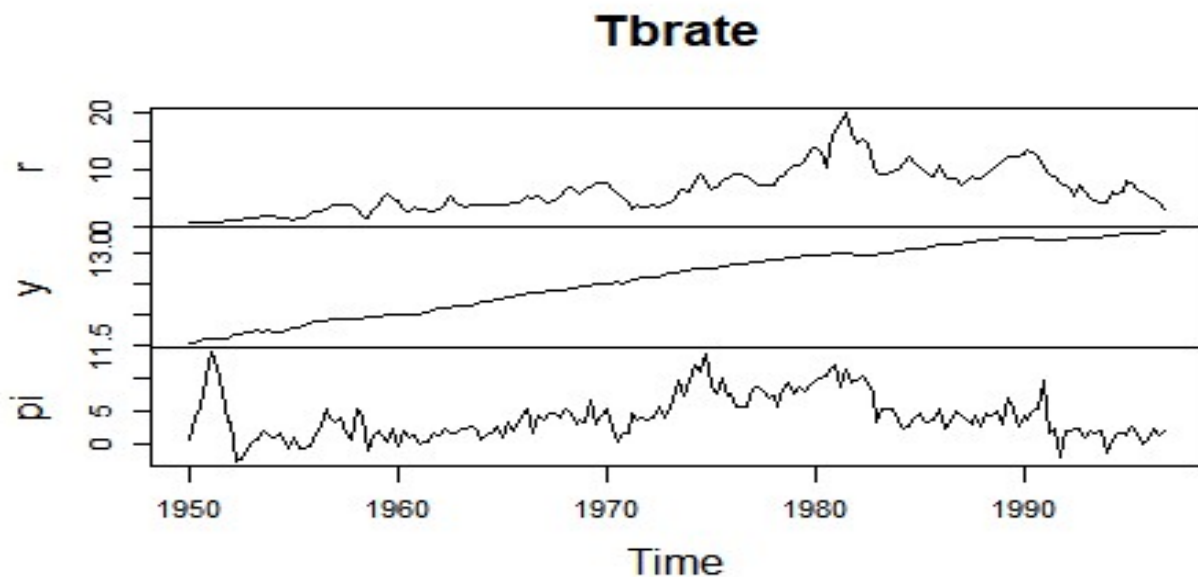# BIA-656 Advanced Data Analytics and Machine Learning

Assignment – 1

1. Describe the signs of nonstationary seen in the time series and ACF plots

Ans:

Tbrate

We use the ADF (Augmented- Dickey-Fuller) Test to check for the stationarity of the time series. After performing the test, we come to conclusion that non-stationarity is seen as there is a unit root

```
ADF: Augmented Dickey-Fuller Test

data:  Tbrate [, 1]
Dickey-Fuller = -1.925, Lag order = 5, p-value = 0.6075
alternative hypothesis: stationary
```

Since the P-value is above the significance level of 0.05 the null hypothesis is not rejected and therefore the time series has unit root which proves the non-stationarity of the time series.

2. Is it necessary to obtain the first or higher difference of the series?

Ans: Yes, it is necessary to obtain higher difference of the series to remove the trend and the seasonal effects in our graphs.
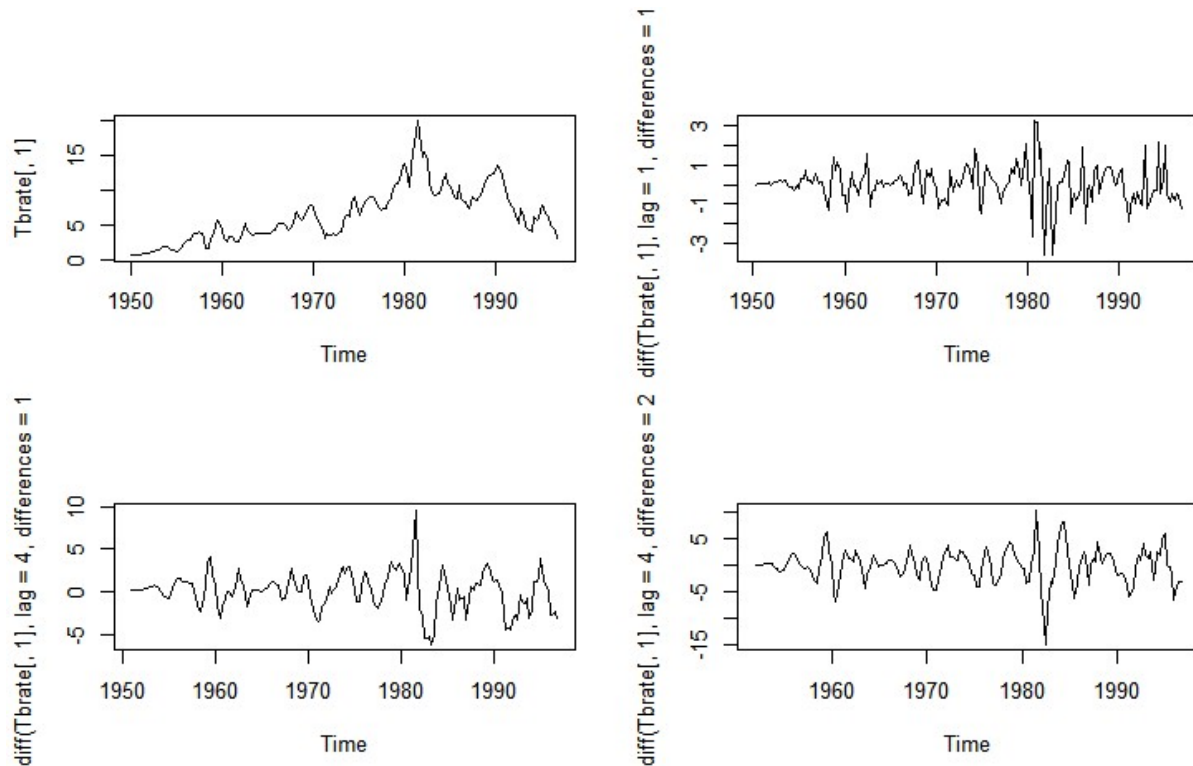
As we can see from the above plots, we used different values for differencing to remove the trend and seasonal values:

1) The 1st graph is before differencing.

2) The 2nd graph has differencing order with lag 1.

3) The 3rd graph has differencing order 1 with lag 4 to show quarterly effects.

4) The 4th graph has differencing order 2 to remove the quadratic trend with lag 4 to show quarterly effects.

3. What order of differencing is chosen?

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

Coefficients:
         ma1    sar1     sma1
      0.3261 0.8281 -0.9516
s.e.  0.0756 0.1750   0.1383

sigma^2 estimated as 0.7831:  log likelihood=-241.85
AIC=491.71    AICc=491.93    BIC=504.63
```

Order 1 of differencing is chosen for the arima plot

4. Does this result agree with your previous conclusions?

Ans: No, this result does not agree with the previous conclusions.

5. What model was chosen by AIC?

Ans: ARIMA (0,1,1) model was chosen by AIC as the value of AIC is less for that model.

6. Which goodness-of-fit criterion is being used here?

Ans:

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

Coefficients:
         ma1    sar1     sma1
      0.3261 0.8281 -0.9516
s.e.  0.0756 0.1750   0.1383

sigma^2 estimated as 0.7831:  log likelihood=-241.85
AIC=491.71    AICc=491.93    BIC=504.63
```

The AIC (Akaike's Information Criterion) is being used as the value of AIC is less than BIC.

7. Change the criterion to BIC. Does the best-fitting model then change?

Ans:

```
Series: Tbrate [, 1]
ARIMA (0,1,1)

Coefficients:
         ma1
      0.3275
s.e.  0.0754

sigma^2 estimated as 0.814:  log likelihood=-245.65
AIC=495.3    AICc=495.37    BIC=501.76
```

No, the best fitting model does not change even after changing the criterion to BIC.

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

Ans: According the Ljung Box test there is no residual autocorrelation

```
        Box-Ljung test

data:  residuals(fit1)
X-squared = 9.925, df = 10, p-value = 0.4471
```

Since the P-value is above the significance level of 0.05 the null hypothesis is not rejected and therefore there is no autocorrelation between the residuals in the model. The spike in the graph may be due to error or some random kind of effect.

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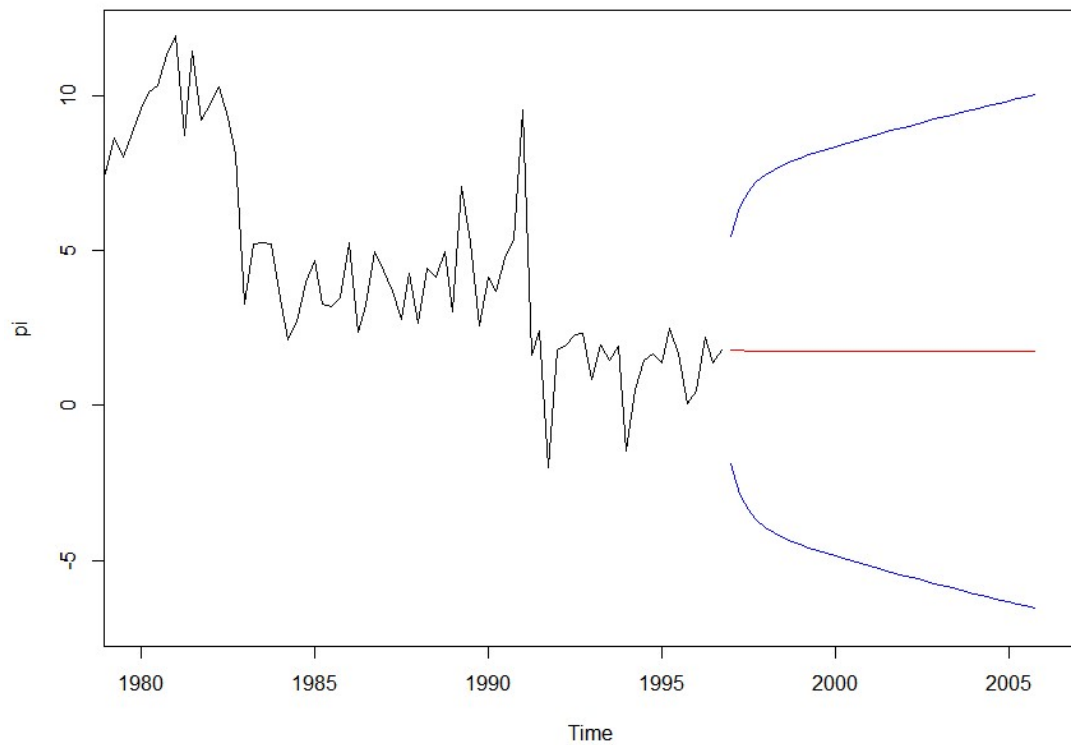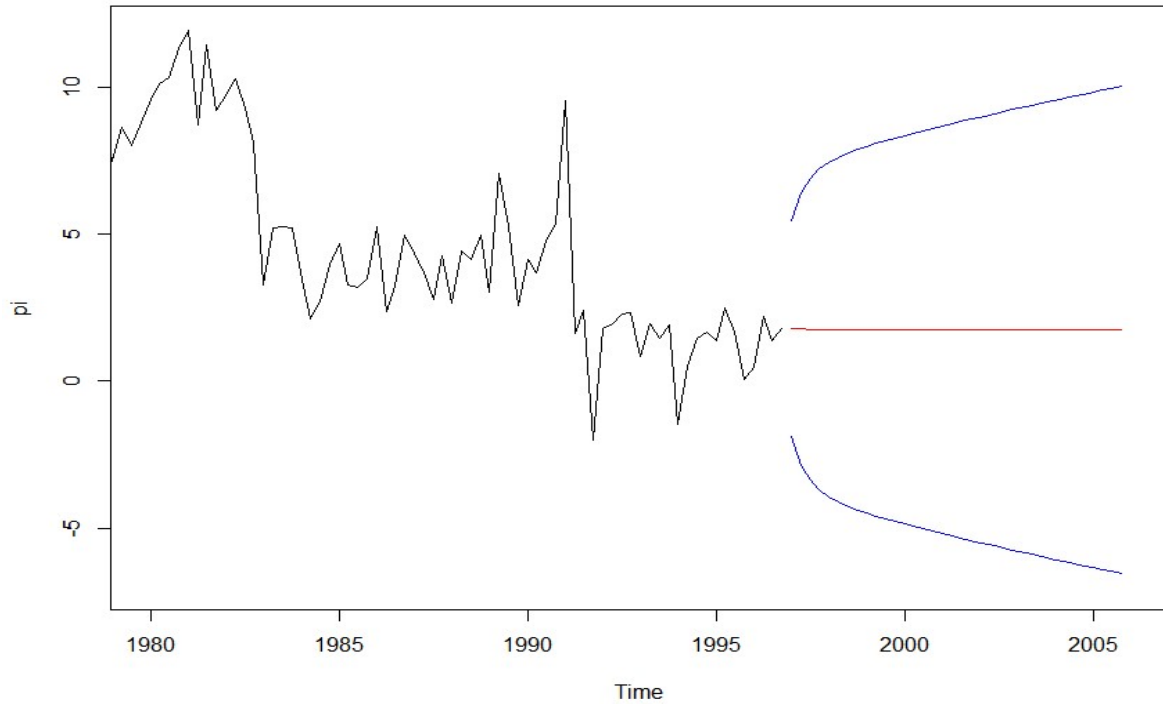

Ans:

The widening of the graph may be due to the forecasting error. As we go forecasting for further periods the forecasting error increases and the shape keeps widening up.

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

Ans:

Further in the forecasting period the forecasting depends on the lags of the previous forecasts but when moving forwards, there comes a time when there is no lag value present for forecasting so it keeps on decreasing until a point when only the average of the forecast is taken as the value for future values. Hence, we see the wiggle.