Problem1

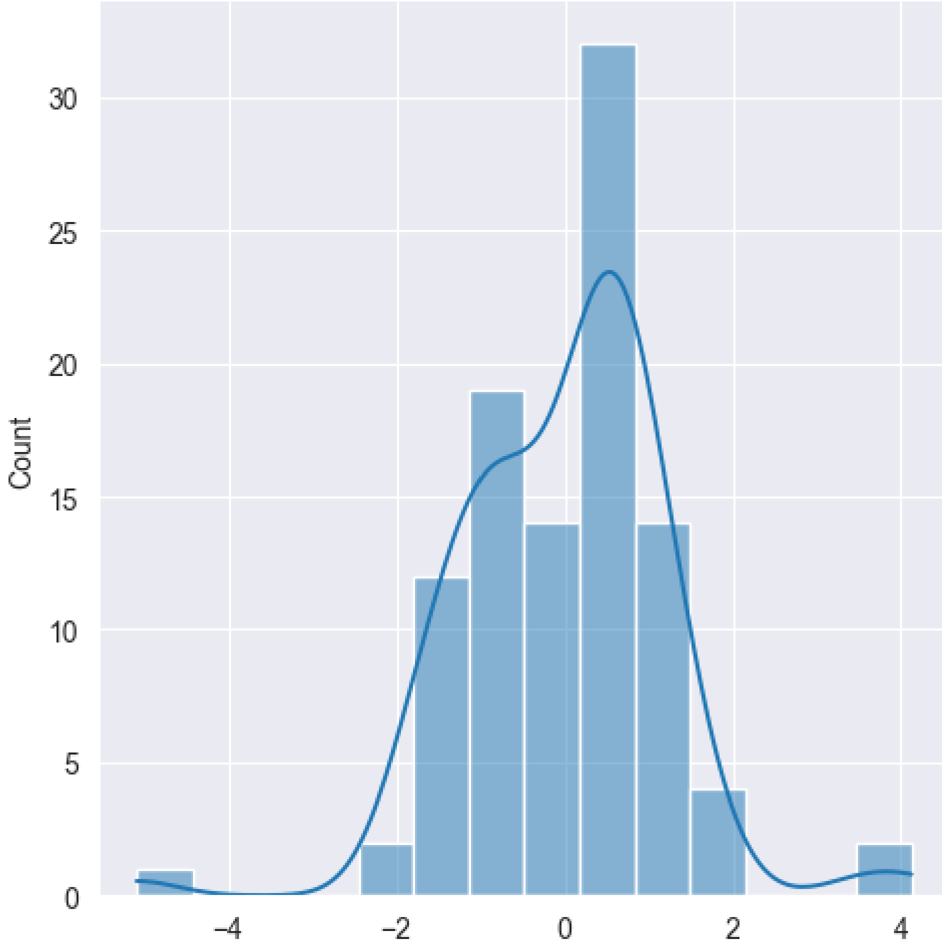
Kurtosis is biased, but skewness can not be proved biased through statistical test.
I used the sample size of 100, and repeated for 1000 times. Each time, I generated 100 random variates from the standard normal distribution and calculated the skewness and kurtosis. Through this process, we can get the sample mean for skewness and kurtosis. I have the null hypothesis that the true mean of kurtosis and skewness should be 0. Then we can have the T statistic and perform the t test. The results for the t test are below.
We can see that the p value for kurtosis is close to 0, while the p value for skewness is high. We can not reject the null hypothesis that skewness is unbiased at 5% significance level.

	Kurtosis	**Skewness**
t statistics	-4.052	1.0201
p value	5.465e-05	0.307

Problem2

Fit data using OLS

First, I fit the data using OLS formula to calculate the beta and then get the error term. (I used both Python included OLS model and formula by hand, and results are similar).
The skewness for error is -0.267, which is close to 0. However, the kurtosis is 3.193, which means the distribution for error term has a fatter tail than the normal one. The distribution plot for error term is below. Thus, it does not completely fit the assumption of normally distributed error term.



Fit data using MLE

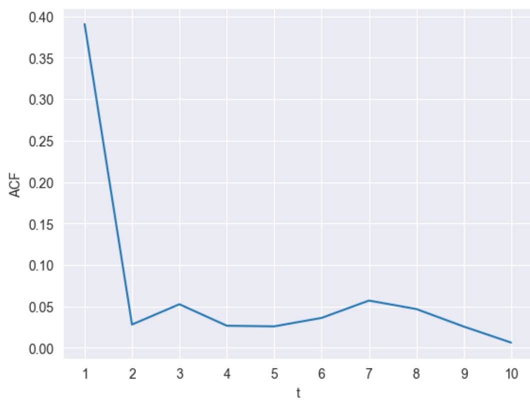
Under MLE method,T distribution assumption for error term fits better than normal assumption. We can see that both AIC and BIC show that T distribution is a better fit.
The fitted parameters for all three methods are in the table below.The MLE method of normal assumption get the same estimator result with OLS. Standard deviation of error is higher in normal assumption than in T assumption, which shows that there is possibility of breaking the assumption of normality. Because the larger the variance for the error term, the more likely that the error term does not have a constant variance as assumed.

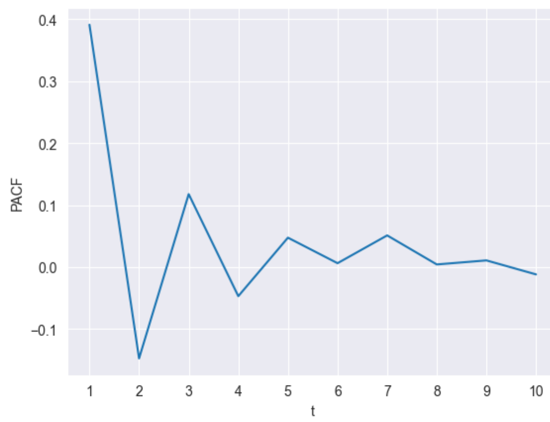
	MLE_Normal	MLE_T	OLS
beta	[0.12,0.61]	[0.14, 0.56]	[0.12,0.61]
std(error)	1.198	0.971	/
ll	-159.992	-155.472	/
AIC	325.984	318.94	323
BIC	333.780	329.37	325.6

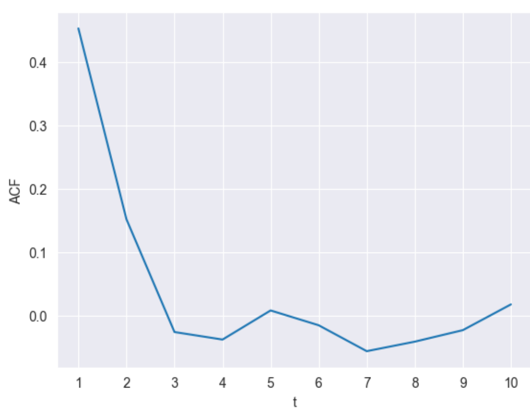
Problem3

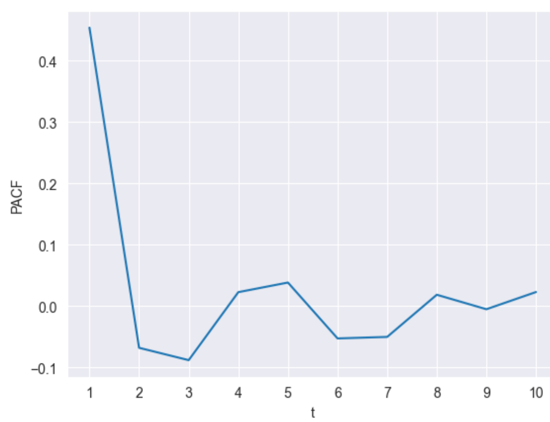
Graphs for MA process are below. From up to down are MA(1),MA(2),MA(3) in order.

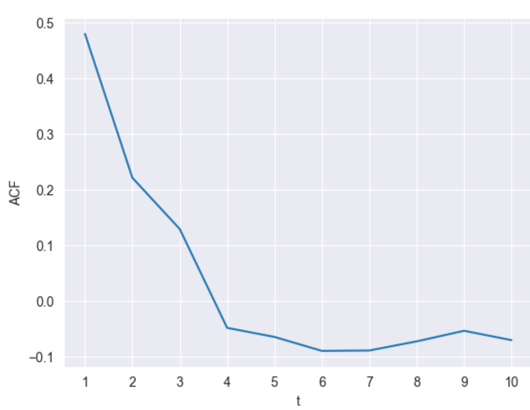
ACF	PACF

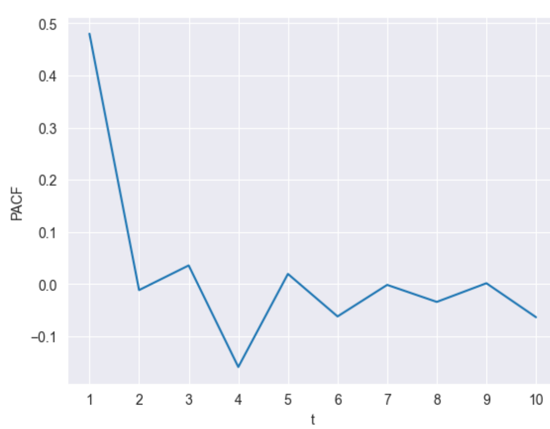




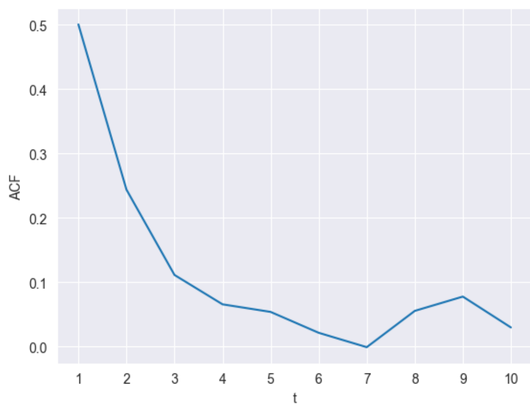
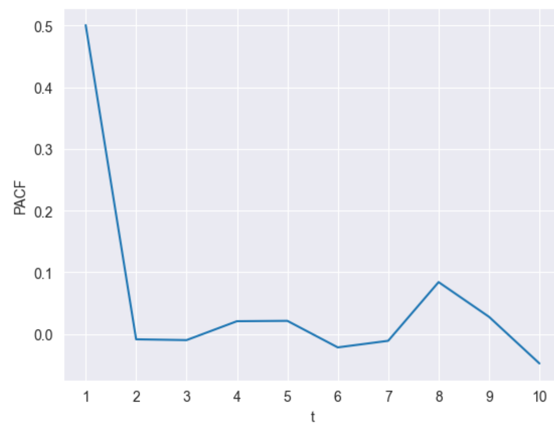
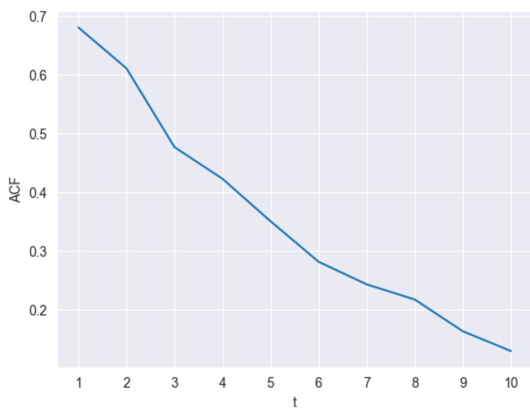
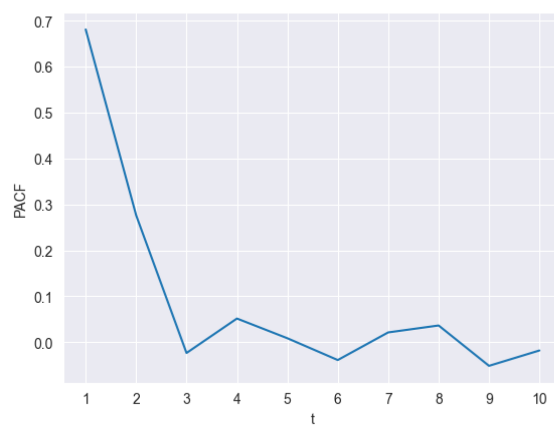
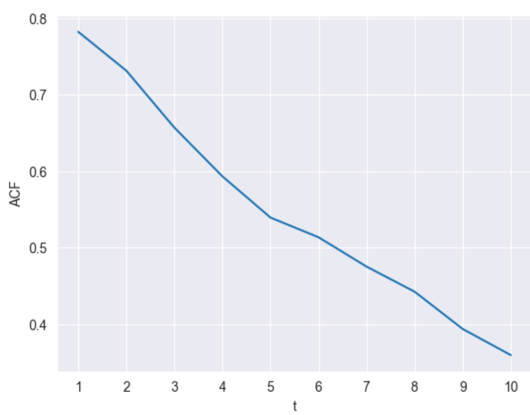
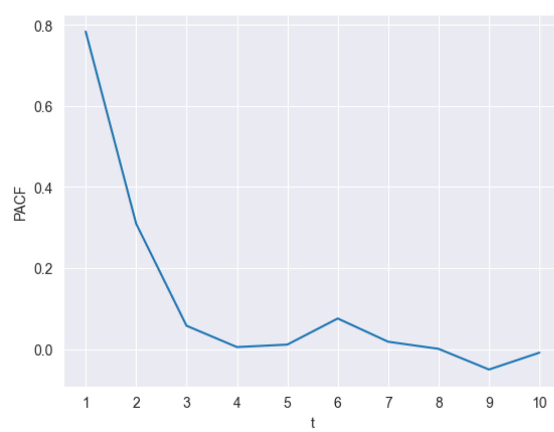








Graphs for AR process are below. From up to down are AR(1),AR(2),AR(3) in order.

ACF	PACF
	
	
	

In general, from both ACF and PACF order can be interpreted at the time lag when it first shrinks to zero or close to zero, which means that the lag terms after that do not have much correlation. Specifically, PACF is more informative for AR while ACF is more informative for MA. Because PACF adjusted for the part that has already been captured by the lags before k-n, it doe not make sense for MA. MA does not has y lags to explain the current y.
For example, in the MA graphs above, we can see that only t below orders have a significant bigger value than 0. We can interpret the order by finding the last t that has a significant postive value. The same applies to the PACF graphs for AR.
We can also tell the difference between AR and MA through ACF graphs, because AR have ACF that declines smoothly, with all the lags postive, while MA have ACF shrinks sharply to 0 at the order. This is because AR has correlation with all the past lags.