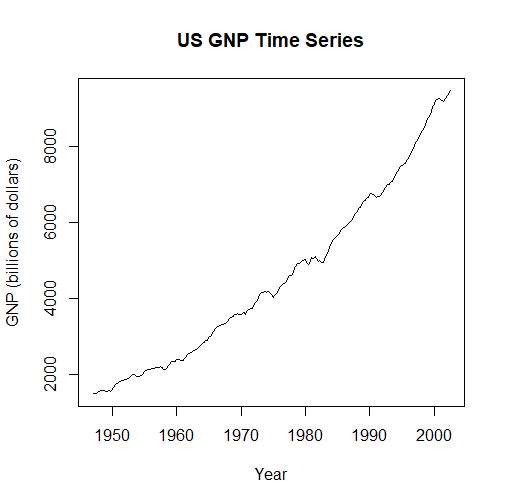
**Analysis of Time Series Data**

***Bideepta Saha***

* **Description of the dataset:**

The data that we are working on is “gnp” available under “astsa” library in R. It is seasonally adjusted quarterly U.S. GNP from 1947(1) to 2002(3).

* **Plotting of the Dataset:**



Observations from the plot:

1. The plot shows an upward trend.
2. Seasonality is also present and it may be noted that it is relatively constant in relation to the trend. Hence an additive model may be used for the purpose of analysis.

* **Decomposing into the respective components :**

An additive model is used. Let Xt bethe value of time series at time point t.

Then,

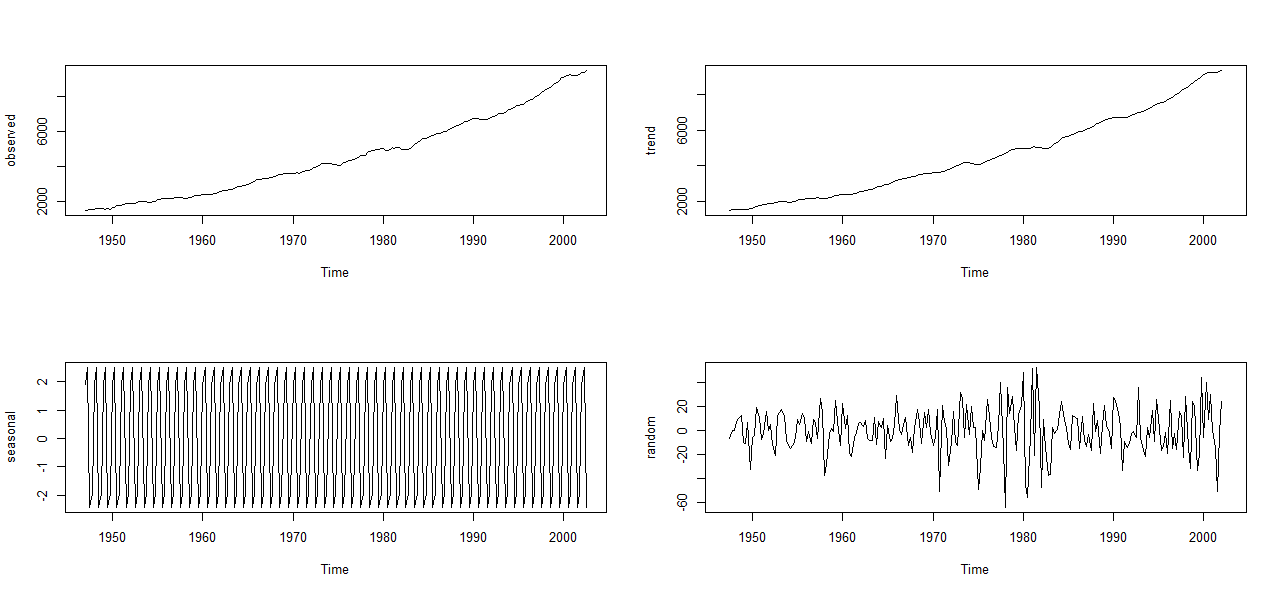
Xt = Tt + St + It

Where Tt : Value of the trend component at time point t

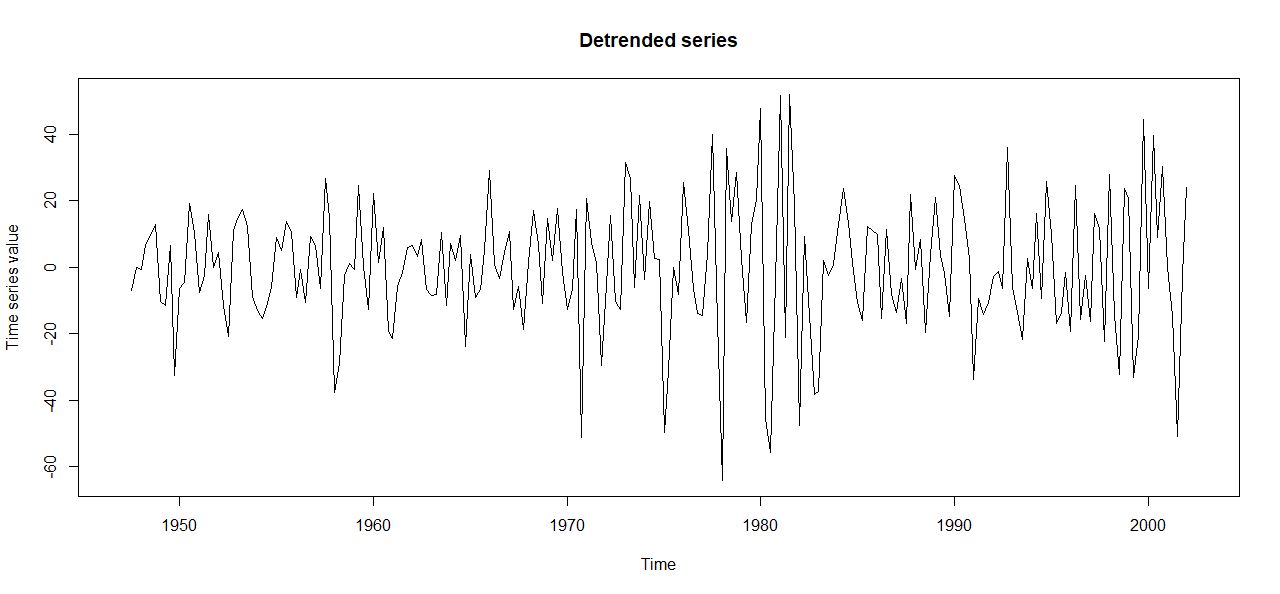
St : Value of the seasonal component at time point t

It : Value of the irregular component at time point t

Using R, we have decomposed the time series data. The various components after decomposition are as follows.



* **An overview of the detrended series:**



* **Augmented Dickey-Fuller Test (ADF Test):**

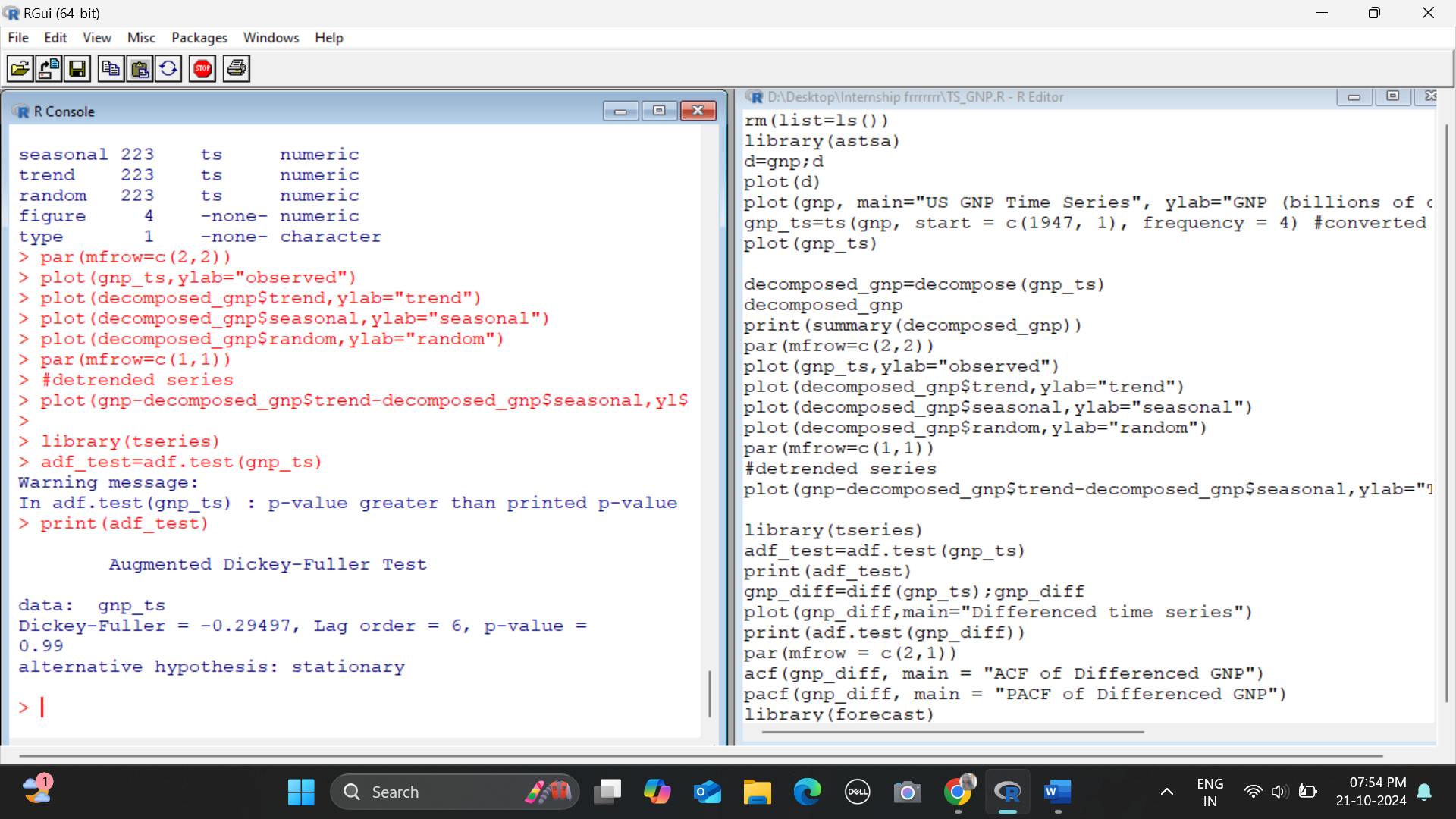
It is a statistical test that is used to determine whether a time series is stationary or not. The ADF test considers the possibility of a linear trend in the time series and can also handle AR terms in the model. It estimates a regression model of the form:

Where is the first difference of the time series, t is a linear trend term and yt-1 is the lagged value of the time series, and for i=1(1)p are the AR coefficients.

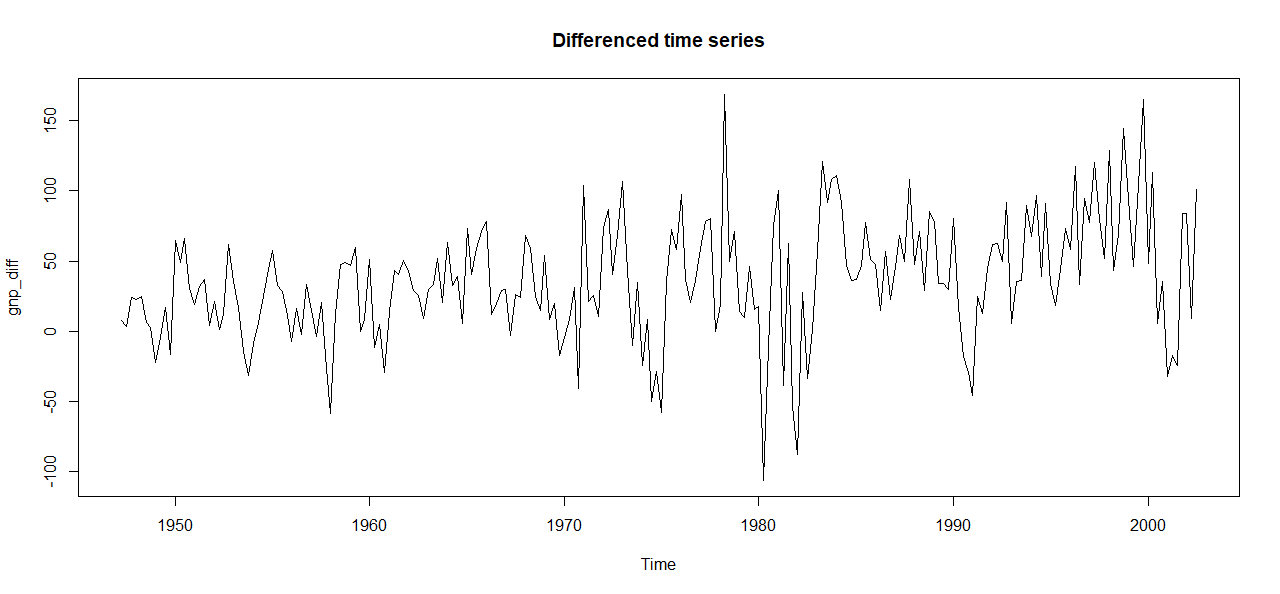
The null hypothesis of ADF is that there is a unit root in time series , indicating that the time series is non-stationary. The alternative is that the time series is stationary. The ADF test statistics is based on t-statistics of the coefficient β. If the absolute value of the test statistic is greater than the critical value at a given significant level, then the null hypothesis is rejected and the time series is considered as stationary. If the absolute value of the test statistic is less than the critical value then the null hypothesis is accepted, indicating that the time series is non-stationary.

* Computations and Results:

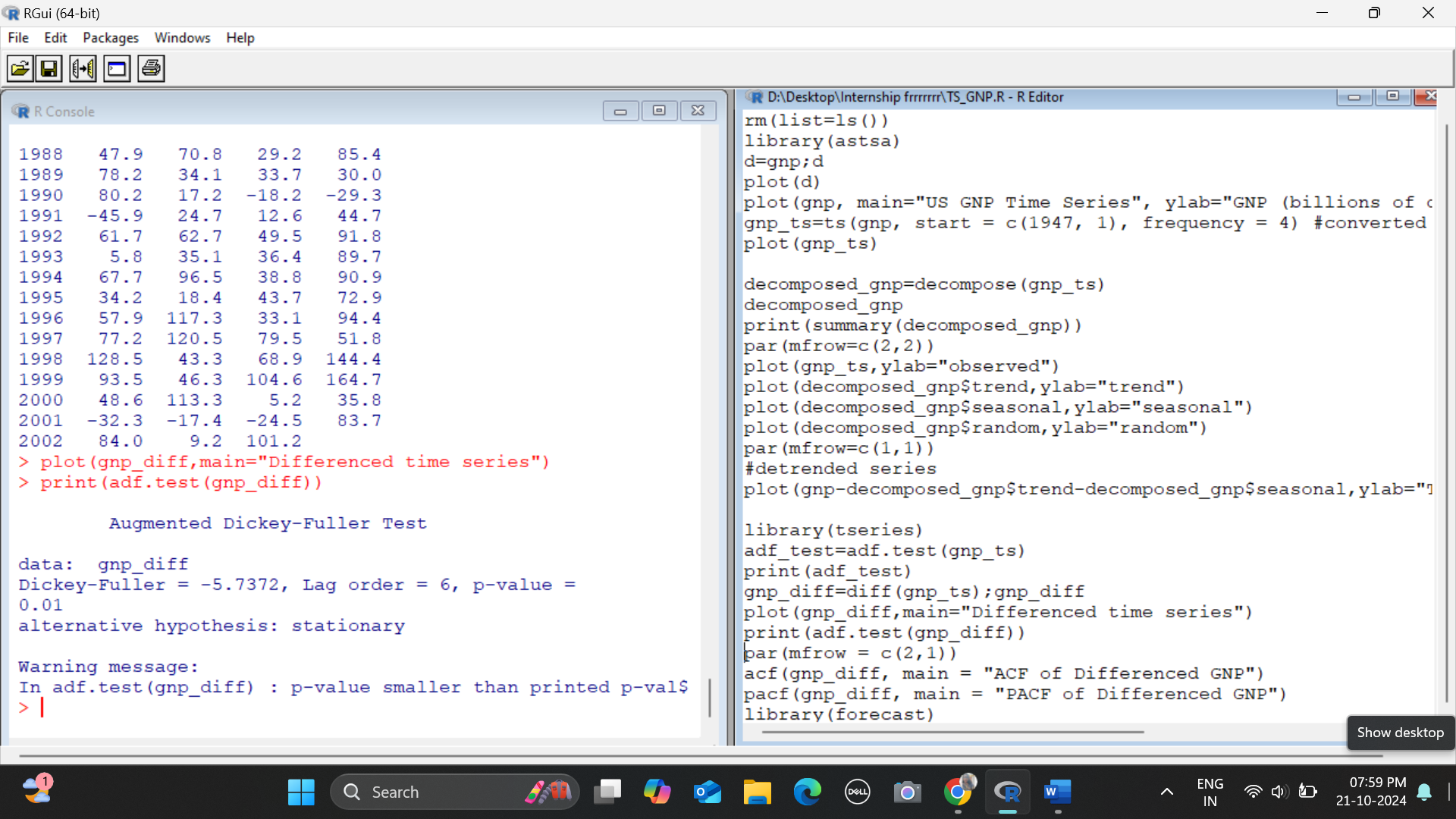
In R, we now perform the ADF test to check the stationarity of the time series. The results obtained are as follows:



We conclude that the time series is not stationary. So, we perform 1st order differencing on the time series and obtain the following:



Performing ADF test on this differenced series yield us a stationary series confirmation.



Now, we will proceed to fit an ARIMA model to the data. But before that let us briefly introduce ARIMA model.

* **ARIMA(Autoregressive Integrated Moving Average):**

This is a widely used statistical model for time series analysis and forecasting. It is designed to capture the temporal dependencies and patterns present in time series data, making it useful for predicting future values based on the past ones. It combines the following concepts:

* + **Autoregression(AR):**

The autoregressive component refers to the dependence of the current observation on the previous observations. It assumes that the value of the time series at any given point is linearly dependent on its past values. The ‘p’ parameters in ARIMA (denoted as ARIMA(p,d,q)) specifies the number of lagged observations to include in the model. A higher value of ‘p’ indicates a stronger dependence on the past observations.

* + **Differencing(I):**

The integrated component represents the differencing of the time series data to make it stationary, i.e., removing trends and seasonality. Stationarity is important for many time series analysis techniques because it ensures that statistical properties like mean and variance remain constant over time. The differencing parameter ‘d’ in ARIMA specifies the number of times differencing is applied to the series to achieve stationarity.

* + **Moving Average (MA):**

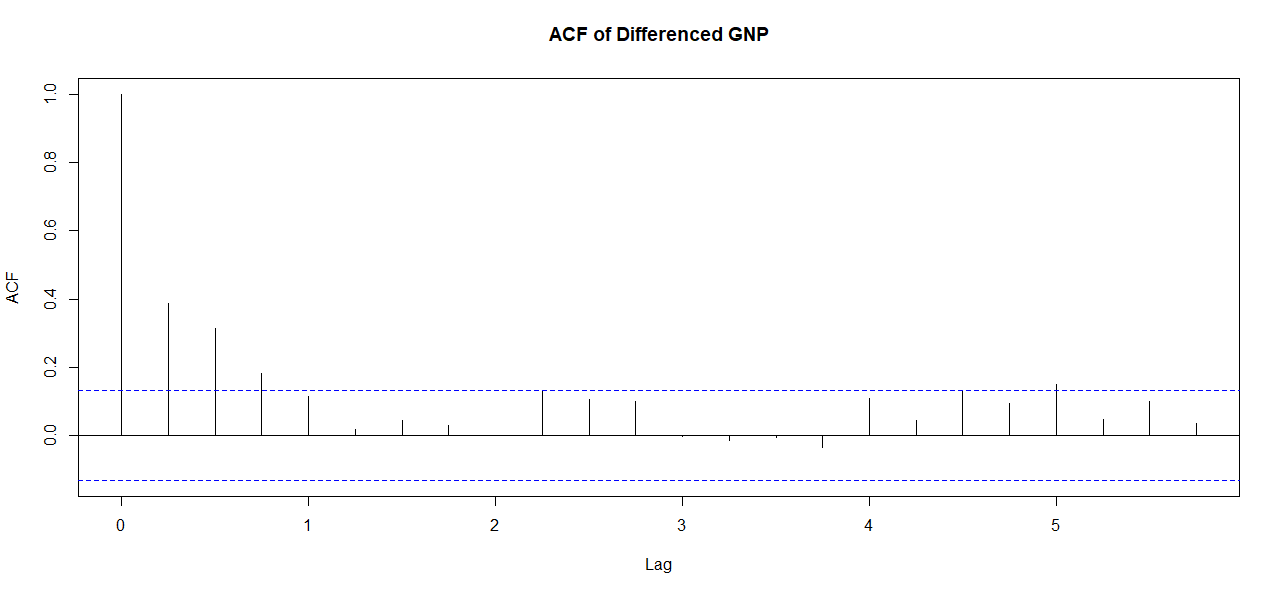
The moving average component accounts for the dependency between an observation and a residual error from a moving average model applied to lagged observations. It models the impact of past forecast errors on future values. The ‘q’ parameter in ARIMA specifies the number of lagged forecast errors to include in the model. A higher value of ‘q’ indicates a longer memory of past forecast errors.

A non-seasonal stationary time series can be modelled as a combination of the past values and the errors which can be denoted as ARIMA(p,d,q) can be expressed as:

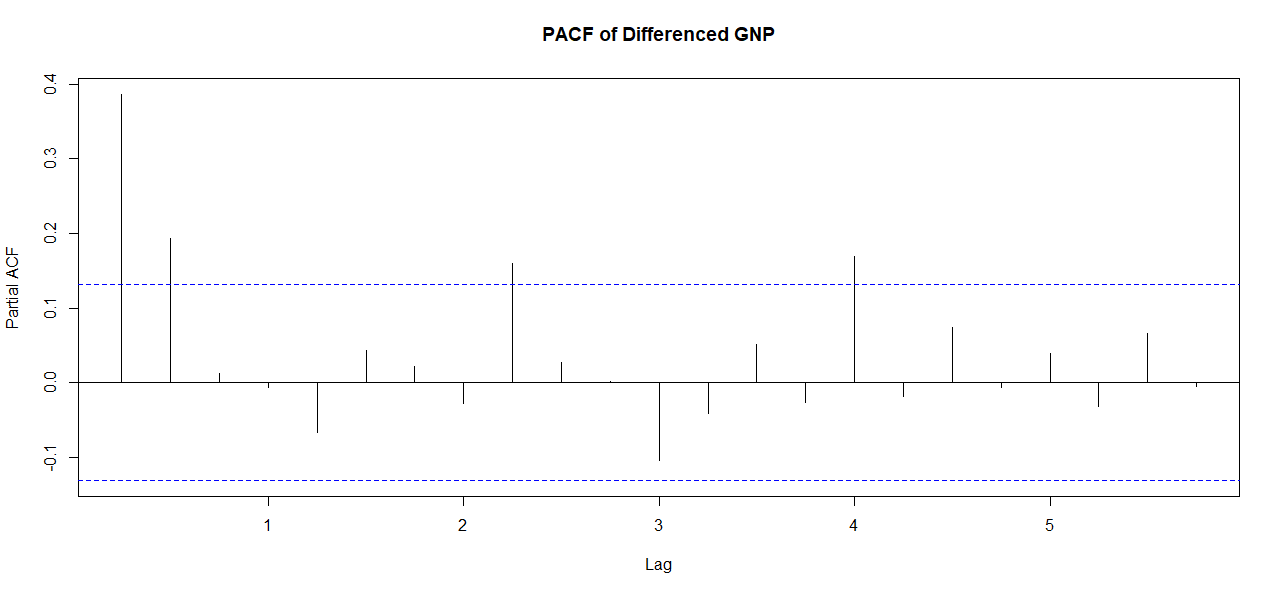
Thus, the value of the parameter d in ARIMA model would be 1 from our ADF test results.

Next, we use the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to identify potential values for p and q. The ACF plot helps us to identify the q parameter (MA order) and PACF plot helps us to identify the p parameter (AR order).

* Computations and Results:



We can take the cut off point for the ACF to be 1. Hence **q=1.**



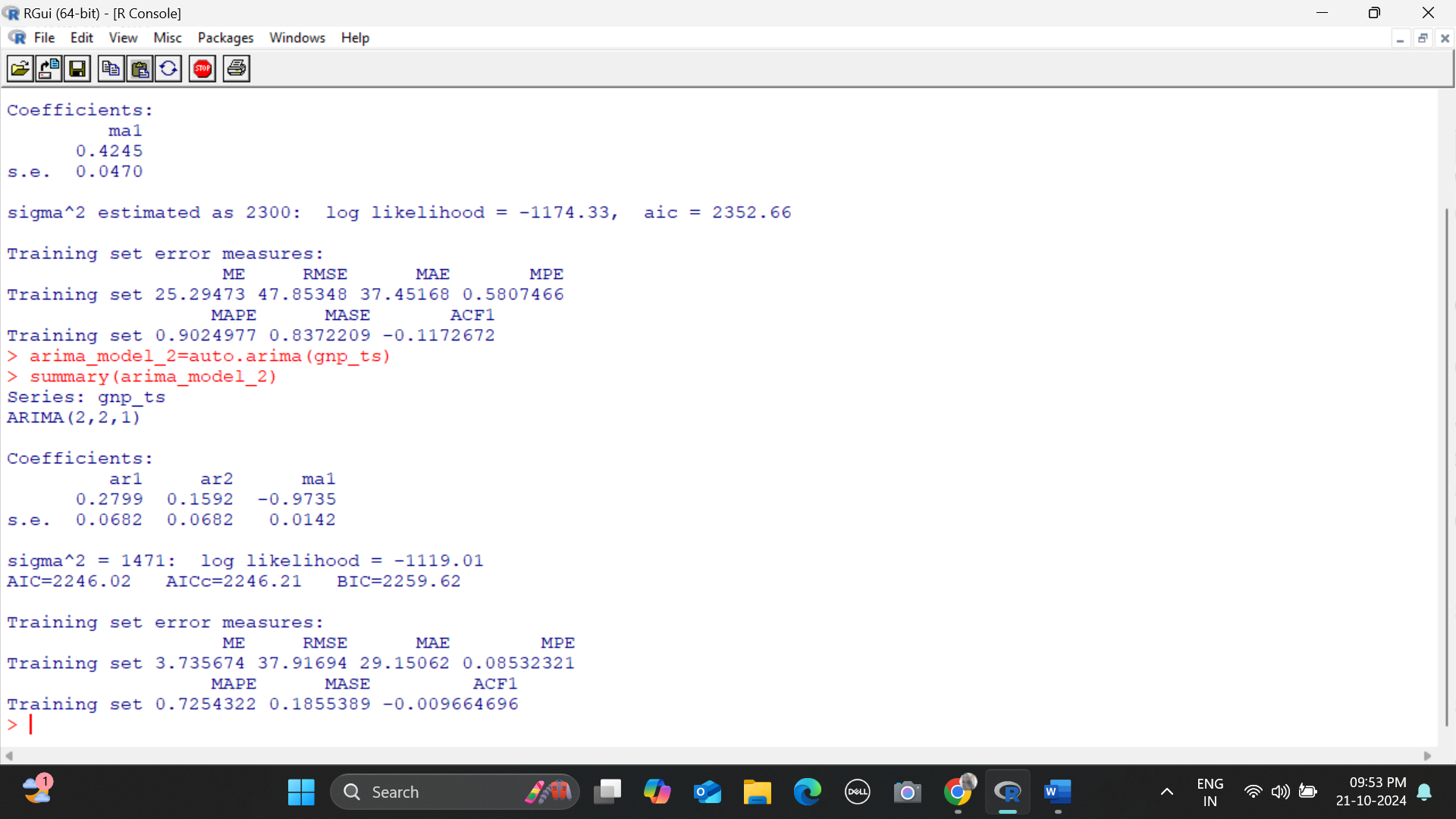
The cut off can be taken as 0. Thus **p=0**.

**Thus, we have the ARIMA model with parameters(0,1,1).**

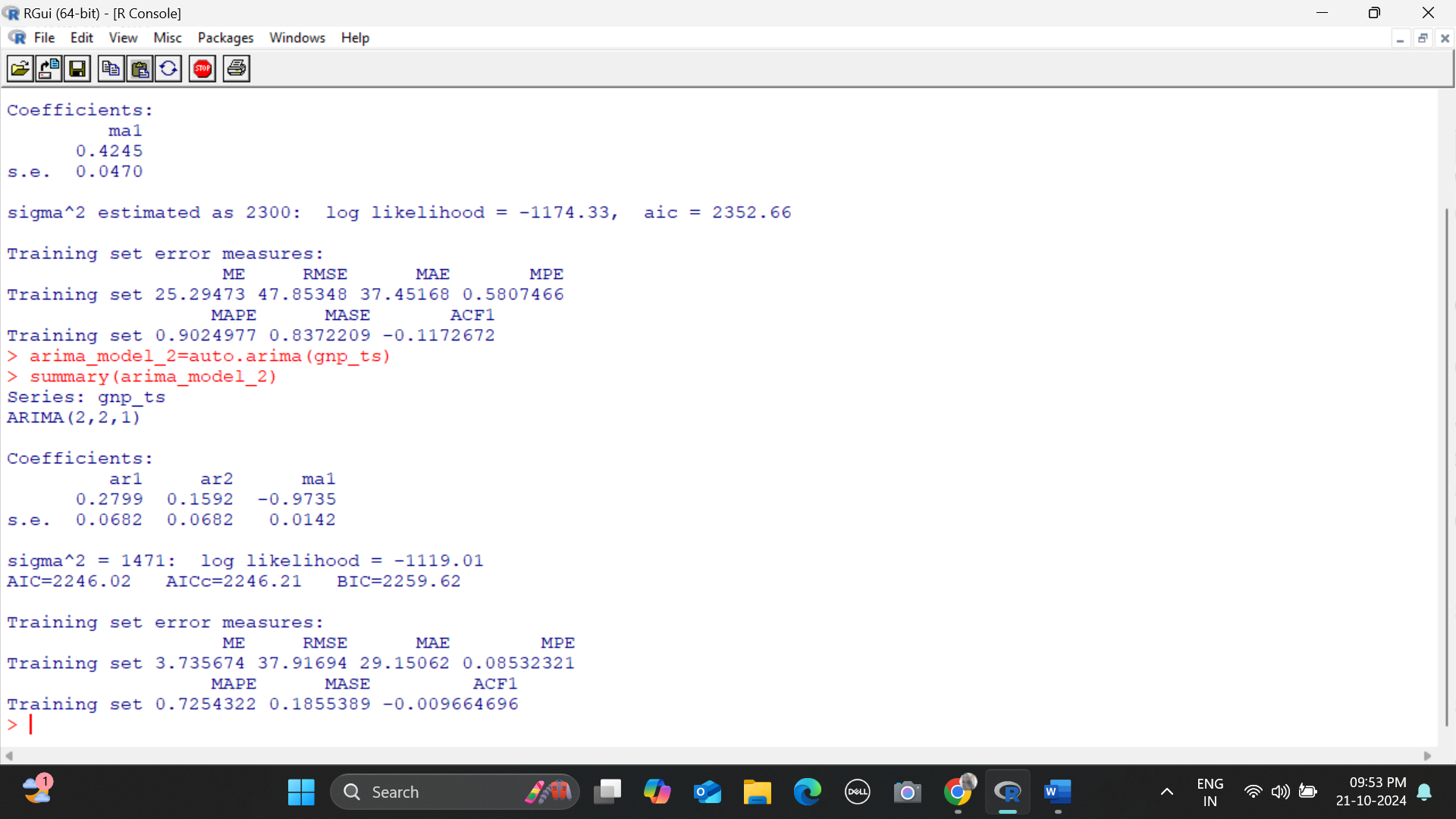
**Using auto arima function , we get the parameters (2,2,1)**

The results are given below:

1. This is for ARIMA model with parameters (0,1,1)



1. This is for ARIMA model with parameters (2,2,1).



Thus, comparing the AIC values for both the models we see that (2,2,1) has lower AIC value than (0,1,1) thus we can say that the former one is a better fit than the latter one.