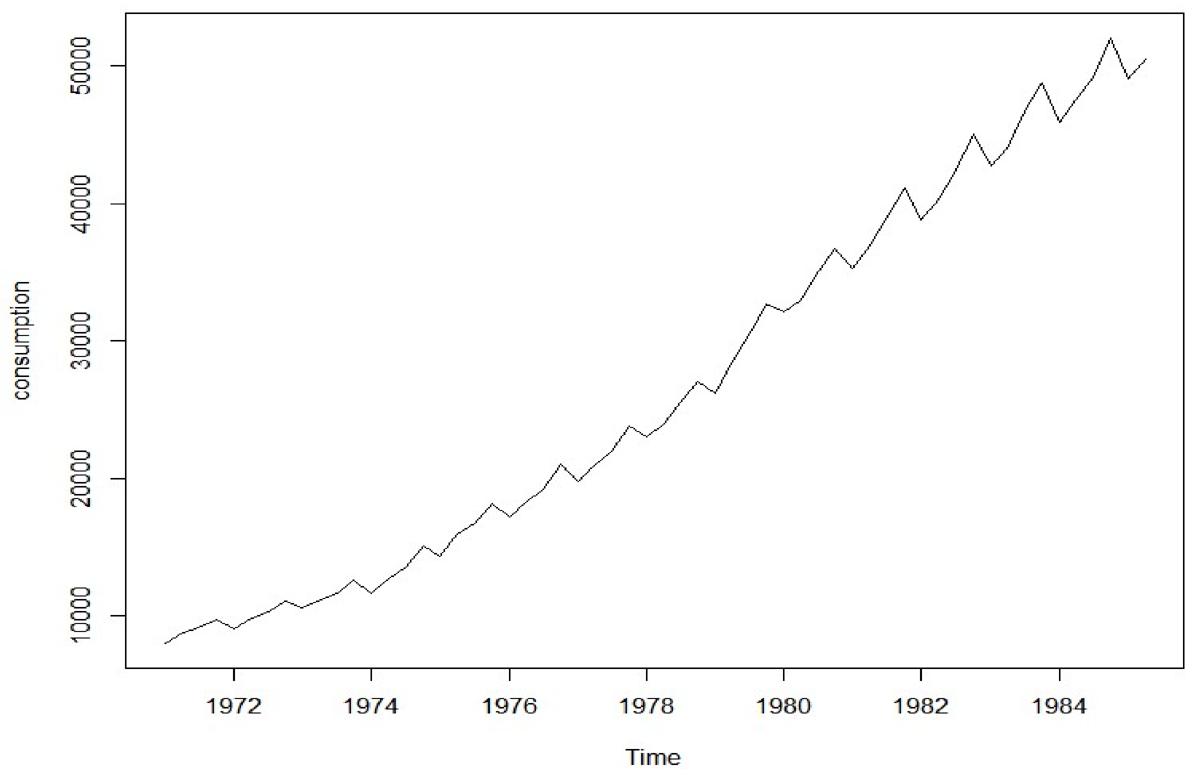
Kavit Shah

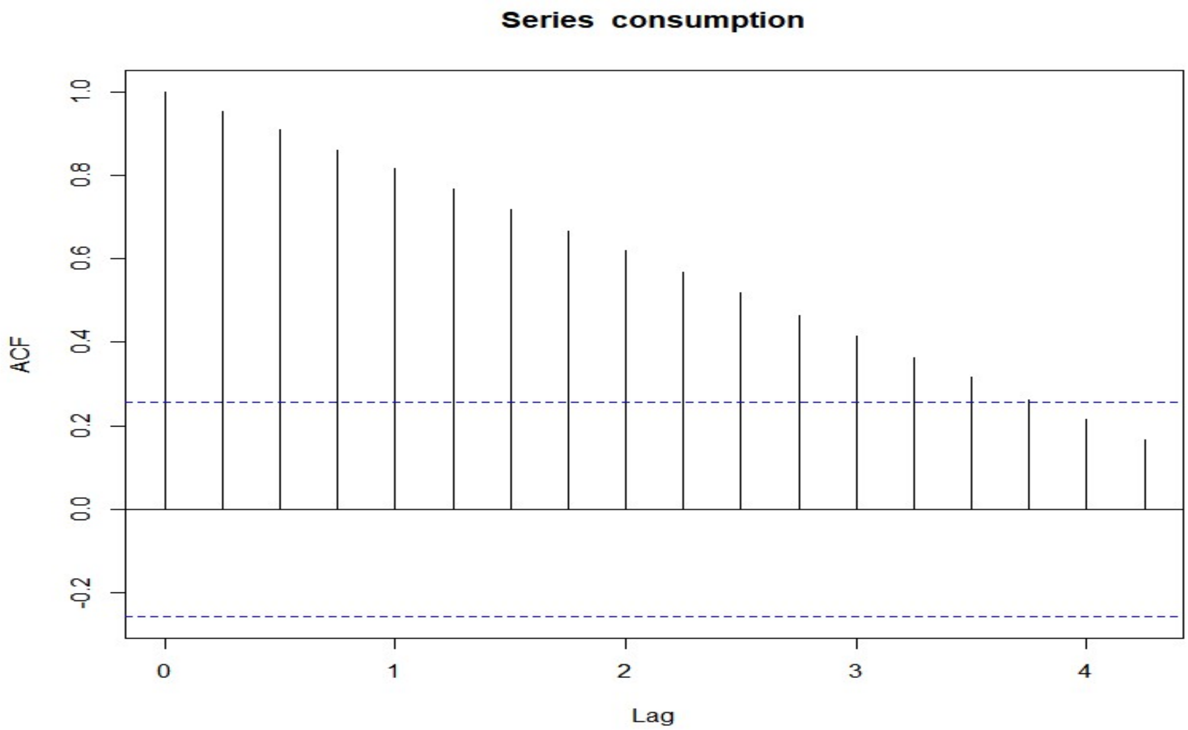
BIA-656 Advanced Data Analytics and Machine Learning

Assignment – 3

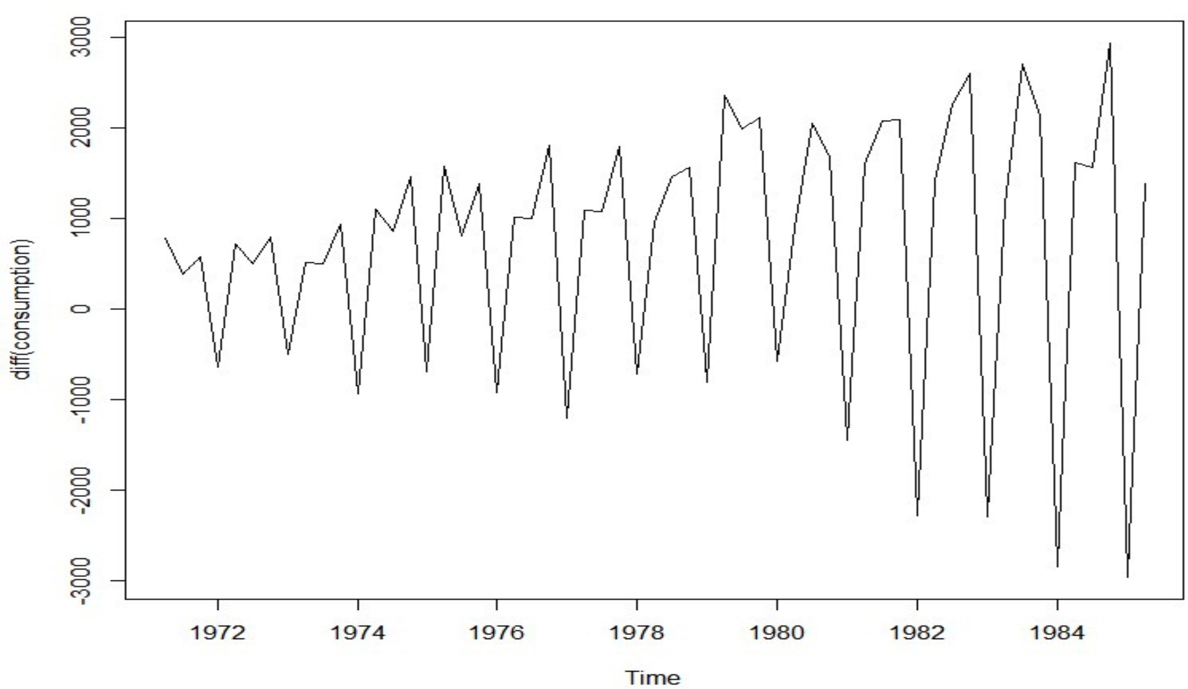
Problem 1 Describe the behavior of consumption. What types of differencing, seasonal, nonseasonal, or both, would you recommend? Do you recommend fitting a seasonal ARIMA model to the data with or without a log transformation? Consider also using ACF plots to help answer these questions.

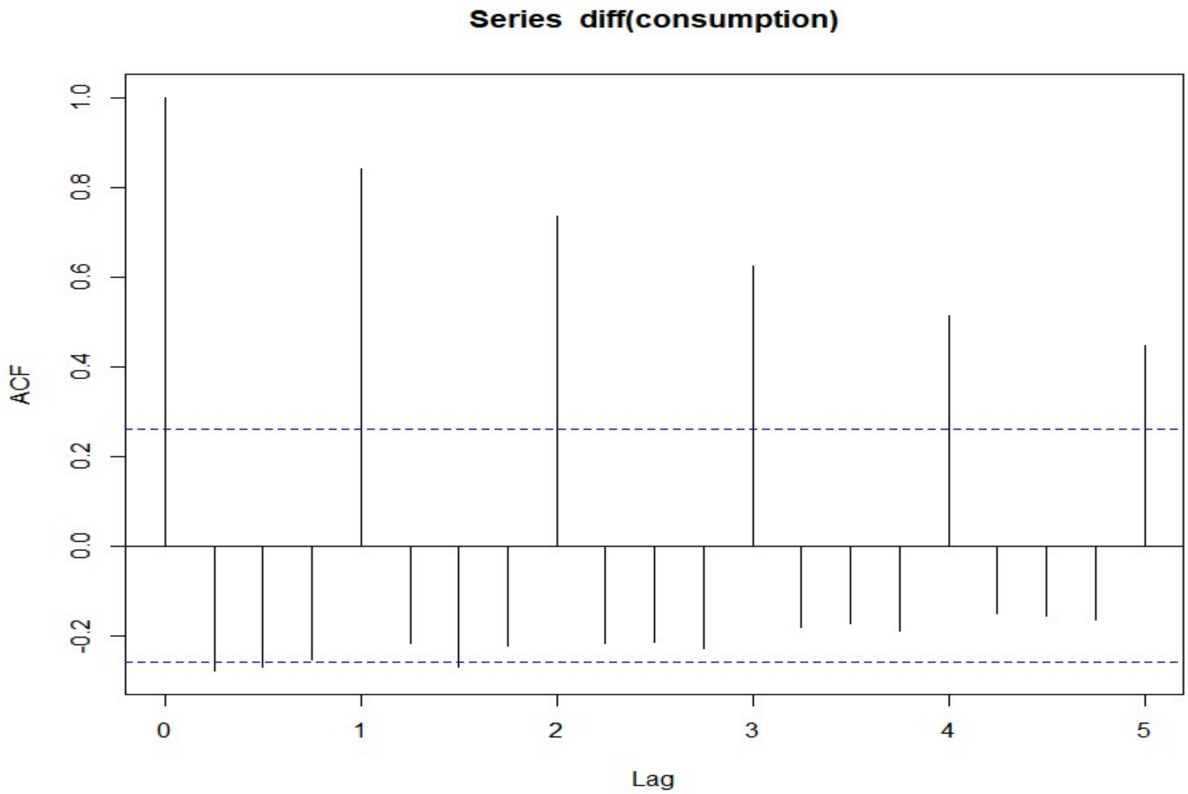
Ans:





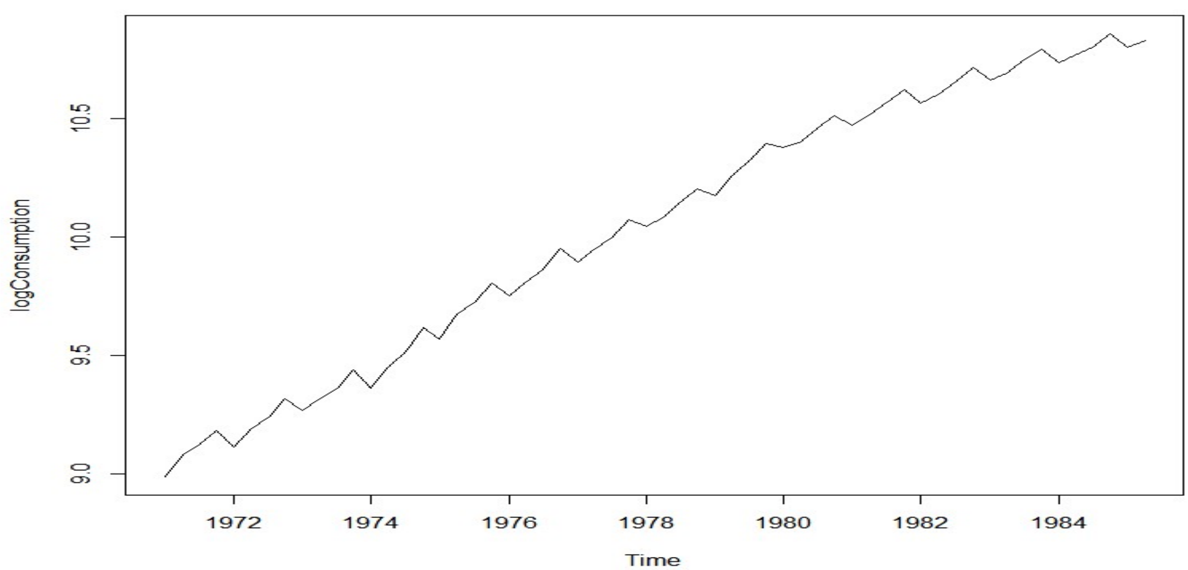
The above plot shows that the consumption time series is non-stationary and the acf plot adds to the confirmation of the non-stationarity of the consumption time series.

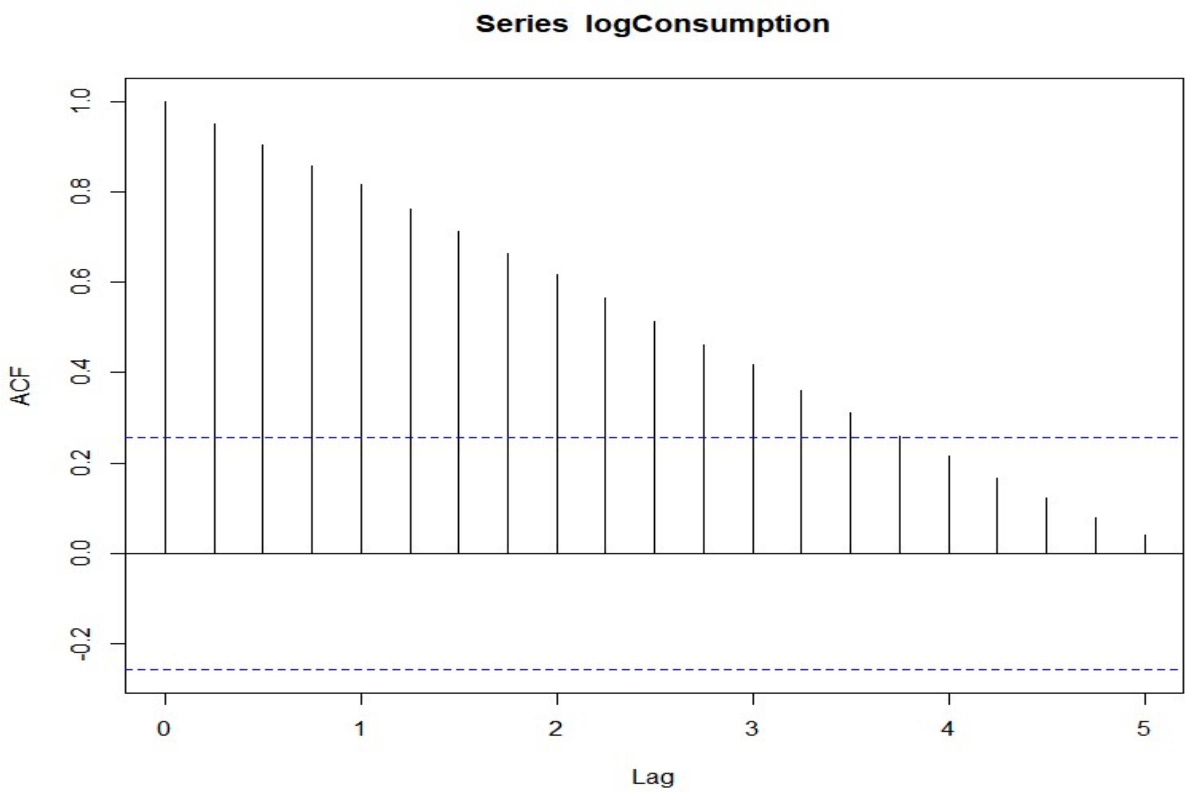




So, after the first differencing of the plot we see a stationary plot but there are spikes at the multiples of 4 in the acf plot which tells us about the seasonal effect of the consumption time series. Therefore, seasonal differencing along with non-seasonal differencing is required.

Plots after log transformation:

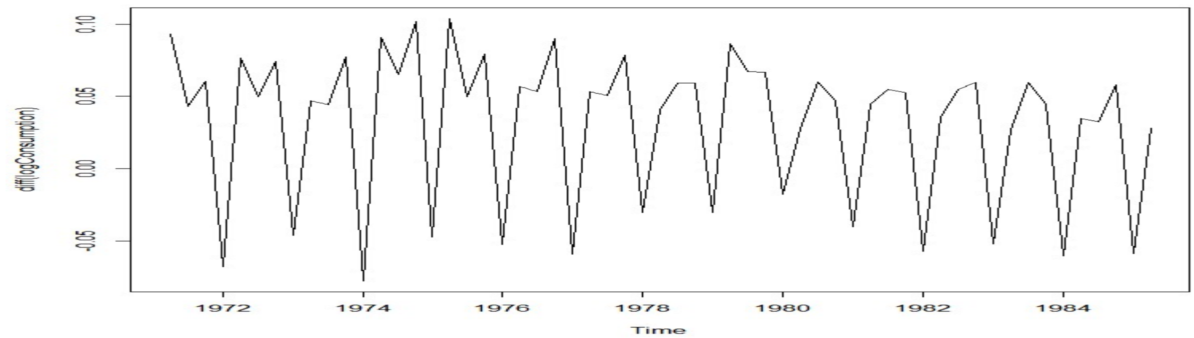


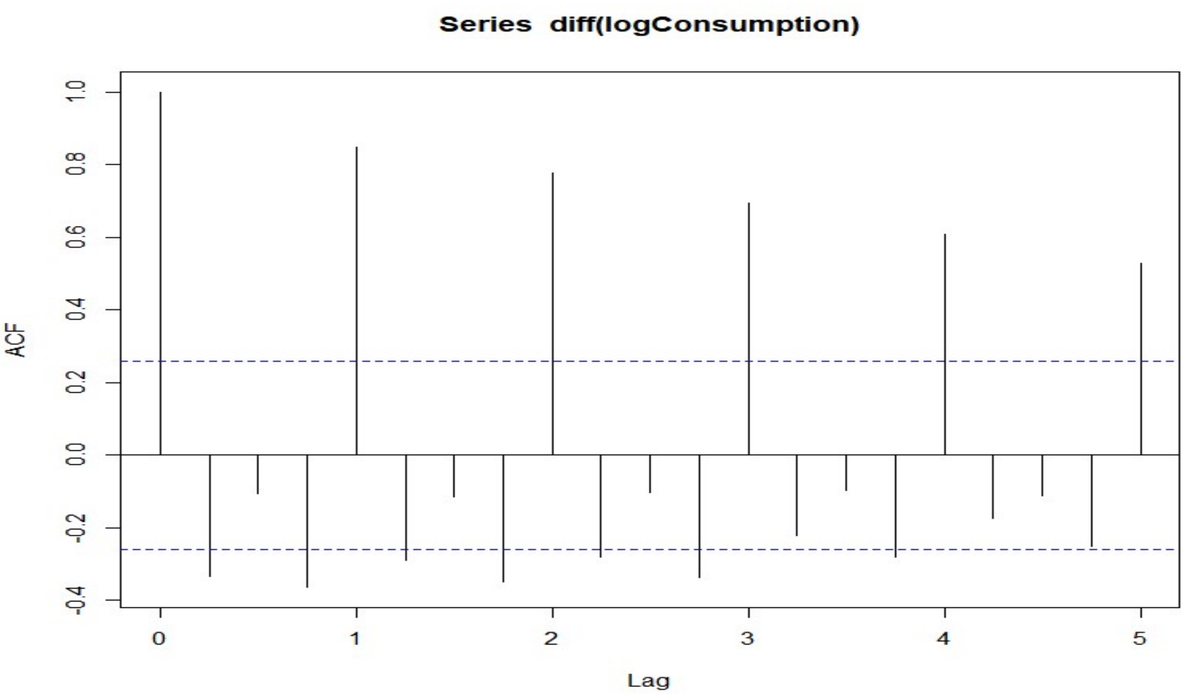


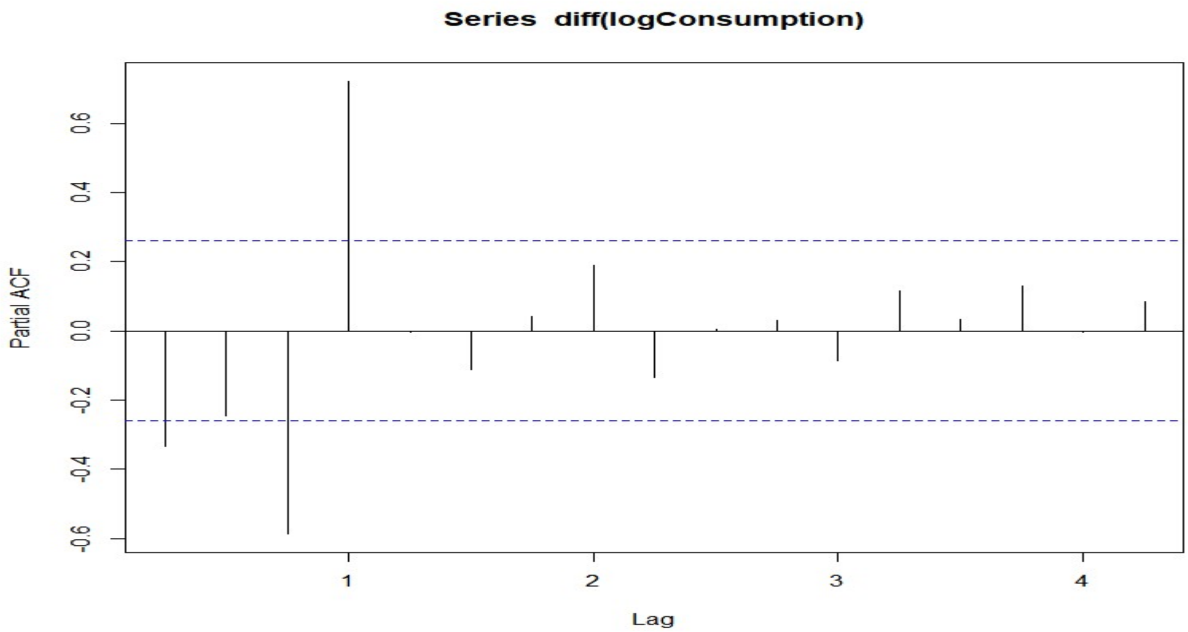
Even after the log transformation we do not see any stationarity and seems non-stationary. Therefore, there was no significant impact of the log transformation on the time series. Therefore, I would recommend fitting the seasonal arima model without the log-transformation.

Problem 2 Regardless of your answers to Problem 1, find an ARIMA model that provides a good fit to log(consumption). What order model did you select? (Give the orders of the nonseasonal and seasonal components.)

Ans: After single differencing, the model looks stationary according to the plot as shown below







But the AR and MA order can be found out after using the auto arima function. The extreme spikes at lags of multiples of 4 give us a hint that the seasonal differencing period might be of the order 4.

```
Call:
arima(x = logConsumption, order = c(0, 1, 0), seasonal = list(order = c(0, 1,
    1), period = 4))

Coefficients:
          sma1
       -0.5348
s.e.    0.1164

sigma^2 estimated as 0.0002923:  log likelihood = 139.77,  aic = -275.55

>
```

The non-seasonal ARIMA has the order of p=0 d=1 q=0

The seasonal ARIMA has the order of p=0 d=1 q=1 (seasonal period=4)

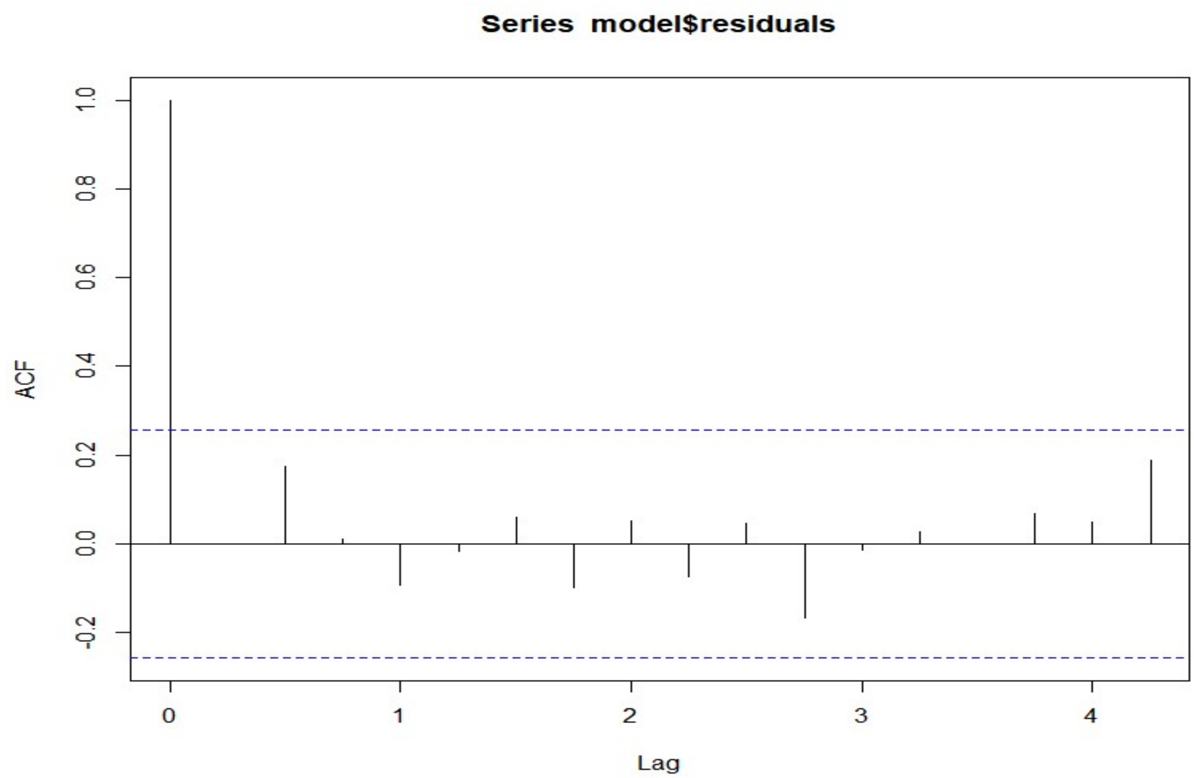
P=AR order

d =Differencing order

Q= MA order

Problem 3 Check the ACF of the residuals from the model you selected in Problem 2. Do you see any residual autocorrelation?

Ans:




```
Box-Ljung test

data:  model$residuals
X-squared = 12.177, df = 20, p-value = 0.9098
```

As per the ACF plot all the correlation is within the significance level which tells us that there is no serial correlation in the model, and it can be further confirmed by the Ljung box test since the p-value is not below the significance level of 0.05 the null hypothesis is not rejected and there is no serial correlation.

Problem 4 Apply auto.arima to log(consumption) using BIC. What model is selected?

Ans:

```
Series: logConsumption
ARIMA (0,1,0) (0,1,1)[4]

Coefficients:
         sma1
      -0.5348
s.e.   0.1164

sigma^2 estimated as 0.0003056:  log likelihood=139.77
AIC=-275.55   AICc=-275.31   BIC=-271.61

Training set error measures:
                       ME       RMSE        MAE          MPE       MAPE       MAS
E         ACF1
Training set -0.002266483 0.01655313 0.0130689 -0.02076354 0.1310031 0.102117
8 -0.001369795
```

The non-seasonal ARIMA has the order of p=0 d=1 q=0

The seasonal ARIMA has the order of p=0 d=1 q=1 (seasonal period=4)

P=AR order

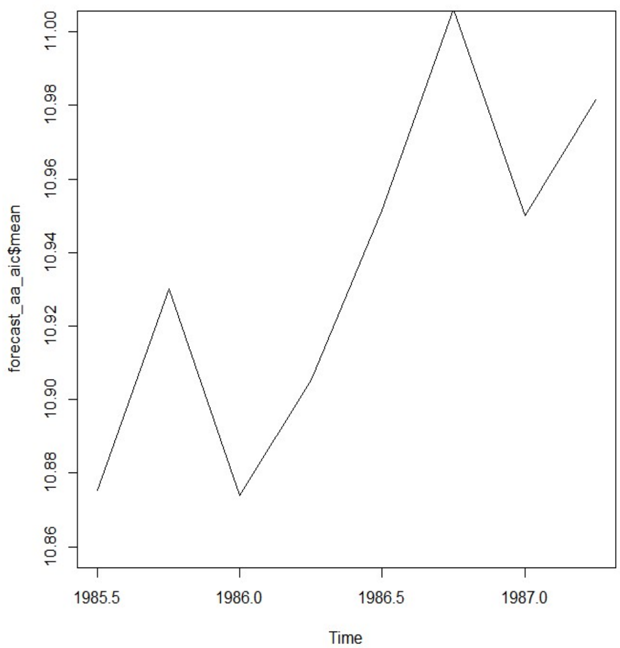
d =Differencing order

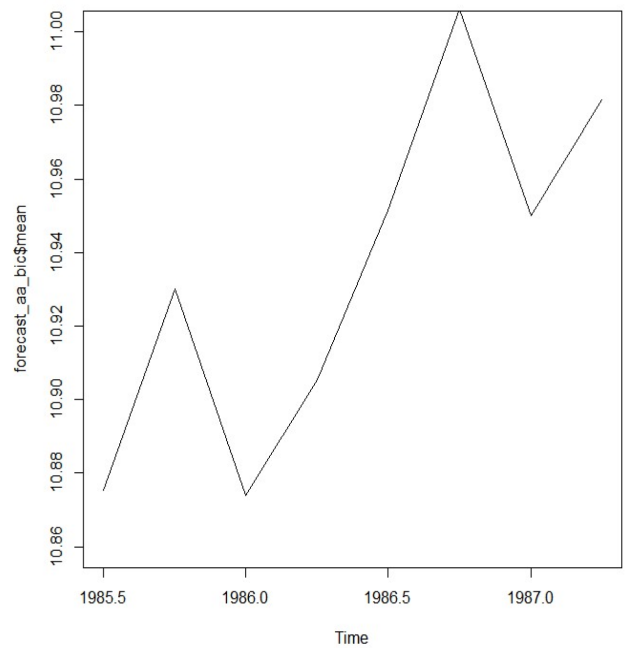
Q= MA order

Problem 5 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.

Ans: There are no notable differences in the plots as we can see below





Using the backshift operator, the equation of the model is:

(1-B) * (1-B^4) * logConsumption$mean(t) = θ(0) + (1- θ(4)(B^4)) ε(t)

Problem 6 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)

Ans: Yes, according to AIC this model is better than the previous model.

```
Series: logConsumption
Regression with ARIMA(1,0,3)(0,1,0)[4] errors

Coefficients:
          ar1     ma1     ma2     ma3   drift    xreg
      -0.6997  1.1131  1.0331  0.8388  0.0117  0.6123
s.e.   0.1553  0.1114  0.1178  0.1089  0.0035  0.1034

sigma^2 estimated as 0.0002673:  log likelihood=147.09
AIC=-280.18   AICc=-277.75   BIC=-266.26
```

We get the AIC value as -280.18 which is better than the previous model where the AIC was -275.55.