Week 02 Project

Fintech 545 Jiwei Xia

Problem 1

QUESTION:

The conditional expectation of the multivariate normal should be equal to the expected value from an OLS regression. Use the data in problem1.csv to prove your answer empirically.

Problem 1

SOLUTION:

To solve this question with "problem1.csv" datasets with two variables x and y, we will need to do following steps.

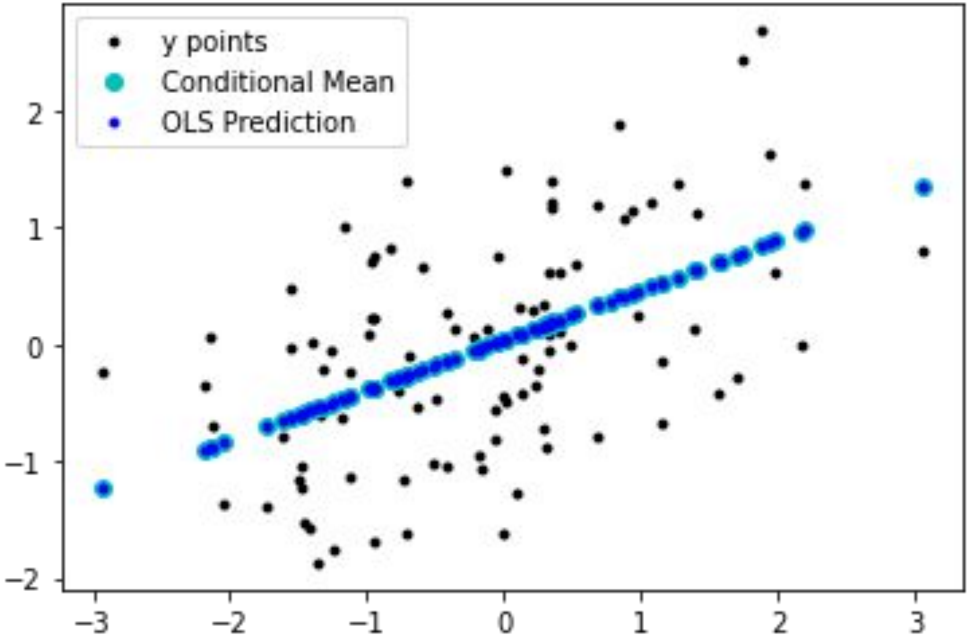
1.1 Given that: $\bar{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$

As we assume x,y variables in our datasets are independent, we calculate their mean and variance and apply it to get the conditional distribution.

1.2 Using the sklearn packages in Python, we can fit the variable x with Ordinary Least Square.

1.3 Compare the result between step1 and step2

Problem 1



Problem 1

ANSWER:

With the plot, we saw the results from Conditional Mean and OLS Prediction are completely overlapped, which empirically indicates that these two are identical.

Problem 2

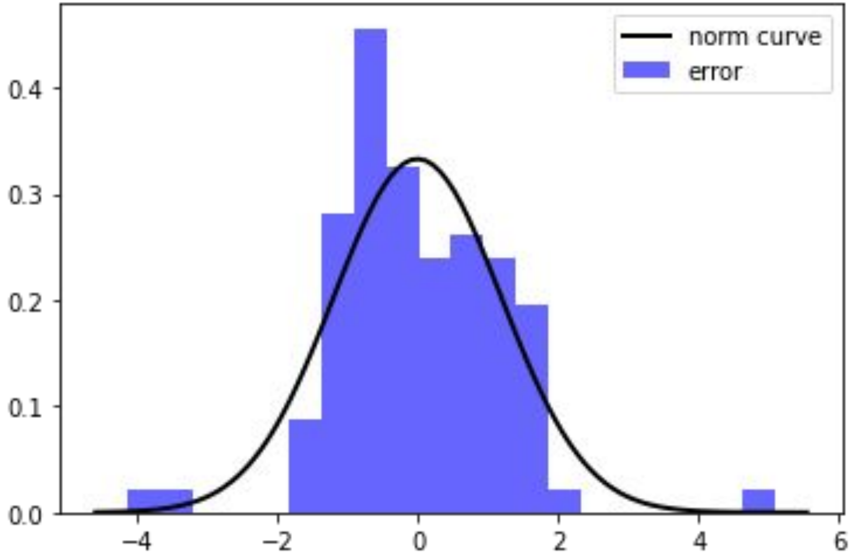
QUESTION:

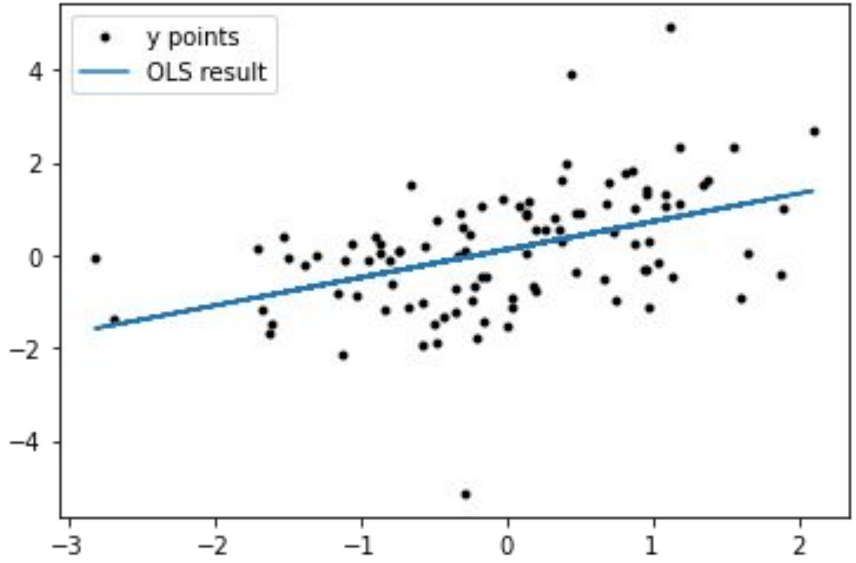
2.1 Fit the data in problem2.csv using OLS and calculate the error vector. Look at its distribution. How well does it fit the assumption of normally distributed errors?

2.2 Fit the data using MLE given the assumption of normality. Then fit the assumption of a T distribution of the errors. What is the best fit?

2.3 What are the fitted parameters of each and how do they compare? What does this tell us about the breaking the normality assumption in regards to the expected values in this case?

Problem 2.1


norm curve
error
0.4
0.3
0.2
0.1
0.0
−4
−2
0
2
4
6


y points
OLS result
4
2
0
−2
−4
−3
−2
−1
0
1
2

Problem 2.1

ANSWER:

As it shown on the graph, the error vector is basically following the normal distribution but not very well. So we tried using the Shapiro-Wilks test and found that the p value is around 0.00015, which is significantly smaller than the commonly applied alpha-level 0.05. As a conclusion, the null hypothesis can be rejected, meaning the errors are not normally distributed.

Problem 2.2

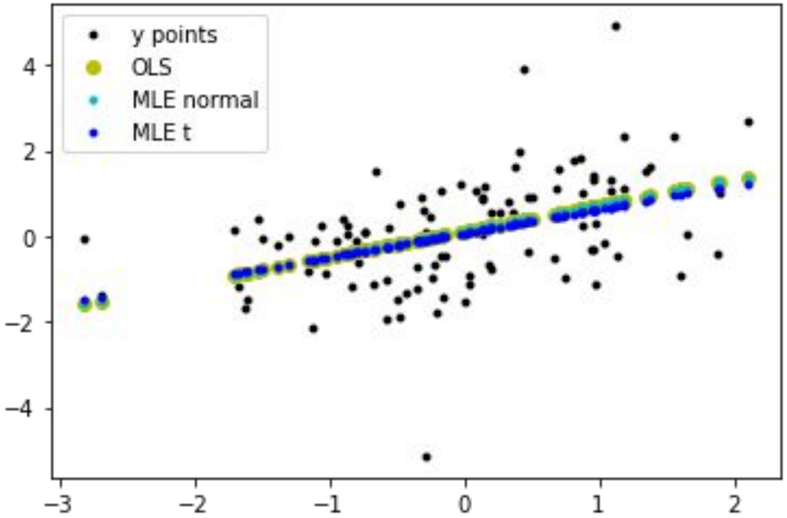
SOLUTION:

Fit the data using MLE with normal distribution/ T distribution assumption.

	SSE	AIC	BIC
Normal Distribution	143.6148	325.9842	333.7997
T Distribution	144.0063	319.0306	329.4513

ANSWER: Given the results, we found that the one with t distribution assumption has a slightly higher SSE but a much lower AIC and BIC. Thus, we conclude that t distribution has a better fit.

Problem 2.3



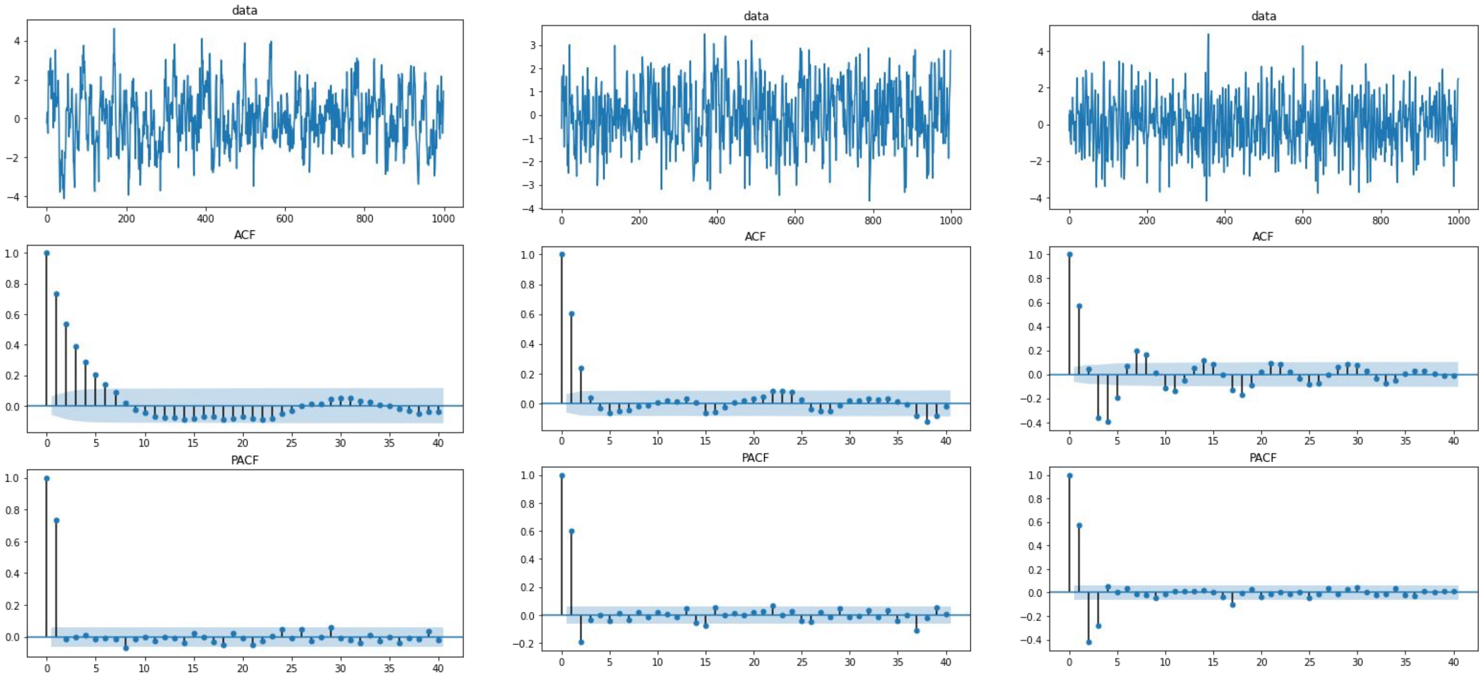
ANSWER: We can see from the plot and the values, that the k and b parameters are similar for both cases, but not the same. This tells us that if the error doesn't follow normal distribution, then expected value of y based on MLE will become different from OLS.

Problem 3

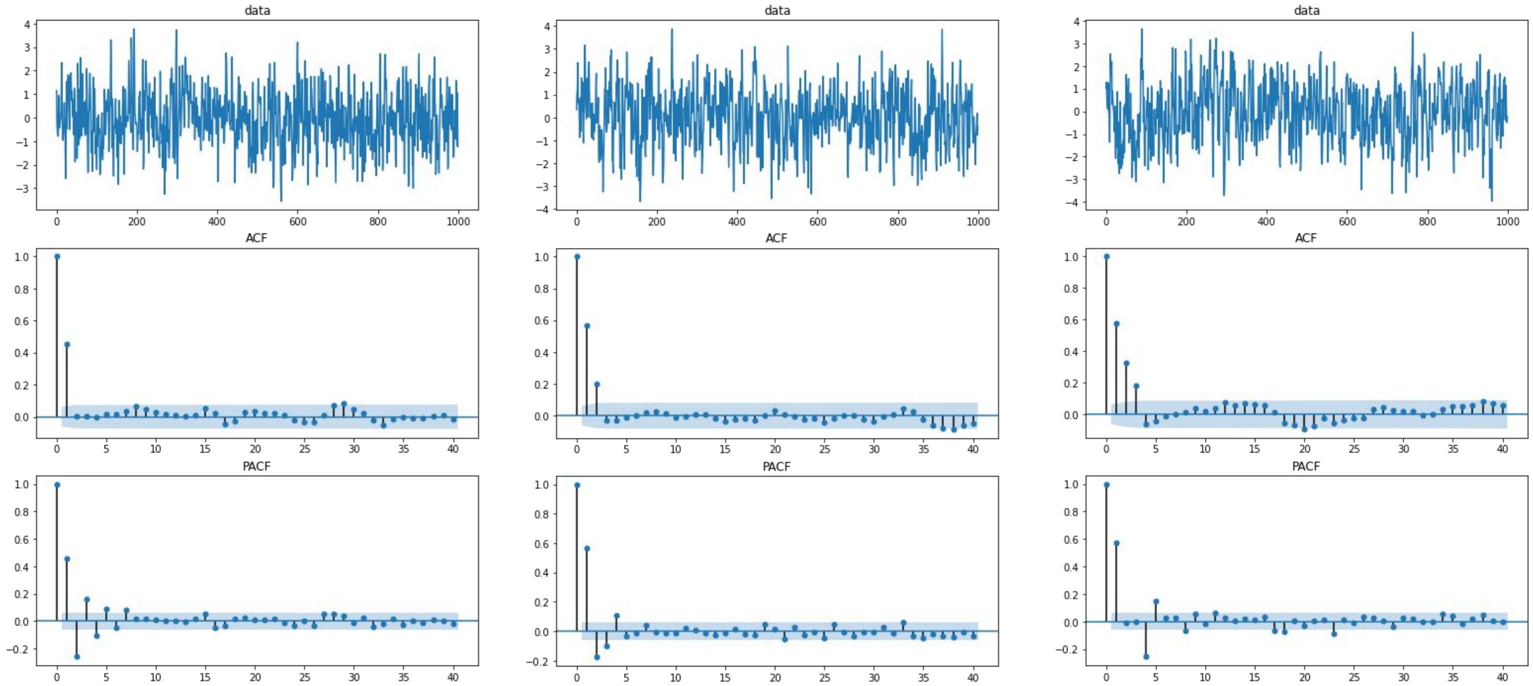
QUESTION:

Simulate AR(1) through AR(3) and MA(1) through MA(3) processes. Compare their ACF and PACF graphs. How do graphs help us to identify the type and order of each process.

Problem 3.1 Simulate AR(1) through AR(3)



Problem 3.2 Simulate MA(1) through MA(3)



Problem 3

ANSWER:

Given the result, we will see that the ACF and PACF plots together help us differentiate the AR/MA process and their term. For instance, we expect to see that in AR process, the ACF plot will gradually decrease while the PACF has a significant drop after the n significant lags. For the MA process, two plots behave the opposite.