Problem 1

A

With the given dataset, the first four moments values are calculated by the following normalized formulas:

$$
\begin{cases}
\widehat{\mu_1} = \frac{1}{n}\sum x \\
\widehat{\mu_2} = \frac{1}{n-1}\sum (x_i - \widehat{\mu_1})^2 \\
\widehat{\mu_3} = \frac{E[(X-\mu)^3]}{\sigma^3} \\
\widehat{\mu_4} = \frac{E[(X-\mu)^4]}{\sigma^4} - 3
\end{cases}
$$

Round to four decimal places, the first four moments are 1.0490, 5.4272, 0.8793 and 23.0700, respectively.

B

There are imported functions in **'scipy.stats'** library for calculating the first four moments(**tmean, tvar, skew, kurtosis**), and the results are 1.0490, 5.5272, 0.8806 and 23.1222, rounded to four decimal digits. The skewness and kurtosis are slightly different with the above calculated result.

C

The samples are generated by re-sampling in Python, and using hypothesis t test to determine if the statistical package 'scipy.stats' is biased. The H0 is that the package is unbiased, which is the same with the result calculated manually through normalized formula. The H1 is the package is biased. After the t-test, the p-value of variance is 0.53. The p-values of skewness and kurtosis are approximately 3.38e-9 and 6.54e-58, respectively, which are much lower than 0.05.

As a result, **the null hypothesis 'the package is unbiased' cannot be rejected for variance calculation**, while **the null hypothesis 'the package is unbiased' can be rejected for the skewness and kurtosis calculations**.

Problem 2

A

To fit the data by the ordinary least squares (OLS), a one column is stacked to X, shown as:

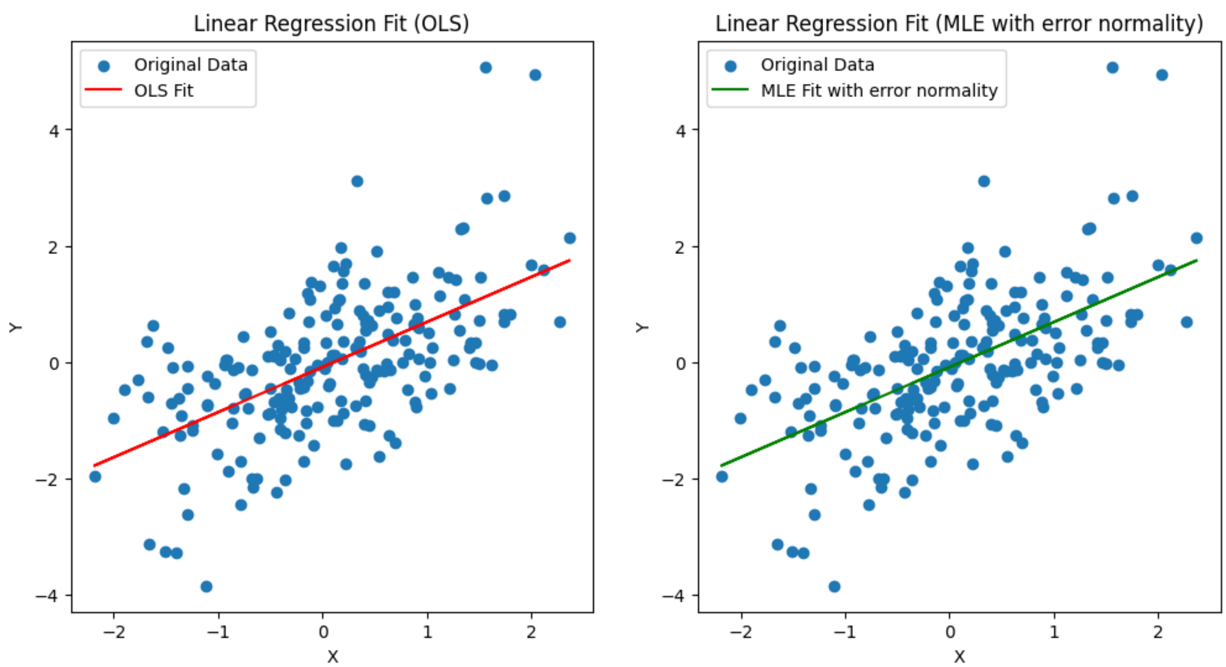
$$X = [1 \ \ X]$$

Then, the intercept and coefficient are -0.0874 and 0.7753, rounded to four decimal digits, calculated by

$$[\hat{\beta}_0, \hat{\beta}_1] = (X'X)^{-1}X'Y$$

Similarly, to fit the data by the maximum likelihood estimation (MLE) with error normality, the negative ll should be maximized. The initial intercept, coefficient and standard deviation are 0.0, 1.0 and 1.0, respectively.

$$ll = -\frac{n}{2}\ln(\sigma^2 2\pi) - \frac{1}{2\sigma^2}\sum \epsilon^2$$



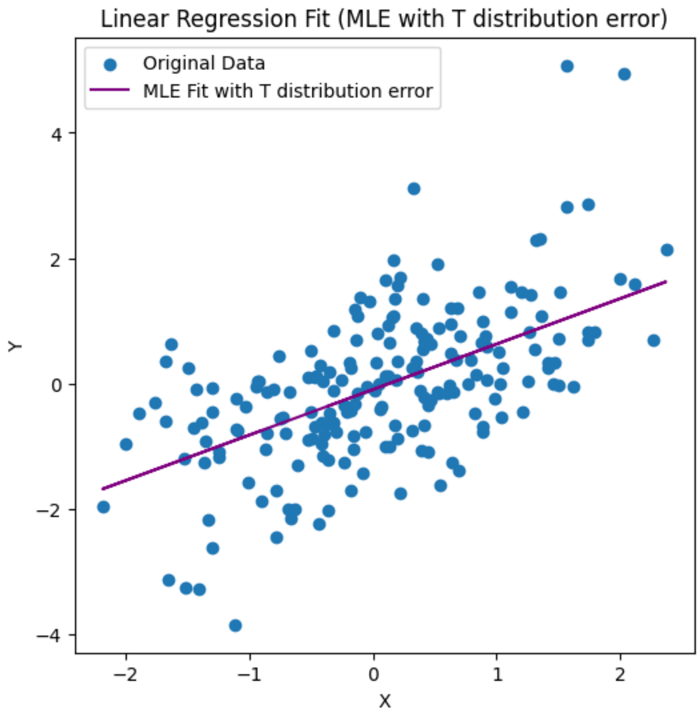
In conclusion, the fitted results are shown in following table:

	OLS	MLE with error normality
Intercept	-0.08738446	-0.08738448
Coefficient	0.7752741	0.7752741
Standard deviation error	1.003756319417732	1.0037563209985647

Where **OLS has slightly better fitness reflected on standard deviation error**.

B

Now, the error distribution is assumed to be T distribution, and the initial number of freedoms is set to 10. Other initial parameters stay the same.



	MLE with error normality	MLE with T distribution error
Intercept	-0.08738448	-0.09619106
Coefficient	0.7752741	0.72658199
Standard deviation error	1.0037563209985647	1.0

Where **the MLE with T distributed error has the best fitness**.

C

When conditional distribution of multivariate normal distribution, there are two sub-variables X1 and X2 (the random values of X1 are given), the distribution of the remaining variable X2 is

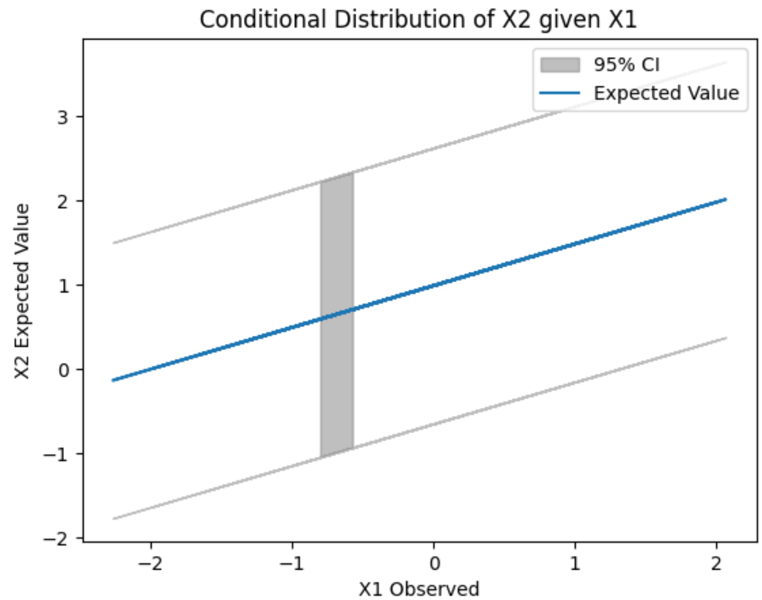
$$X_2 \sim N(\bar{\mu}, \bar{\Sigma})$$

Given X1 = a

$$\bar{\mu} = \mu_2 + \Sigma_{21}\Sigma_{11}{}^{-1}(a - \mu_2)$$

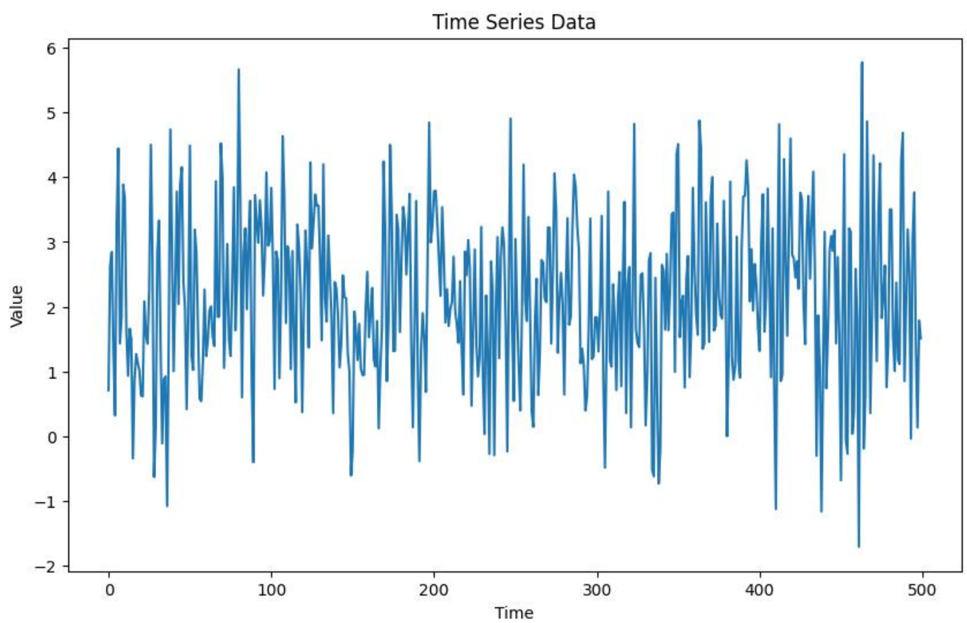
$$\bar{\Sigma} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}{}^{-1}\Sigma_{21}$$

In addition, there are 100 observed X1 samples. Therefore, the expected values of X2 are calculated with 95% confidence interval by the normal distribution. The line chart of the result is shown below:

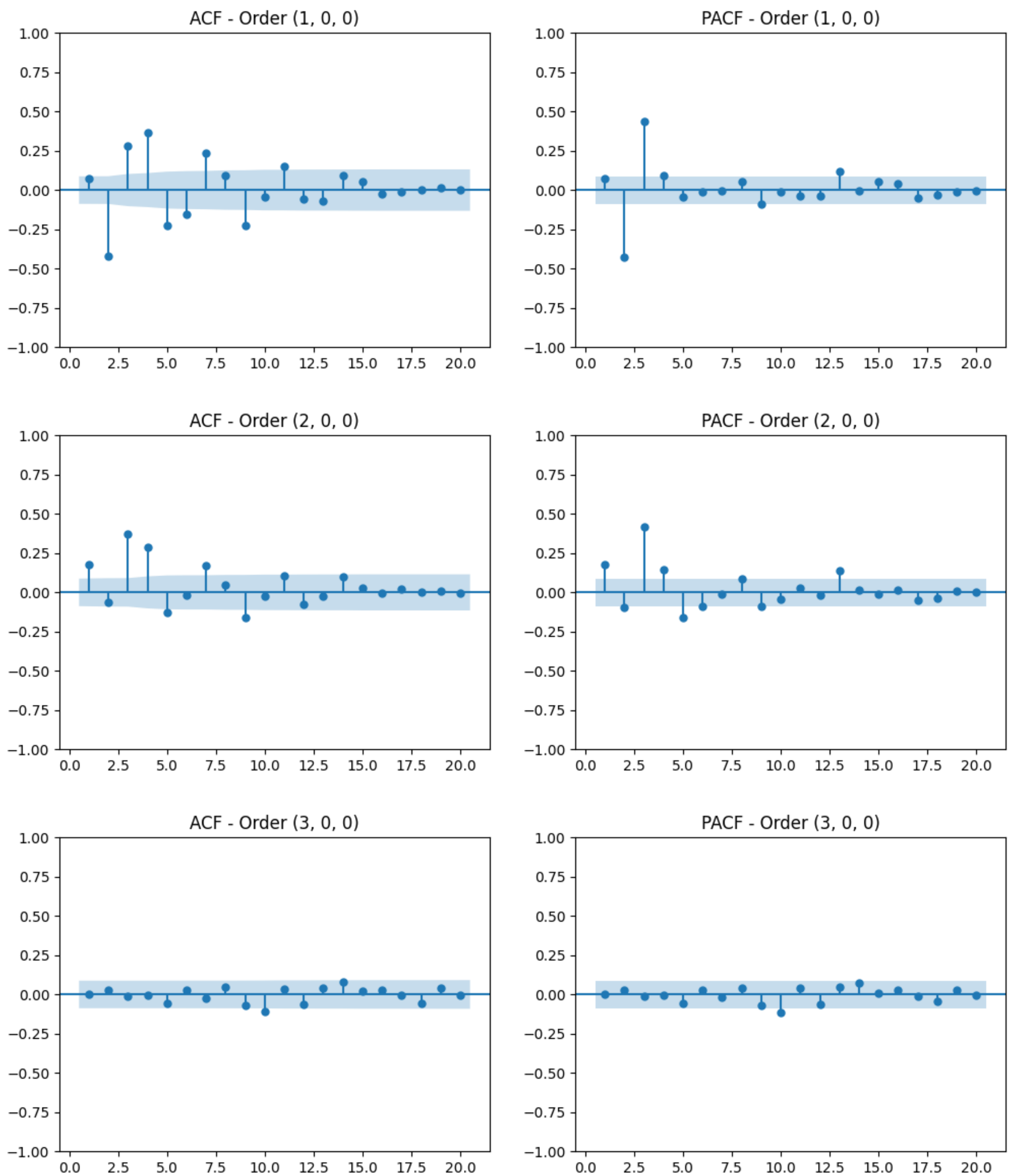


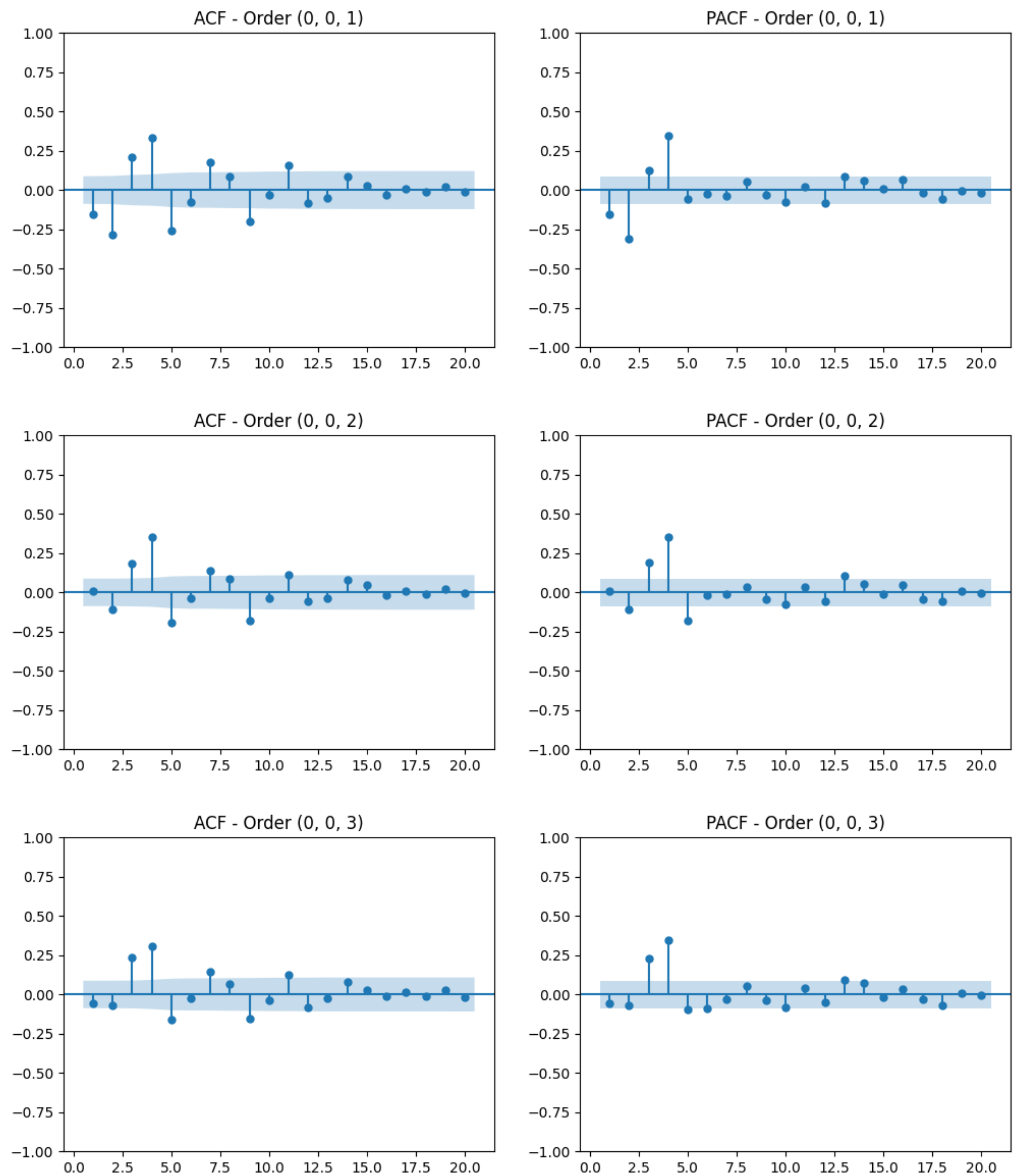
Problem 3

The original data is shown below:



With the use from AR(1) to AR(3), from MA(1) to MA(3), the ACF and PACF figures are shown below:


ACF - Order (1, 0, 0)
PACF - Order (1, 0, 0)
ACF - Order (2, 0, 0)
PACF - Order (2, 0, 0)
ACF - Order (3, 0, 0)
PACF - Order (3, 0, 0)



It is clear that the adoption of **AR(3) provides the most stable fit**, and the information criteria (AICs and BICs) tell **AR(3) has the best fitness with the lowest AIC and BIC**.

	AIC	BIC
AR(1)	1644.655505	1657.299329
AR(2)	1581.079266	1597.937698
AR(3)	1436.659807	1457.732847

MA(1)	1567.403626	1580.047451
MA(2)	1537.941206	1554.799639
MA(3)	1536.867709	1557.940749