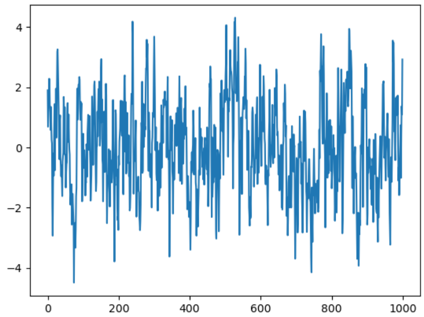
Lab2

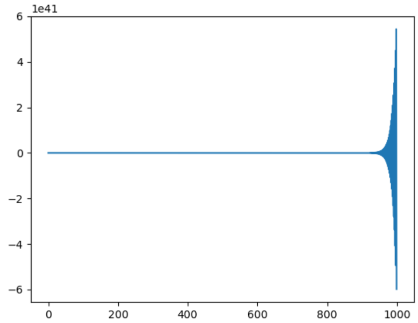
Conglei Xiang Yuqi Sun

1. Task 1 Stationarity of AR models

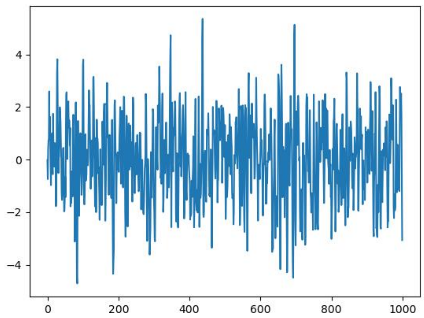
1.1. Generate four time series according to the above AR models, one for each. Draw a line plot for each time series.



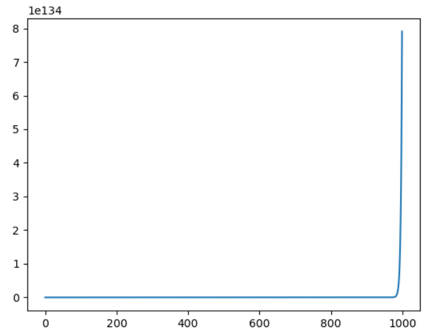
Line plot of (1)



Line plot of (2)



Line plot of (1)



Line plot of (2)

1.2. Judge whether each time series is stationary or not, by visual inspection.

Time series 1 & 3 are stationary, while 2 & 4 are not.

1.3. Judge whether each time series is stationary or not, by its model's auto-regressive coefficient values.

For p = 1: $-1 < \phi 1 < 1$ => the time series is stationary, so 1 is stationary and 2 is not.

For p = 2: $-1 < \phi 2 < 1$; $\phi 2 + \phi 1 < 1$; $\phi 2 - \phi 1 < 1$ => the time series is stationary, so 3 is stationary and 4 is not.

1.4. Call the following statsmodel function to judge if each AR process is stationary.

```
print(arma_process1.isstationary)
print(arma_process2.isstationary)
print(arma_process3.isstationary)
print(arma_process4.isstationary)
```

```
True
False
True
False
```

The results verify the answers in 1.2 and 1.3.

1.5. Use the unit-root based Augmented Dickey-Fuller (ADF) test to check if each time series generated by the AR models is stationary or not. Do the results match your visual inspection?

```
ADF Statistic: -11.315957
p-value: 0.000000
Critical Values:
	1%: -3.437
	5%: -2.864
	10%: -2.568
```

p-value < 0.05, then reject the null hypothesis and consider that the series is stationary

```
ADF Statistic: -291109596645929984.000000
p-value: 0.000000
Critical Values:
	1%: -3.437
	5%: -2.864
	10%: -2.568
```

p-value < 0.05, then reject the null hypothesis and consider that the series is stationary

```
ADF Statistic: -22.988026
p-value: 0.000000
Critical Values:
	1%: -3.437
	5%: -2.864
	10%: -2.568
```

p-value < 0.05, then reject the null hypothesis and consider that the series is stationary

```
ADF Statistic: 118464253970018512.000000
p-value: 1.000000
Critical Values:
	1%: -3.437
	5%: -2.864
	10%: -2.568
```

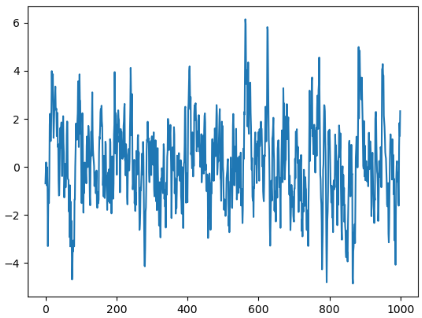
p-value > 0.05, so accept the null hypothesis, and consider that the series is not stationary.
The result of time series 2 doesn't match the expected result.

[Questions & Answers]

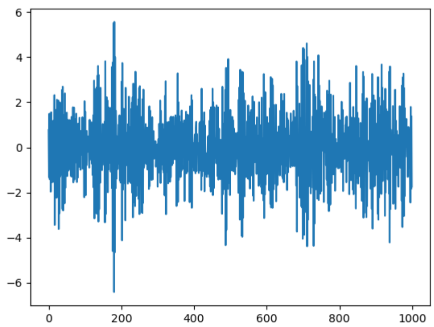
- **With visual inspection, how do you identify if a time series is stationary or not?**
  It can be identified in a line plot that there is no significant upward or downward trend. This illustrates the mean remains constant. The data points are evenly distributed on both sides of the mean, which illustrates the variance remains constant. There is no obvious patterns or cycles over time.
- **How do you judge the stationarity of time series using the unit-root method? Does it always give correct results?**
  By judging the p-value: The p-value is much less than the significance level of 0.05 and hence we can reject the null hypothesis and judge that the series is stationary. It doesn't always give correct results when the sample size is small, or with inappropriate model.
- **What is the role of component ϵt in the model? Why is it important?**
  The component ϵt represents the random error term or the residuals at time t. It is an unobservable component that captures the portion of the time series data that cannot be explained by the autoregressive relationship with its past values.
- **To have an AR(p) model be stationary, is there any requirement on the auto-regressive coefficients? List the constraints for AR(1) and AR(2) models.**
  For $p = 1$, when $-1 < \phi 1 < 1$, the AR model is stationary. For $p = 2$, when $-1 < \phi 2 < 1$, $\phi 2 + \phi 1 < 1$, and $\phi 2 - \phi 1 < 1$, the AR model is stationary.

2. Task 2 ACF, PACF of AR models

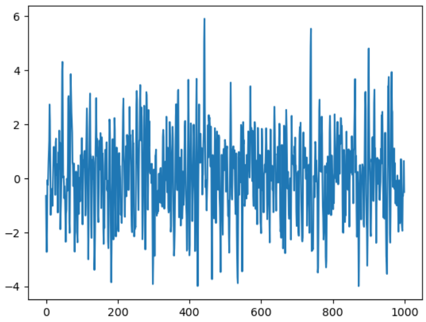
2.1. Generate four time series according to the above AR models, one for each. Are all of them stationary? Draw a line plot for each time series.

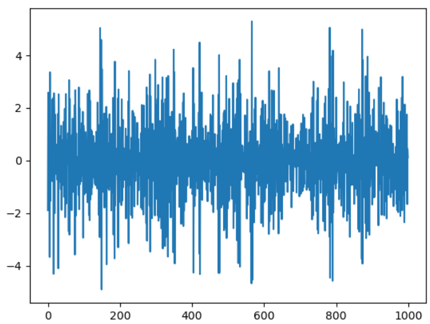


Line plot of (1)



Line plot of (2)

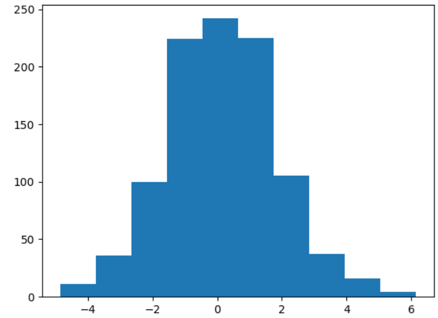




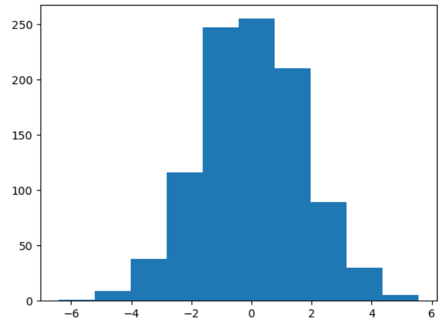
Line plot of (3)

Line plot of (4)

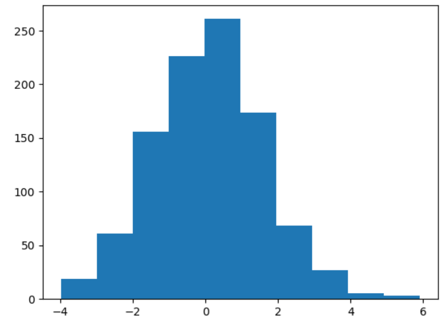
2.2. Draw histogram, density plot, and box plot for each time series with 1000 data points. Are there any outliers? Why?



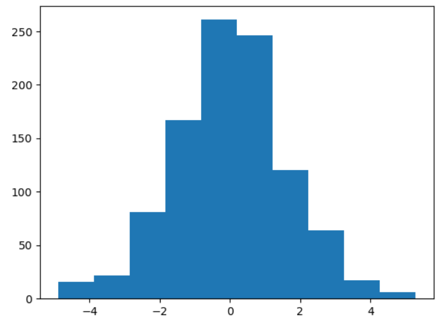
Histogram of (1)



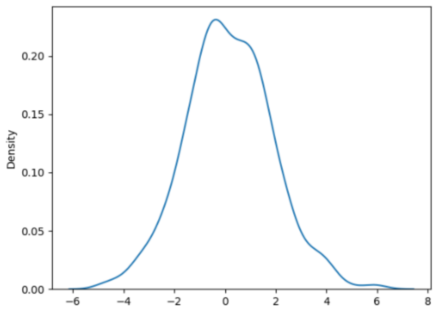
Histogram of (2)



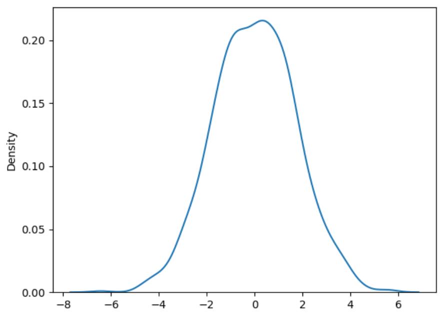
Histogram of (3)



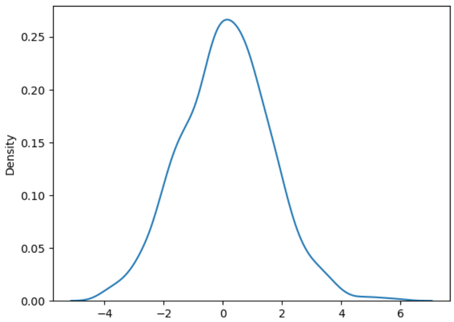
Histogram of (4)



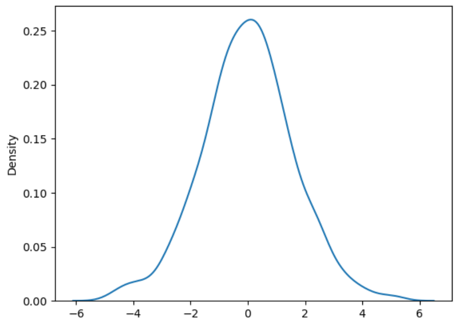
Density plot of (1)



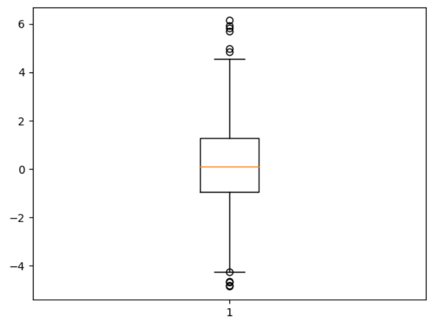
Density plot of (2)



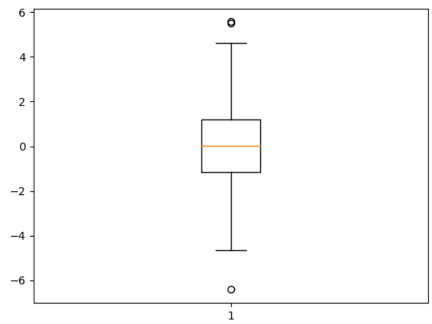
Density plot of (3)



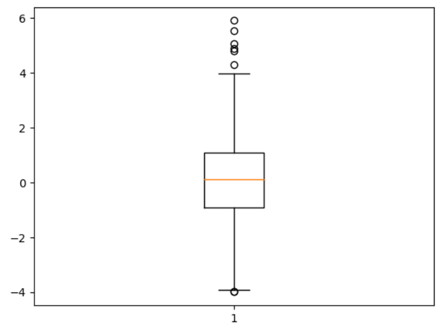
Density plot of (4)



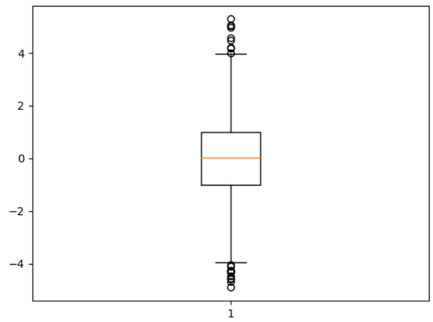
Box plot of (1)



Box plot of (2)



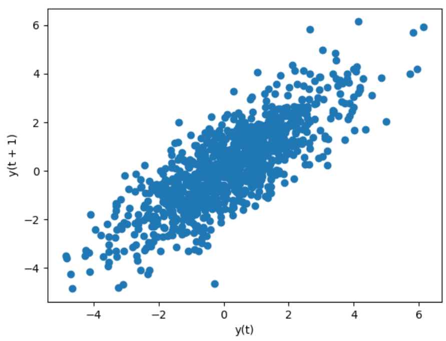
Box plot of (3)



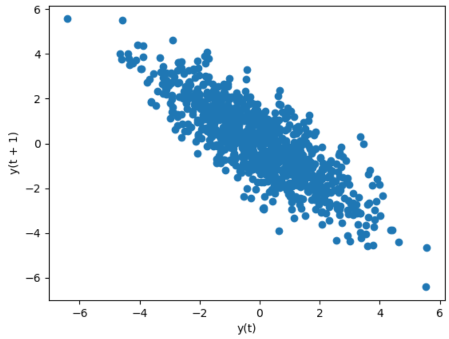
Box plot of (4)

There are some outliers according to the box plot because of the Gaussian noise.

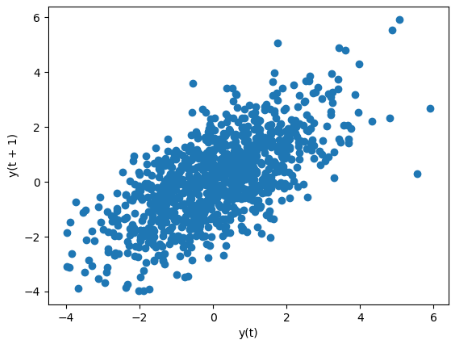
2.3. Draw lag-1 and lag-2 plots for each time series. Do you observe any auto-correlation from the lag plots?



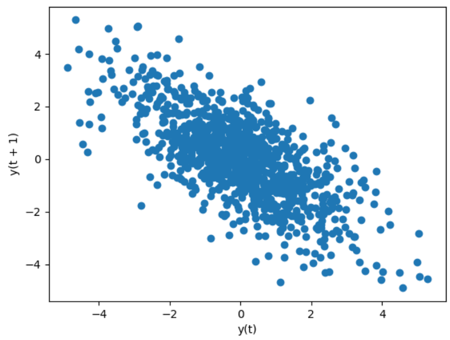
Lag-1 plot of (1)



Lag-1 plot of (2)



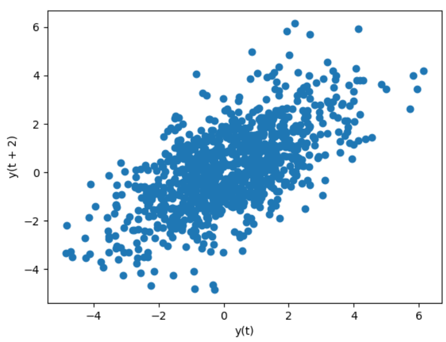
Lag-1 plot of (3)



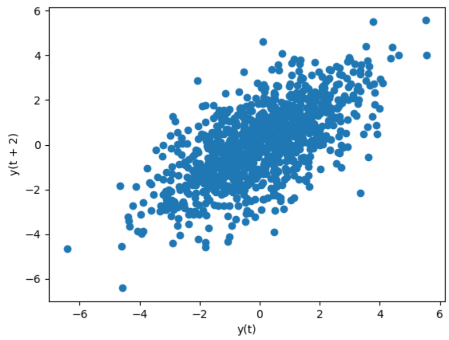
Lag-1 plot of (4)

The plots of time series 1 & 3 indicate that past observations are positively correlated with future observations of 1 time unit later.

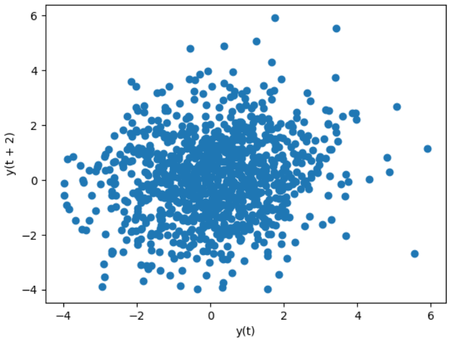
The plots of time series 2 & 4 indicate that past observations are negatively correlated with future observations of 1 time unit later.



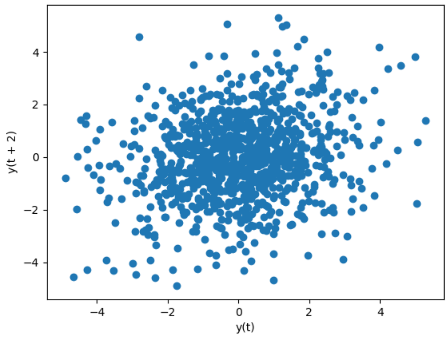
Lag-2 plot of (1)



Lag-2 plot of (2)



Lag-2 plot of (3)

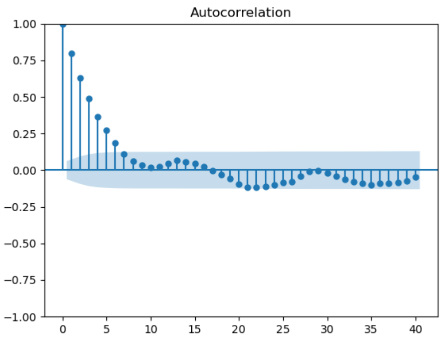


Lag-2 plot of (4)

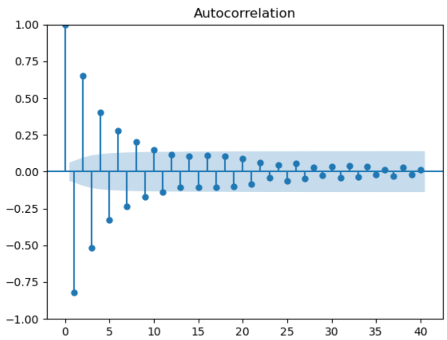
The plots of time series 1 & 2 indicate that past observations are positively correlated with future observations of 2 time unit later.

There are no obvious autocorrelation in time series 3 & 4.

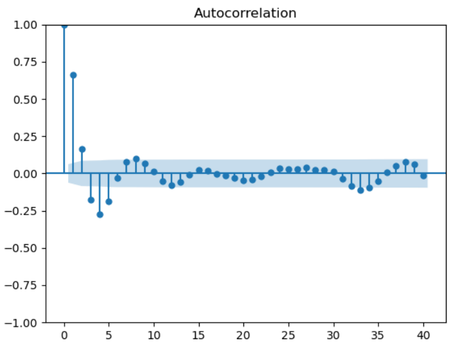
2.4. Draw ACF graph for each time series.



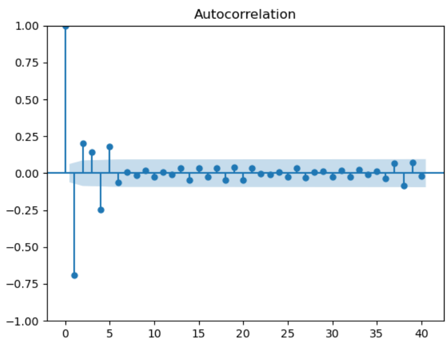
ACF graph of (1)



ACF graph of (2)

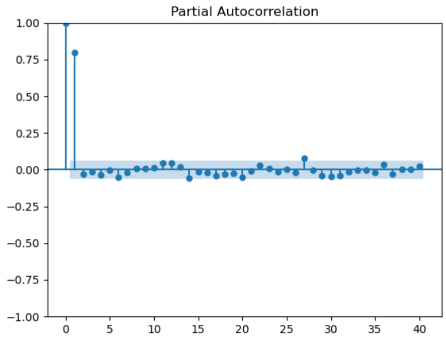


ACF graph of (3)

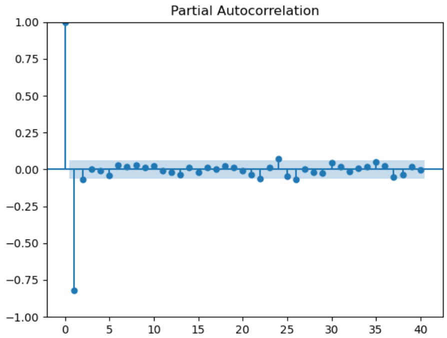


ACF graph of (4)

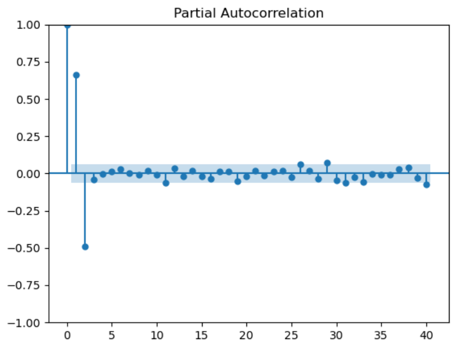
2.5. Draw PACF graph for each time series.



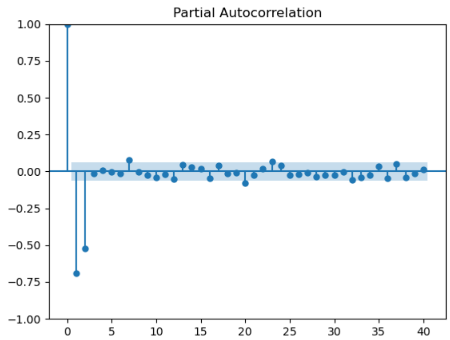
PACF graph of (1)



PACF graph of (2)



PACF graph of (3)



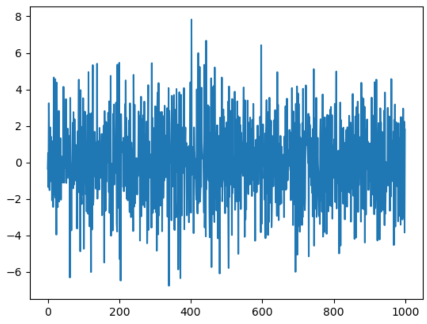
PACF graph of (4)

[Questions & Answers]

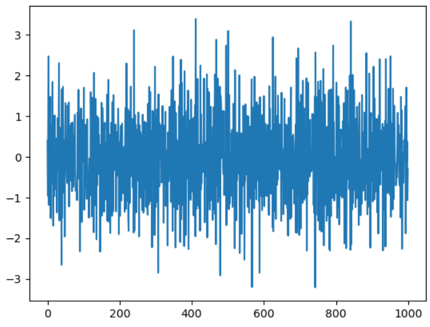
- **What characteristics can you observe from the ACF graphs of the AR(p) models?**
  In an AR(p) model, the ACF plot shows a slow decay in values as the lag increases and show significant spikes at the first few lags. It also shows fluctuation for lags beyond p.
- **What characteristics can you observe from the PACF graphs of the AR(p) models?**
  In an AR(p) model, the PACF plot shows a significant partial autocorrelation coefficient at lag p and the lags that are less than p, and shows a sudden cut-off to zero or approximately close to zero after lag p.

3. Task 3 Invertibility, ACF, PACF of MA models

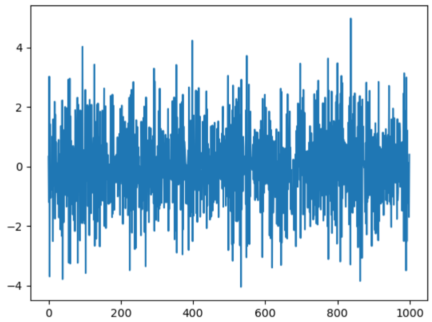
3.1. Generate four time series according to the above MA(q) models, one for each. Draw a line plot for each time series.



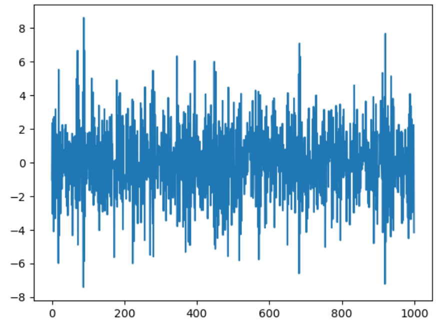
Line plot of (1)



Line plot of (2)



Line plot of (3)



Line plot of (4)

3.2. Judge whether each time series is invertible or not, by visual inspection.

It is not sufficient to determine the invertibility of the series just by its visual inspection of the line plot.

3.3. Judge whether each time series is invertible or not, by its models's auto-regressive coefficient values.

For q = 1, when $-1 < \theta1 < 1$, MA models are invertible. For q = 2, when $-1 < \theta2 < 1$, $\theta2 + \theta1 > -1$, $\theta1 - \theta2 < 1$, MA models are invertible. Thus, time series 1 and 4 are not invertible, while time series 2 and 3 are invertible.

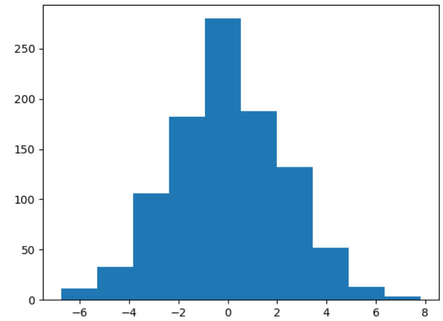
3.4. Call the statsmodel function to judge if each MA process is invertible.

```
print(arma_process3_1.isinvertible)
print(arma_process3_2.isinvertible)
print(arma_process3_3.isinvertible)
print(arma_process3_4.isinvertible)
```

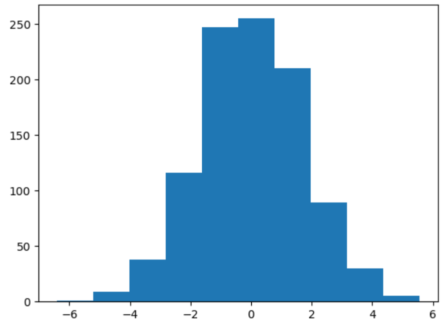
```
False
True
True
False
```

It shows that time series 1 and 4 are not invertible while 2 and 3 are invertible.

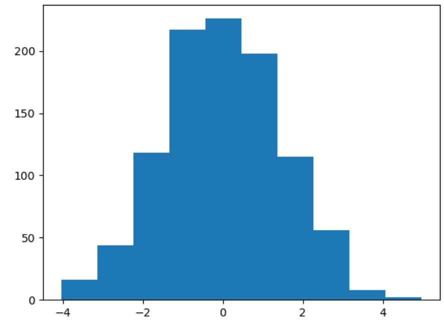
3.5. Draw histogram, density plot, and box plot for each time series with 1000 data points. Are there any outliers? Why?



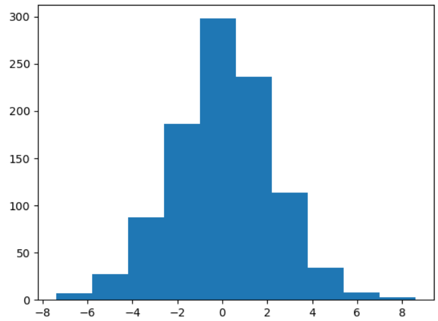
Histogram of (1)



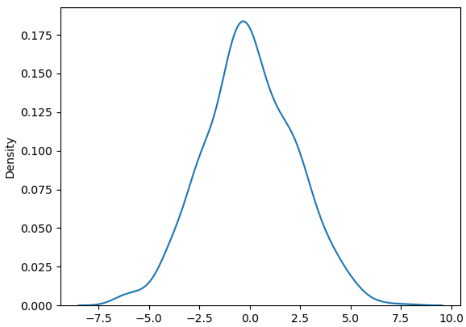
Histogram of (2)



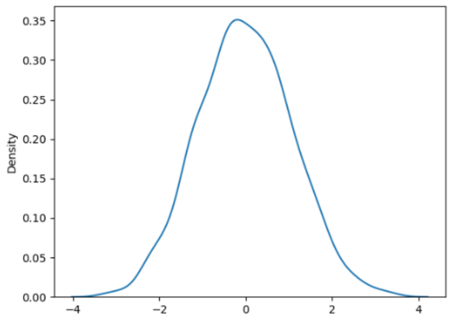
Histogram of (3)



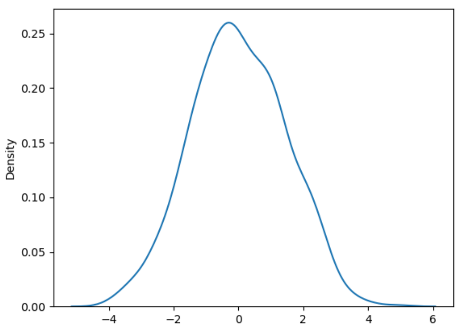
Histogram of (4)



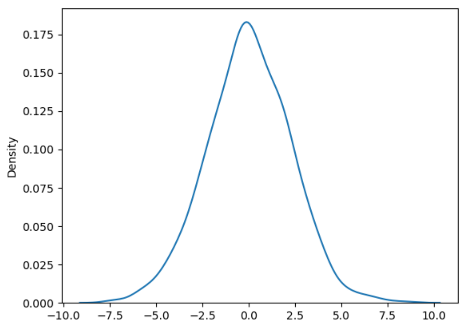
Density plot of (1)



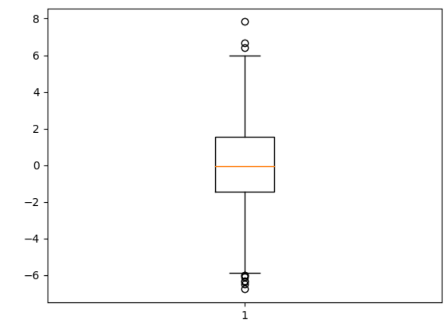
Density plot of (2)

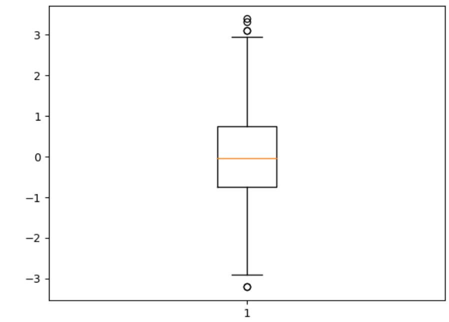


Density plot of (3)

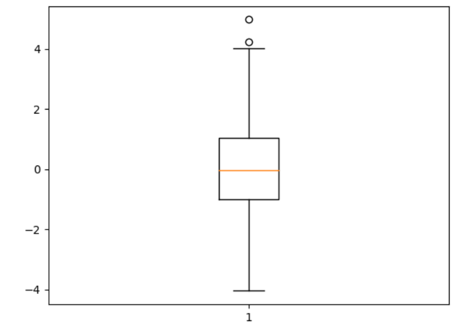


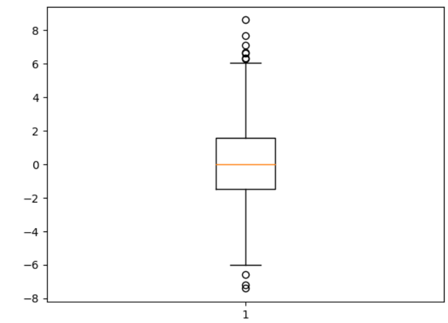
Density plot of (4)





Box plot of (1) Box plot of (2)

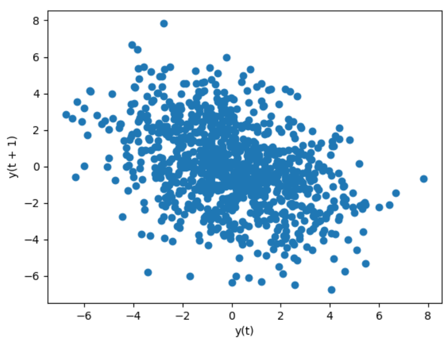


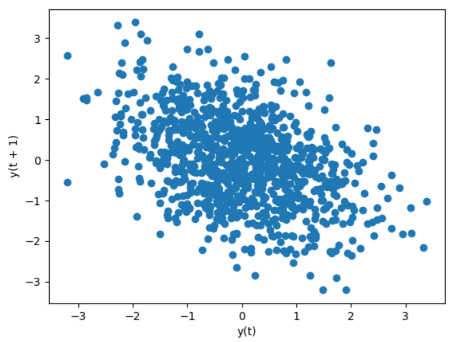


Box plot of (3) Box plot of (4)

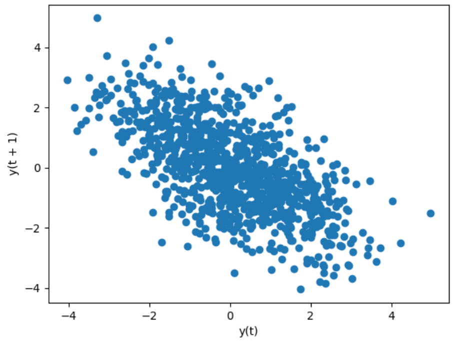
There are some outliers according to the box plot.

3.6. Draw lag-1 and lag-2 plots for each time series. Do you observe any auto-correlation from the lag plots?

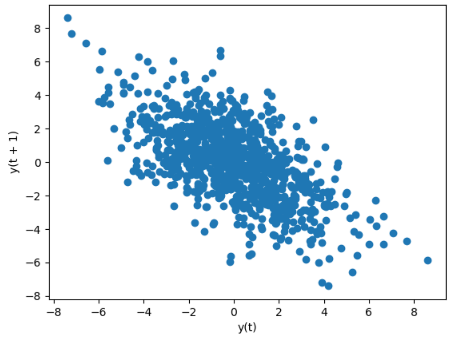




Lag-1 plot of (1) Lag-1 plot of (2)

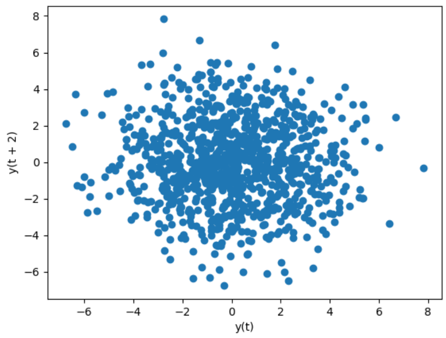


Lag-1 plot of (3)

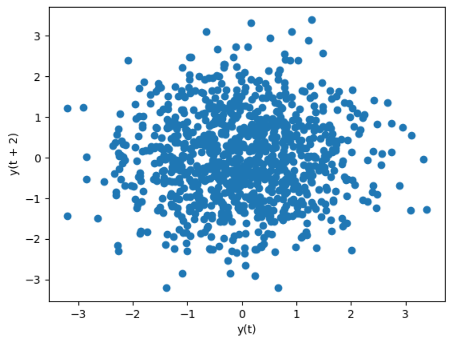


Lag-1 plot of (4)

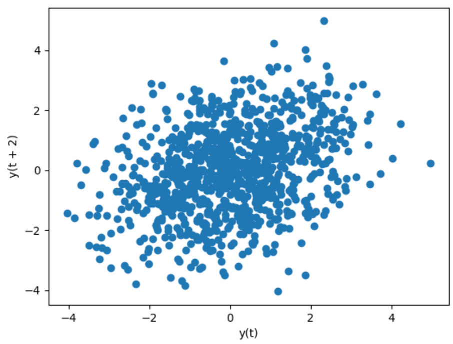
The plots of all the time series indicate that past observations are negatively correlated with future observations of 1 time unit later.



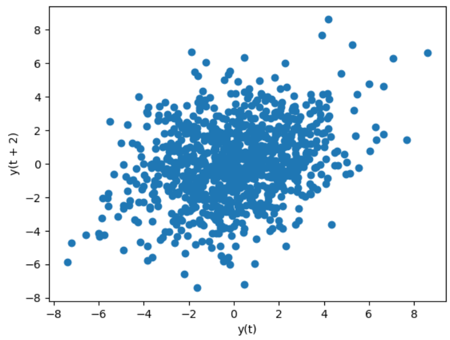
Lag-2 plot of (1)



Lag-2 plot of (2)



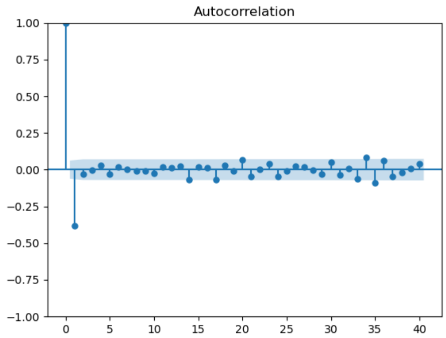
Lag-2 plot of (3)

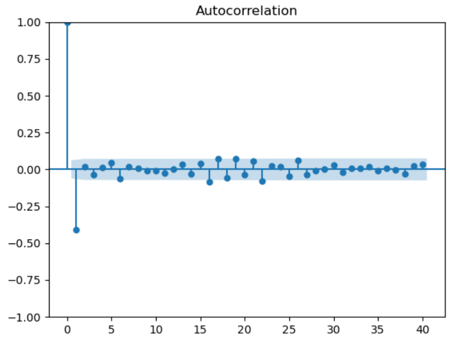


Lag-2 plot of (4)

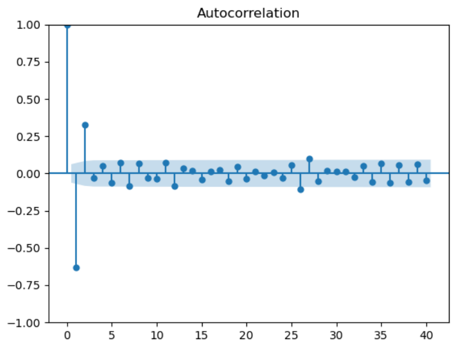
There are no obvious autocorrelation in lag-2 plot.

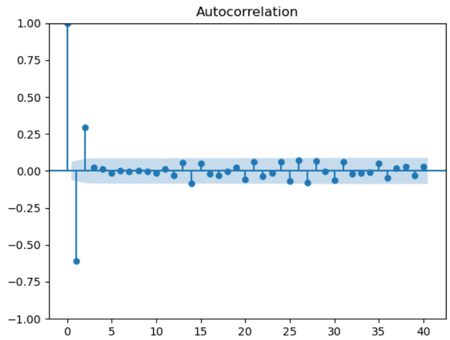
3.7. Draw ACF graph for each time series.





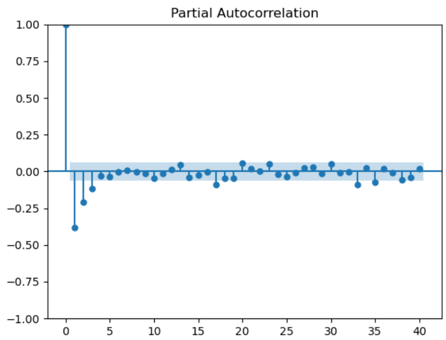
ACF graph of (1) ACF graph of (2)

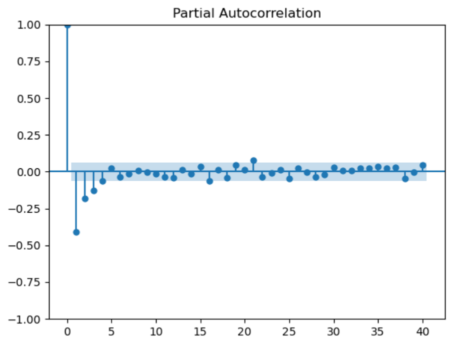




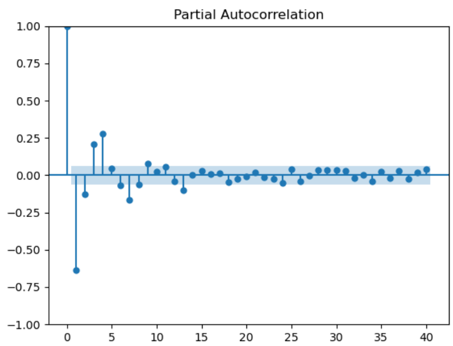
ACF graph of (3) ACF graph of (4)

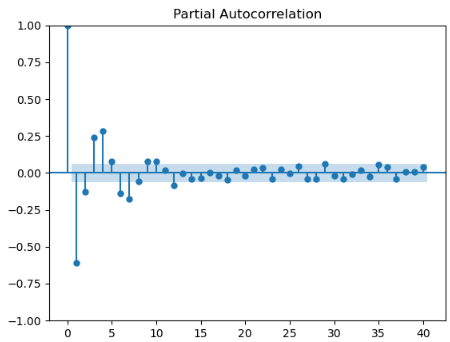
3.8. Draw PACF graph for each time series.





PACF graph of (1) PACF graph of (2)





PACF graph of (3) PACF graph of (4)

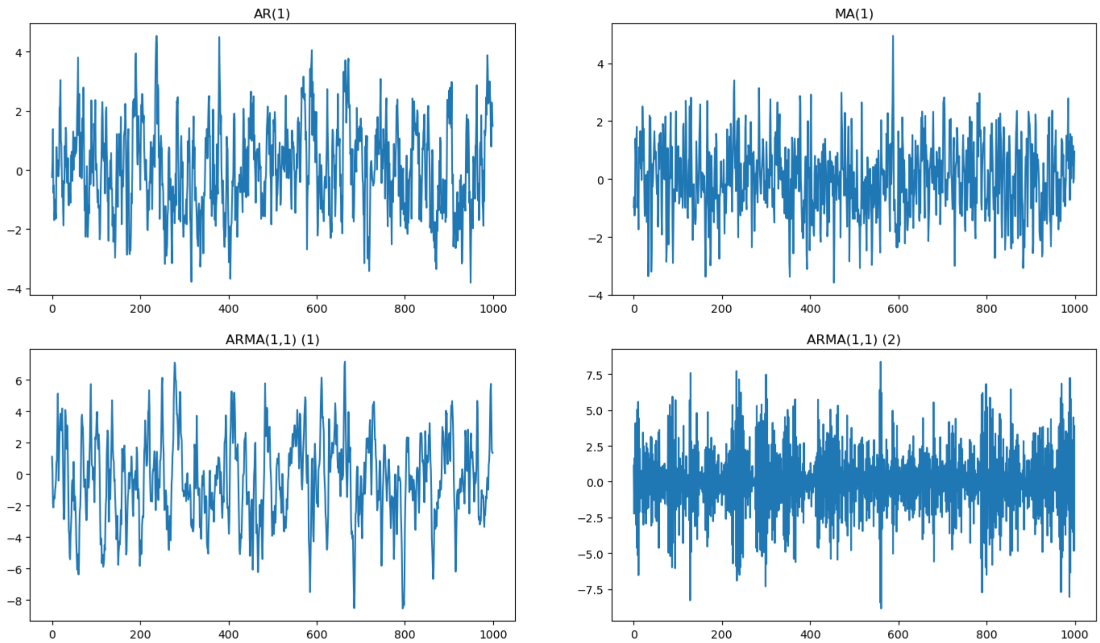
[Questions & Answers]

- **Are all the MA models invertible? If not, which ones are invertible and which ones are not invertible?**
  Not all the MA models are invertible. For q = 1, when $-1 < \theta 1 < 1$, MA models are invertible. For q = 2, when $-1 < \theta 2 < 1$, $\theta 2 + \theta 1 > -1$, $\theta 1 - \theta 2 < 1$, MA models are invertible. Thus, time series 1 and 4 are not invertible, while time series 2 and 3 are invertible.
- **What characteristics can you observe from the ACF graphs of the MA(q) models?**
  The ACF plot for an MA(q) model shows a significant autocorrelation coefficient at lag q and the lags that are less than q, and shows a sudden drop-off to zero or approximately close to zero after lag q.
- **What characteristics can you observe from the PACF graphs of the MA(q) models?**
  In an MA(q) model, the PACF plot shows significant spikes at the q-th lag, and shows a sharp drop in values to zero after the q-th lag.
- **To have an MA(q) model be invertible, is there any requirement on the auto-regressive coefficients? List the constraints for MA(1) and MA(2) models.**
  Yes. For q = 1, when $-1 < \theta 1 < 1$, MA models are invertible. For q = 2, when $-1 < \theta 2 < 1$, $\theta 2 + \theta 1 > -1$, $\theta 1 - \theta 2 < 1$, MA models are invertible.

Task 4: Stationarity, ACF and PACF of ARMA models

Complete the following tasks:

1. Generate three time series according to the above ARMA(p, q) models, one for each. Draw a line plot for each time series.



2. Judge whether each time series is stationary or not, by visual inspection.

By visual inspection, all of four time series are stationary.

3. Use the ADF test method to judge whether each time series is stationary or not. Do the results match your visual inspection?

AR(1):
ADF Statistic: -11.759239439450408
p-value: 1.1586164648583269e-21
Critical Values:
1%: -3.4369127451400474
5%: -2.864437475834273
10%: -2.568312754566378
The series is stationary.

MR(1):
ADF Statistic: -8.859035718274493
p-value: 1.5002890758191018e-14
Critical Values:
1%: -3.436979275944879
5%: -2.8644668170148058
10%: -2.5683283824496153
The series is stationary.

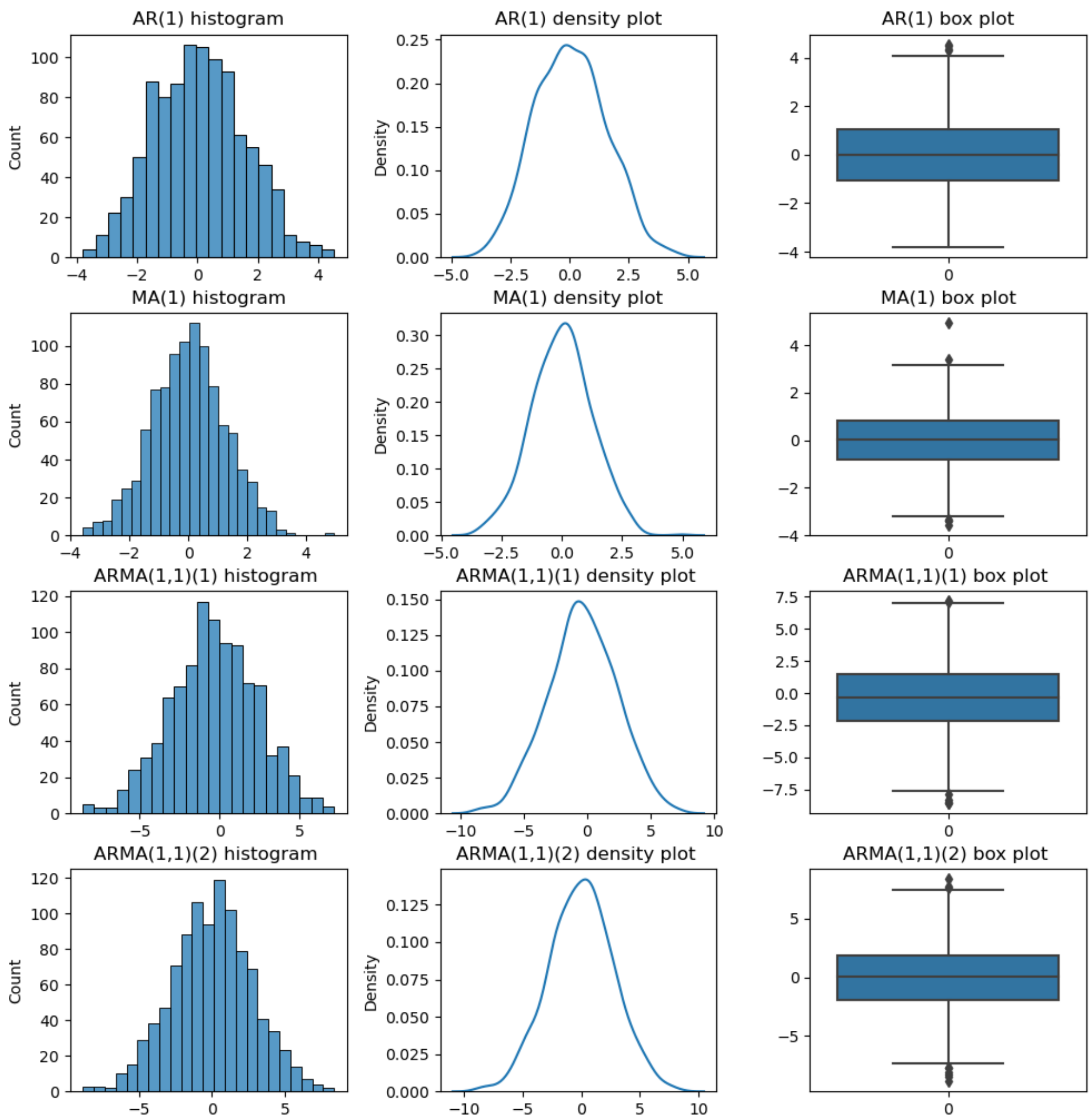
ARMR(1,1)(1):
ADF Statistic: -8.826983622060299
p-value: 1.812378848570518e-14
Critical Values:
1%: -3.436959175494265
5%: -2.8644579524531975
10%: -2.568323660940752
The series is stationary.

ARMR(1,1)(2):
ADF Statistic: -16.987813437609734
p-value: 8.914602341098802e-30
Critical Values:
1%: -3.436959175494265
5%: -2.8644579524531975
10%: -2.568323660940752
The series is stationary.

4. Call the corresponding statsmodels functions to judge if each ARMA process is stationary and invertible.

All ARMA processes are stationary and invertible.

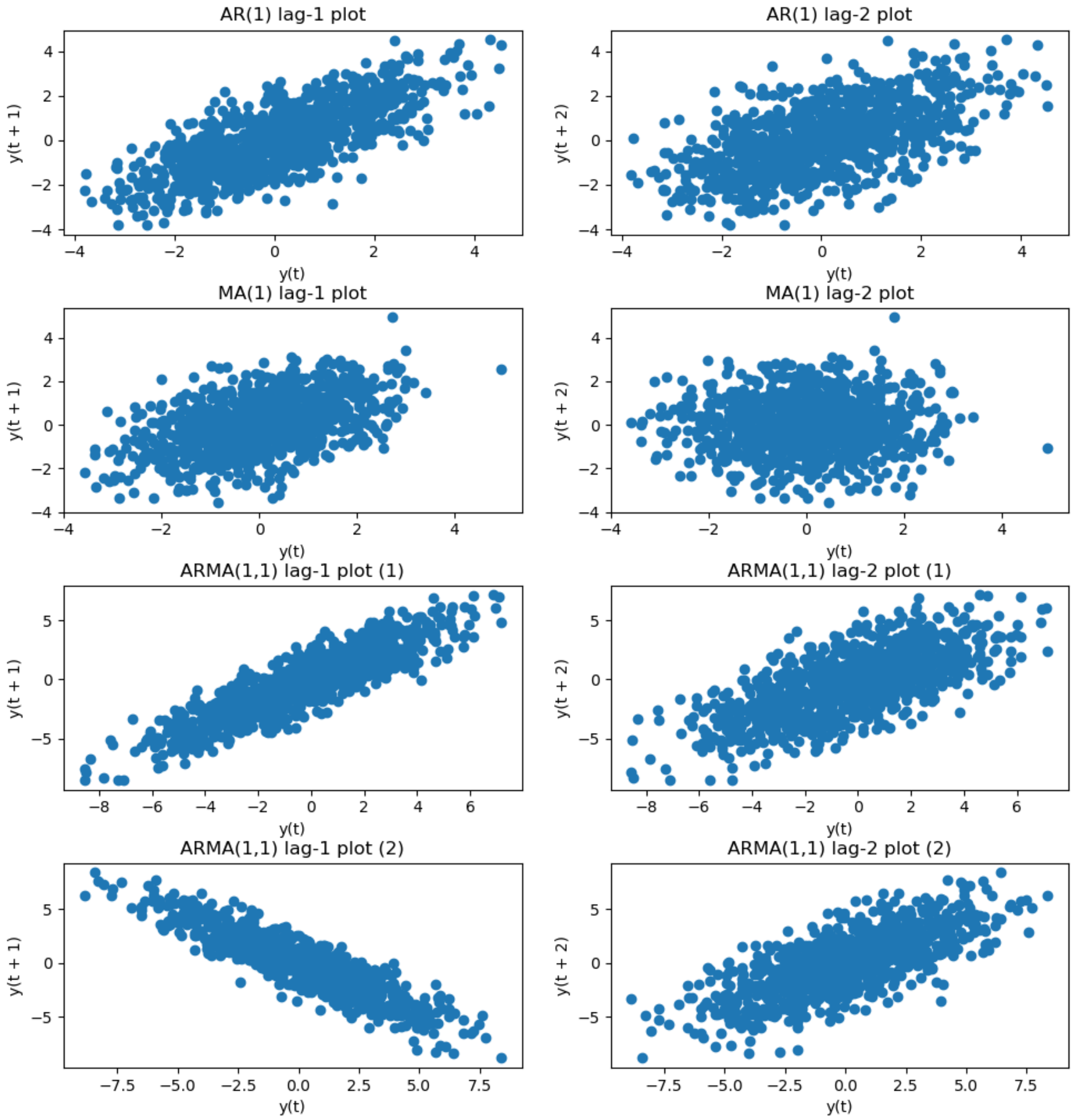
5. Draw histogram, density plot, and box plot for each time series with 1000 data points. Are there any outliers? Why?



According to the plots, there are some outliers.

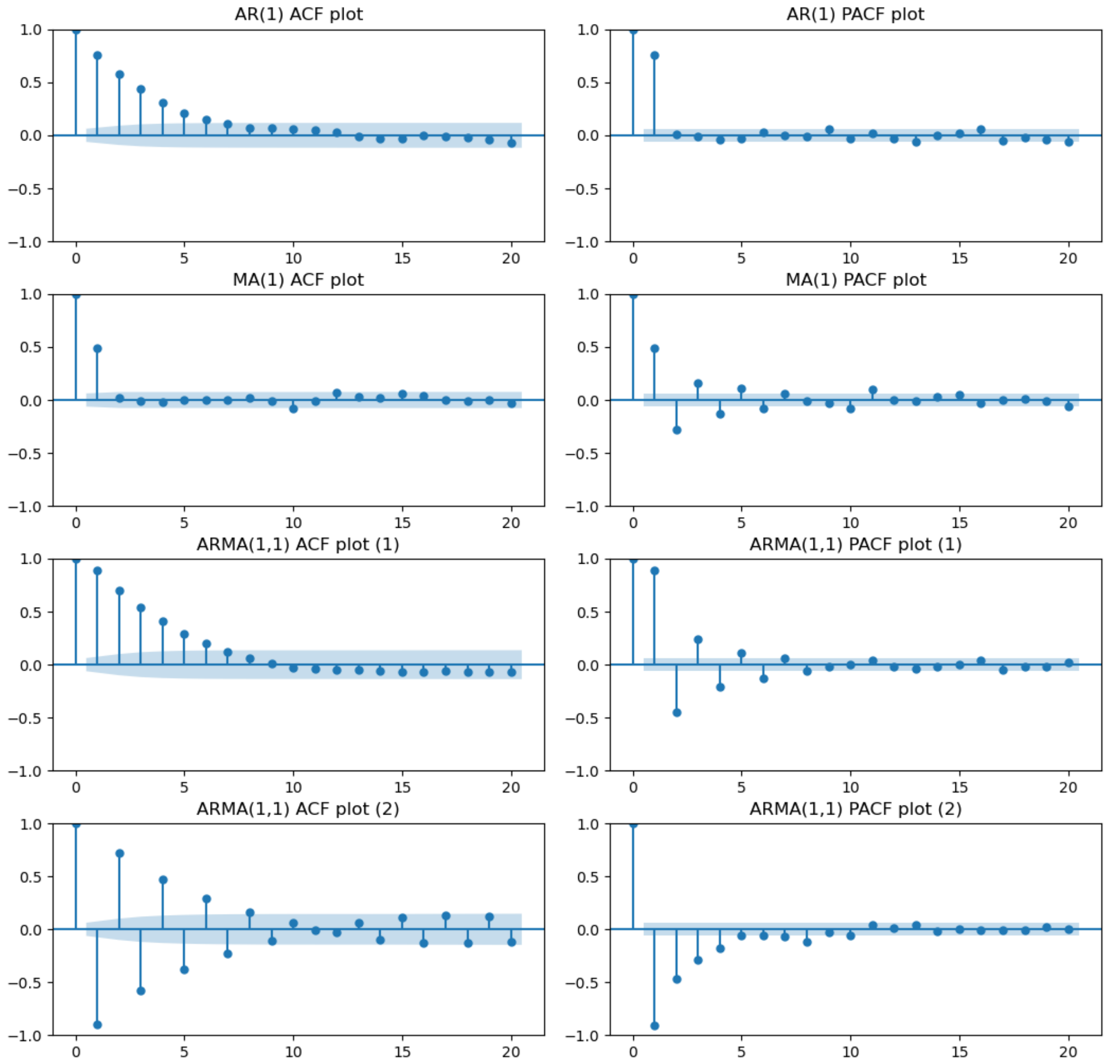
6. Draw lag-1 and lag-2 plots for each time series. Do you observe any auto-correlation from the lag plots?

According to the plots, I can observe that the lag-1 plot and lag-2 plot of AR(1) are positively correlated, the lag-1 plot and lag-2 plot of MA(1) aren't auto-correlated, the lag-1 plot and lag-2 plot of ARMA(1) are positively correlated, the lag-1 plot of ARMA(2) are negatively correlated and lag-2 plot of ARMA(2) are positively correlated.



7. Draw ACF graph for each time series.

8. Draw PACF graph for each time series.



Answer the following questions:

• What characteristics can you observe from the ACF, PACF graphs of the AR(p) model?

The ACF will decay exponentially or gradually decrease, indicating that the autocorrelation is high for the first few lags and then decreases as the lags increase.

The PACF will have significant spikes at the first p lags and then will be close to zero or have no spikes for the remaining lags, indicating that each observation is highly dependent on the previous p observations.

• What characteristics can you observe from the ACF, PACF graphs of the MA(q) model?

The ACF will have significant spikes at the first q lags and then will be close to zero or have no spikes for the remaining lags, indicating that each observation is highly dependent on the previous q errors.

The PACF will decay exponentially or gradually decrease, indicating that the partial correlation between the observations decreases as the lags increase.

• What characteristics can you observe from the ACF, PACF graphs of the ARMA(p, q) model?

The ACF will have significant spikes at the first q lags, followed by gradual decay or no spikes

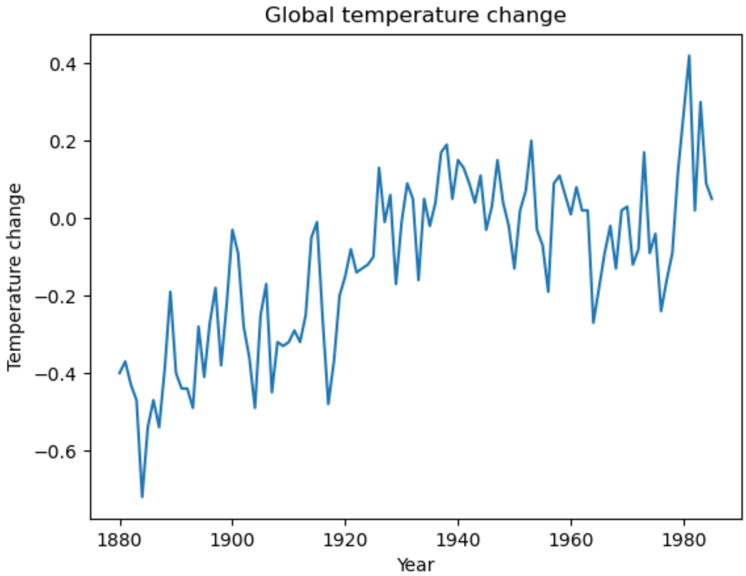
until the p-th lag, and then a sharp drop to zero, indicating that the autocorrelation between observations is high for the first few lags and then decreases slowly.
The PACF will have significant spikes at the first p lags, followed by gradual decay or no spikes until the q-th lag, and then a sharp drop to zero, indicating that each observation is highly dependent on the previous p observations and the previous q errors.

Model	ACF	PACF
AR(p)	gradually decrease	have significant spikes at the first p lags and then will be close to zero or have no spikes for the remaining lags
MA(p)	have significant spikes at the first q lags and then will be close to zero or have no spikes for the remaining lags	gradually decrease
ARMA(p,q)	have significant spikes at the first q lags, followed by gradual decay or no spikes until the p-th lag, and then a sharp drop to zero	have significant spikes at the first p lags, followed by gradual decay or no spikes until the q-th lag, and then a sharp drop to zero

Task 5: ARIMA modeling and prediction

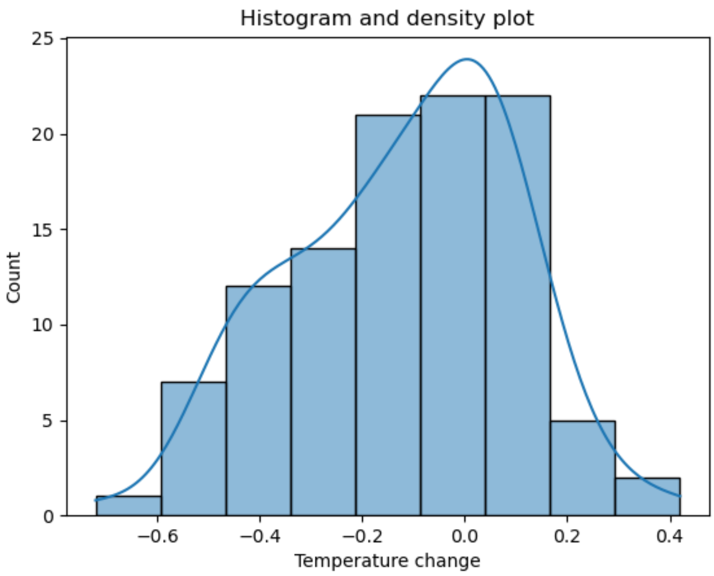
Perform relevant activities and answer the following questions:

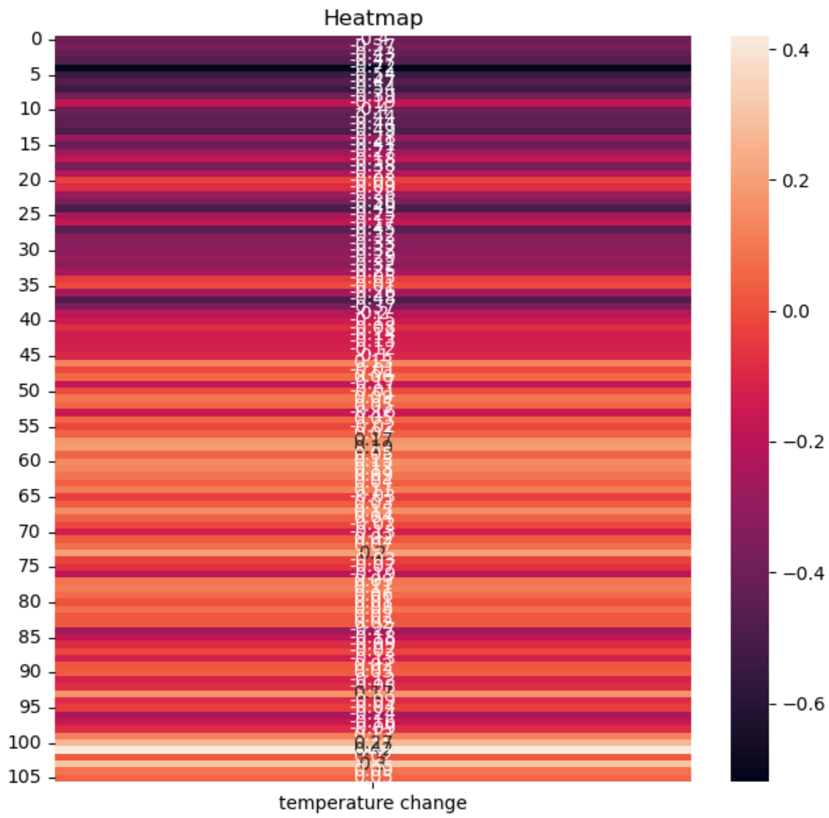
1. Draw a line plot for the time series data. Do you observe any trend, season in the data?

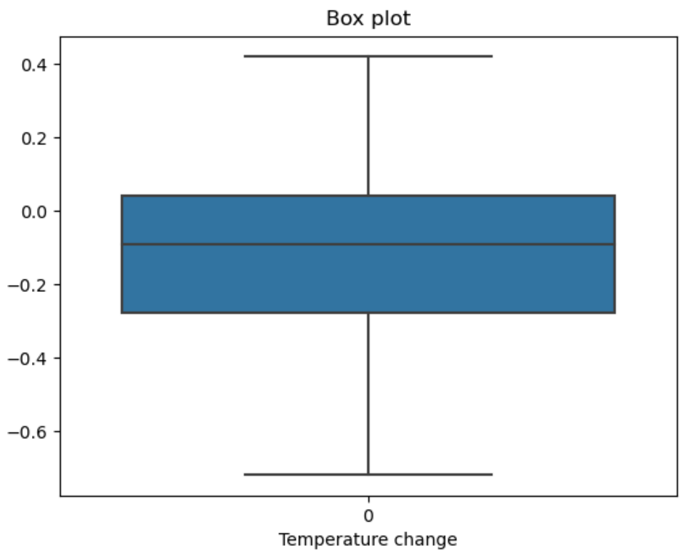


This time series data generally increases.

2. Draw histogram, density plot, heat map, and box plot for the time series data. Are there any outliers? Why?

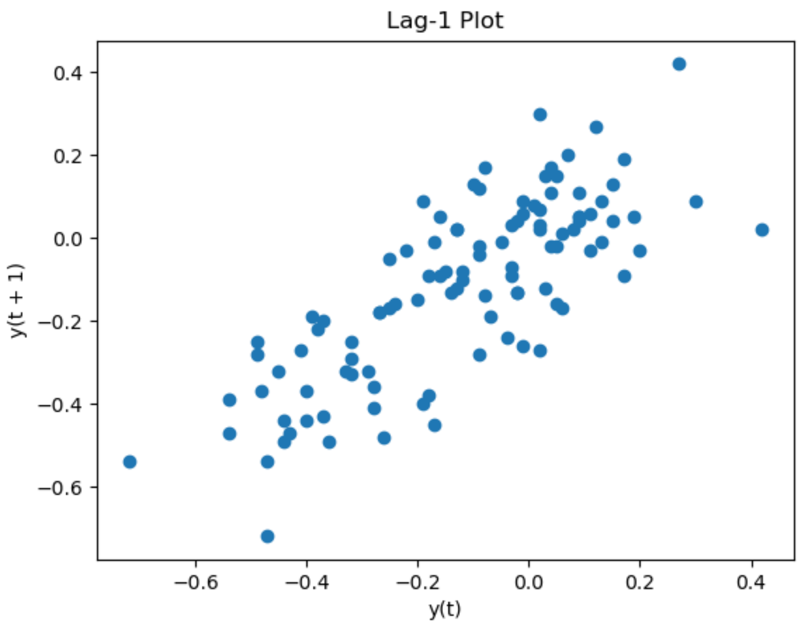

Histogram and density plot
Count
Temperature change
25
20
15
10
5
0
−0.6
−0.4
−0.2
0.0
0.2
0.4

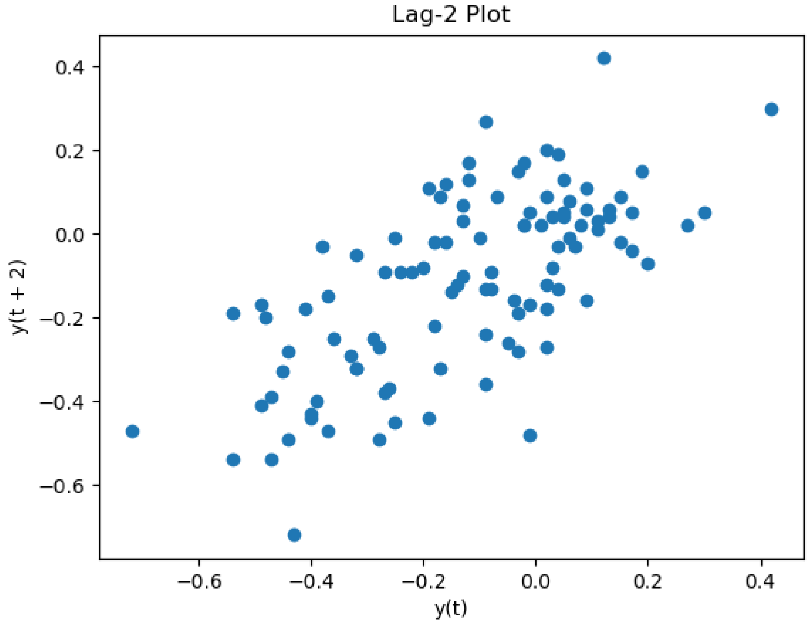

Heatmap
temperature change
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
0.4
0.2
0.0
−0.2
−0.4
−0.6



According to the plots, there are no outliers.

3. Draw lag-1 and lag-2 plots for the time series data. Do you observe any auto-correlation from the lag plots?



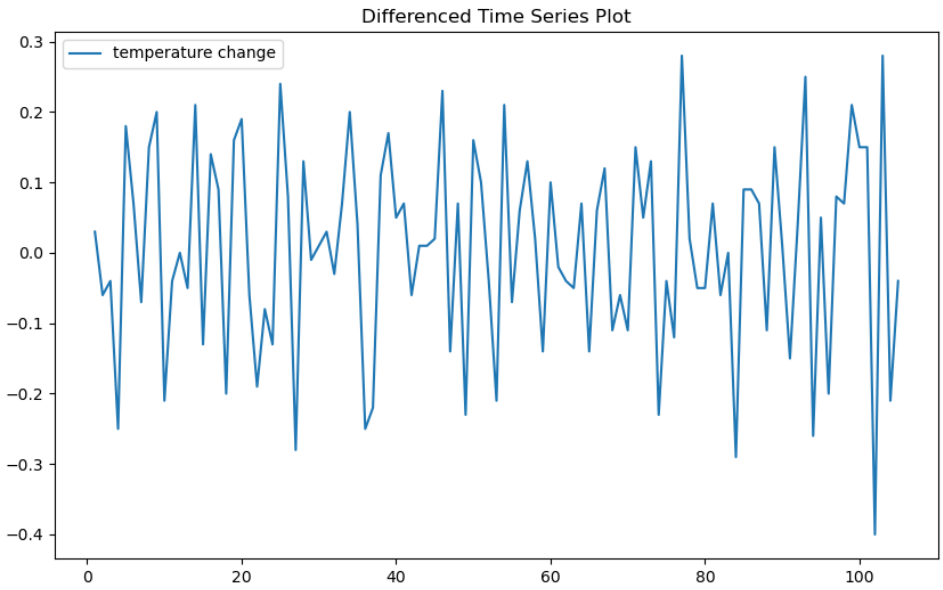


According to the plots, I can observe that both the lag-1 plot and lag-2 plot are positively correlated.

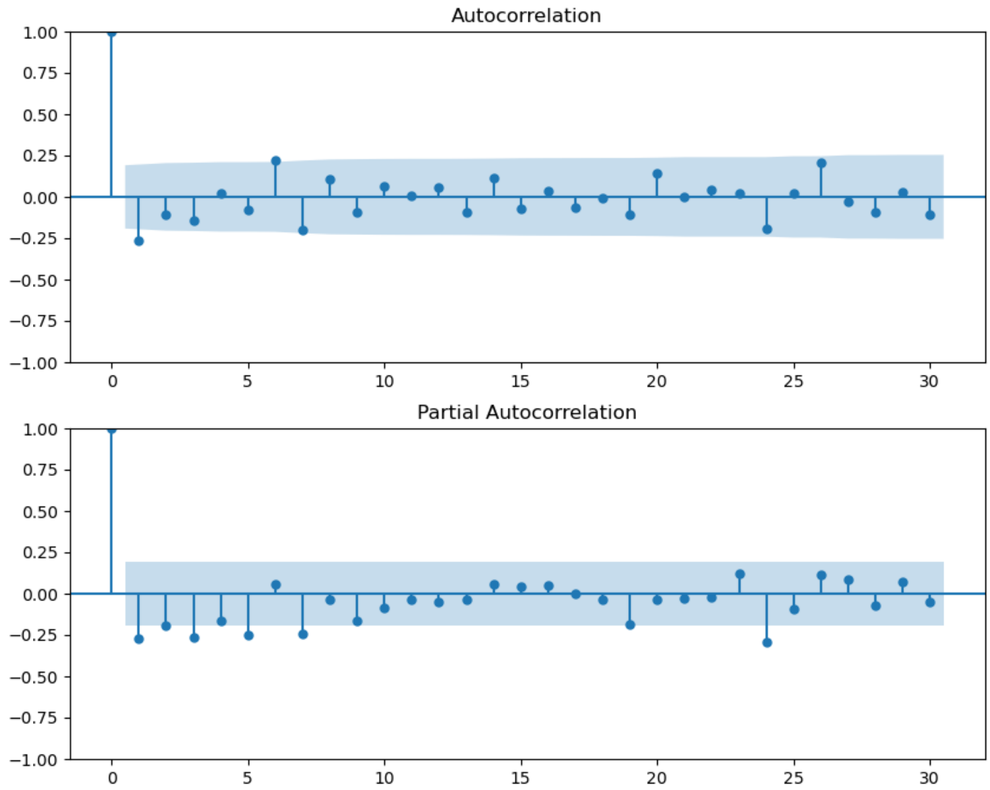
4. Is the series random? How do you check it? Are the three methods (line plot, lag-1 plot, and Ljung-Box test) give the same results?
From the line plot, we can see that there is a clear increasing trend in the temperature change over time, as well as some variability around this trend.
From the lag-1 plot, we can see that the points tend to cluster around a diagonal line, suggesting that there is some auto-correlation in the data.
The resulting p-values for the Ljung-Box test are:
p_value = 6.033530e-68
From the p-values, we can see that they are all very small, indicating that there is significant evidence of auto-correlation in the data.
In summary, the three methods (line plot, lag-1 plot, and Ljung-Box test) all suggest that the time series is not random, but rather exhibits some level of auto-correlation and non-stationarity.

5. Is the series stationary? Try with visual inspection and ADF test. Do they give the same results?
For visual inspection, there is an obvious increase, the series is likely non-stationary.
The resulting p-values for the ADF test are:
p_value = 0.3176856214820573
Since the p-value is much larger than 0.05, we fail to reject the null hypothesis and conclude that the series is non-stationary.
In conclusion, they give the same results.

6. If the series is not stationary, how do you make it stationary? If you use the differencing operation, how do you decide a proper order of differencing without under-differencing/over-differencing?
The series is not stationary, so I use the differencing operation to make it stationary. Since the original time series has a clear trend, I use first differencing.

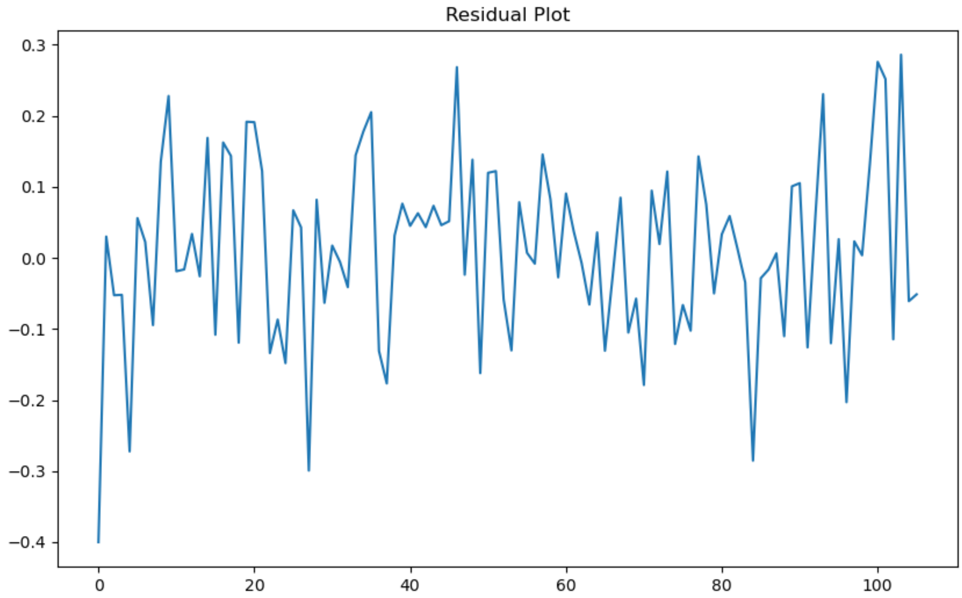


7. What (p, d, q) values do you use? How do you determine them?



I use (1,1,7). I determine p =1 according to ACF plot and determine q =7 according to PACF plot. Besides, because it is first differencing, d is 1.

8. After model fitting, is the reminder series (in-sample prediction) considered to be white noise?



The resulting p-values for the Ljung-Box test are:

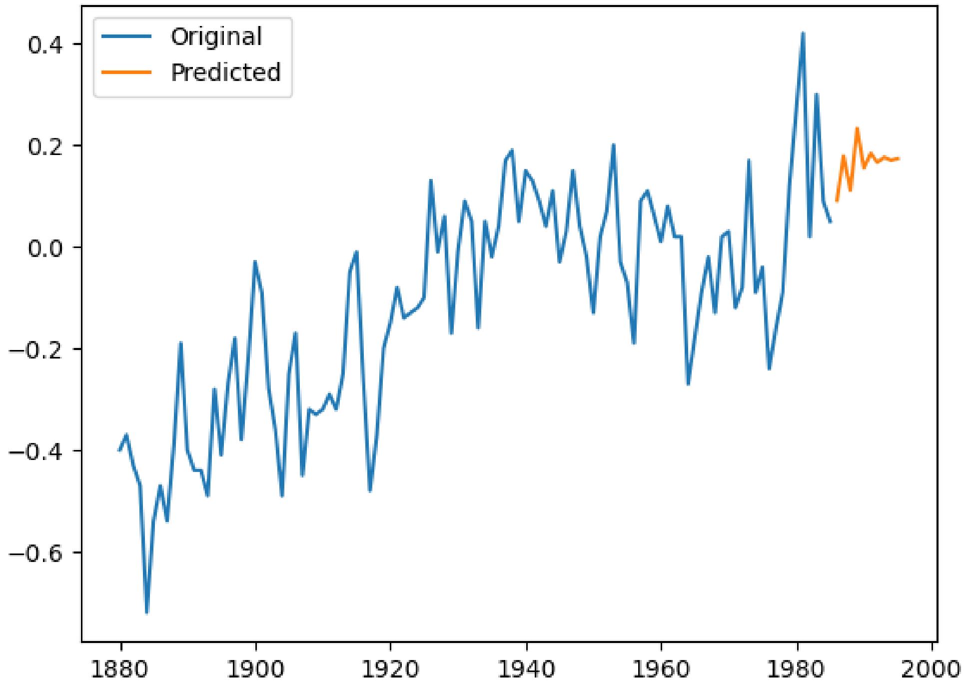
p_value = 0.997225

The p-value is greater than 0.05, so the remainder series is white noise.

9. What is the MSE of the fitted model for the data?

MSE = 0.016219454449012131

10. For out-of-sample prediction, do the predicted values (10 steps) reflect the trend and fluctuation of the series?



Task 6: Series transformation

1. What are the common transformation techniques applicable to turn a non-stationary series into a stationary series?

Differencing: taking the difference between consecutive observations.

Log transformation: taking the natural logarithm of the observations.

Box-Cox transformation: transforming the data using a power function.

2. What is the Box-Cox transform? Give its definition and explain its generality.

$\mathrm{Y}(\lambda) = \frac{Y^{\lambda}-1}{\lambda}$, if $\lambda \neq 0$

$\log(\mathrm{Y})$, if $\lambda = 0$

where Y is the original data, lambda is a parameter that can take any value, and Y(lambda) is the transformed data. The Box-Cox transform is a general transformation method because it can handle a wide range of data distributions, from highly skewed to normal.

3. Can a differencing operation remove a linear trend? Give an example by generating a synthetic series, and draws the series before differencing and after differencing.

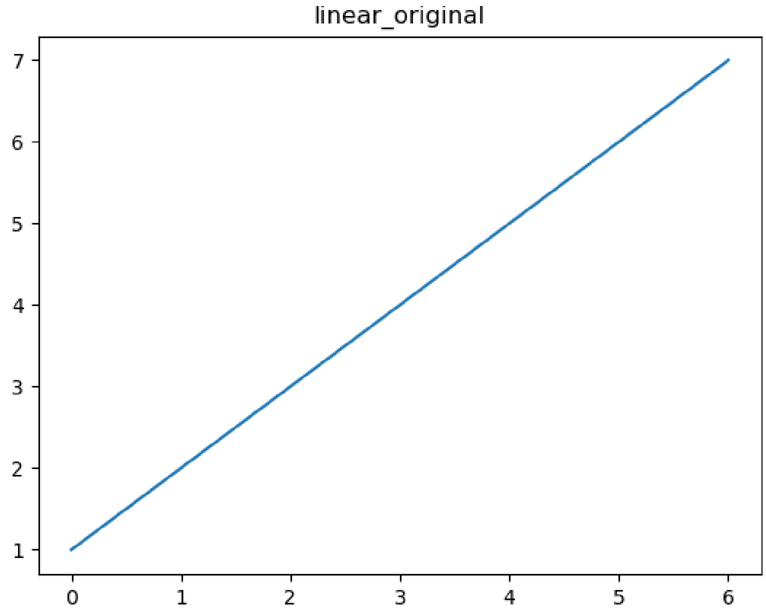
Yes, a differencing operation can remove a linear trend. For example, consider the following synthetic series:

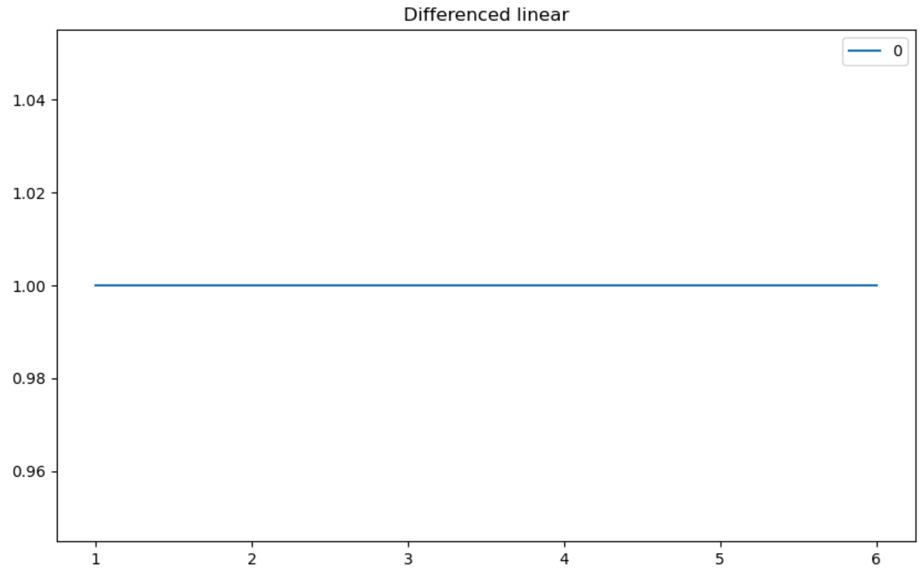
[1, 2, 3, 4, 5, 6, 7]

The trend is clearly linear. The first differences of this series are:

[1, 1, 1, 1, 1, 1]

which removes the linear trend.





4. Can a differencing operation remove an exponential trend? If not, which additional transformation needs to be taken? Give an example by generating a synthetic series, and plots the series before transformation and after transformation, before differencing and after differencing.

No, a differencing operation cannot remove an exponential trend. An additional transformation, such as a log transformation, is needed. For example, consider the following synthetic series:

[1, 2, 4, 8, 16, 32, 64]

The trend is clearly exponential. Taking the first differences of this series gives:

[1, 2, 4, 8, 16, 32]

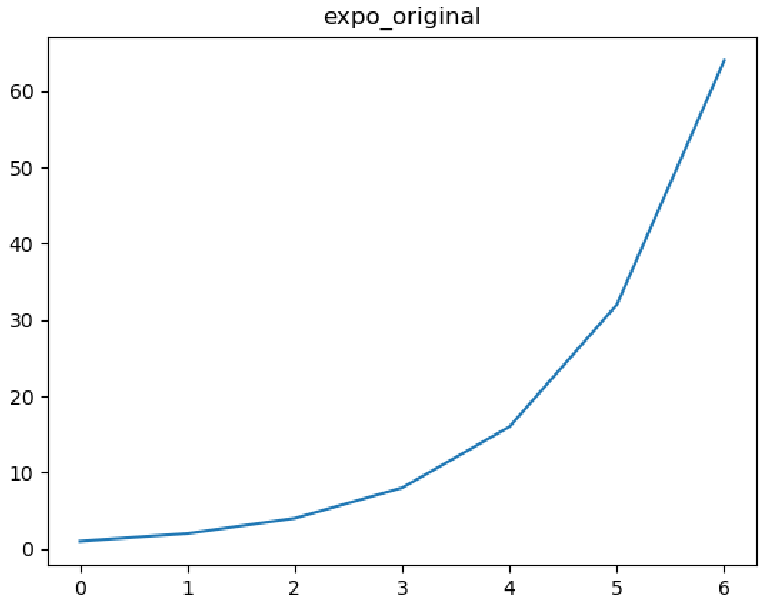
which does not remove the trend. However, taking the logarithm of the series first gives:

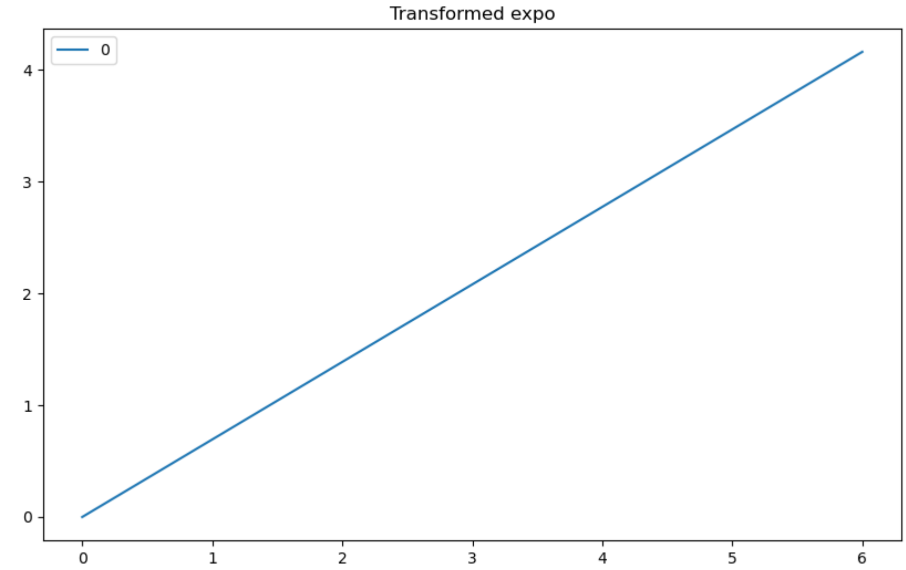
[0, 0.693, 1.386, 2.079, 2.773, 3.468, 4.162]

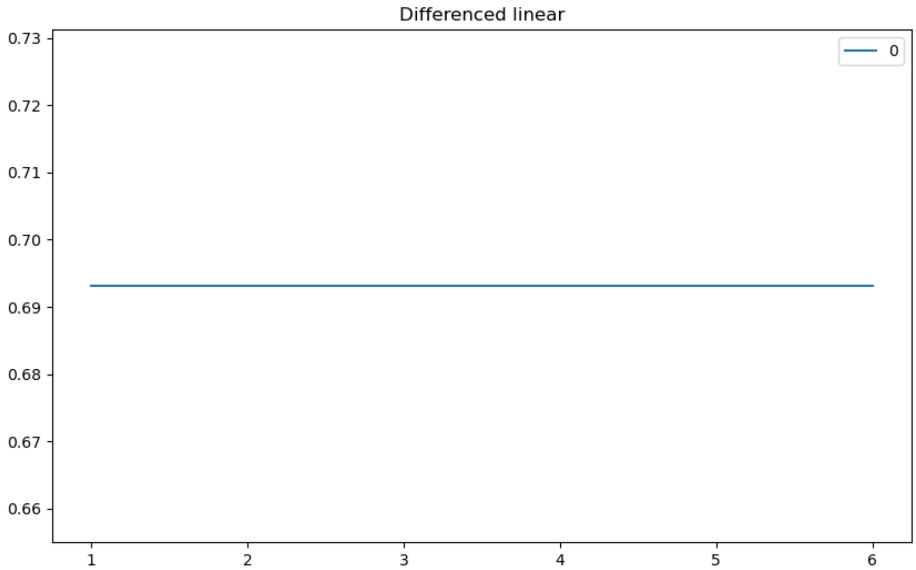
and then taking the first differences gives:

[0.693, 0.693, 0.693, 0.693, 0.695, 0.694]

which removes the trend.

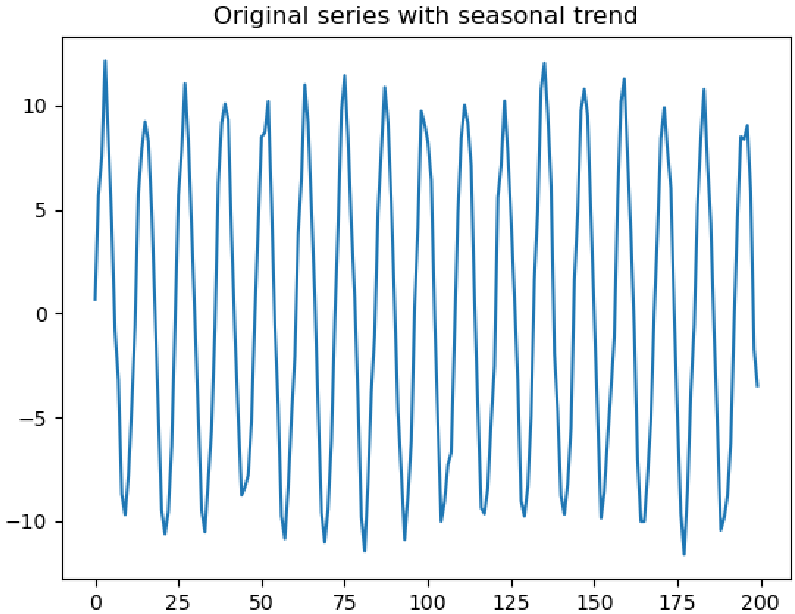


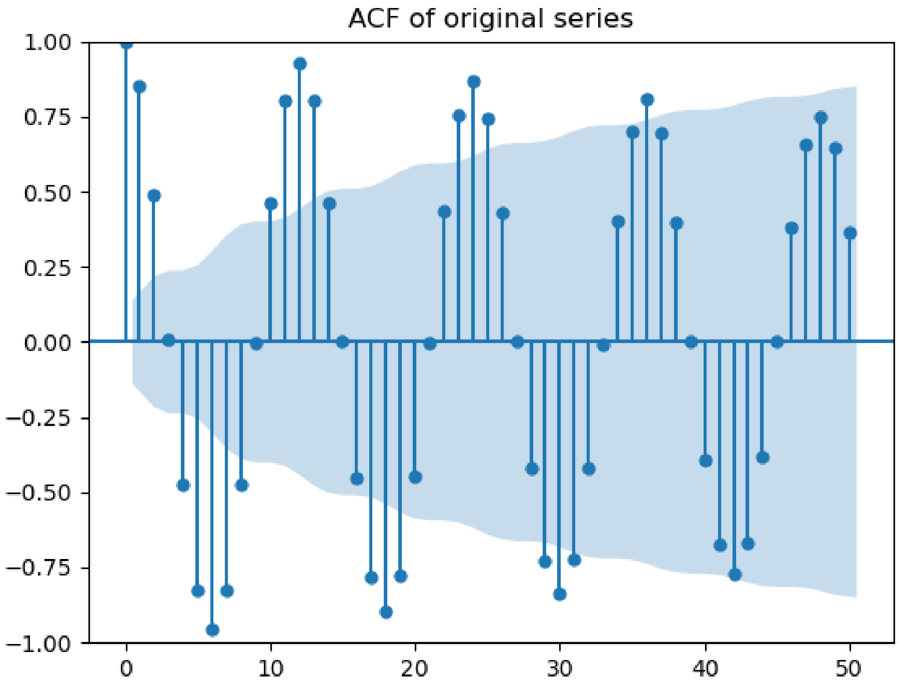


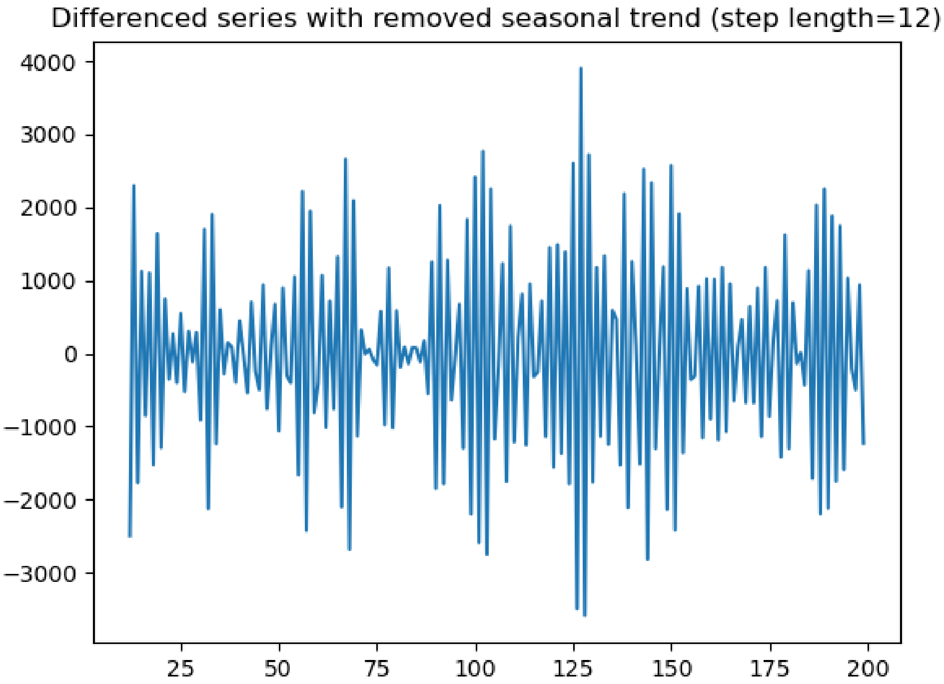


5. Can a differencing operation remove a seasonal (periodic) trend? If yes, under what condition? Give an example by generating a synthetic series, draw its ACF, and draws the series before differencing and after differencing with different step length.

In this example, we generated a synthetic series with a seasonal trend of period 12, meaning that the pattern repeats every year. We plotted the ACF of the series to identify the seasonal period, which is 12 in this case. Then, we applied differencing with step length 12 to remove the seasonal trend. The resulting differenced series shows no apparent seasonal fluctuations, indicating that the trend has been removed.


Original series with seasonal trend
10
5
0
−5
−10
0
25
50
75
100
125
150
175
200


ACF of original series
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40
50



Differenced series with removed seasonal trend (step length=12)
4000
3000
2000
1000
0
−1000
−2000
−3000
25
50
75
100
125
150
175
200