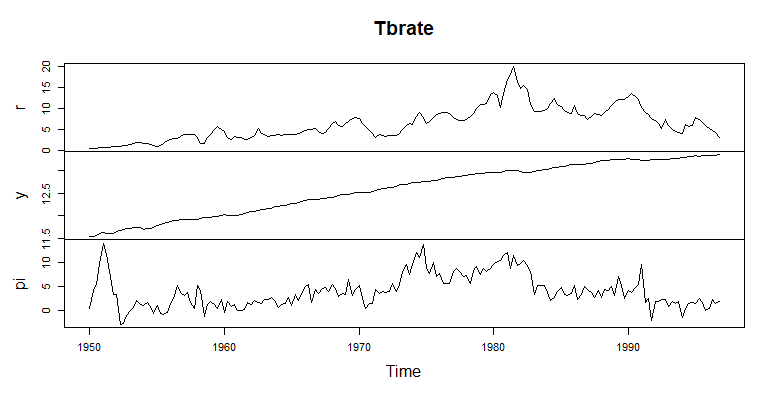
**BIA 656**

**Advance Data Analytics and Machine Learning**

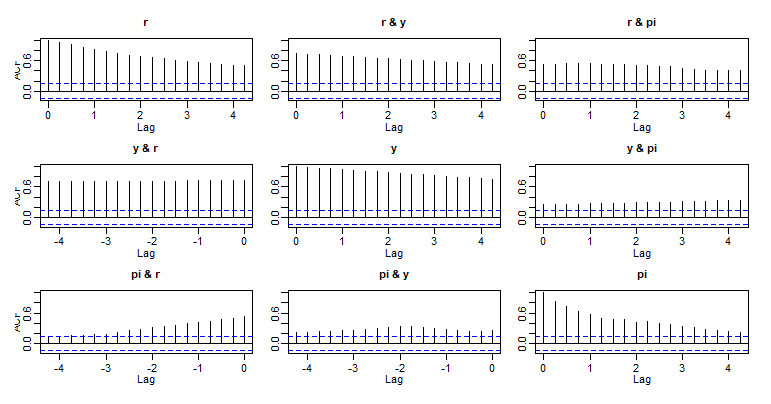
Assignment : Lab1

Shrey Kshatriya

**Q1.) Describe the signs of nonstationarity seen in the time series and ACF plots.**



Here in the time series plot, we can see that ‘y’ has seasonality, i.e. it has linear change and the is clearly non-stationary as the mean, variance and co-variance tend to change over time. Also, it might seem that r and pi are stationary, but they also have noticeable trends and changing level.

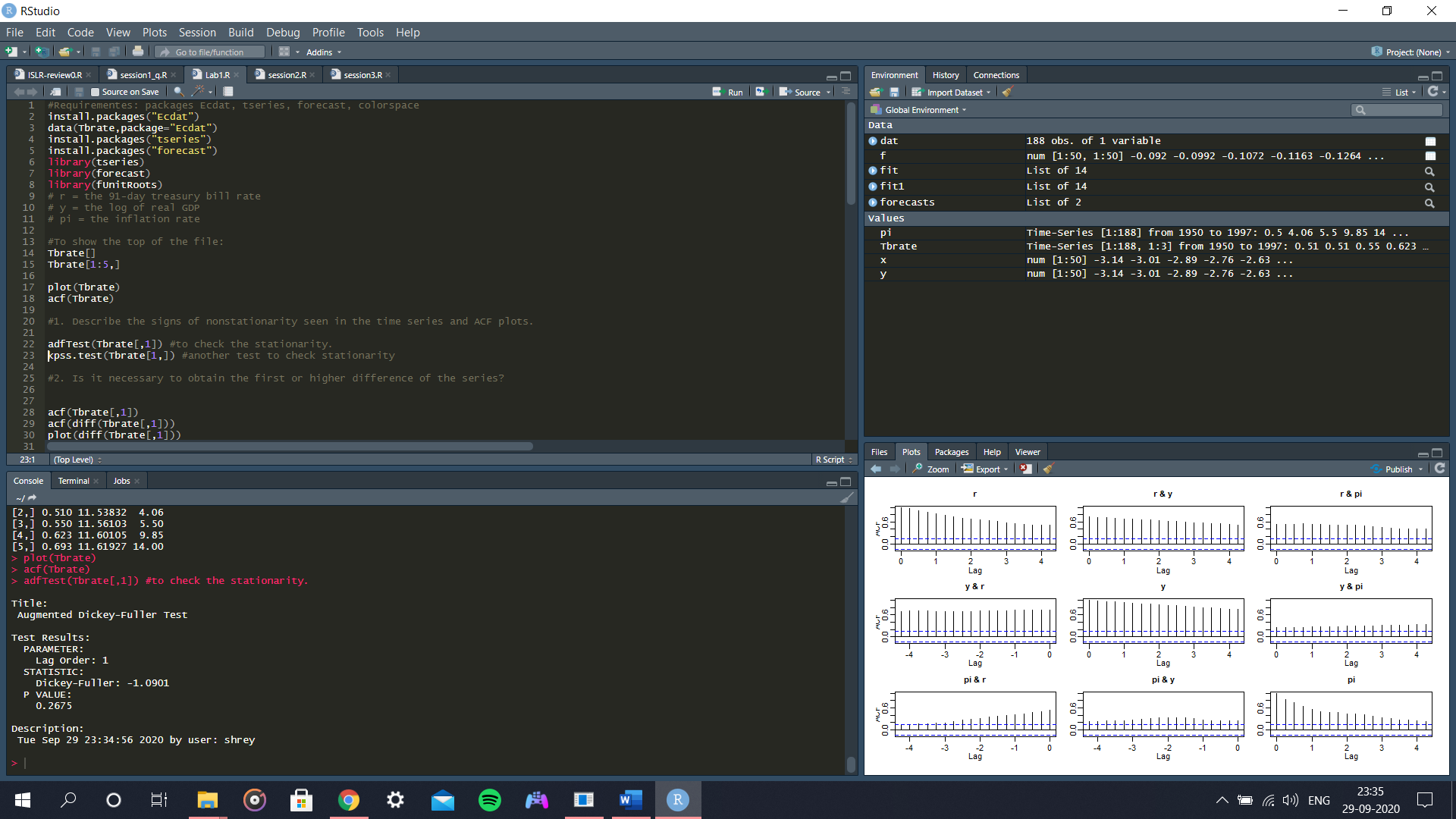


Now here we have plotted the ACF graphs. The ACF has the tendency to show sudden decay to 0 when the graph is stationary. However, here we observe that the decay to 0 is very slow and hence it can also be stated as non-stationary.

To support our observations, we can perform unit-root testing. Once such test that we can perform can be the Augmented Dickey Fuller test.

It states that if the p-value is greater than 0.05, then we can accept the Null Hypothesis. And in this case the null hypothesis taken is that the graphs are non-stationary.

We run the command : ‘adfTest(Tbrate)’



Here we see that the lag order is 1 and the p-value is greater than 0.05. Hence, we can accept the Null Hypothesis. Hence, the graph plots are non-stationary.

**Q2.) Is it necessary to obtain the first or higher difference of the series?**

Yes, it is necessary to obtain the first or the higher difference of the series. As observed above, the graphs are non-stationary. Hence, to convert non-stationary graphs to stationary to build models, we need to do differentiation. Hence it is necessary to obtain the order of differentiation. Sometimes the first differentiation is not enough, so use the second or a higher differentiation order.

**Q3.) What order of differencing is chosen?**

The auto ARIMA model chooses the differencing order of 1.

**Q4.) Does this result agree with your previous conclusions?**

Yes, this result agrees with our previous conclusions. We previously stated that first order of differentiation is important. And the auto ARIMA model also gives us the differencing order of 1. Hence our conclusions are satisfied.

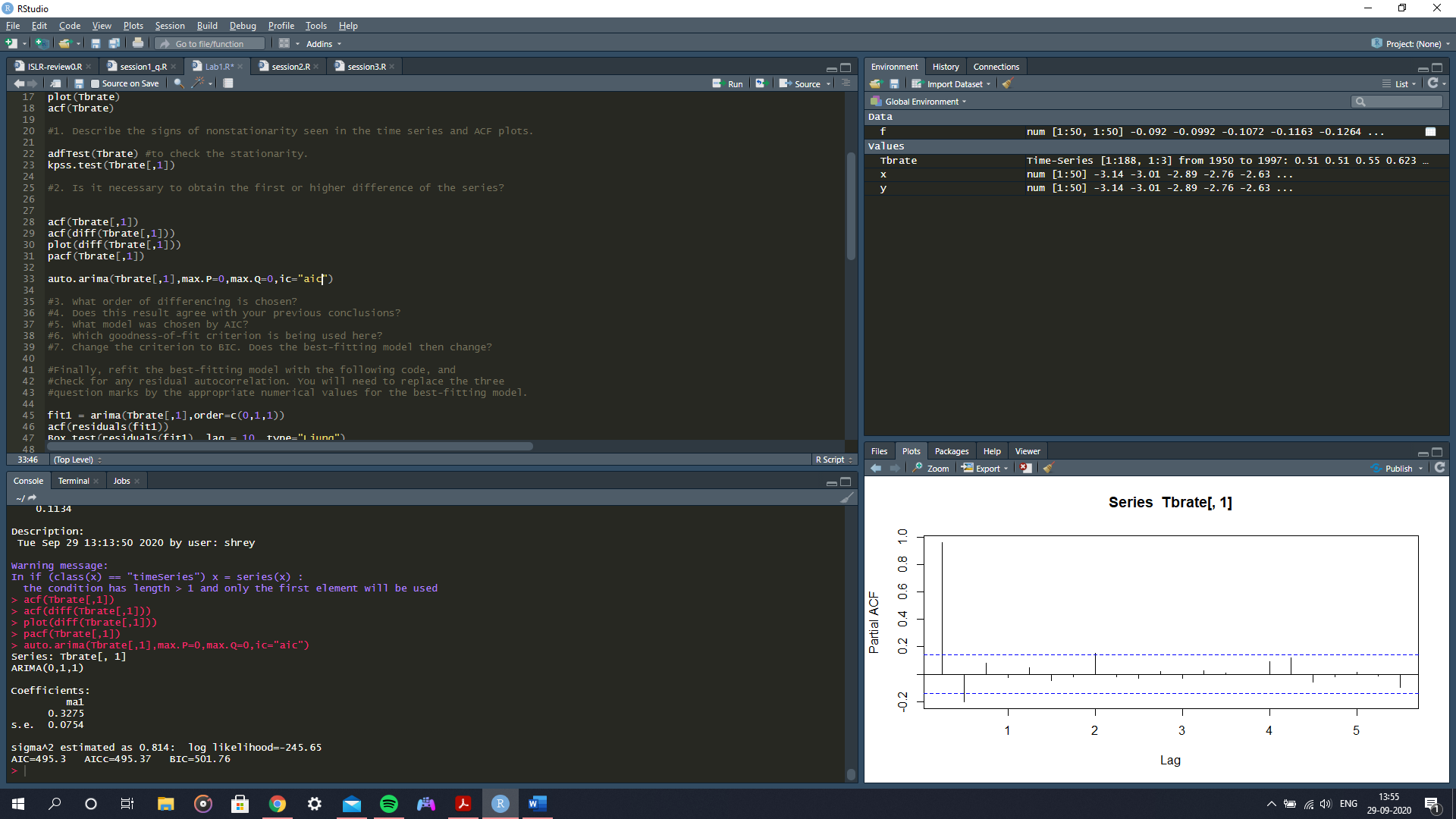
**Q5.) What model was chosen by AIC?**

The model chosen by AIC is ARIMA(0,1,1)

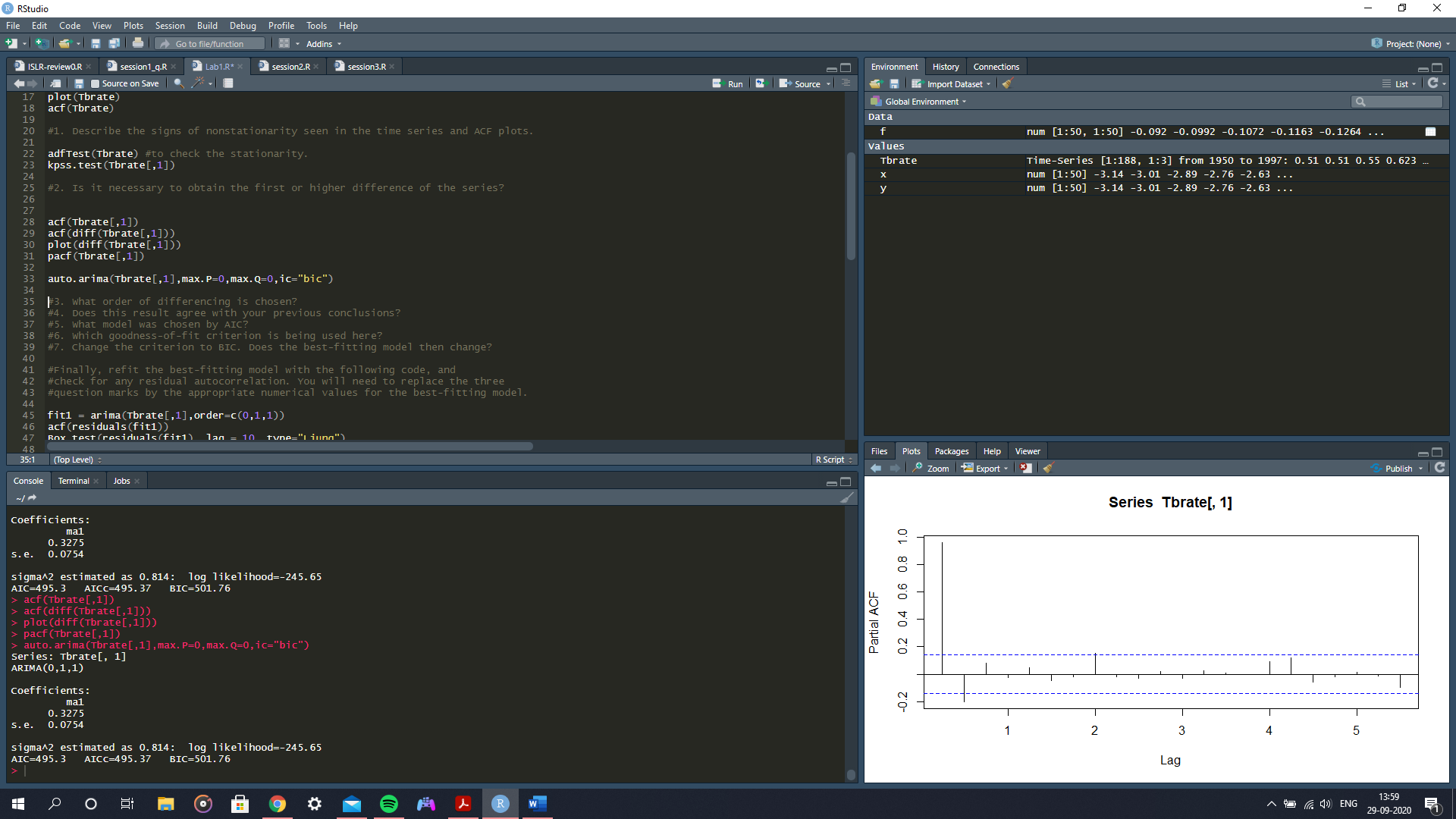
**Q6.) Which goodness-of-fit criterion is being used here?**

The goodness of fir criterion chosen here is AIC since the value of AIC is less than BIC and it satisfies the 2 criterions for good-of-fit. They are as follows:

1. Good fit
2. Less Complexity



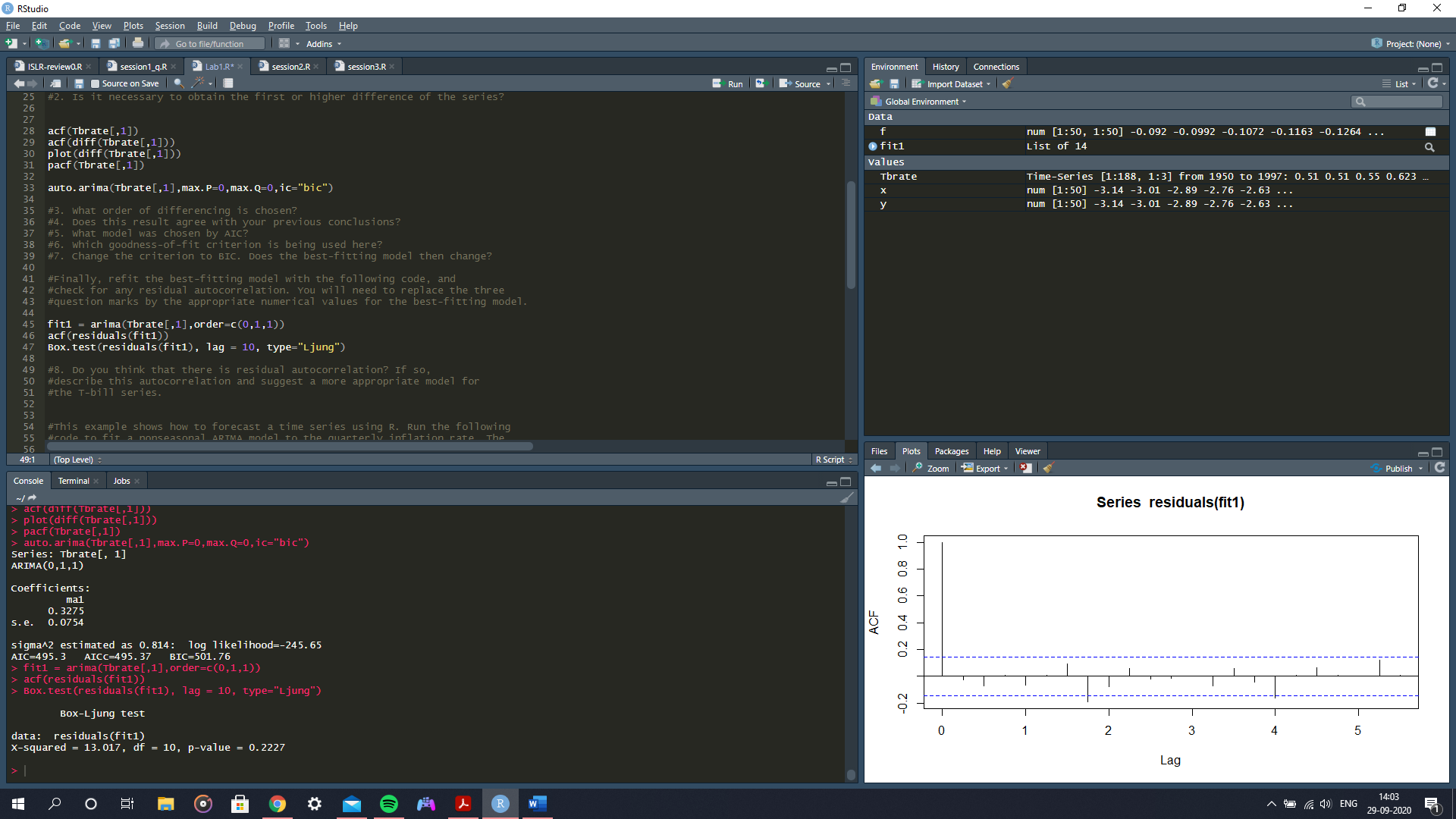
**Q7.) Change the criterion to BIC. Does the best-fitting model then change?**

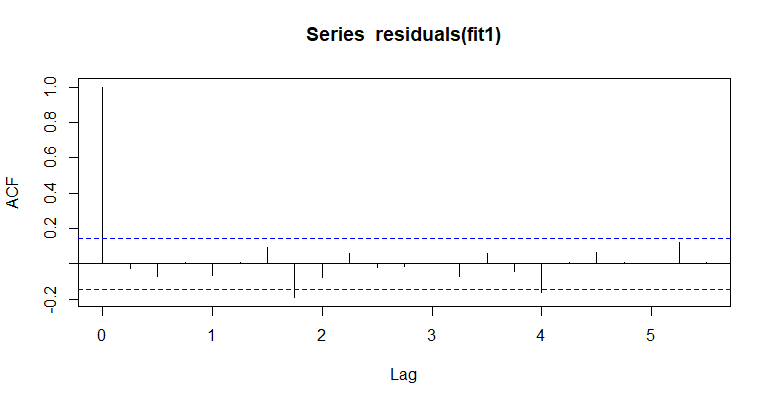


When the criterion is changed to BIC, the AIC still gives a better value, hence the best-fitting model does not change.

**Q8.) Do you think that there is residual autocorrelation? If so, describe this autocorrelation and suggest a more appropriate model for the T-bill series.**

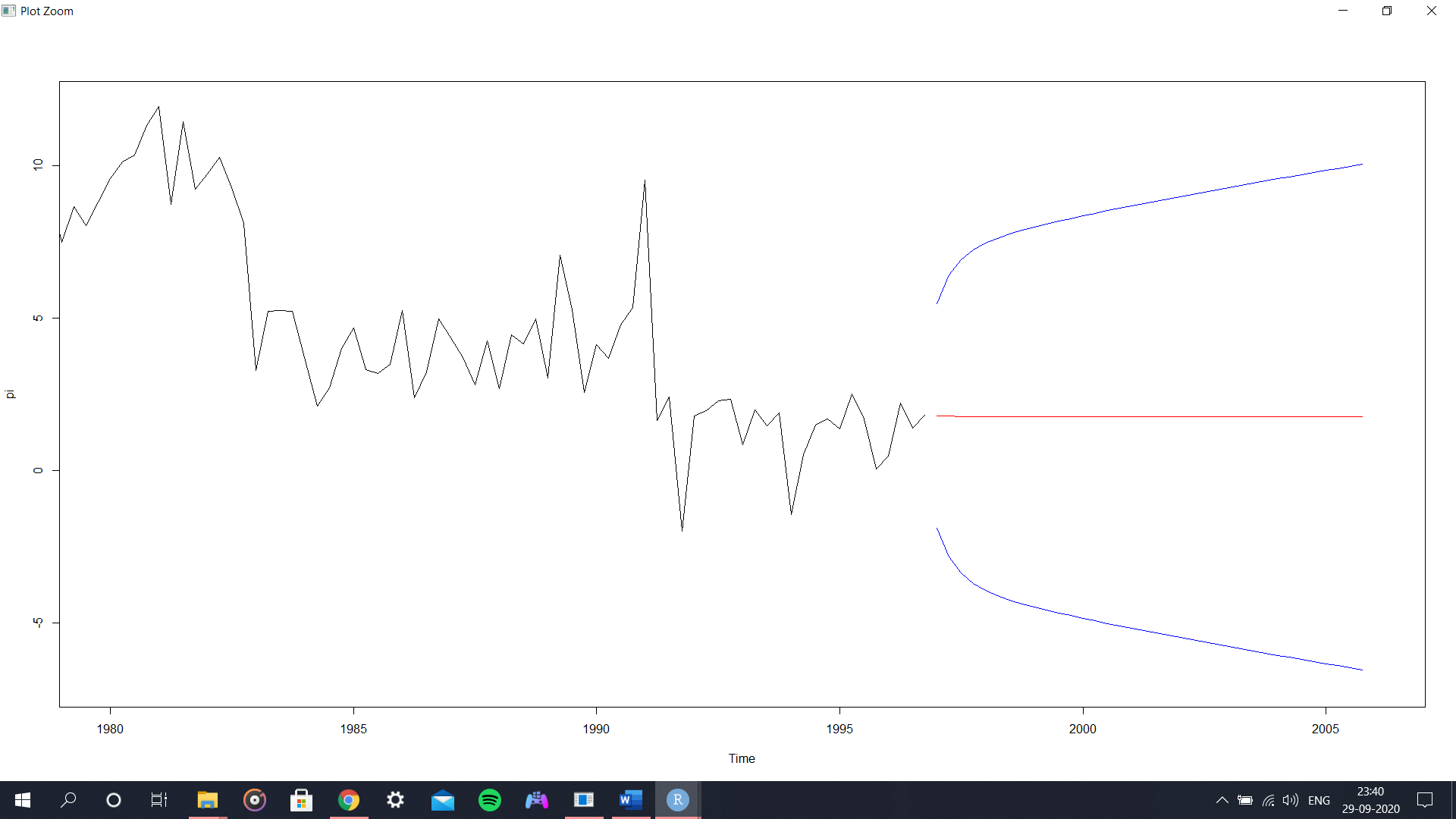
When we perform the Box-Ljung test, we find that the p-value is 0.05. This satisfies the null hypothesis which we know is true if p-value is greater than 0.05. The null hypothesis chosen here is that there is 0 residual autocorrelation. Since, the null hypothesis is satisfied, there is no residual autocorrelation.





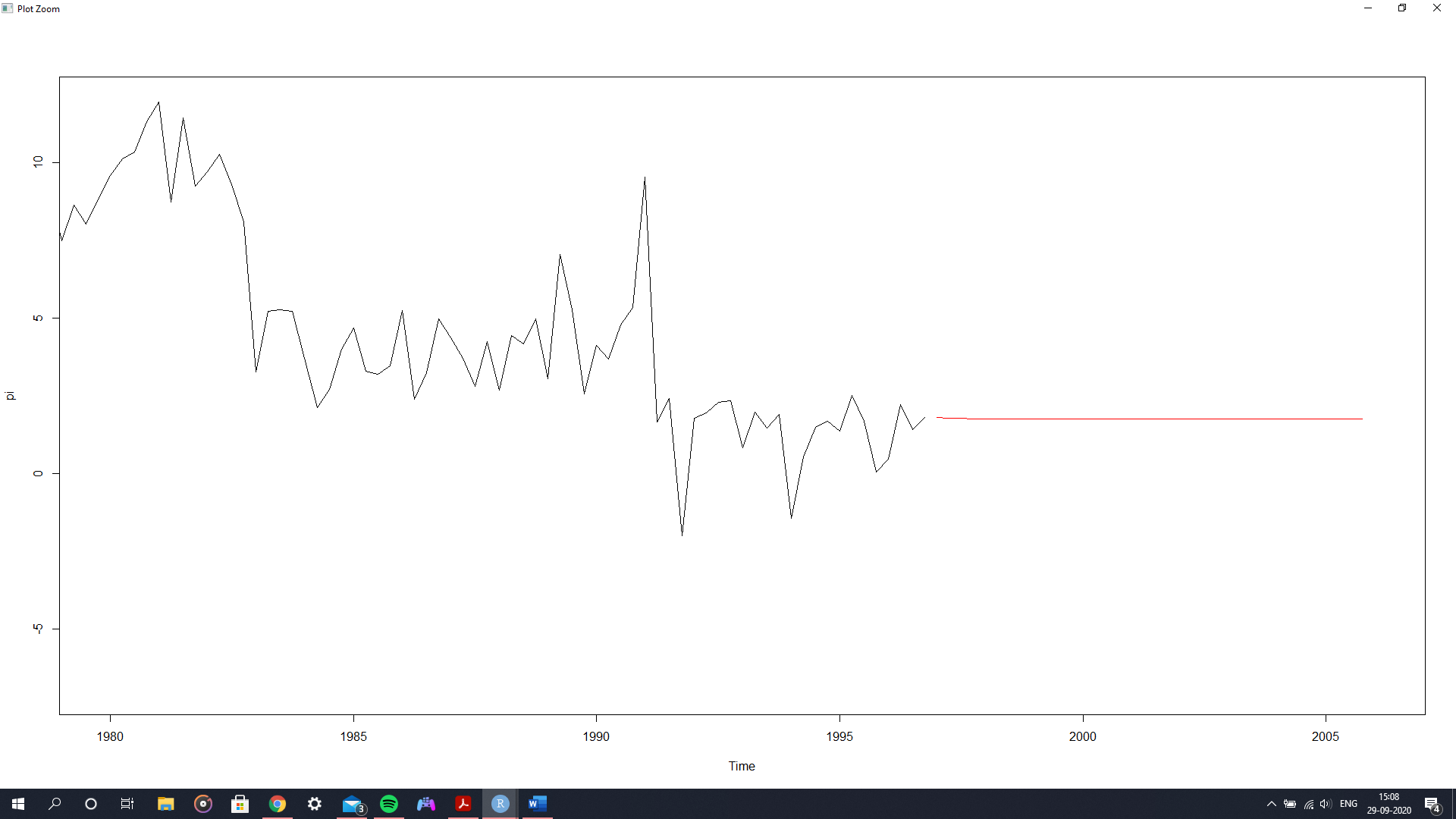
The above graph may show that there is some relevant lag, however we can consider those be outliers and anomalies.

**Q9.) Why do the prediction intervals (blue curves) widen as one moves farther into the future?**



The shock component is forecasting is not known; it is random. Hence there is a difference in the forecasted and actual value which is known as the forecasting error. As we move farther into the future, the forecasting error also increases, and this curve keeps widening as we go ahead.

**Q10.) What causes the predictions (red) and the prediction intervals to wiggle initially?**



As we do more forecasting, the accuracy of the forecasting value keep decreasing up and until a point where the model gives out the mean of the series as the forecasting value. This is due to the decreasing dependency on the previous lags. Hence, the values between the initial value and the value till when we get the mean of the series as forecast values, we see the initial wiggle.