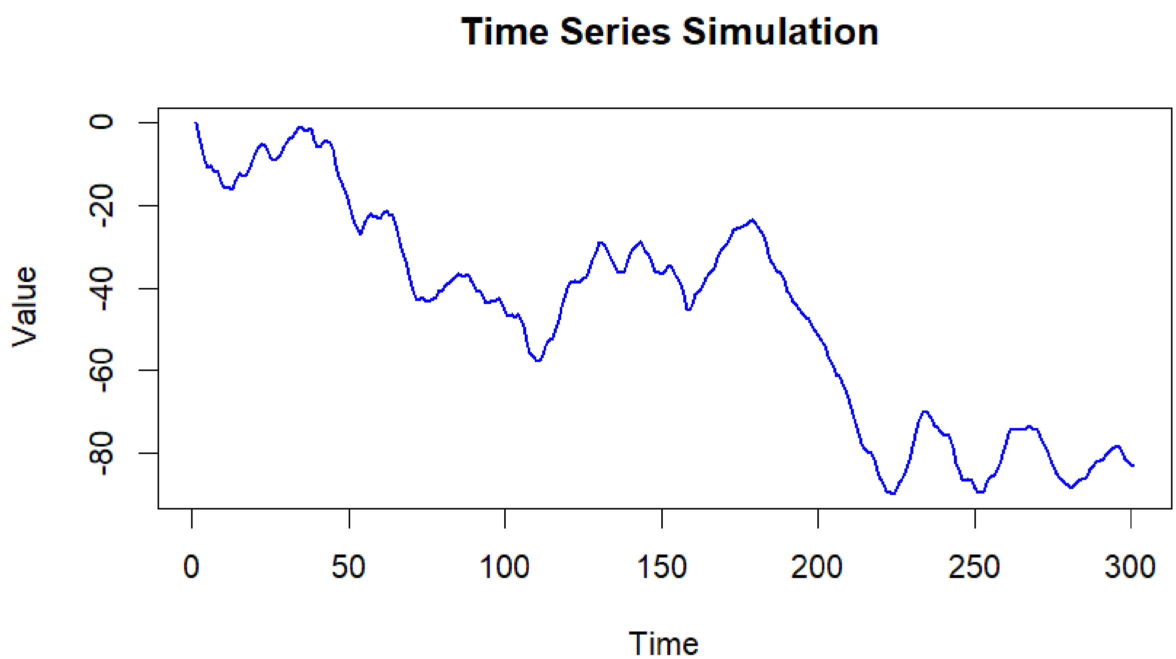
**15794**
**Athukorala W.A.A.D.N.D.**

**3009 – Take Home Assignment**

(i).

**a. Plot the time series giving appropriate labels for each axis and paste the chart into your answer.**

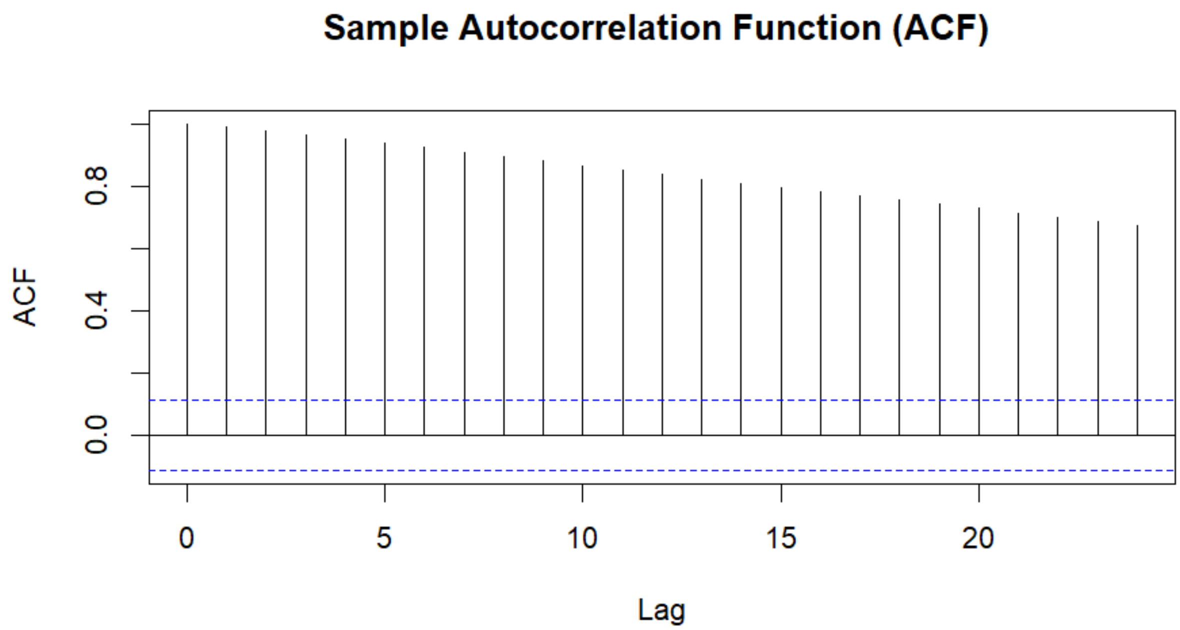


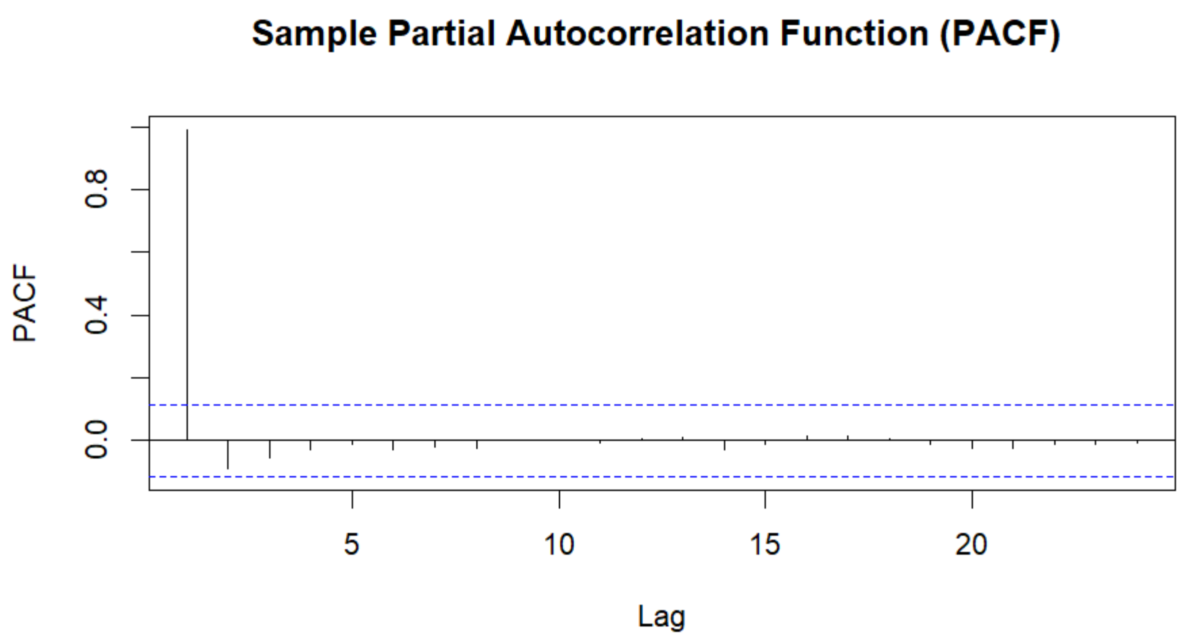

**b. comment on the general features of your chart.**

The overall pattern of the time series shows a downward trend, with values decreasing as time progresses. There don't appear to be any seasonal patterns in the data. However, the plot does exhibit some cyclic behavior, with the series rising and falling over various intervals. Additionally, the time series is not stationary, as it does not have constant variance.

(ii)

**a. Plot the sample Autocorrelation function (ACF) and sample Partial Autocorrelation function (PACF) of the original data, giving appropriate labels for each axis and paste the charts into your answer.**



c)**Comment, by visually inspecting these plots, on the possible modelling strategy which could be adopted**.

The ACF plot shows significant autocorrelation at multiple lags, starting high and gradually decreasing, which suggests the presence of a downward trend in the data. This time series is a non-stationary time series. The slow decay in autocorrelation values indicates that the data may have a unit root, further confirming non-stationarity. This suggests that differencing the data is necessary to remove the trend and achieve stationarity.

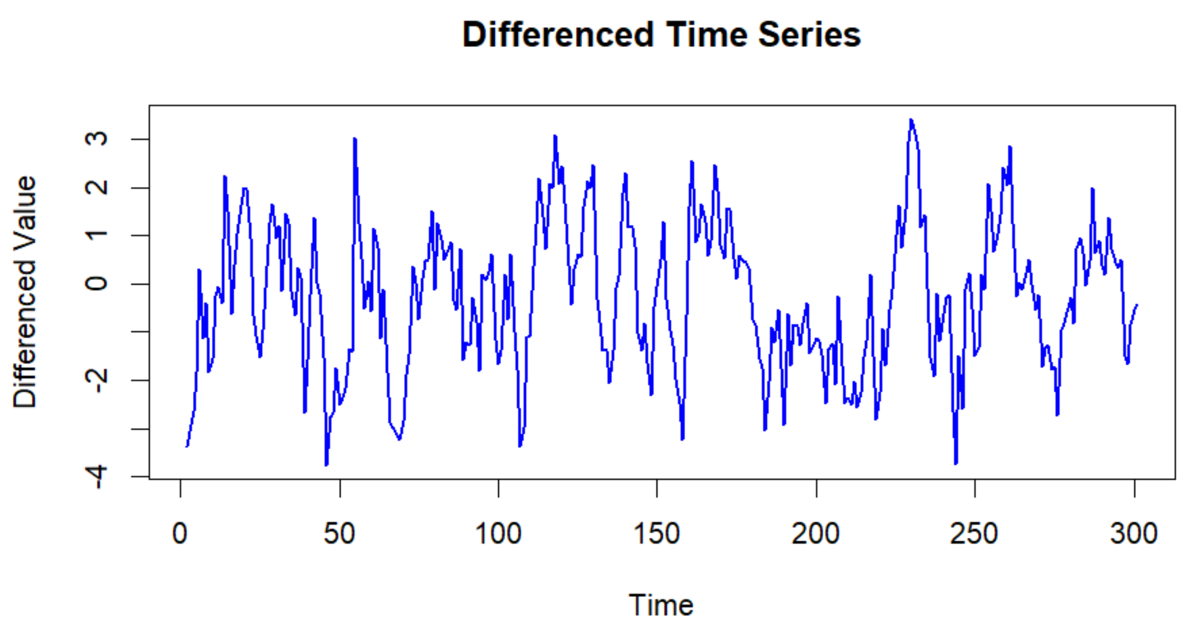
The PACF shows a significant spike at lag 1, which is much larger than the others. This indicates that the first lag has a strong partial autocorrelation with the time series, which is expected in an AR(1) process. The AR(1) model predicts that the current value is primarily influenced by the immediately preceding value.

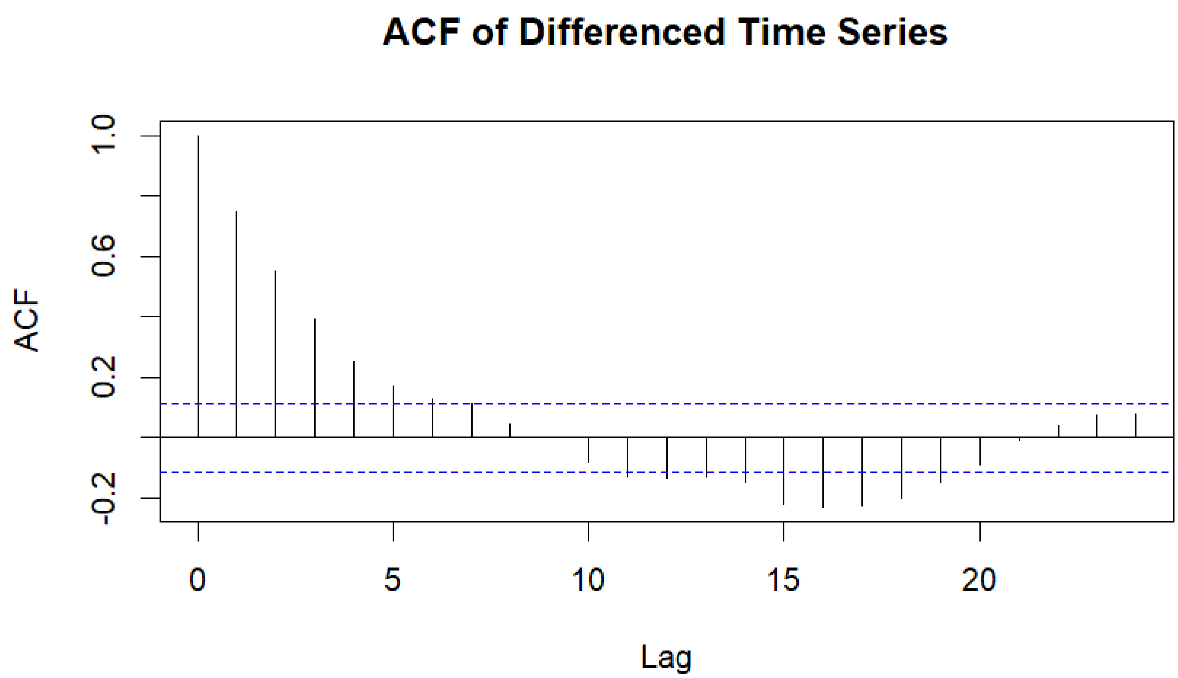
possible modelling strategy

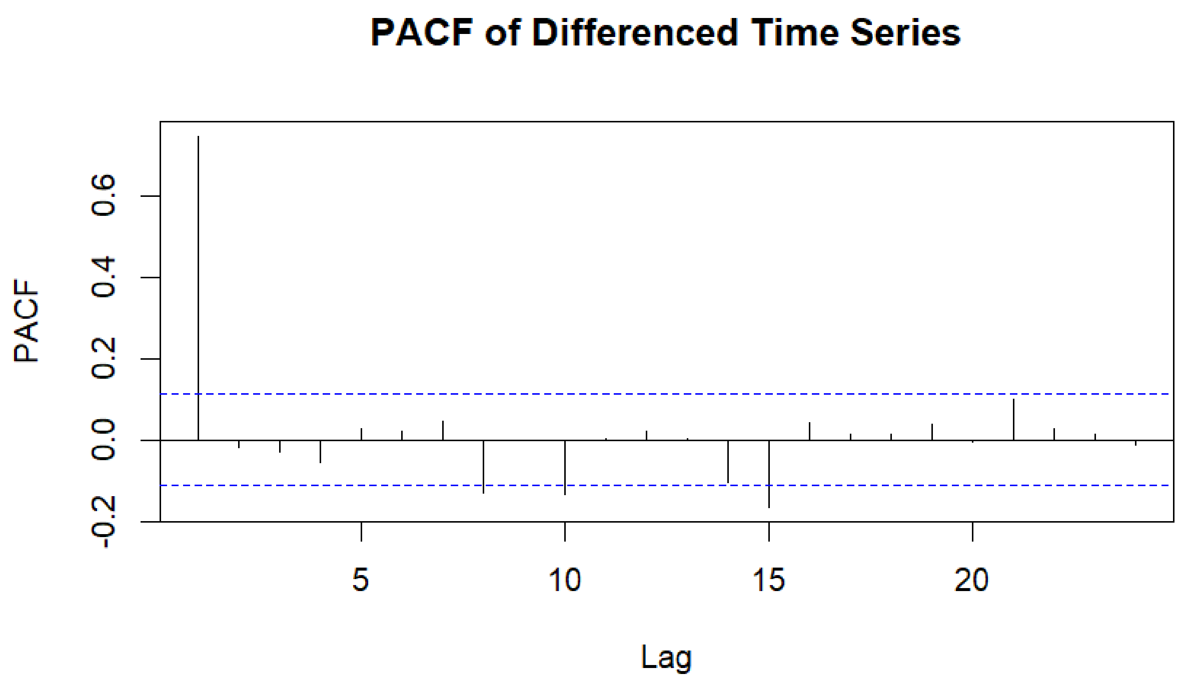
Since the ACF shows a slow decay, applying first differencing needed to remove the trend and make the series stationary. Then test a stationarity with ADF test or KPSS test. After differencing and test stationarity we can adopt ARIMA model to be forecast this time series.

(iii)

a. **Perform an appropriate transformation to the data such that a stationary model is possible, pasting any relevant charts into your answer.**




ACF of Differenced Time Series
ACF
1.0
0.6
0.2
-0.2
0
5
10
15
20
Lag


PACF of Differenced Time Series
PACF
0.6
0.4
0.2
0.0
-0.2
5
10
15
20
Lag

**b. Comment on your answer to part (iii)(a).**

The 1st differencing plot suggests that the differenced series is stationary since there is no obvious trend or seasonal pattern. The values seem to fluctuate around a constant mean (close to zero). The differenced time series does not show any significant outliers, meaning that there are no data points that deviate from the rest of the series.

The ACF plot gradual decline suggests that the differencing has helped remove some of the trends and non-stationarity in the data. . The spike at lag 1 suggests a remaining autoregressive (AR) component, and the quick decline afterward indicates that differencing has achieved stationarity.

The PACF plot shows that after first differencing, the time series exhibits the expected behavior of an AR(1) process, with significant partial autocorrelation only at lag 1 and no significant patterns at higher lags, further confirming the stationarity of the series.

**c.Perform a suitable test to check the stationarity of the data and interpret the results.**

```
        Augmented Dickey-Fuller Test

data:  y_diff
Dickey-Fuller = -4.6349, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

**Null Hypothesis (H0)**: The time series has a unit root, which is non-stationary
**Alternative Hypothesis (H1)**: The time series does not have a unit root, which is stationary.

Here the p-value (0.01) is less than 0.05, so that we can reject the null hypothesis, concluding that the series is stationary.

(iv).

**a. Propose an appropriate model for the transformed data.**

ARIMA(1,1,0) model.

The plot aligns with what we would expect for an ARIMA(1,1,0) model after differencing. The spike at lag 1 suggests a remaining autoregressive (AR) component, and the quick decline afterward indicates that differencing has achieved stationarity.

b. **Write the model equation using backshift operator**.

$$y_t-(1+\phi 1)y_{t-1}+\phi 1 y_{t-2}=\epsilon_t$$

$\phi$ is the autoregressive parameter (in this case, estimated from the data).
$y_t$ is the time series value at time t
$\epsilon_t$ is the white noise error term.

c. **Justify the choice of model in part (iv)(a) by performing an appropriate diagnostic procedure and comparisons with alternative models**.

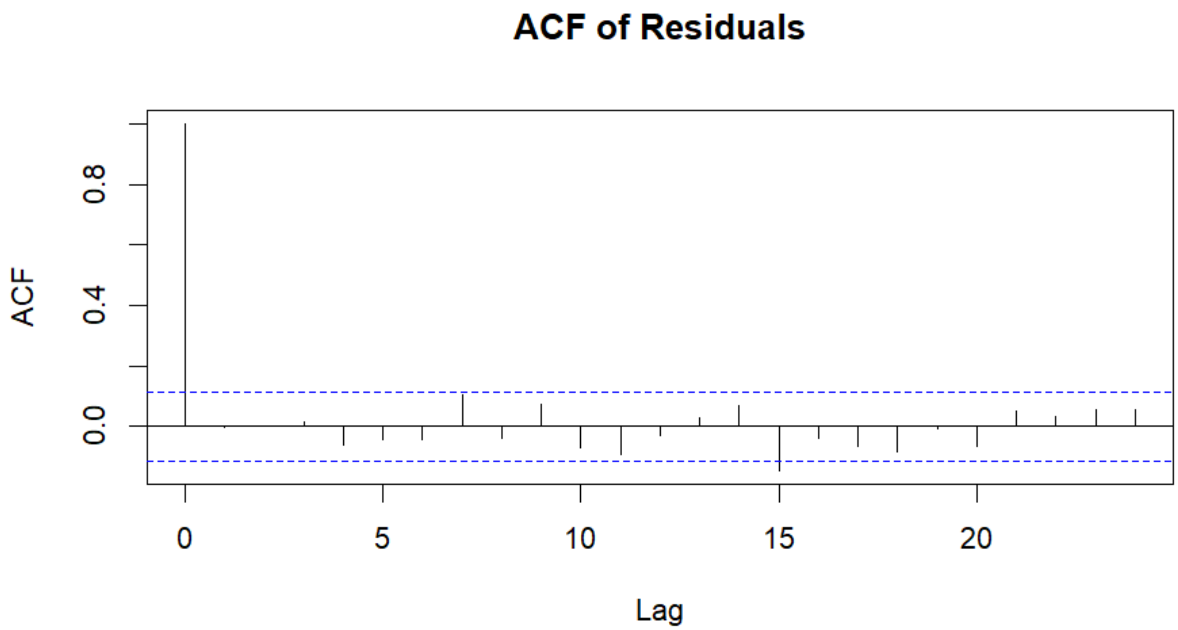
```
>
> cat("AIC for ARIMA(1,1,0):", aic_110, "\n")
AIC for ARIMA(1,1,0): 849.2572
> cat("AIC for ARIMA(1,1,1):", aic_111, "\n")
AIC for ARIMA(1,1,1): 851.2142
> cat("AIC for ARIMA(0,1,1):", aic_011, "\n")
AIC for ARIMA(0,1,1): 943.3026
> cat("AIC for ARIMA(1,1,6):", aic_116, "\n")
AIC for ARIMA(1,1,6): 858.1058
>
> cat("BIC for ARIMA(1,1,0):", bic_110, "\n")
BIC for ARIMA(1,1,0): 856.6647
> cat("BIC for ARIMA(1,1,1):", bic_111, "\n")
BIC for ARIMA(1,1,1): 862.3256
> cat("BIC for ARIMA(0,1,1):", bic_011, "\n")
BIC for ARIMA(0,1,1): 950.7102
> cat("BIC for ARIMA(1,16):", bic_116, "\n")
BIC for ARIMA(1,16): 887.7361
>
```

**AIC-** A lower AIC value generally indicates a better model

- ARIMA(1,1,0) has the lowest AIC (849.2572), making it the best model according to this criterion.

**BIC**- Lower BIC values are preferable.

- ARIMA(1,1,0) also has the lowest BIC (856.6647), further supporting it as the preferred model.




The ACF of residuals should ideally fall within the blue confidence intervals (typically at 95% confidence). Most of the ACF values for lags greater than 0 fall within these confidence intervals, indicating that the residuals have no significant autocorrelation. This is a good sign, as it suggests that the model has captured the underlying pattern in the data well.

```
        Box-Ljung test

data:  residuals(model)
X-squared = 26.588, df = 20, p-value = 0.1473
```

**Null Hypothesis (H0)**: The residuals are independently distributed, meaning there are no significant autocorrelations at the specified lags

**Alternative Hypothesis (H1)**: The residuals exhibit significant autocorrelations at one or more of the specified lags, indicating that the residuals are not independent.

Since the p-value (0.1473) is greater than the typical significance level (0.05), null hypothesis cannot be rejected. This suggests that there is no significant evidence of autocorrelation in the residuals of the model and hence follow a white noise process.

**d. Get the forecast for next 6 months from the fitted model.**

```
> print(forecast_result)
    Point Forecast      Lo 80     Hi 80     Lo 95     Hi 95
302      -83.41310 -84.67981 -82.14638 -85.35037 -81.47583
303      -83.66484 -86.23498 -81.09471 -87.59553 -79.73416
304      -83.85754 -87.79159 -79.92349 -89.87415 -77.84093
305      -84.00504 -89.30169 -78.70838 -92.10557 -75.90450
306      -84.11793 -90.74406 -77.49181 -94.25173 -73.98414
307      -84.20435 -92.11116 -76.29754 -96.29677 -72.11193
>
```

**e. Draw the actual and forecast in the same plot.**