Project Week 02

Yuemin Tang

Fintech 545

Problem 1

Remember from last week we discussed that skewness and kurtosis functions in statistical packages are often biased. Is your function biased? Prove or disprove your hypothesis.

To test if skewness and kurtosis functions are biased, I generated random data from a standard normal distribution and repeated for 1000 times. Each time, I used the function kurtosis() or skew() to calculate the kurtosis or skewness. Then the hypothesis test is conducted. The null hypothesis H0 is that the function kurtosis or skewness is unbiased, which means that the mean of kurtosis or skewness for the sample should be 0. The significance level of the test is 0.05. If the p-value of the test is less than 0.05, the null hypothesis will be rejected and we can say the function is biased. Else, the null hypothesis cannot be rejected and we cannot conclude that the function is biased. I conducted the hypothesis test on different sample sizes of 100, 1000, and 10000. The p-value results are shown in Table 1:

Table 1

	size=100	size=1000	size=10000
Kurtosis	**0.0019**	0.1686	0.5539
Skewness	0.6718	0.6720	0.9232

We can see that the p-value for the kurtosis function when having a 100 sample size is smaller than 0.05, and it's greater than 0.05 for other cases. Therefore, the kurtosis function can be proved to be biased when the sample size is 100, and as the sample size increases, the bias effect decreases. For the skewness function, we cannot show that it's biased under the current data sample size and significance level.

Problem 2

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?

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

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

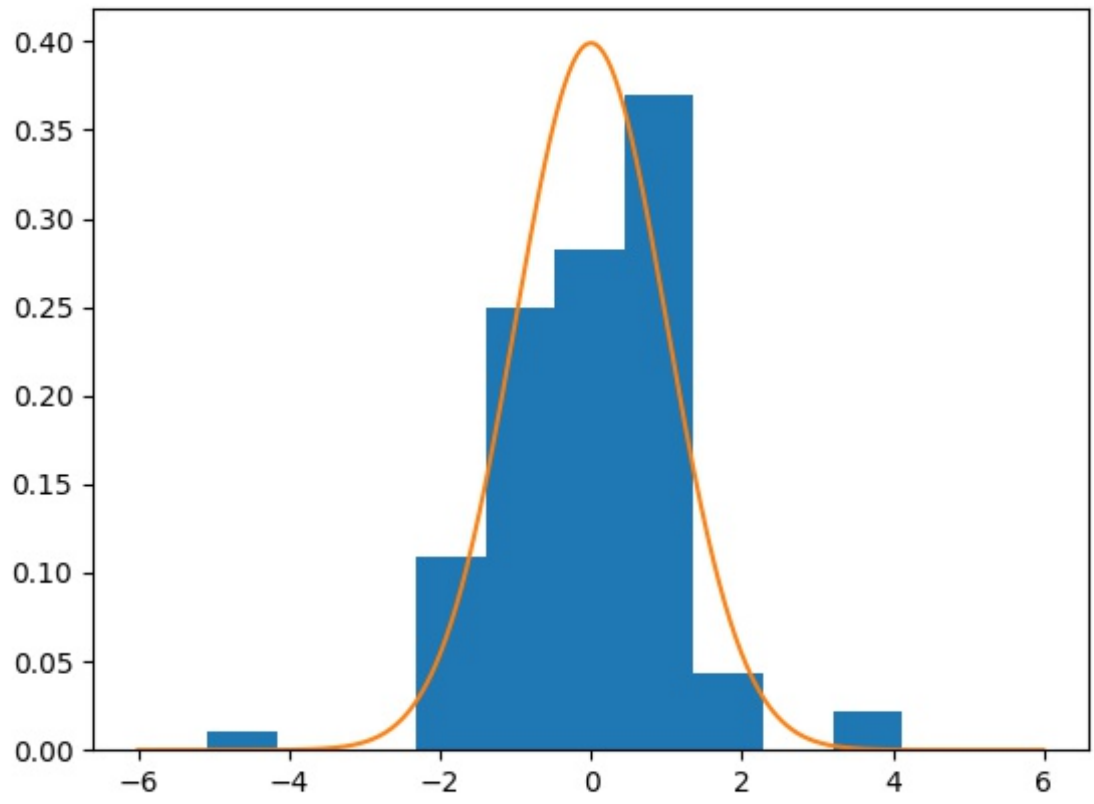
First, read the data and fit the data using the OLS model. The summary of the model is shown in Figure 1:

Figure 1

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.195
Model:                            OLS   Adj. R-squared:                  0.186
Method:                 Least Squares   F-statistic:                     23.68
Date:                Fri, 27 Jan 2023   Prob (F-statistic):           4.34e-06
Time:                        23:41:32   Log-Likelihood:                -159.99
No. Observations:                 100   AIC:                             324.0
Df Residuals:                      98   BIC:                             329.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.6052      0.124      4.867      0.000       0.358       0.852
const          0.1198      0.121      0.990      0.325      -0.120       0.360
==============================================================================
Omnibus:                       14.146   Durbin-Watson:                   1.885
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               43.673
Skew:                          -0.267   Prob(JB):                     3.28e-10
Kurtosis:                       6.193   Cond. No.                         1.03
==============================================================================
```

The error vector can be calculated by using the .resid in python. To see if the errors are normally distributed, I calculated the skewness and kurtosis of the error vector. The skewness is -0.2673 and the kurtosis is 3.1931. These values have big differences with the value of skewness and kurtosis for normal distribution, which should both be 0. Also, I plotted the histogram of errors and compare it with the normal distribution curve. The graph is shown in Figure 2:

Figure 2



It can be seen that the error distribution has a fat tail and it is not normally distributed.

For MLE, first fit the data using MLE in both normal distribution and T distribution. We can get the betas for each model, as well as their log-likelihood. These data are shown in Table 2:

Table 2

	MLE for Normal	MLE for T
Beta 0 (Const)	0.1198	0.1426
Beta 1	0.6052	0.5576
LL (log-likelihood)	−159.9921	−155.4730

To see which model fit the data better, I calculated the R^2 and Information Criteria for each model and compare them. The value of R^2, AIC, and BIC of normal distribution and T distribution is shown in Table 3:

Table 3

	MLE for Normal	MLE for T
R^2	0.1946	0.1934
AIC	323.9842	314.9459
BIC	329.1945	320.1563

We can see that the value of R^2, AIC, and BIC are all smaller for T distribution than for normal distribution. Although the values of R^2 for the two models are close, the difference between their information criteria is obvious. This shows that the model of T distribution is a better fit for the data.

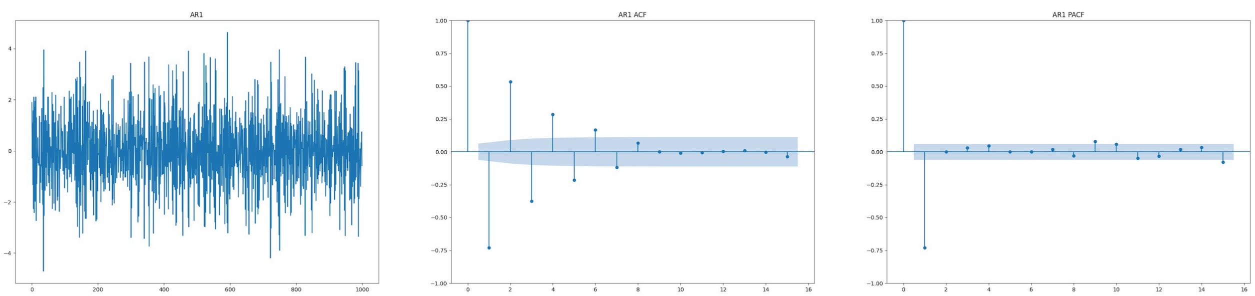
The result shows that the breaker of normality assumption happens a lot and brings a limit for us to use the OLS model. In this situation, we can use MLE to fit the data to get a better model.

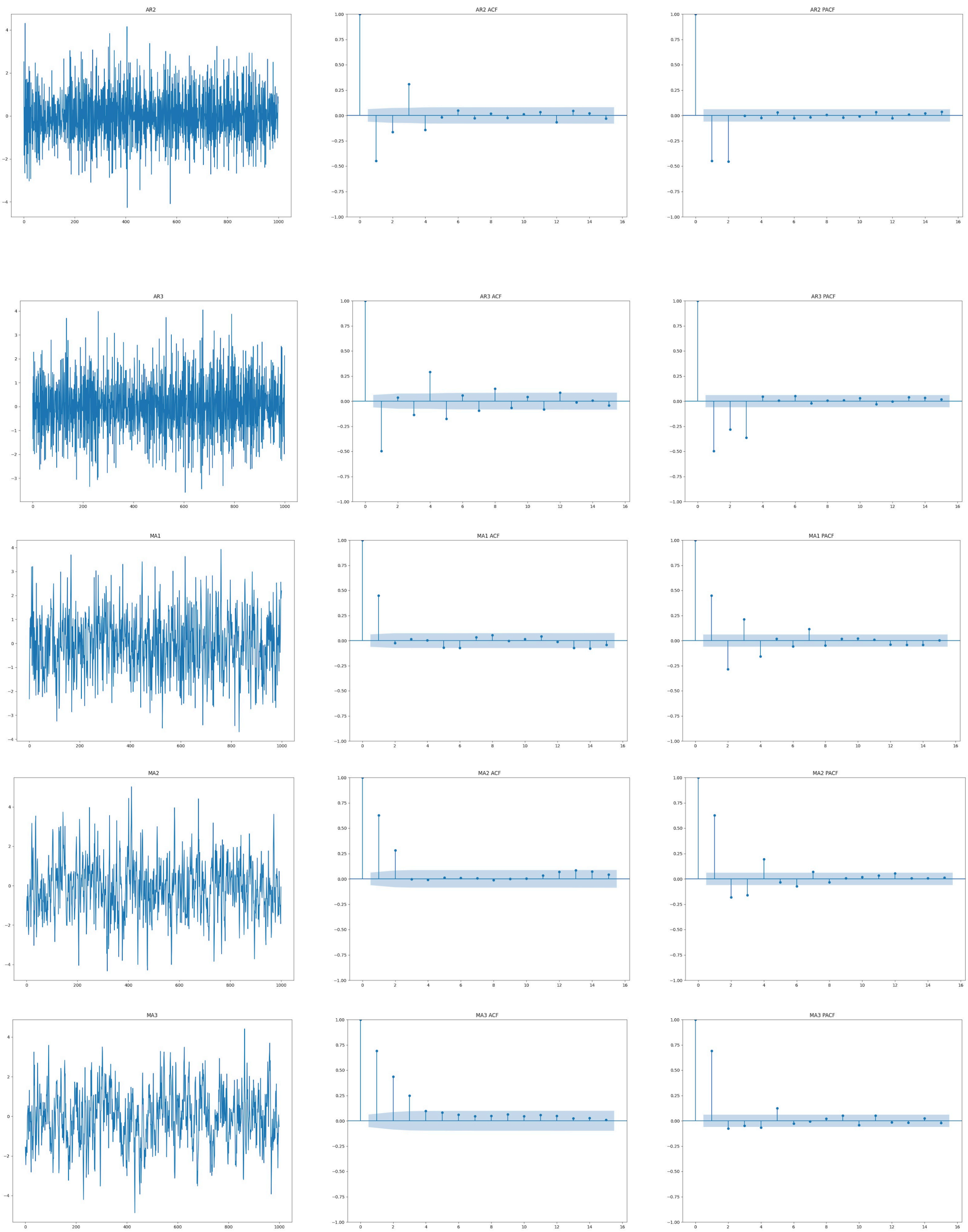
Problem 3

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

First, simulate AR(1), AR(2), AR(3), MA(1), MA(2), MA(3) and plot their ACF and PACF. These graphs are shown in Figure 3:

Figure 3





From these graphs, we can see that for AR processes, the PACF has a sharp cutoff while the ACF shows a smoother decreasing trend. And for MA processes, it's the opposite. Their ACF has a sharp cutoff while their PACF shows a smoother decreasing trend. Also, AR processes all have a negative value at lag = 1, while for MA at lag 1, the values of autocorrelation are all positive. This can be used to identify the type and order of each process. These characteristics can help us to identify whether the process is AR or MA.