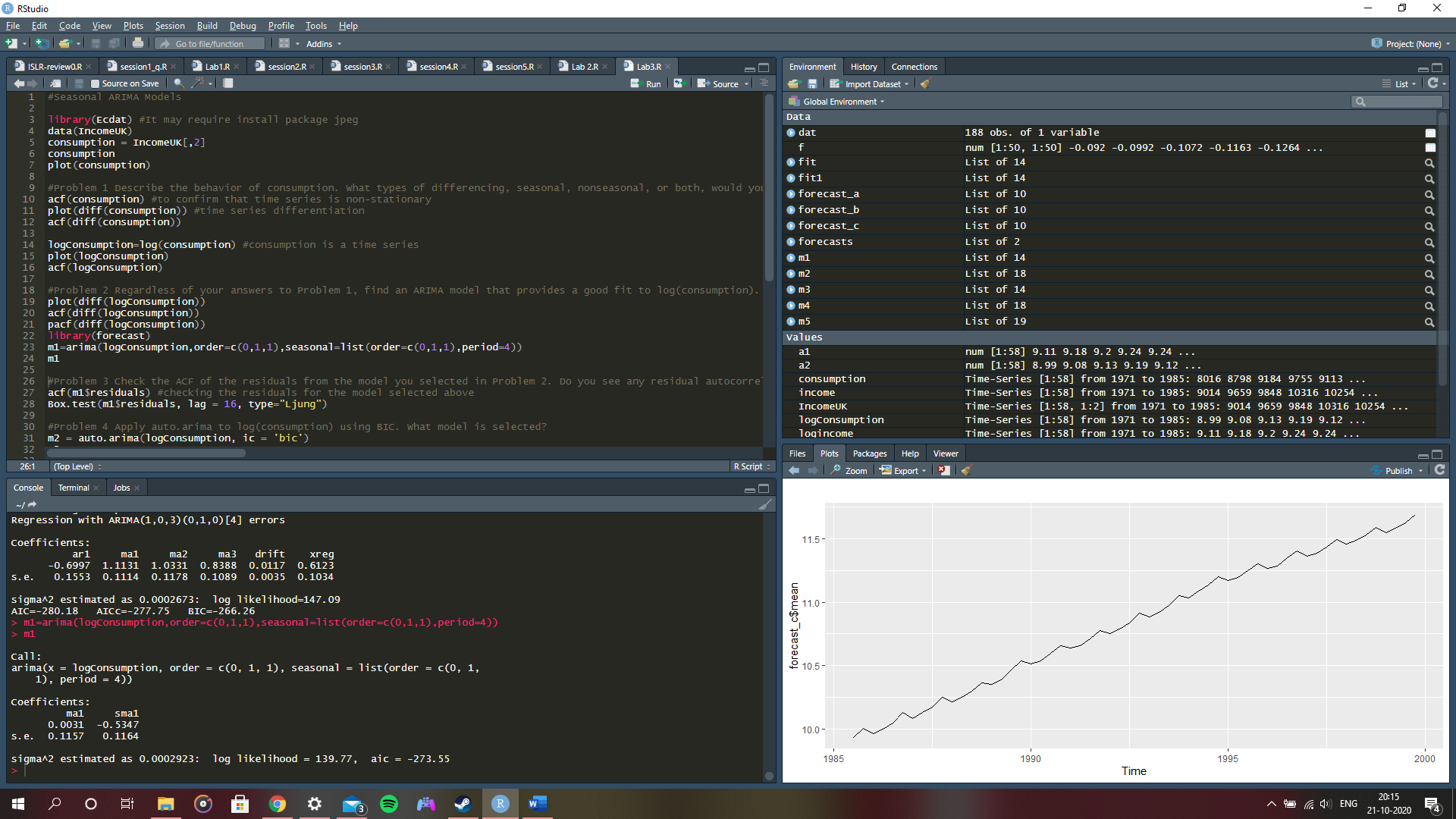
**Q2. Regardless of your answers to Problem 1, find an ARIMA model that provides a good fit to log(consumption).**

A2.





The above ACF plot shows us that for first difference, the order of 1 is enough to make the time series stationary. Also, the spikes at the multiple of 4 indicate that the seasonal differencing could be of order 4.

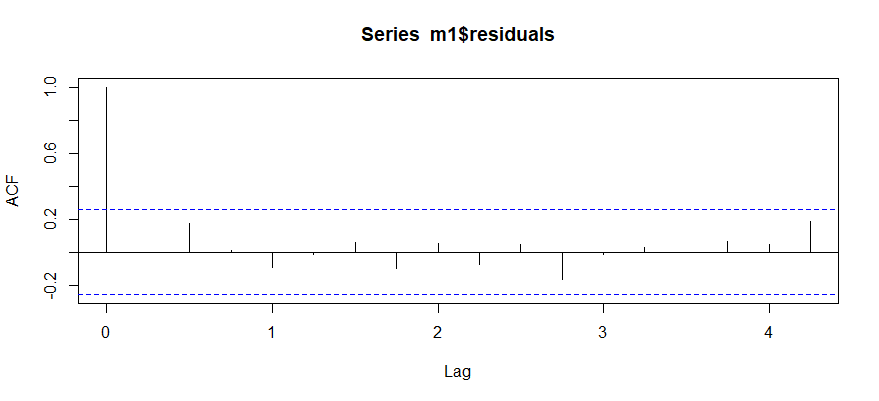


The fitted ARIMA model is of order (0,1,1)(0,1,1).

This means ARIMA order for non-seasonal components is p = 0, d = 1, q = 0

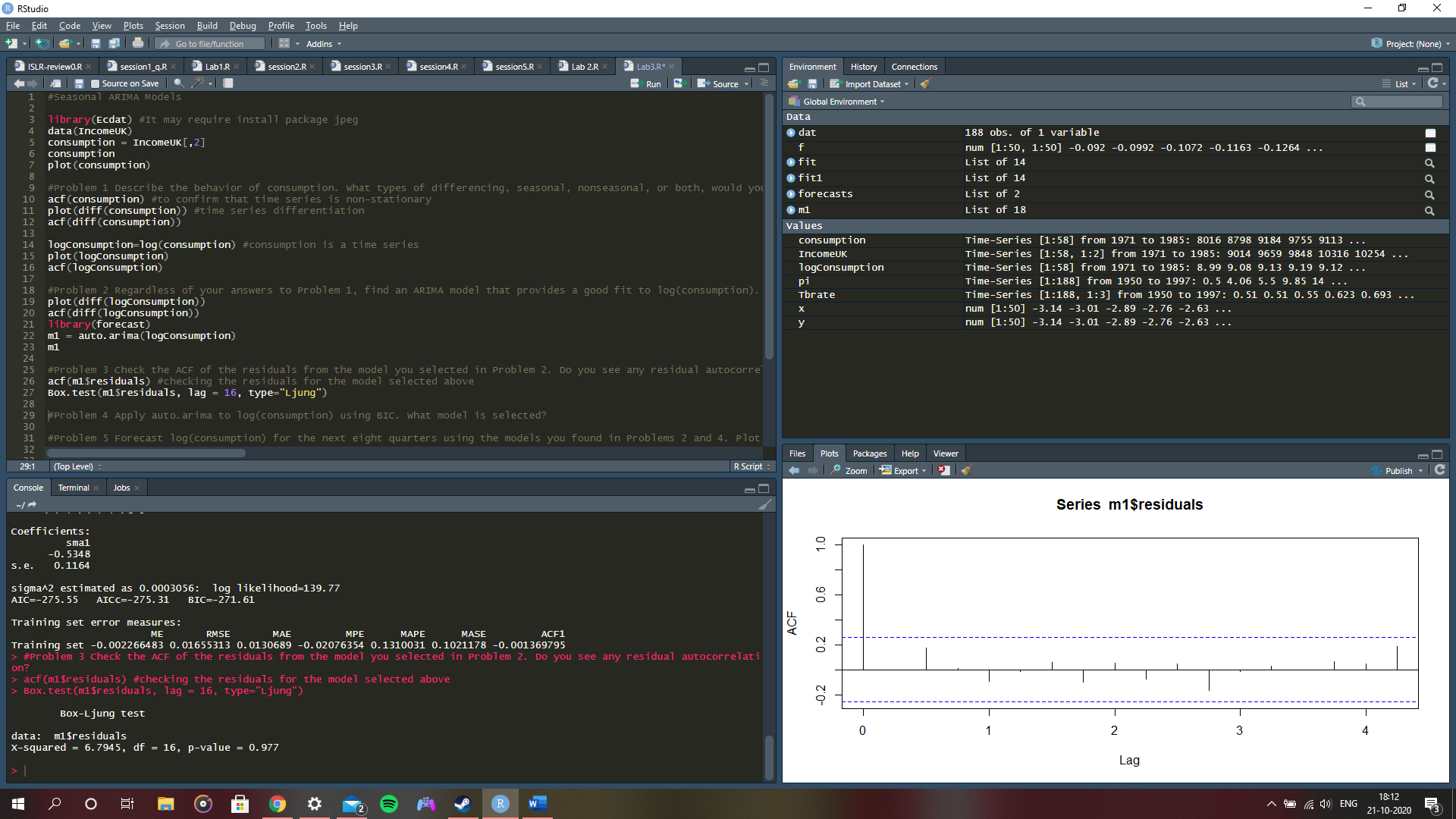
And ARIMA order for seasonal components is P = 0, D = 1, Q = 1 with a seasonal period of 4.

**Q3. Check the ACF of the residuals from the model you selected in Problem 2. Do you see any residual autocorrelation?**

A3. 

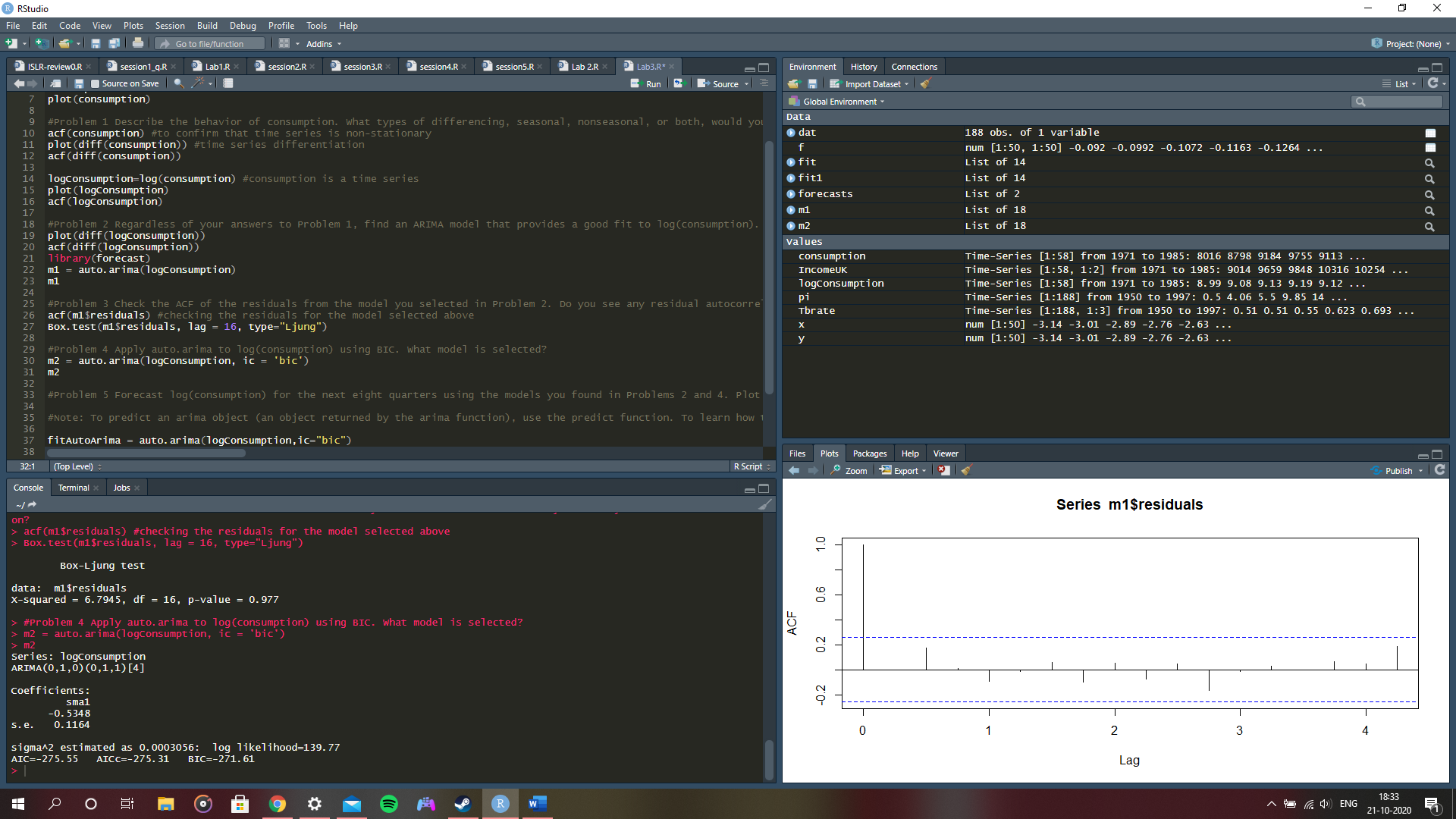
Above shown is the ACF of residuals for model selected in previous question.

To check any correlation, we perform Ljung Box test.



As seen above, the p-value is greater than 0.05 and we cannot reject the null hypothesis of no autocorrelation. Hence, by performing Ljung Box test, we realize that all correlations are within the significance level and there seems to be no residual correlation.

**Q4. Apply auto.arima to log(consumption) using BIC. What model is selected?**



Applying auto.arima to the logconsumption using BIC, we observe that the same model is selected.

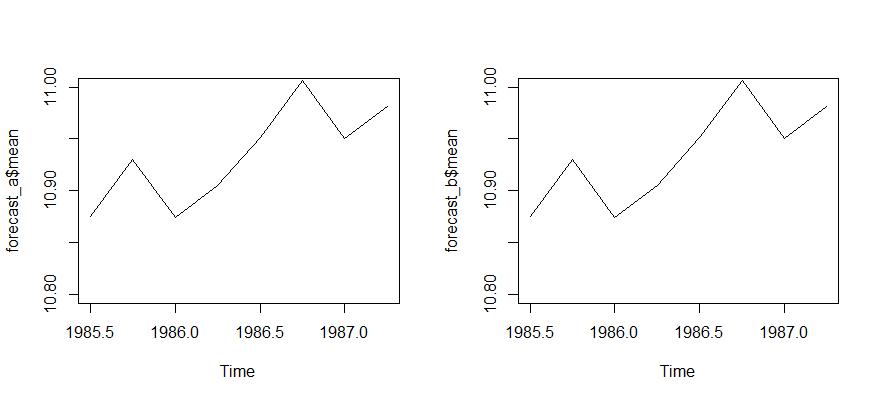
The fitted ARIMA model is of order (0,1,0)(0,1,1).

This means ARIMA order for non-seasonal components is p = 0, d = 1, q = 0

And ARIMA order for seasonal components is P = 0, D = 1, Q = 1 with a seasonal period of 4.

**Q5. Forecast log(consumption) for the next eight quarters using the models you found in Problems 2 and 4. Plot the two sets of forecasts in side-by- side plots with the same limits on the x- and y-axes. Describe any differences between the two sets of forecasts. Using the backshift operator, write the models you found in problems 2 and 4.**

A5.



From the above graph comparison, we can see that there is no visual difference in the graphs.

Forecasting logconsumption for the next eight quarters.

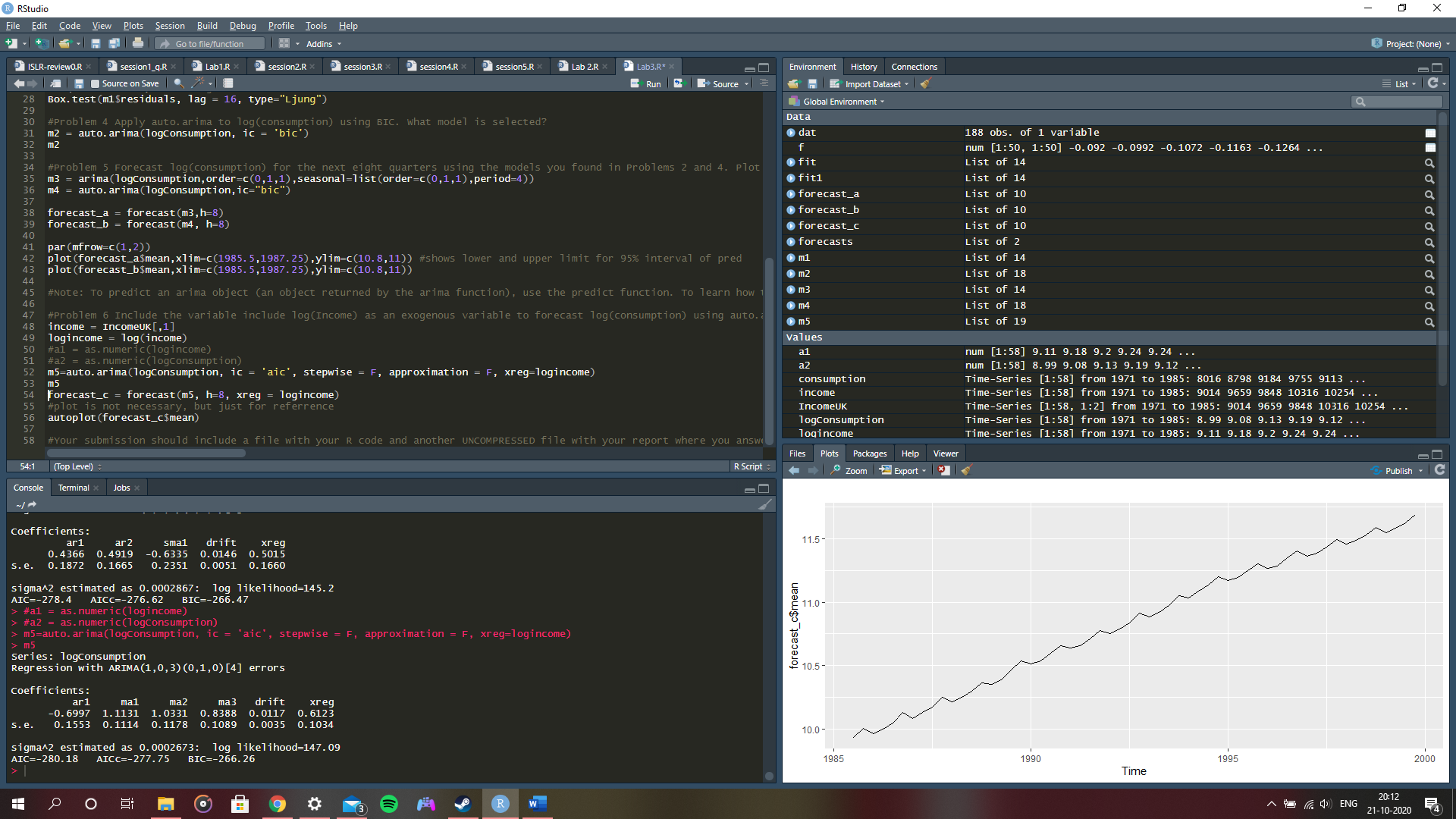
Using the backshift operator –

**(1-β) \* (1-β4) \* logConsumption$mean(t) = θ (0) + (1- θ4)( β4)) ε(t)**

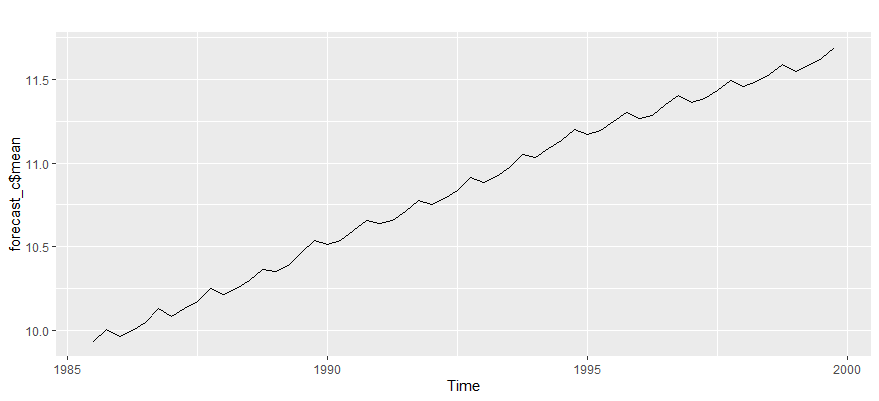
**Q6. Include the variable include log(Income) as an exogenous variable to forecast log(consumption) using auto.arima. According to the AIC, is this model better than the previous models? (Hint: use xreg to include exogenous variables in arima and auto.arima)**

A6.

When we include exogenous variable, i.e. xreg, the AIC value is better than the previous models.



The previous AIC value was -275.55 which is not better than the current AIC value -280.18.



The above is the forecast plot. It has not been asked in the question; however, it is done for reference purpose.