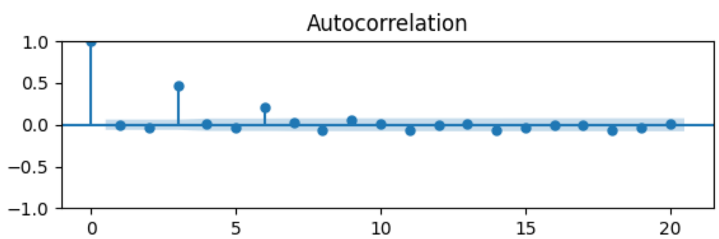
Time Series Homework5

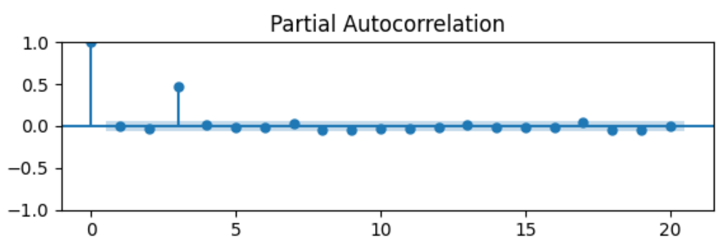
Example 1:

Step2:

ARIMA$(1,0,0)_3$ = y(t) – 0.5y(t-3) = e(t)

Step 3 and 4:

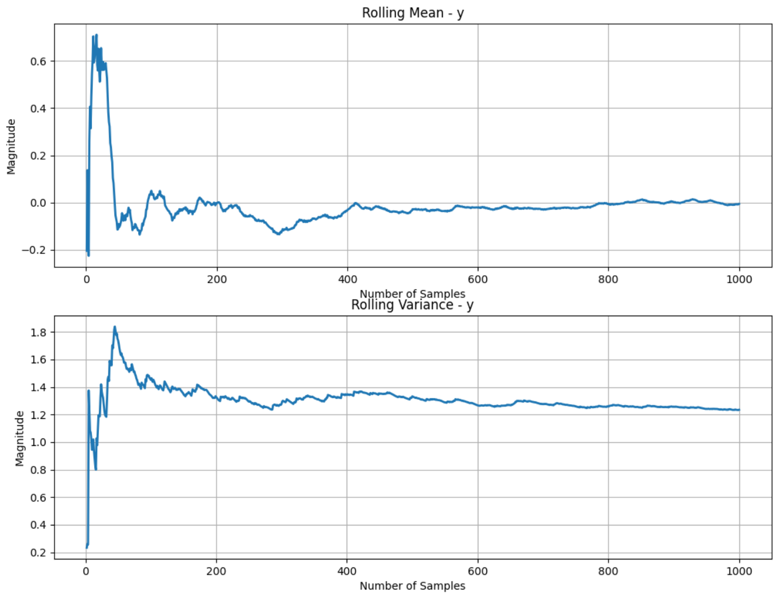




Step 5:

ADF Statistic: -11.095933
p-value: 0.000000
Critical Values:

    1%: -3.437
    5%: -2.864
    10%: -2.568



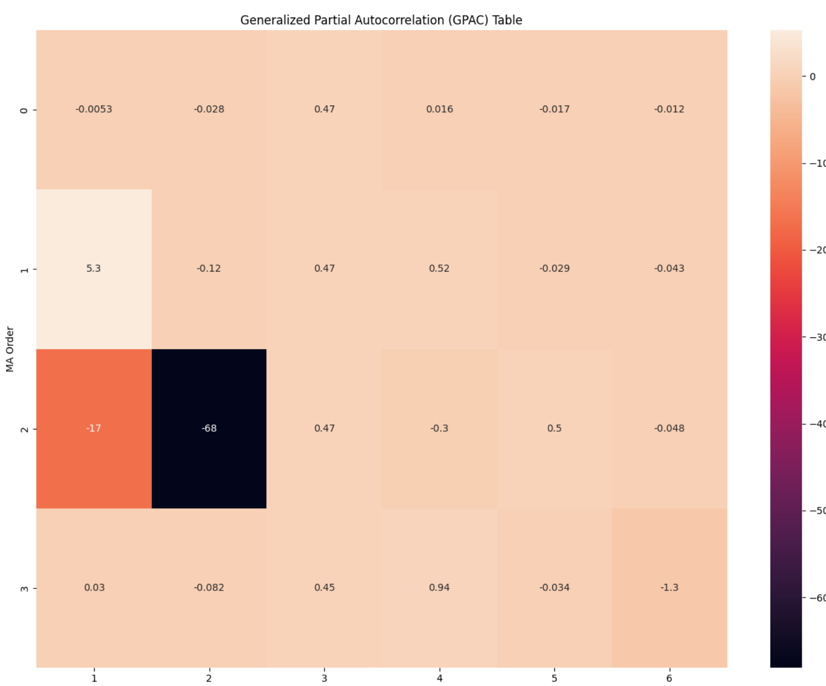
Raw dataset is stationary.

Step 6 and 7:

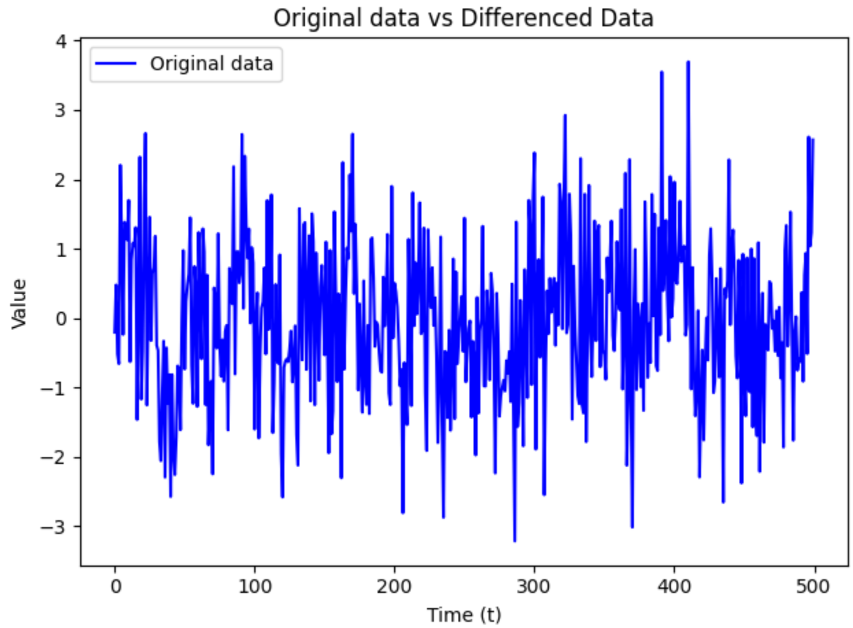
Based on the ADF-test, **the raw data is found to be stationary**. Hence, these two steps are skipped.

Step 8:

GPAC of the raw data



Step 9:

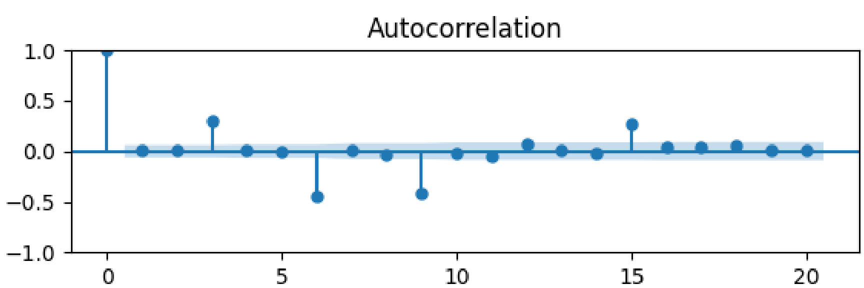


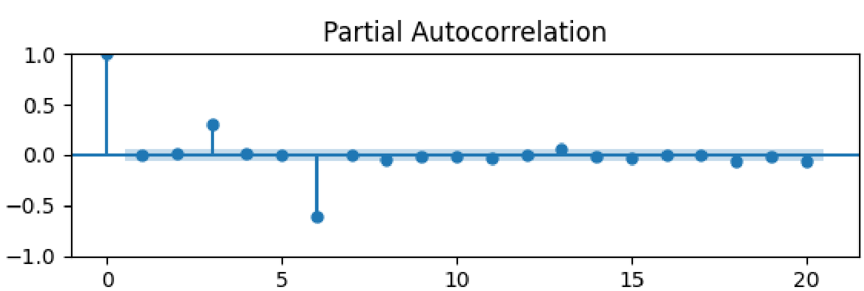
Since the raw data was found to be stationary, no differencing was done. The title differenced data is a static one, and can be ignored safely.

Question 2:

ARIMA(2,0,0)$_3$ = y(t) – 0.5y(t-3) + 0.6y(t-6) = e(t)

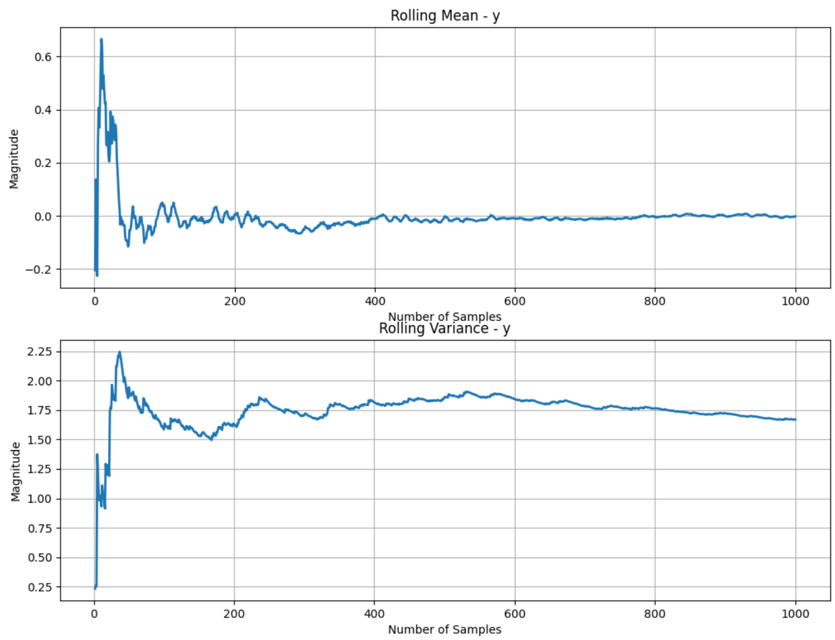
Step 3 and 4:





Step 5:

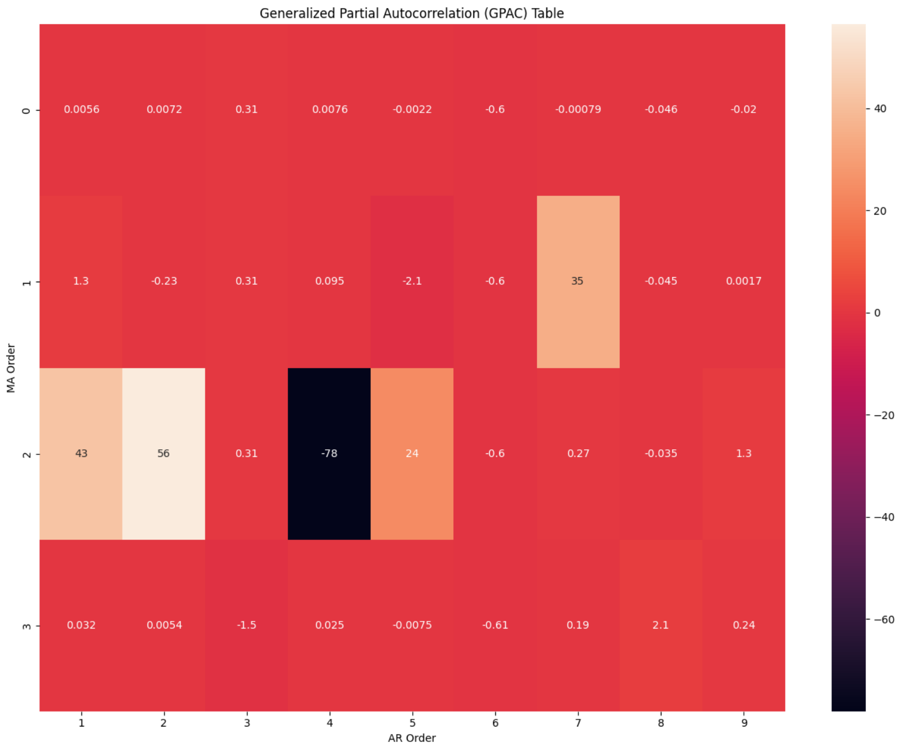
ADF test on raw dataset:
ADF Statistic: -21.387126
p-value: 0.000000
Critical Values:
    1%: -3.437
    5%: -2.864
    10%: -2.568



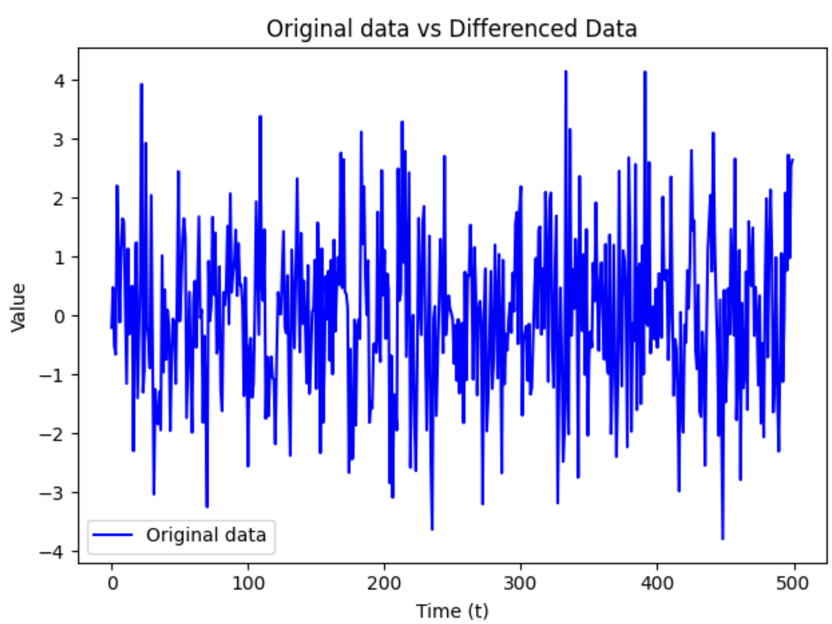
Step 6 and 7:

Based on the ADF-test, the **raw data is found to be stationary**. Hence, these two steps are skipped.

Step 8:
GPAC of the raw data



Step 9:



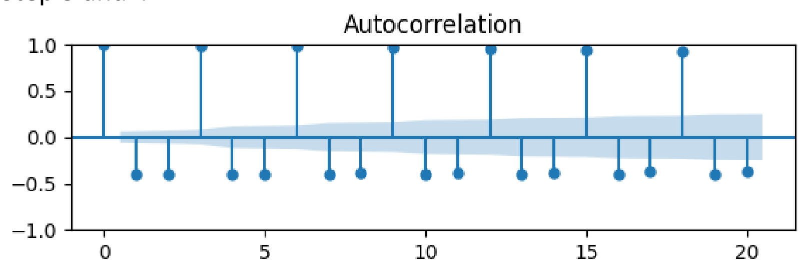
Since the raw data was found to be stationary, no differencing was done. The title differenced data is a static one, and can be ignored safely.

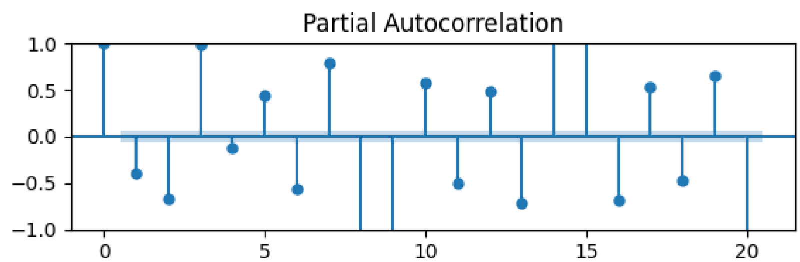
Example 3:

Step 2

ARIMA(1,1,0)$_3$ = y(t) – 1.5y(t-3) + 0.5y(t-6) = e(t)

Step 3 and 4





Step 5

ADF test on raw dataset:

ADF Statistic: -2.759749

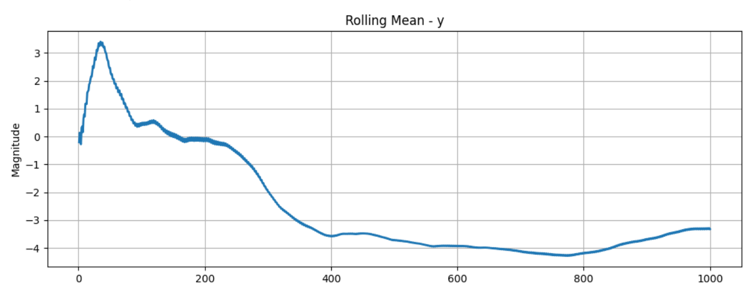
p-value: 0.064257

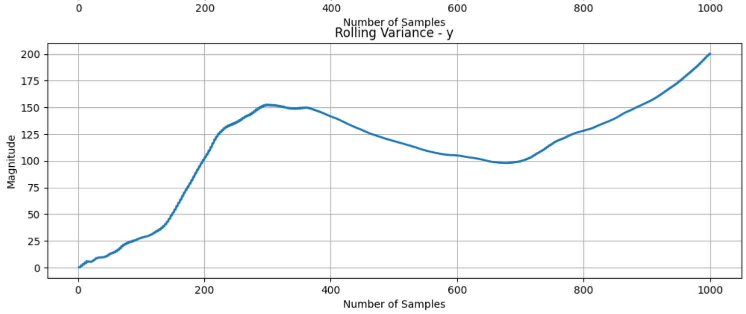
Critical Values:

1%: -3.437

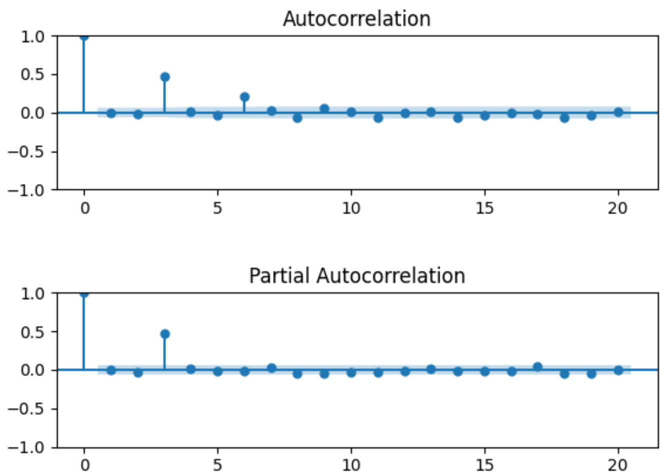
5%: -2.864

10%: -2.568





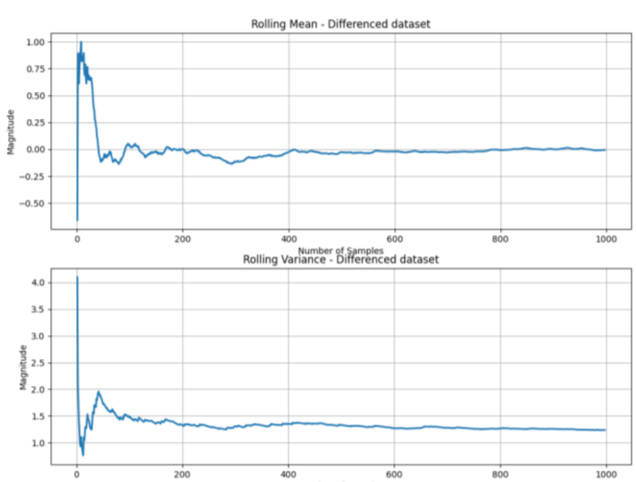
The raw dataset is not stationary.
Step 6



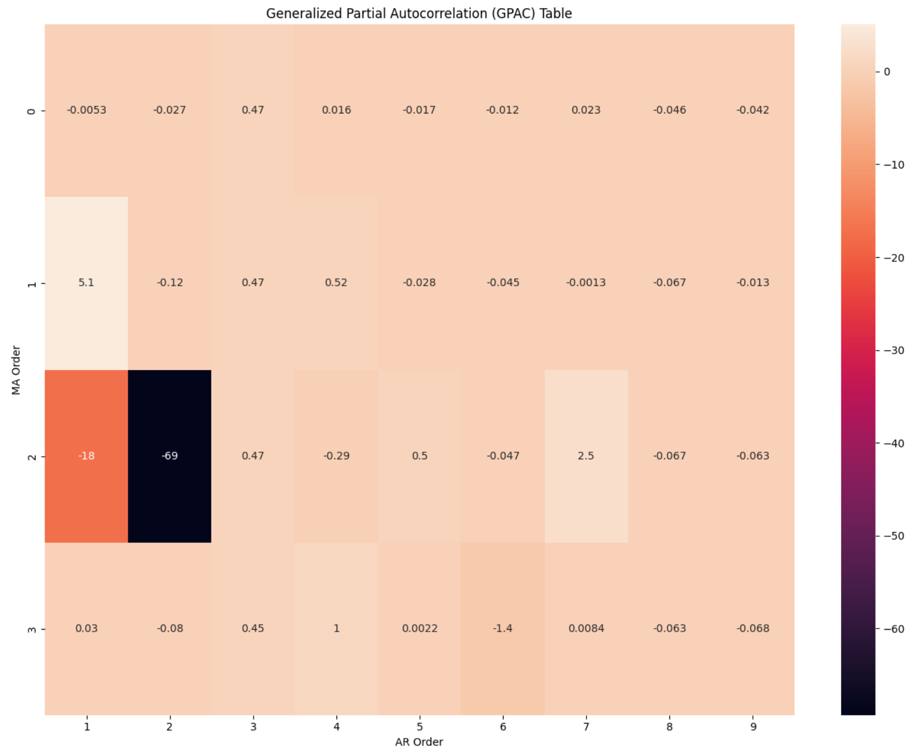
Once the data is differenced, the ACF and PACF shows what is pending, in the sense, order that is not differenced yet.

Step 7
ADF on the differenced dataset
ADF Statistic: -11.115992
p-value: 0.000000
Critical Values:
- 1%: -3.437
- 5%: -2.864
- 10%: -2.568

The dataset is stationary given the p-value is less than 0.01 for 99% confidence level, and 0.05 for 95% confidence level.

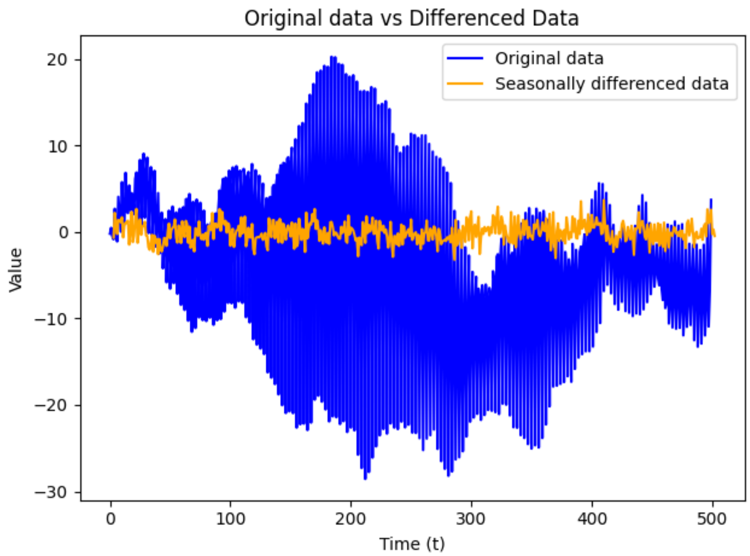


Step 8



From the GPAC table above, I can see there is a column (column=3) of constants with a row of zeros adjacent to it. Hence the order of AR is Na = na/seasonality_index = 3/3 = 1, and MA is 0. Here, na, which is the non-seasonal AR order, is 3 and is retrieved from the GPAC. Since it took 1 seasonally differencing (whose index retrieved from the coefficients) to gain stationarity, then the discovered order of the process is ARIMA$(1,1,0)_3$.

Step 9



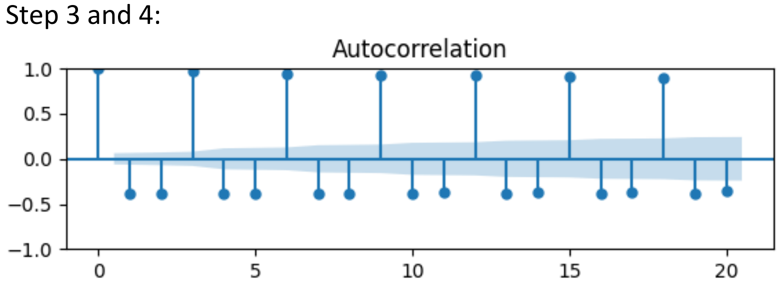
Example 4:

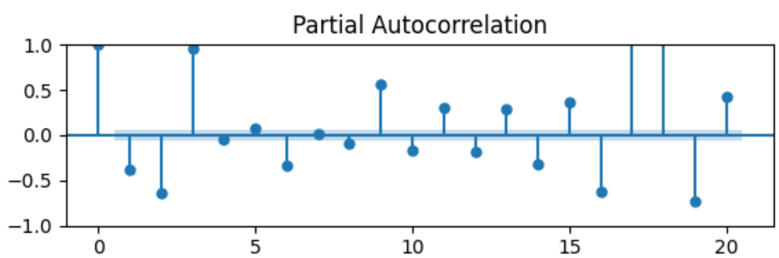
ARIMA$(2,1,0)_3$ =

Step 2:

$y(t) - 1.5y(t-3) + 1.1y(t-6) - 0.6y(t-9) = e(t)$

Step 3 and 4:





Step 5:

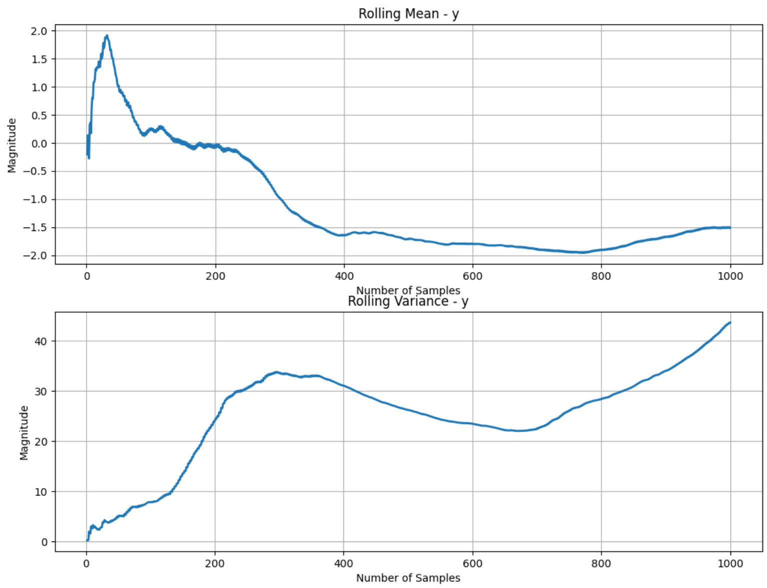
ADF test on raw dataset:

ADF Statistic: -2.993179

p-value: 0.035536

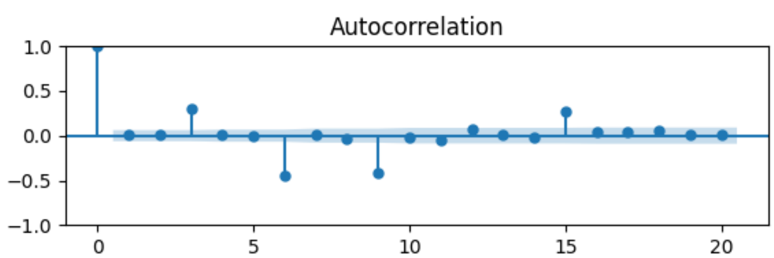
Critical Values:

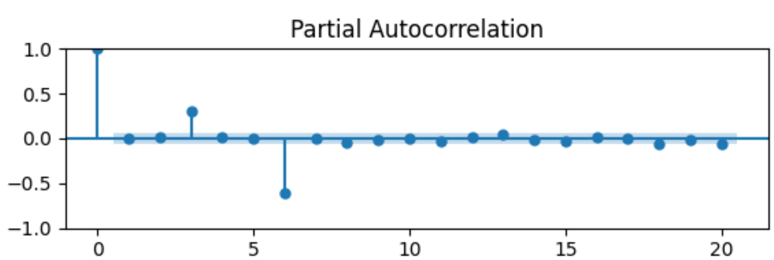
1%: -3.437
5%: -2.864
10%: -2.568



The raw data is not stationary.

Step 6:





Once the data is differenced, the ACF and PACF shows what is pending, in the sense, order that is not differenced yet.

Step 7

ADF on the differenced dataset

ADF Statistic: -21.373656

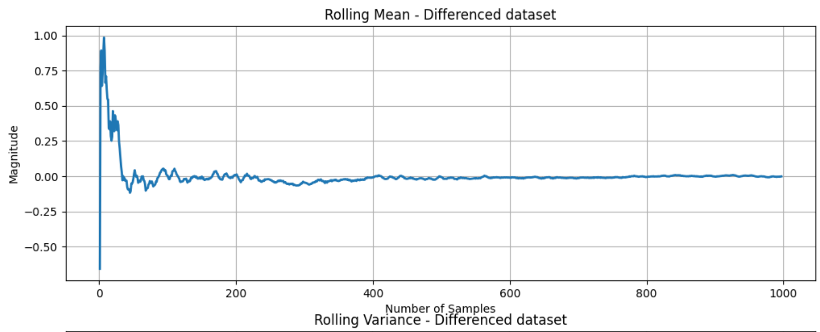
p-value: 0.000000

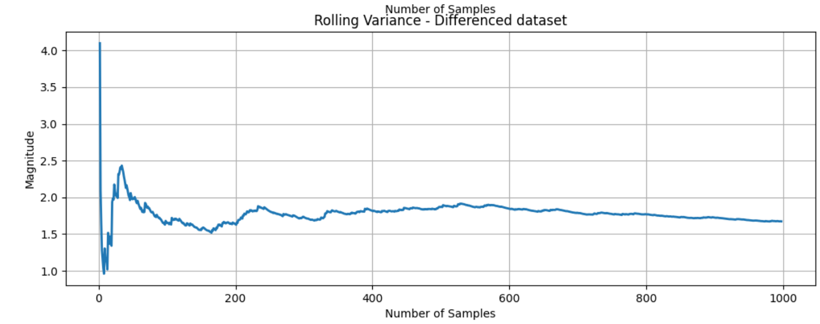
Critical Values:

 1%: -3.437

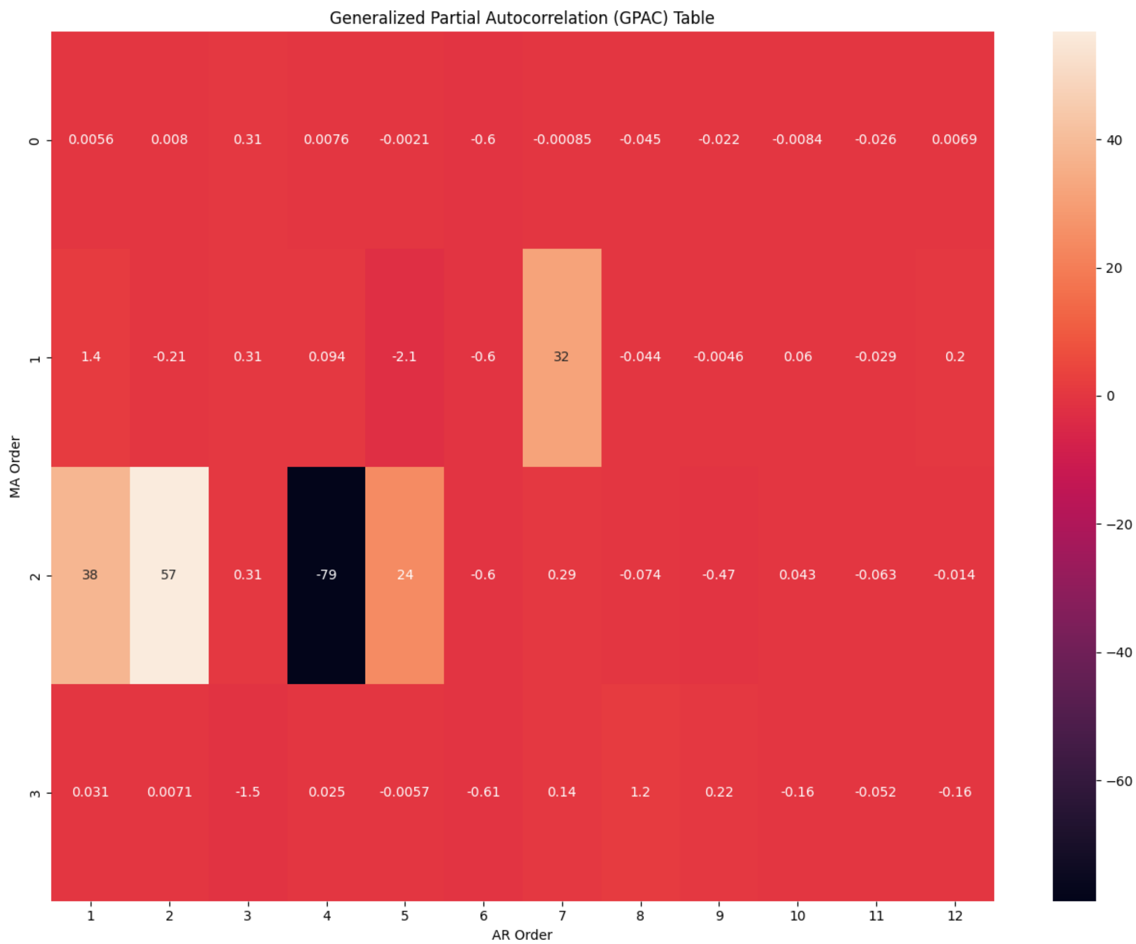
 5%: -2.864

 10%: -2.568



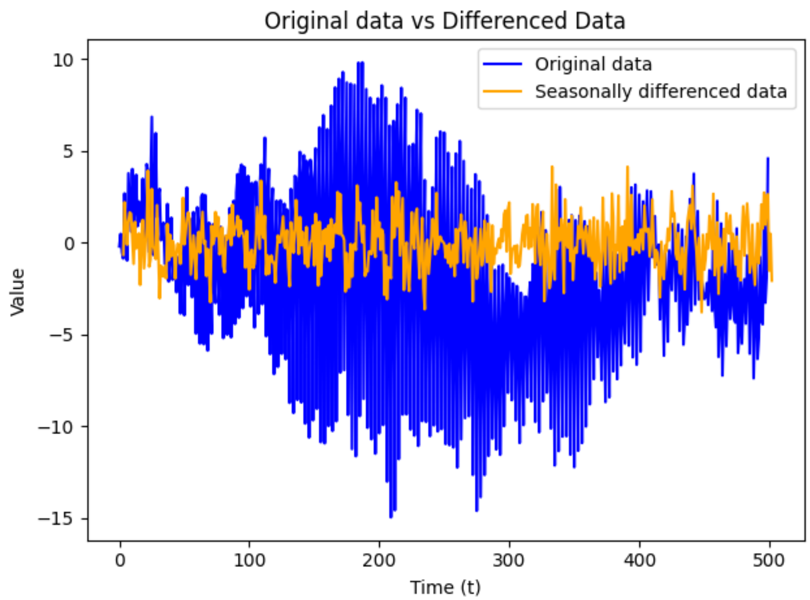


Step 8



For this example, the program did 1 seasonal differencing, whose index is 3. From the GPAC table given above, it can be seen that there is a column (6) of constants (~-0.6) placed adjacent to a row of approximately zeros. Hence, the na is 6, and the Na = na/s = 6/3 = 2. Hence the discovered order is ARIMA(2,1,0)$_3$.

Step 9



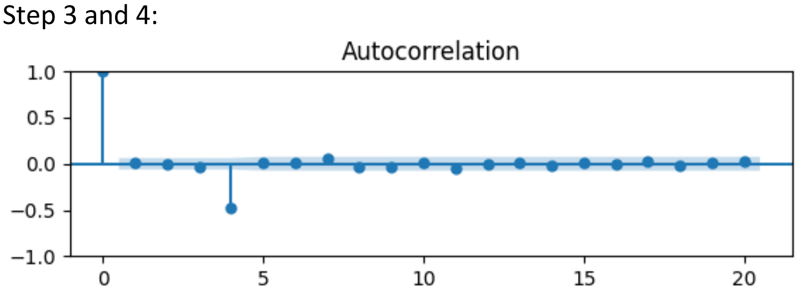
Example 5

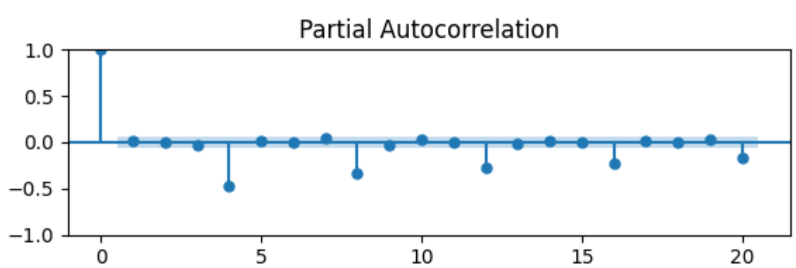
ARIMA$(0,0,1)_4$

Step 2

$y(t) = e(t) – 0.9\ e(t-4)$

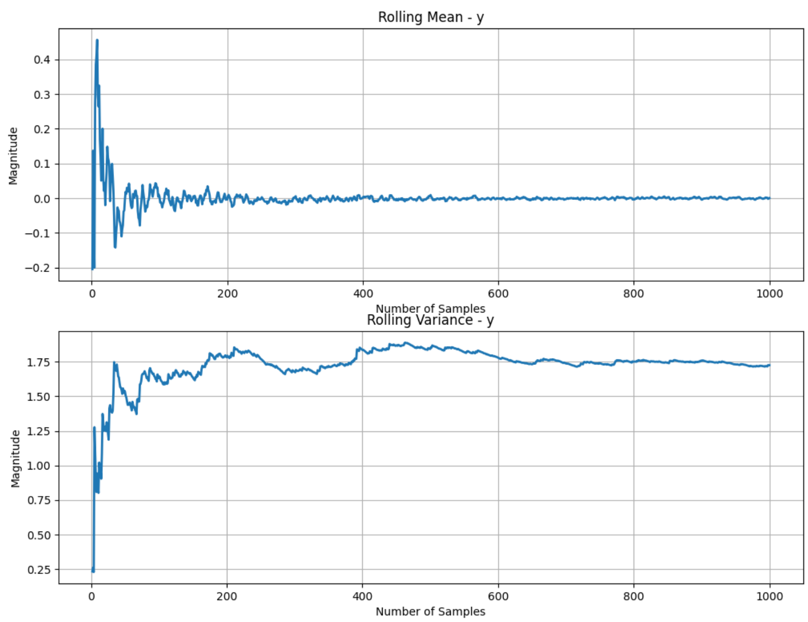
Step 3 and 4:





Step 5

ADF test on raw dataset:
ADF Statistic: -10.723059
p-value: 0.000000
Critical Values:
1%: -3.437
5%: -2.864
10%: -2.568



The raw data is stationary.

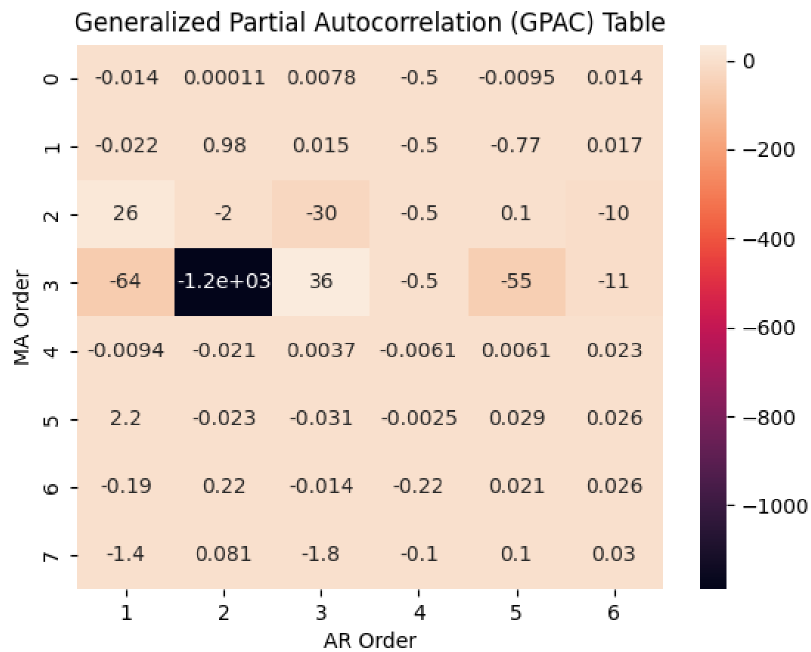
Step 6

The generated data is stationary, and hence, this step is skipped.

Step 7

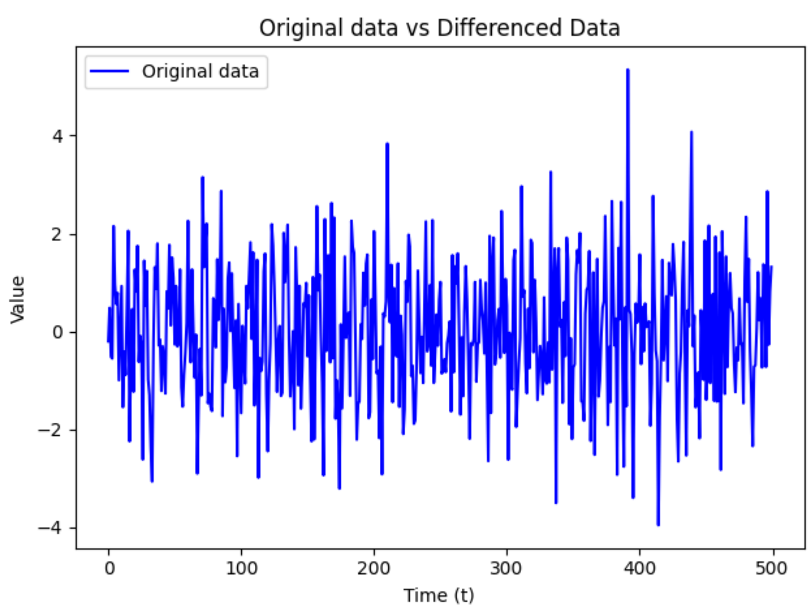
The generated data is stationary, and hence, this step is skipped.

Step 8



Here the discovered order is Na = na/seasonality_index = 0, and Nb = nb/seasonality_index = 4/4 = 1. The written python code was able to detect the seasonality index from the ARMA coefficients and from the above GPAC table, it can be observed that there is a row (4) of approximately zeros that indicate the order of MA process (Nb). Hence the process is ARIMA(0,0,1)$_4$.

Step 9

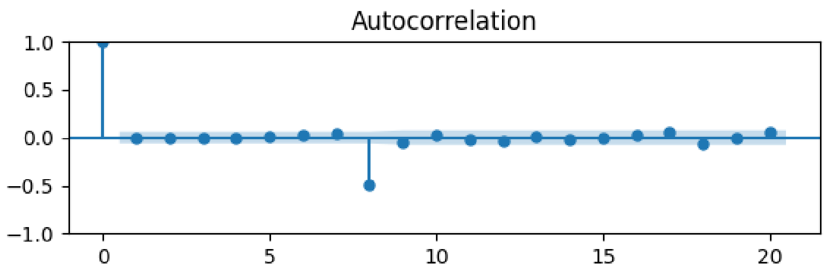


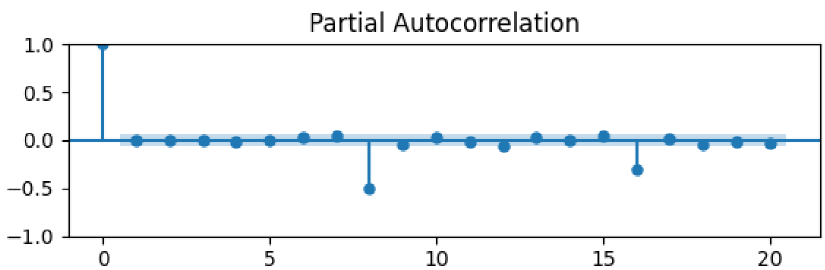
Example 6

ARIMA(0,0,2)$_4$

Step 2
y(t) = e(t) – 0.2 e(t-4) - 0.8 e(t-8)

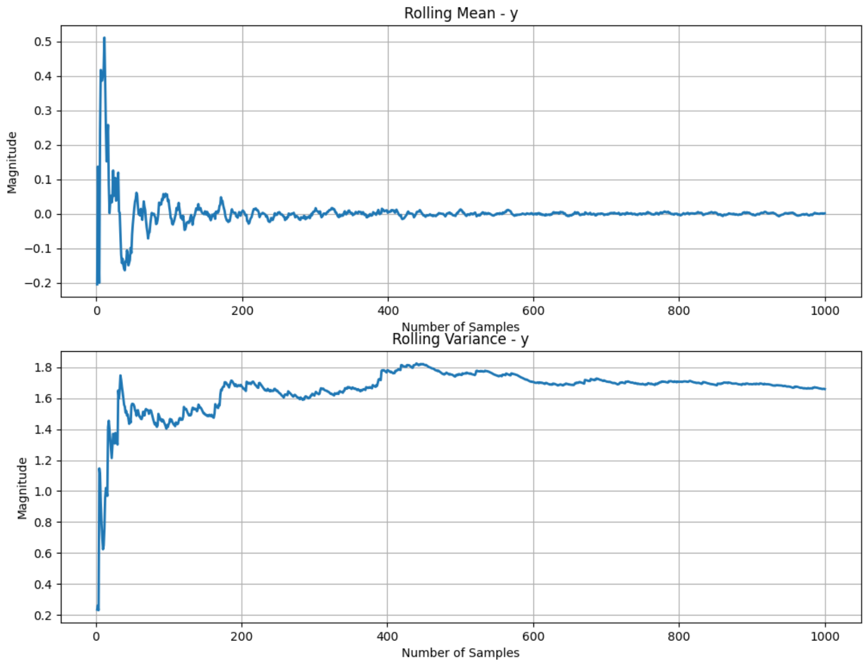
Step 3 and 4





Step 5
ADF test on raw dataset:
ADF Statistic: -12.889068
p-value: 0.000000
Critical Values:

1%: -3.437
5%: -2.864
10%: -2.568

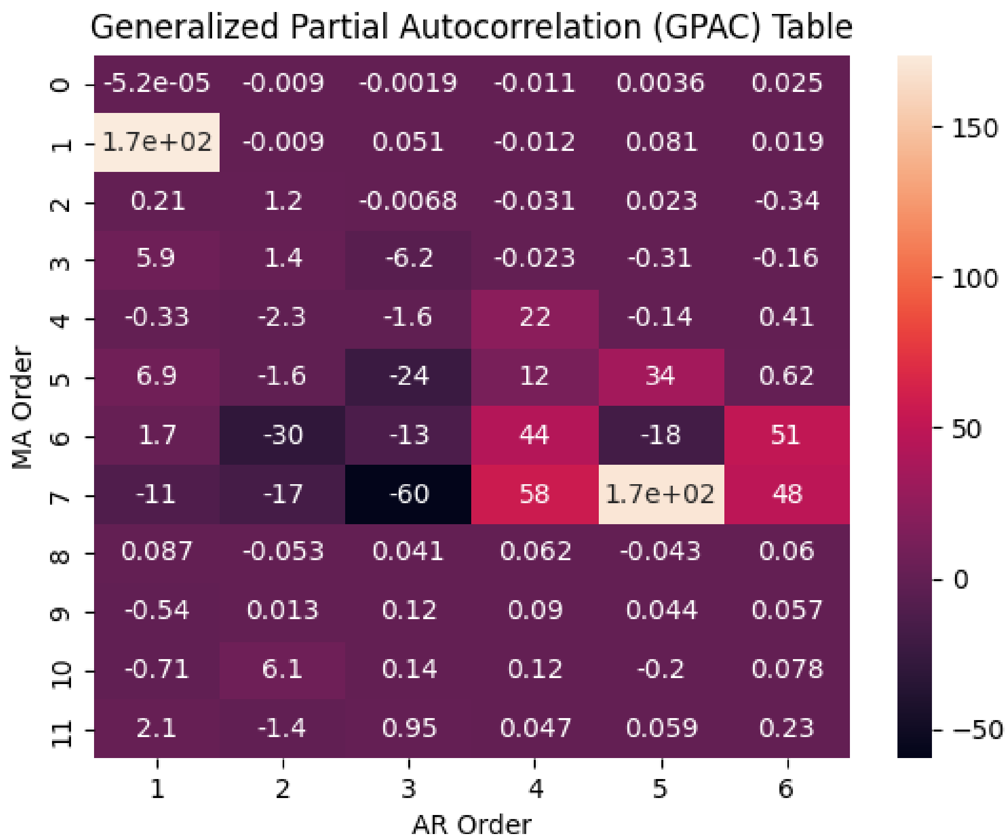


The raw data is stationary.

Steps 6 and 7

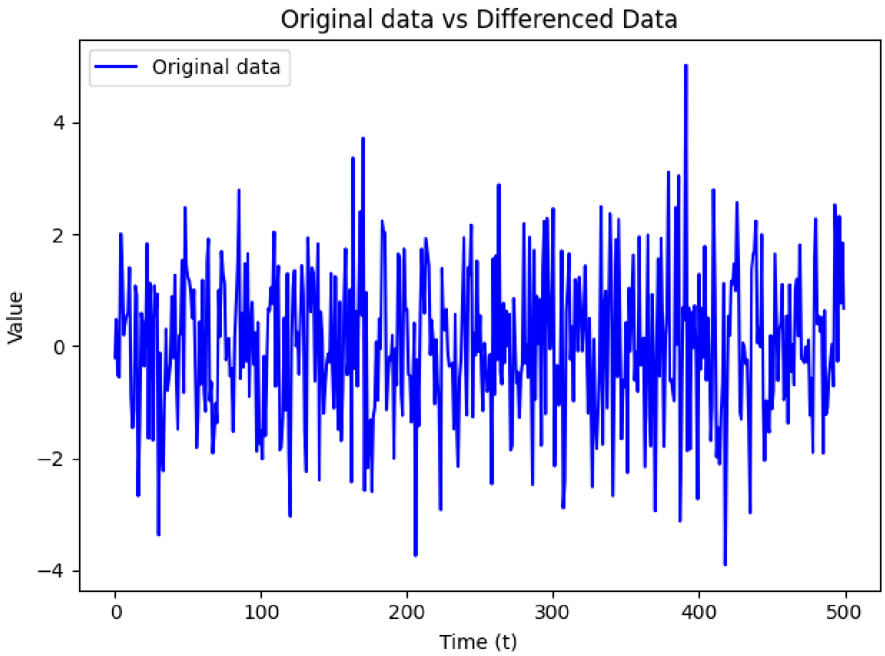
These steps are skipped since **the raw data is found to be stationary**.

Step 8



To be honest here, there are three rows whose cells are populated with values approximately zero. Row -0, 1, and 8. If we observe with known information, we may go with MA order nb=8, but being objective, it is difficult to infer the nb only using the GPAC. Given this is a pure MA process, I believe it would make more sense to make use of the ACF plot to infer the order. With real datasets, I would use all three plots – ACF, PACF, and the GPAC to conclude the order of the process that generated the data.

Step 9



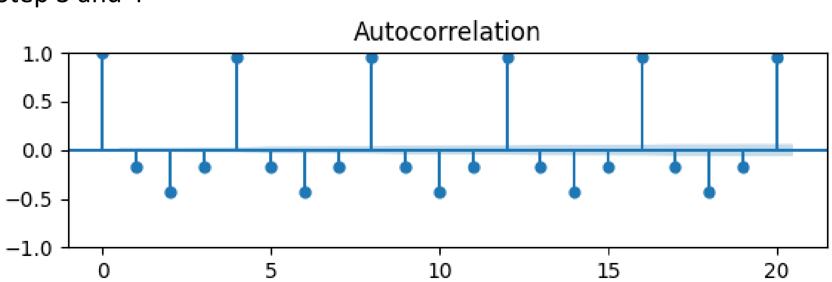
Example 7

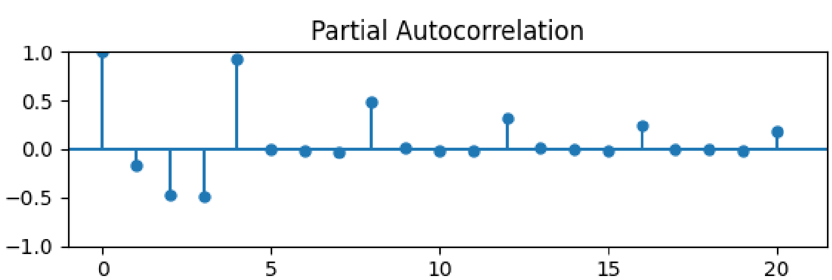
ARIMA$(0,1,1)_4$

Step 2

y(t) – y(t-4) = e(t) – 0.9 e(t-4)

Step 3 and 4

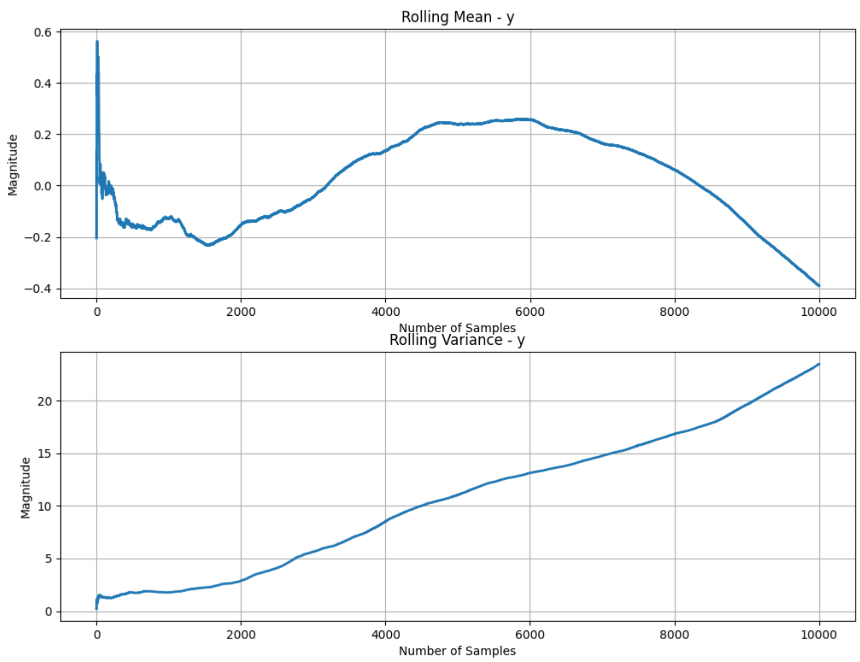




Step 5

ADF test on raw dataset:
ADF Statistic: -2.075303
p-value: 0.254560
Critical Values:

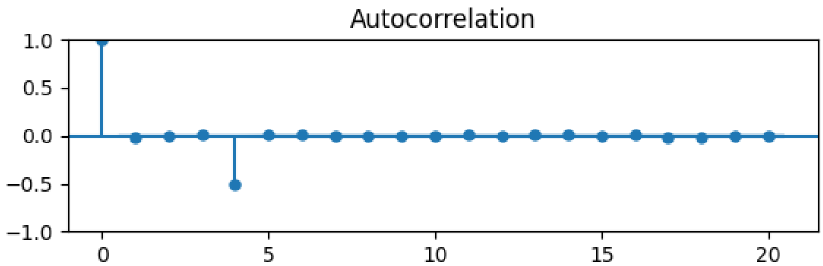
1%: -3.431
5%: -2.862
10%: -2.567

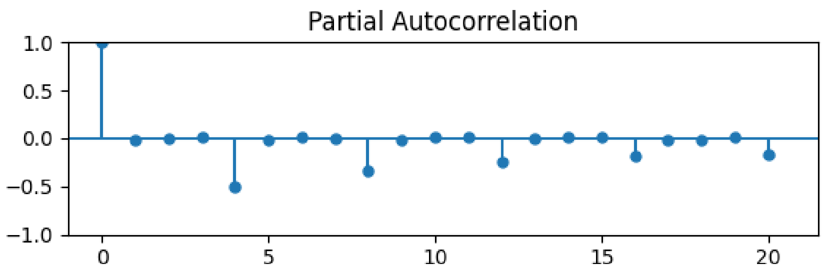


Note: I had to slightly increase the number of samples from 1000 to 10000 (a number from one of our previous labs/homework) to enable the computation be more accurate. It is observed that when the number of samples is lesser, much of the results become inaccurate leading to wrong conclusions. Increasing the T value allowed the ADF-test to return a p-value that is greater then 0.01 for 99% confidence level. Hence, respective differencing will be done to make it stationary.

The raw data is not stationary.

Step 6



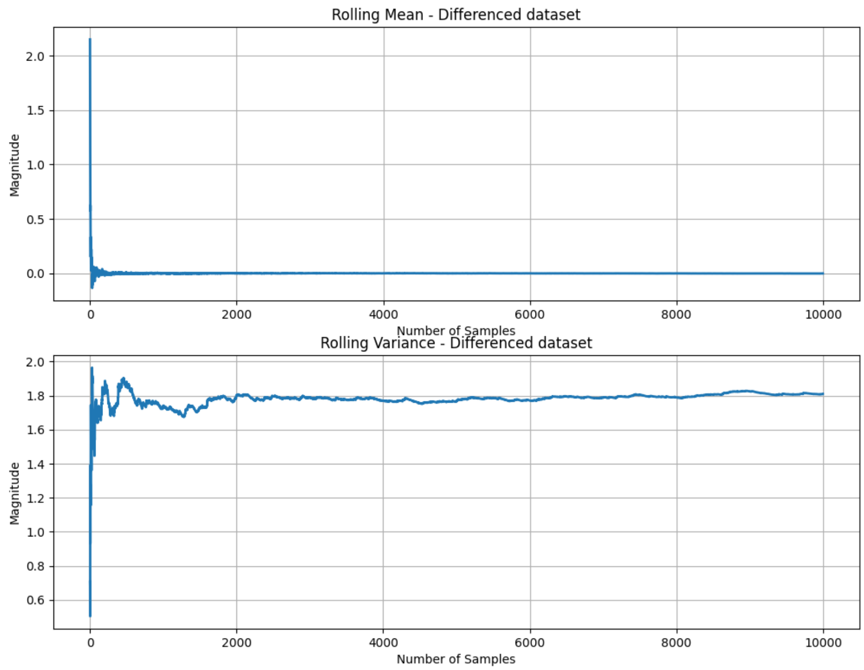


Once the data is differenced, the ACF and PACF shows what is pending, in the sense, order that is not differenced yet. Here, the ACF shows cut-off while the PACF shows tail-off which means we are dealing with an MA process.

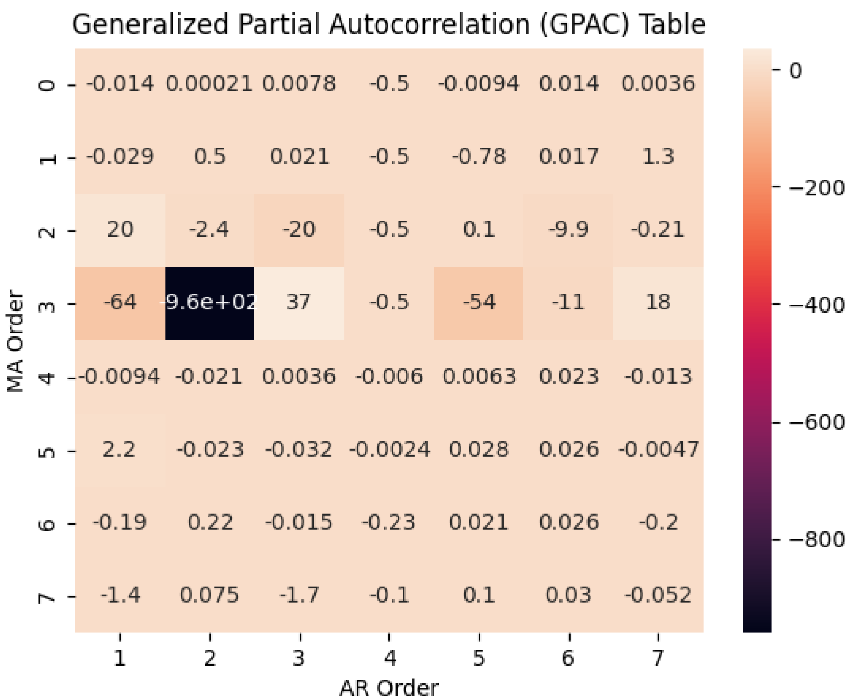
Step 7

ADF on the differenced dataset
ADF Statistic: -27.847192
p-value: 0.000000
Critical Values:

1%: -3.431
5%: -2.862
10%: -2.567

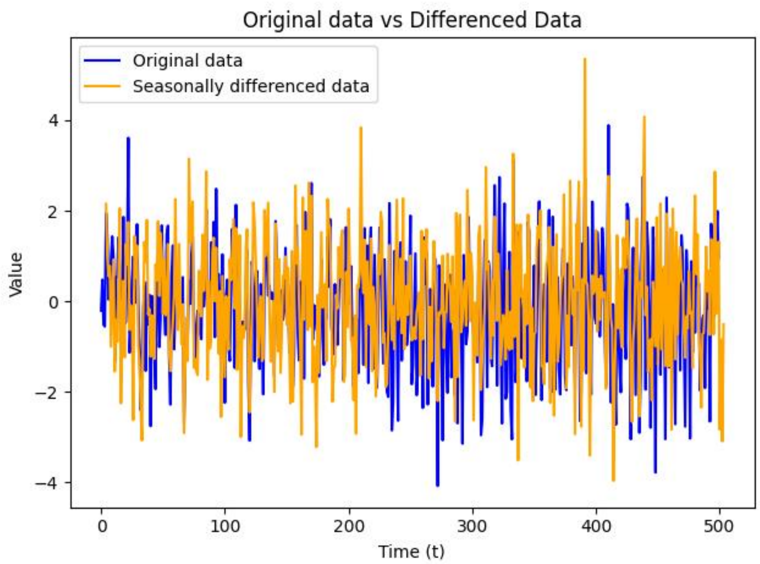


Step 8



Although it is tempting to say that non-seasonal AR has an order of 4, looking carefully at the table at 4^{th} column, after the 3^{rd} row, the values start to vary and is no longer a column of constants. Hence, I would say there is only a row (nb=4) that is filled with zeros. Hence, Nb = s/nb = 4/4 = 1. Since the program detected the raw data to be non-stationary, and it took 1 seasonal differencing (index=4 detected from the AR coefficients I supplied, post transforming backward operator to time), the ARMA order would be $(0,1,1)_4$.

Step 9

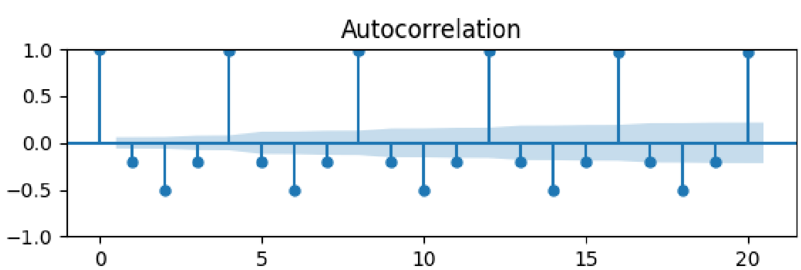


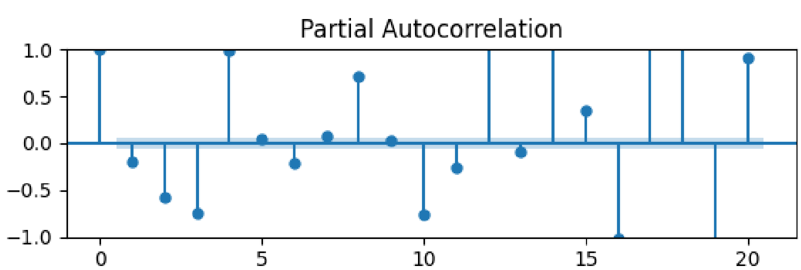
Example 8 - ARIMA$(0,1,2)_4$

Step 2

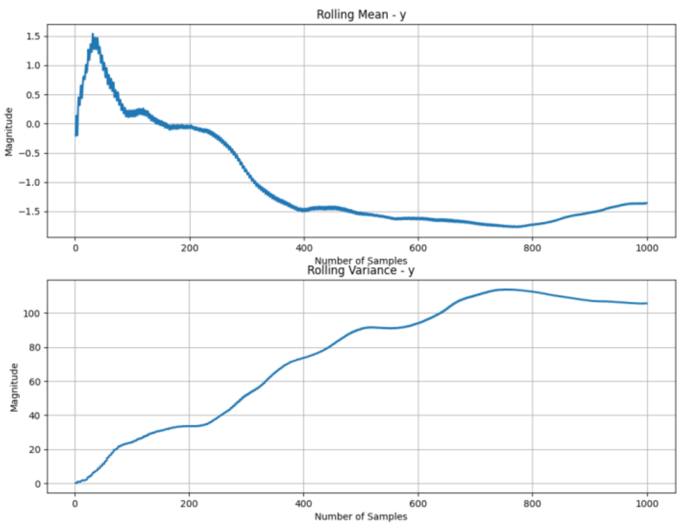
y(t) – y(t-4) = e(t) – 0.5 e(t-4) + 0.6 e(t-8)

Step 3 and 4



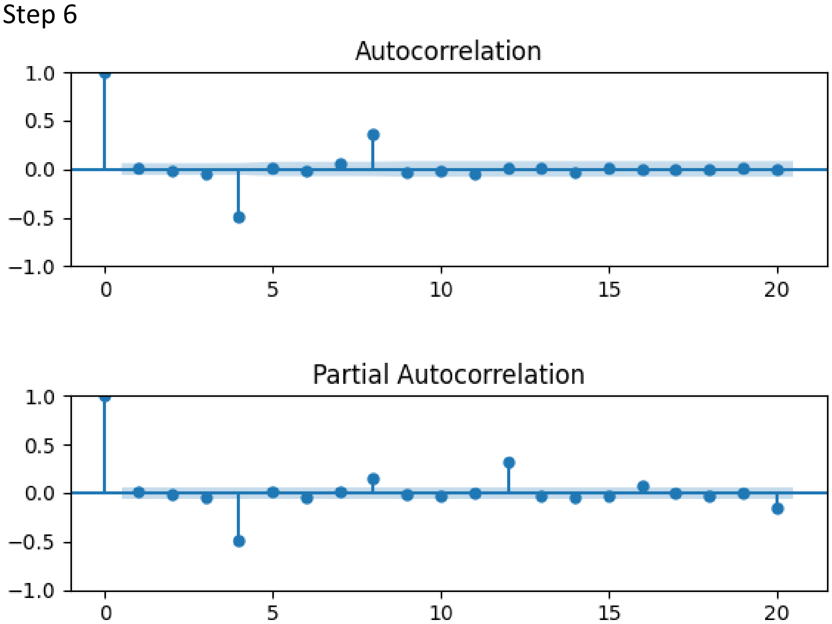


Step 5
ADF test on raw dataset:
ADF Statistic: -3.028289
p-value: 0.032336
Critical Values:
 1%: -3.437
 5%: -2.864
 10%: -2.568



The raw data is not stationary

Step 6



Once the data is differenced, the ACF and PACF shows what is pending, in the sense, order that is not differenced yet.

Step 7

ADF on the differenced dataset

ADF Statistic: -7.487499

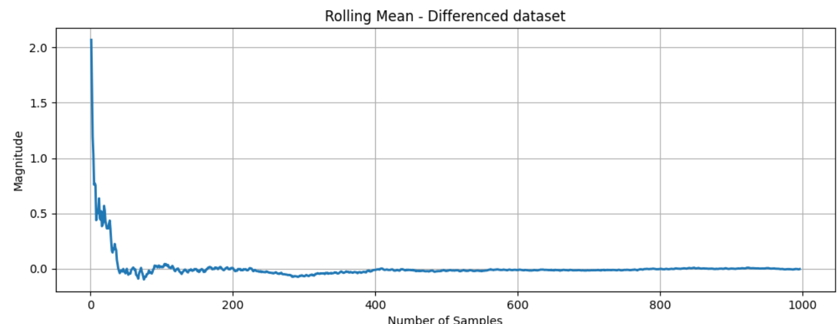
p-value: 0.000000

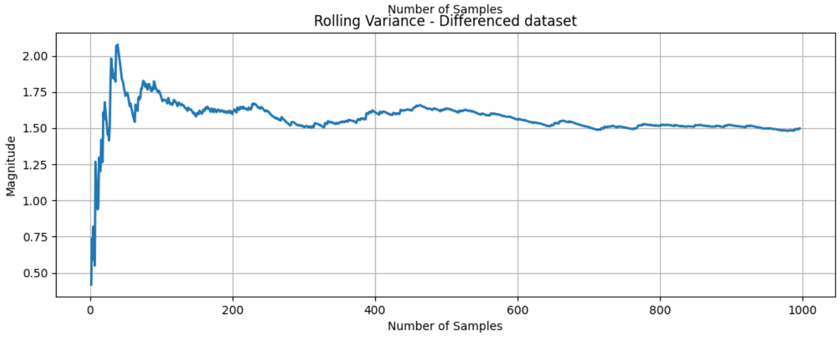
Critical Values:

1%: -3.437

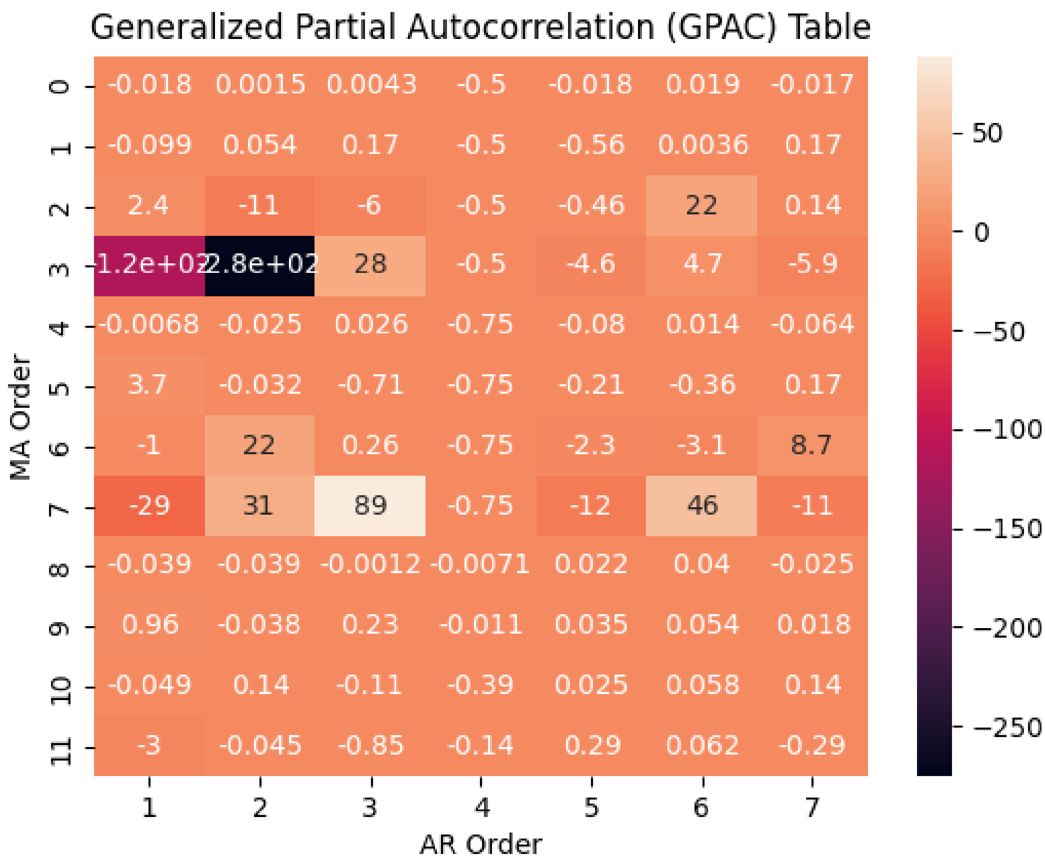
5%: -2.865

10%: -2.568



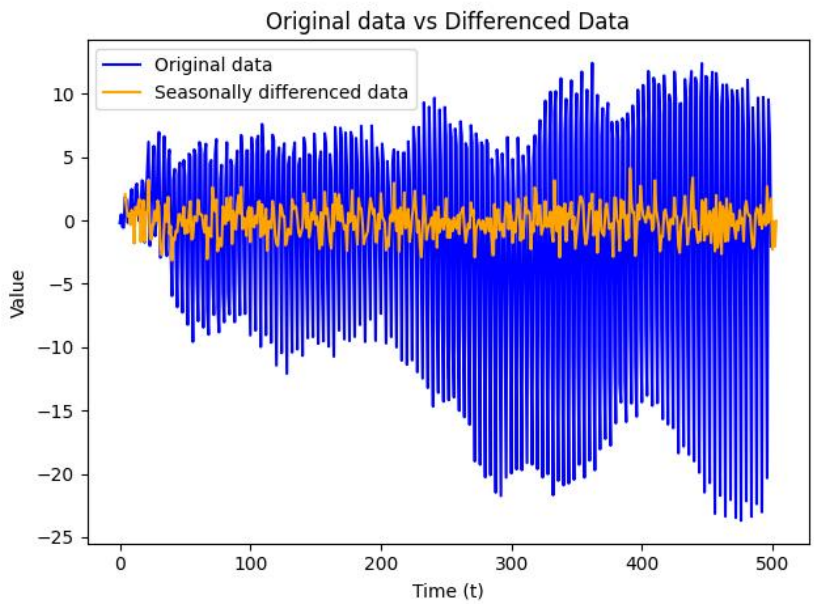


Step 8



Once again, similar to one of the previous examples, where the GPAC did not allow me to extract the accurate order of the ARIMA process, here, although the 4th column starts with a constant value -0.5, it changes after the 3rd row, and the behavior repeats. Ignoring that column, and looking for a row of zeros, there are multiple rows 0, 4, and 8. One thing I can observe here is that all three indices are multiples of 4, and the appearance of zeros stops post the 8th row. I believe this indicates that the nb is 8. Hence Nb = nb/s = 8/4 = 2, and the ARIMA order is: $(0,1,2)_4$. Also, the program did 1 seasonal differencing, whose index was estimated to be 4 by the Python code I had written.

Step 9



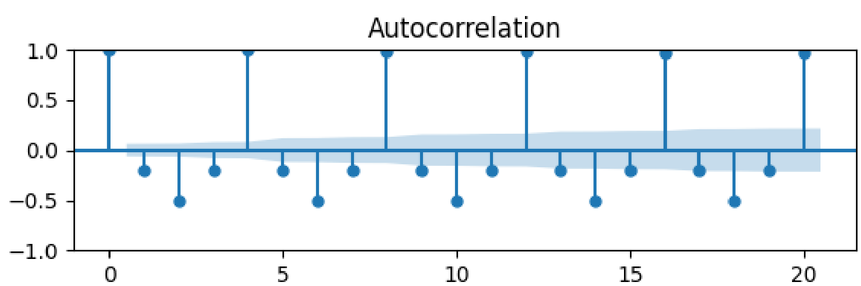
Example 9

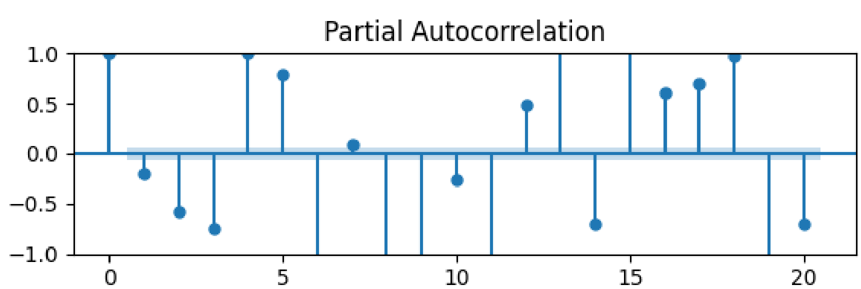
ARIMA$(1,1,1)_4$.

Step 2

$y(t) - 1.5\, y(t-4) + 0.5\, y(t-8) = e(t) - 0.2\, e(t-4)$

Step 3 and 4





Step 5

ADF test on raw dataset:

ADF Statistic: -2.617116

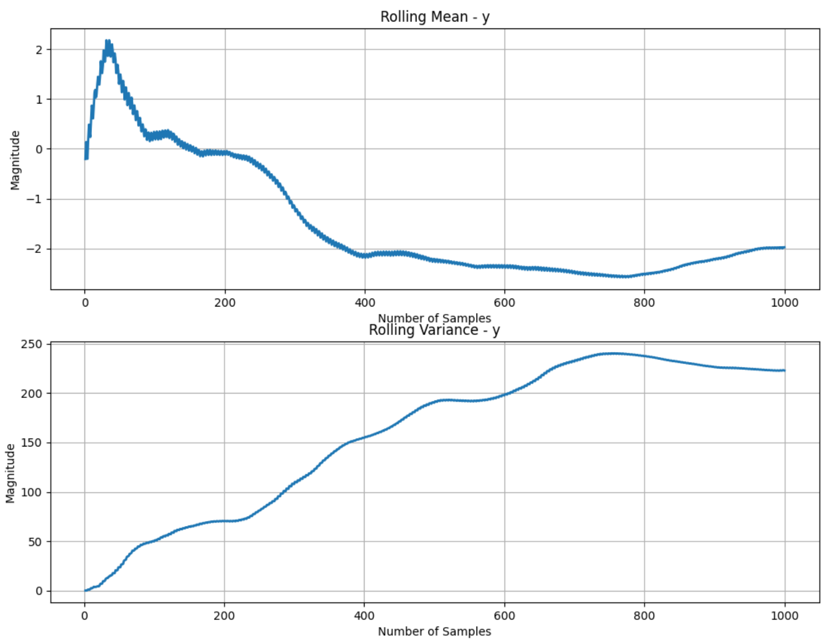
p-value: 0.089511

Critical Values:

1%: -3.437

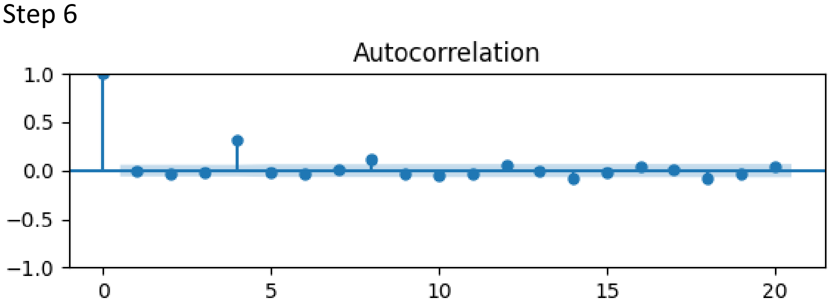
5%: -2.864

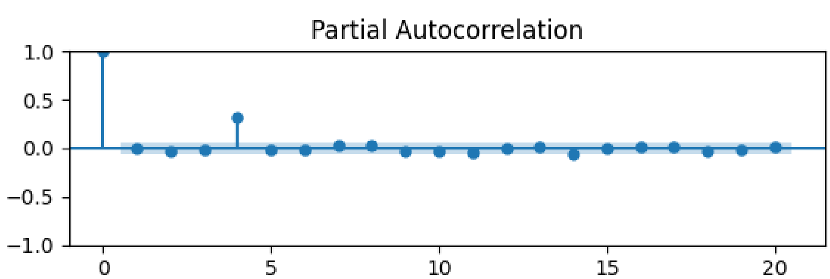
10%: -2.568



The raw data is not stationary.

Step 6

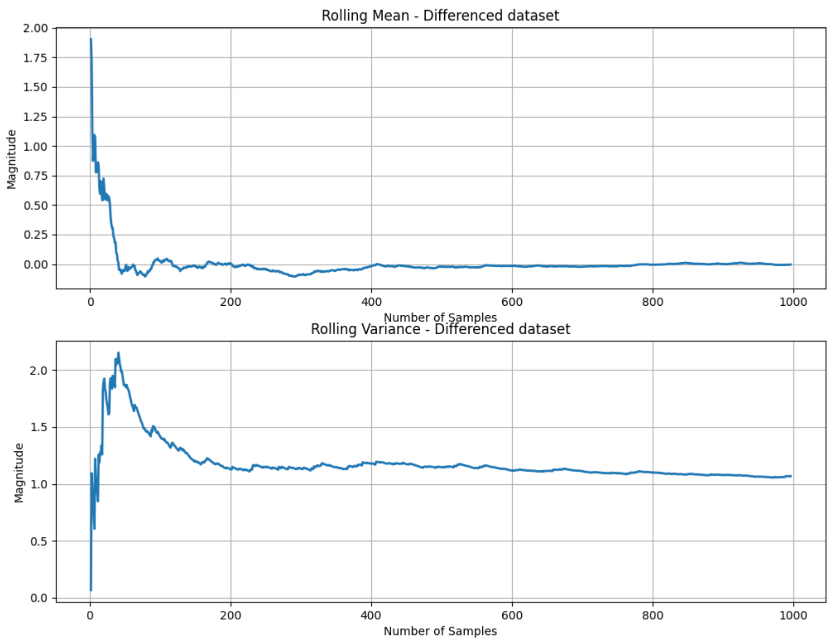




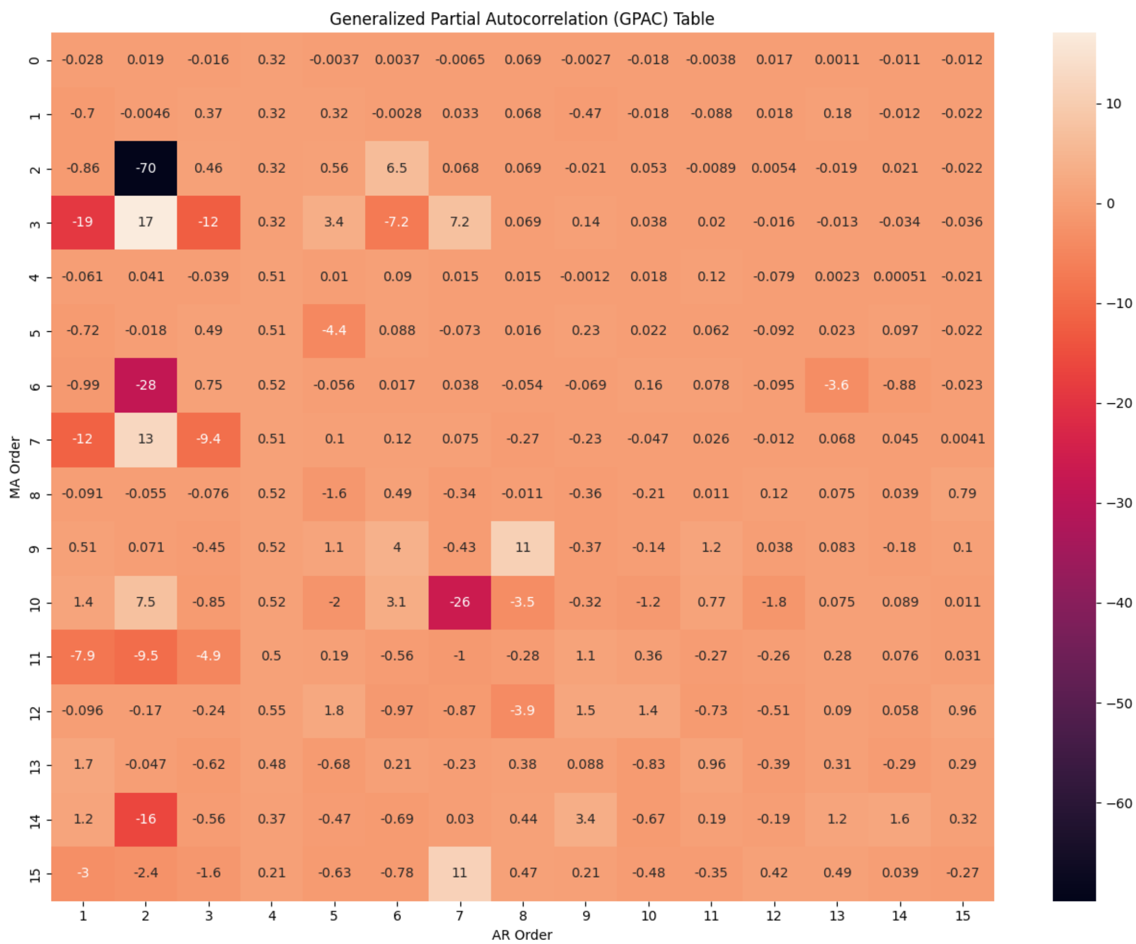
Once the data is differenced, the ACF and PACF shows what is pending, in the sense, order that is not differenced yet.

Step 7
ADF on the differenced dataset
ADF Statistic: -11.779248
p-value: 0.000000
Critical Values:

1%: -3.437
5%: -2.864
10%: -2.568

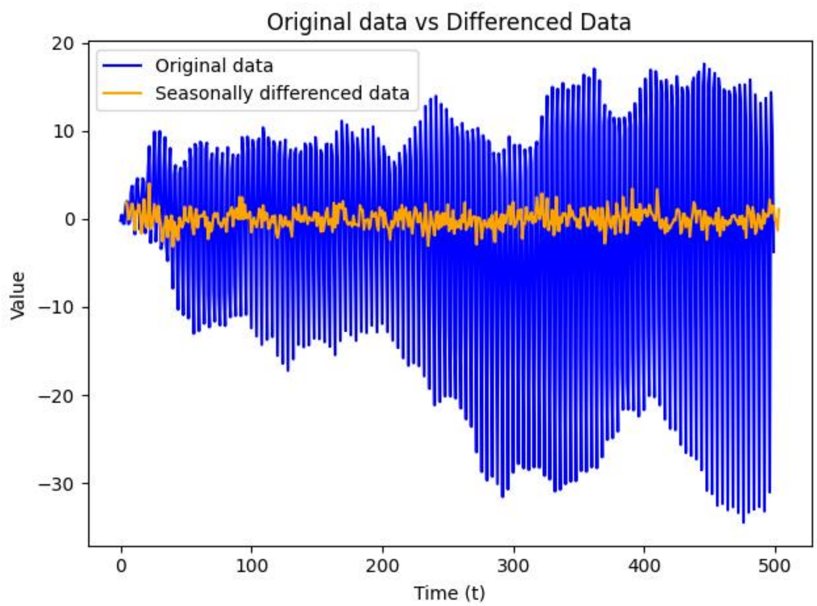


Step 8



From the above GPAC, it can be seen that the 4th column has constant values approximately ~0.52 with an adjacent row at 4th position. Only at 4th row, the 4th column begins to show value ~0.51. Hence the order is Na = na/s = 4/4 = 1, and Nb = 4/4 = 1. Since the original data was non-stationary, the program did one seasonal differencing to make it stationary whose index is 4. Hence the order extracted would be ARIMA$(1,1,1)_4$.

Step 9

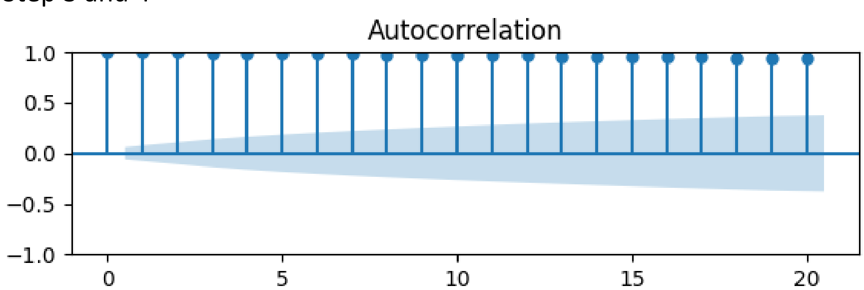


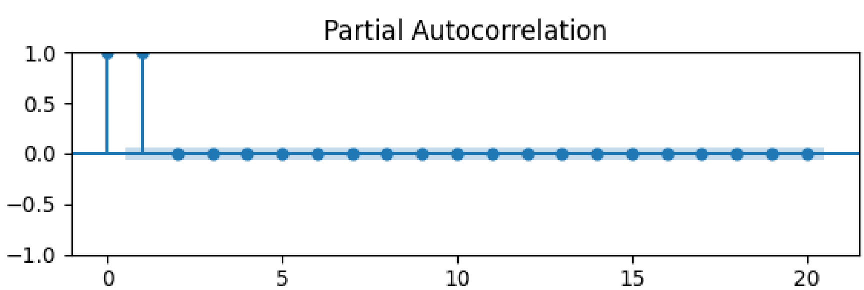
Example 10

Step 2

y(t) – 2y(t-1) + y(t-2) – y(t-3) + 2y(t-4) -y(t-5) = e(t) – 0.2e(t-1) + 0.35e(t-3) – 0.07e(t-4)

Step 3 and 4





Step 5

ADF test on raw dataset:

ADF Statistic: -0.977642

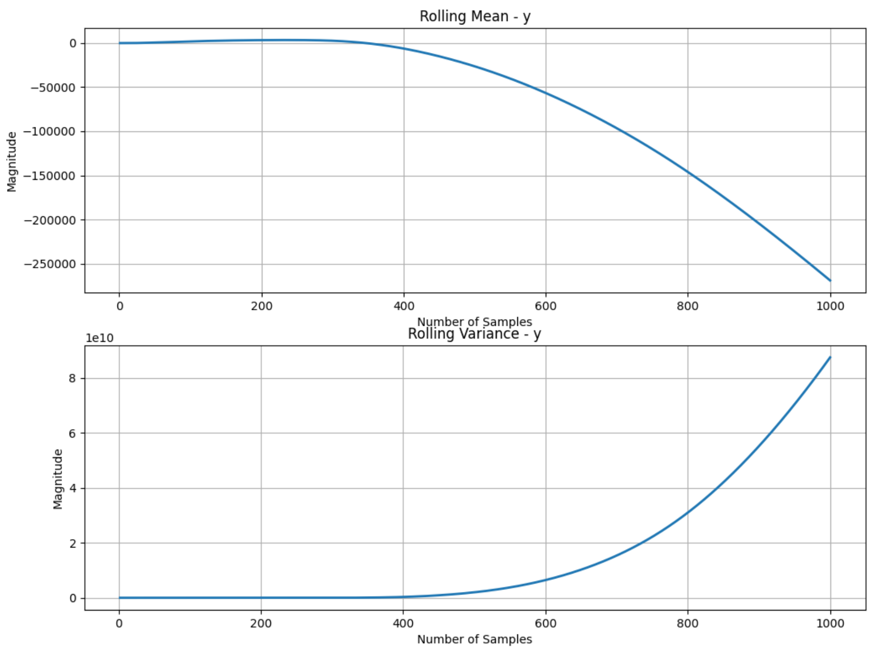
p-value: 0.761331

Critical Values:

    1%: -3.437

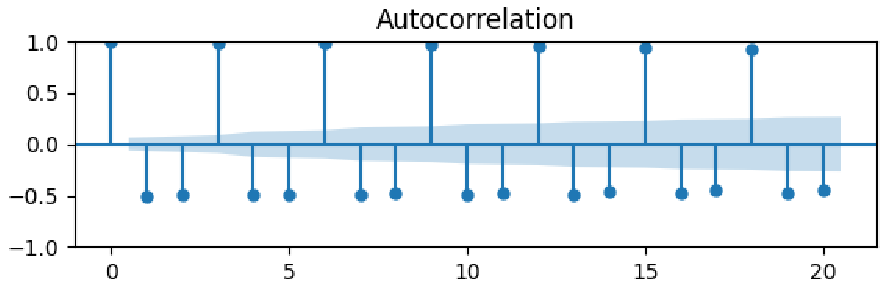
    5%: -2.864

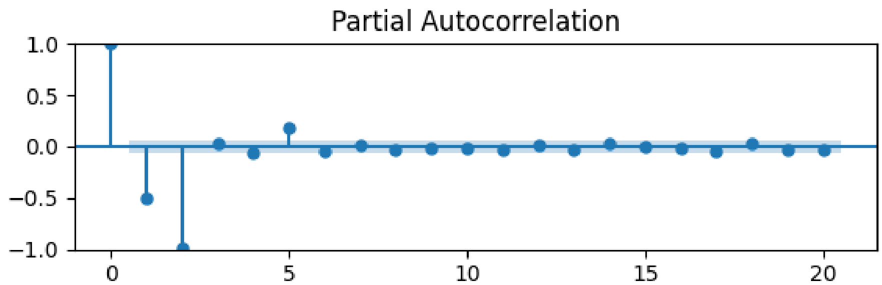
    10%: -2.568



The raw data is not stationary.

Step 6

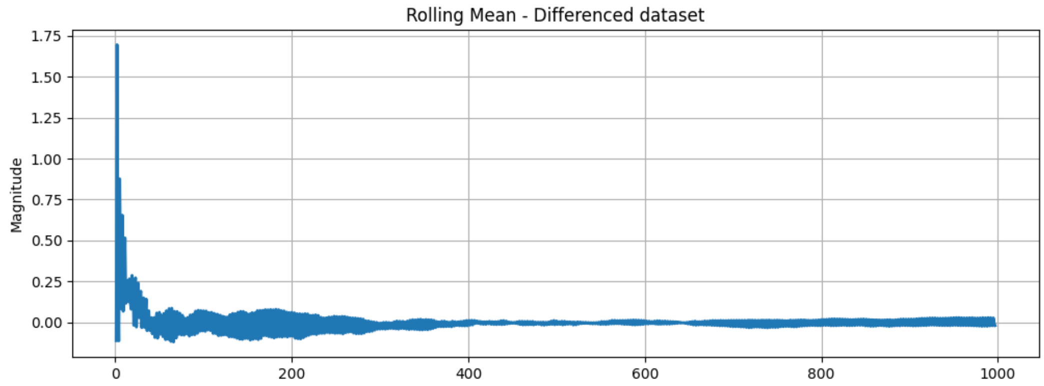


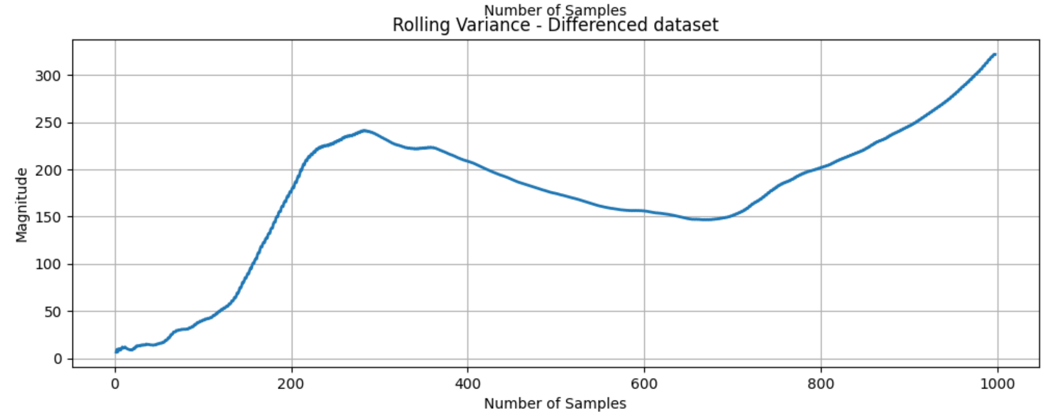


Once the data is differenced, the ACF and PACF shows what is pending, in the sense, order that is not differenced yet.

Step 7
It took 3 non-seasonal differencing to make the data stationary.

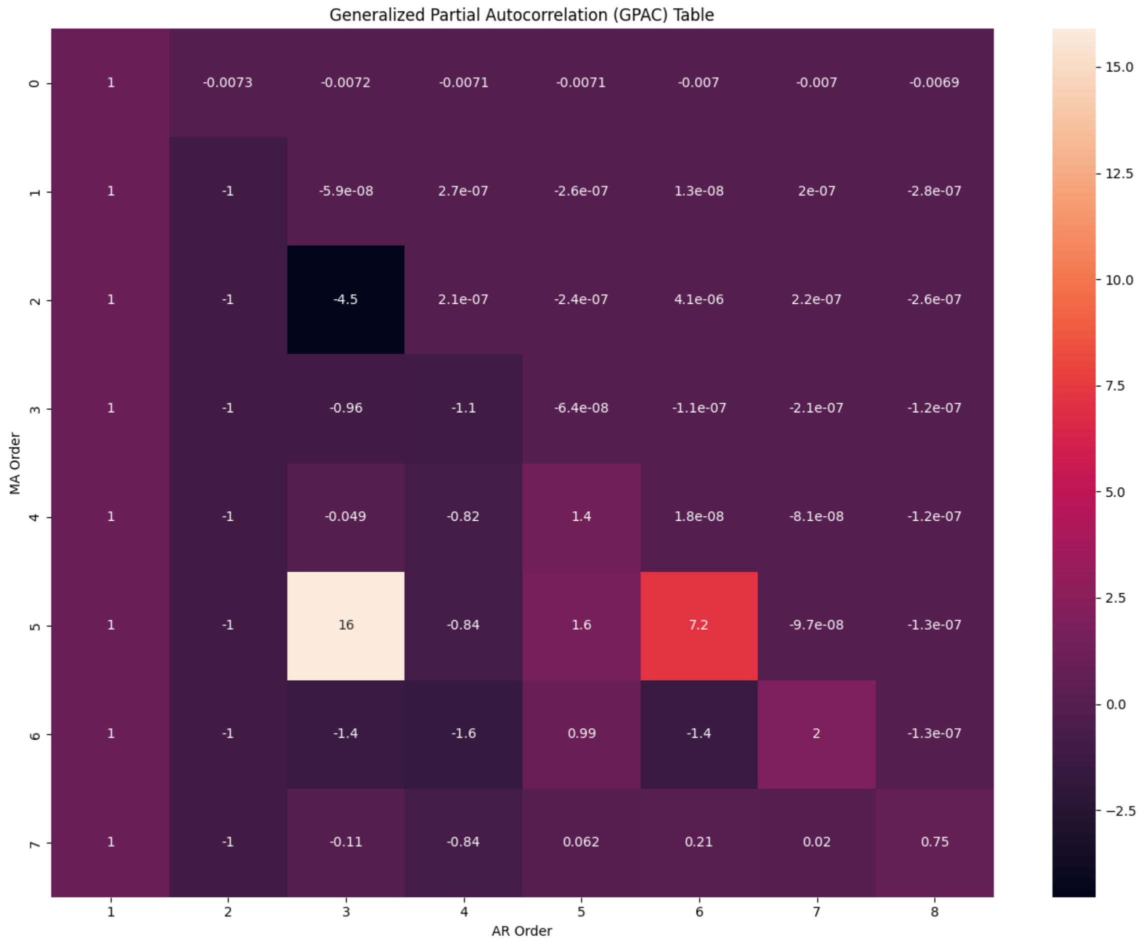
ADF on the differenced dataset
ADF Statistic: -12.406231
p-value: 0.000000
Critical Values:
    1%: -3.437
    5%: -2.864
    10%: -2.568





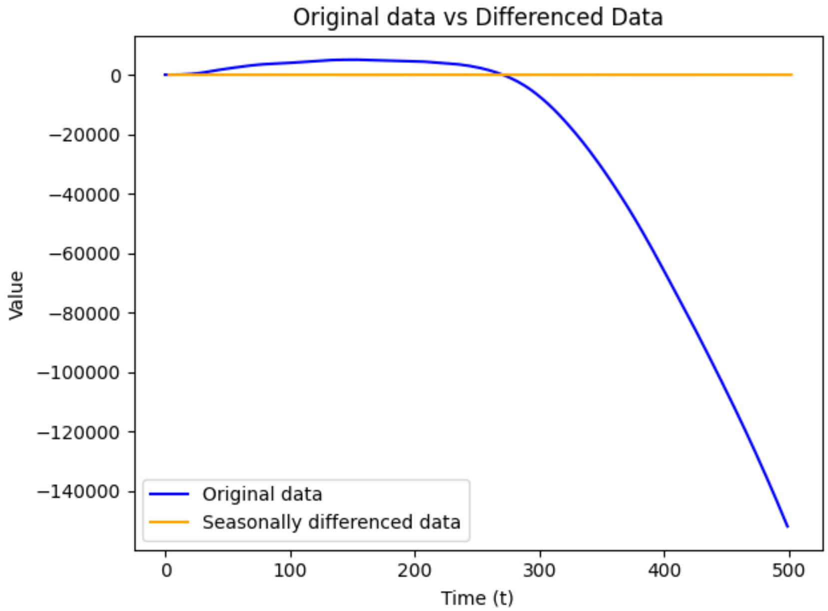
But the rolling statistics doesn't show the data being stationary. However, the ADF-test produces a p-value of 0, which means the data is stationary.

Step 8

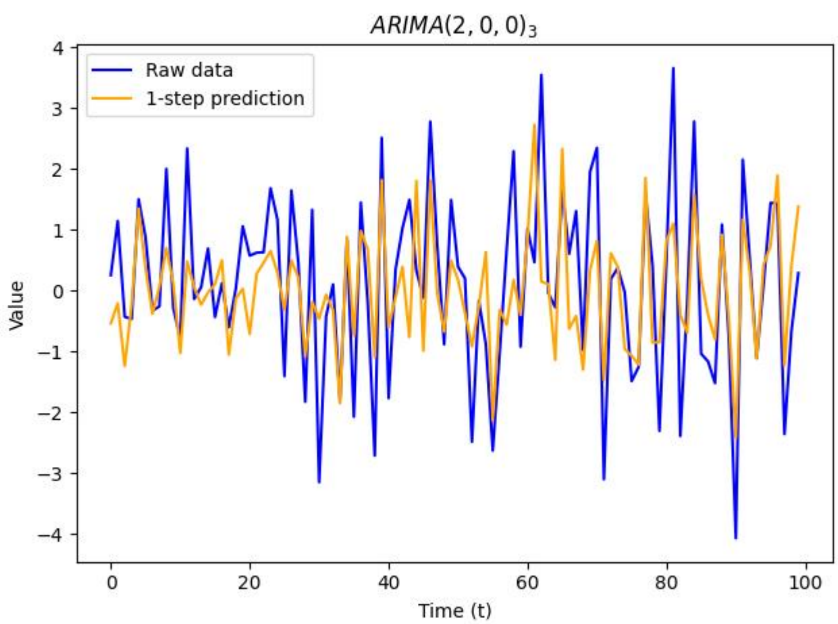


I believe this is one fine case where I have personally observed the GPAC not being able to show the correct order, not including cases that are subject to zero-pole cancellation. The fundamental problem is, my code that automatically extracts the seasonality from the coefficients was unable to extract for this particular example that caused the GPAC to produce this.

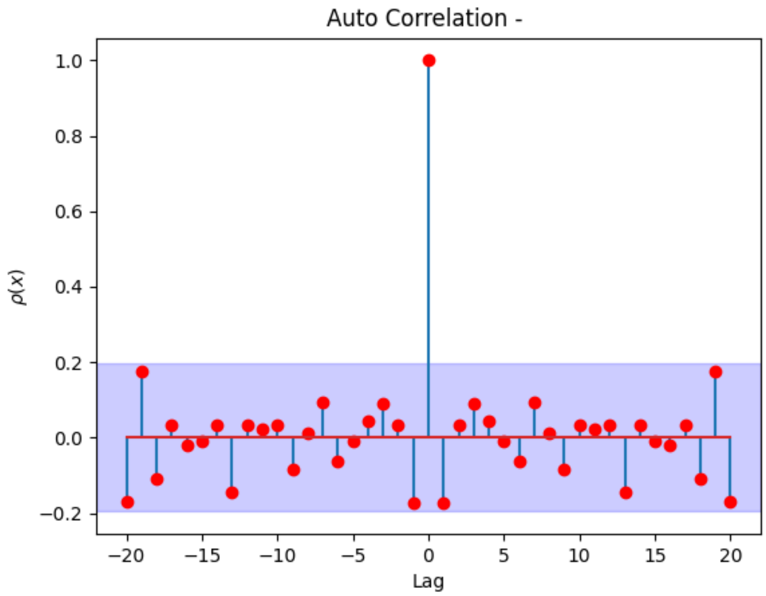
Step 9



Questions 12 and 13



ACF of the residuals

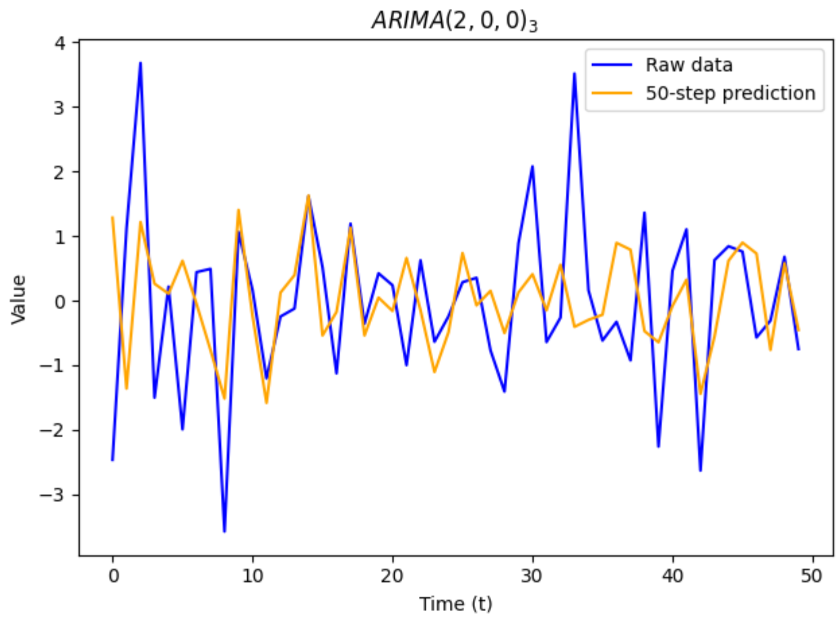


Question 14

```
>>> sm.stats.acorr_ljungbox(residual, lags=[20])
      lb_stat  lb_pvalue
20  18.634202   0.545705
```

The Ljung-Box test returned a p-value of ~0.54 which does not allow us to reject the null hypothesis. In other words, we accept the null hypothesis and conclude that the autocorrelation of residuals is indeed white. Hence, the residual error is white.

Question 15



In terms of ratio of variance of test set vs predicted, I was unable to bring it near 1.

```
>>> print(f"Variance of test set vs predicted {np.var(y_out[952:1002])/np.var
 (y_hat_50)}")
Variance of test set vs predicted 3.200025166463083
```