Zhiqiang Tian Fintech 545 09/10/2023

Week 02 Project

Problem 1:

Both of the functions are biased. There are two ways to justify the conclusion:

(1) The code computes the kurtosis and skewness of multiple random samples, each drawn from a normal distribution. It then conducts a hypothesis test to assess whether the calculated kurtosis or skewness values display bias. If the p-value obtained from the test falls below 0.05, it concludes that the built-in kurtosis function exhibits bias; otherwise, it determines the function to be unbiased.

The results indicate that the kurtosis function is biased. However, the sensitivity of skewness is relatively lower, and the stability of the t-test results is poor. Another method for assessment can also be employed.

(2) By inspecting the source code of the two functions within the Scipy package, it becomes evident that the Scipy package controls bias by using a "biased" variable, which defaults to "True." Thus, the function is biased when a "biased" value is not explicitly passed as a parameter.

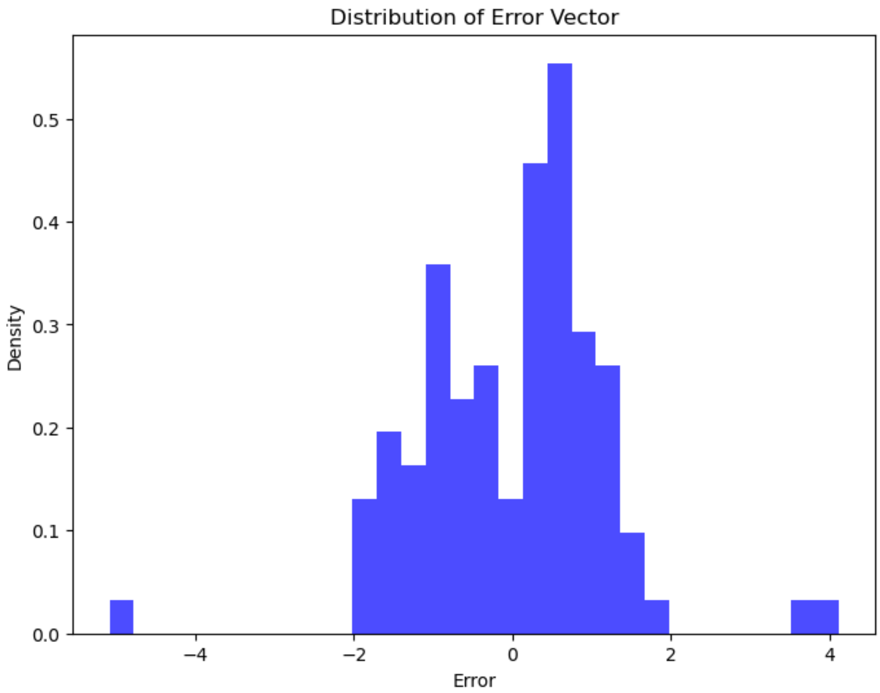
Problem 2:

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?

(1) The distribution graph of the error vector is shown as below:



It is challenging to determine whether it conforms to the assumption of normally distributed errors. Therefore, I computed the excess kurtosis and skewness. The results indicate that the excess kurtosis is significantly higher than 0 (~3.19), suggesting that it may not adhere to the assumption.

(2) The result of normal distribution is:

```
MLE Results (Normal Distribution):
b0: 0.11983620833636628
b1: 0.6052048090554638
error: 1.1983941257586068
AIC: -313.98419337832456
BIC: -306.1686828203603
```

The result of T distribution is:

```
MLE Results (t-Distribution):
b0: 0.23430842746282665
b1: 0.5656446626396994
error: 1.0
df: 1.0354843259541995
AIC: -343.66596447828823
BIC: -333.2452837343359
```

We observe that when assuming a t-distribution, we obtain smaller AIC and BIC values. Therefore, the t-distribution is the best fit.

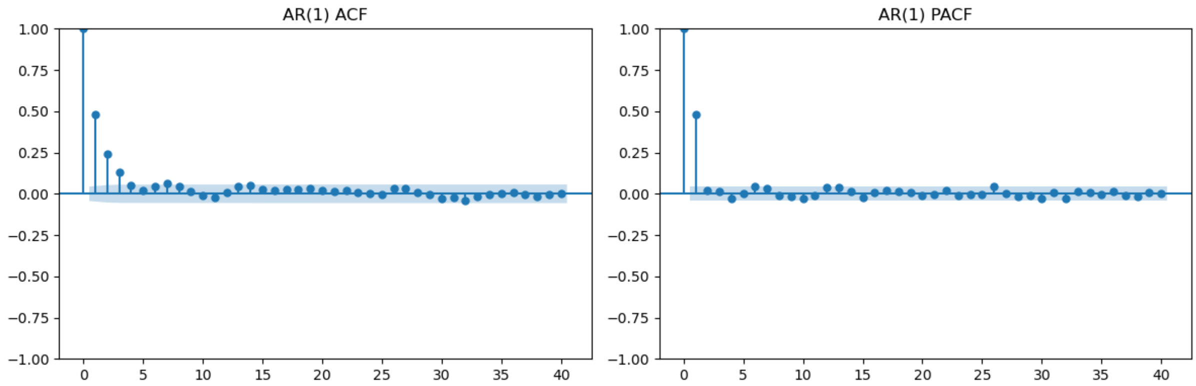
(3) The parameters of the two distributions, as presented above, collectively indicate that the parameter estimates from the t-distribution model offer a superior fit to the data and align more closely with the distributional characteristics of the actual dataset. Therefore, it can be concluded that the t-distribution model is better suited for

addressing scenarios where the assumption of normality is not met. However, it's important to note that the limited sample size (100) may not be sufficient to form a robust conclusion, and the results remain open to further discussion.

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?

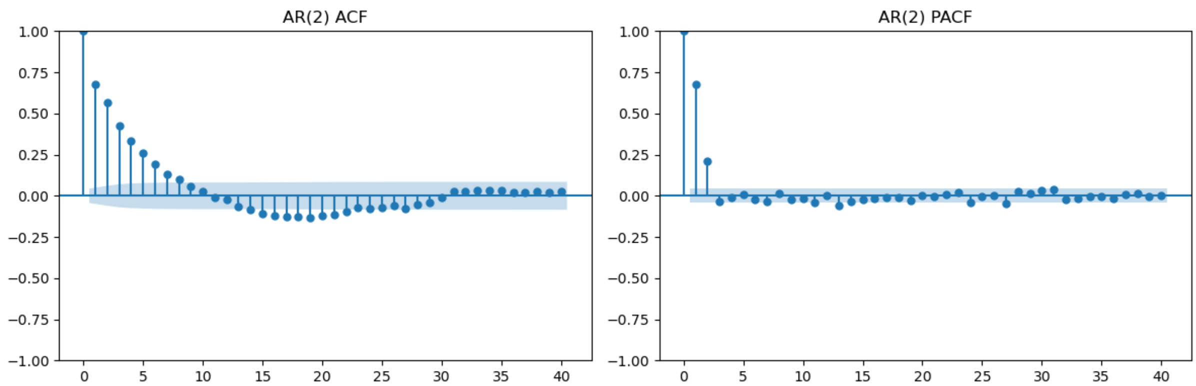
AR(1):



ACF plot: The ACF plot of the AR(1) process shows exponential decay with a distinct peak only at lag 1 and then rapidly decaying to zero. This indicates that the autocorrelation of the AR(1) process exists only at lag 1.

PACF plot: The PACF plot has a distinct peak at lag 1 and then rapidly decays to zero. The PACF values at all other lags are close to zero. This indicates that the biased autocorrelation of the AR(1) process exists only at lag 1, which is a first-order AR process.

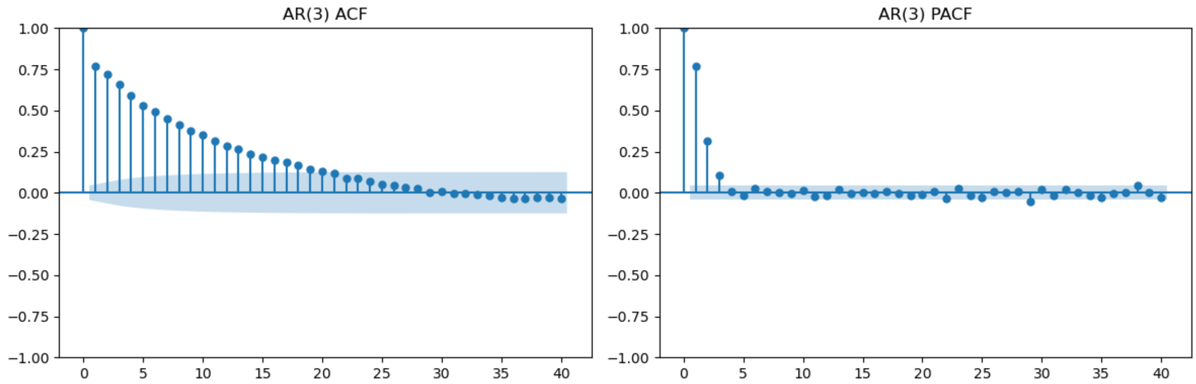
AR(2):



ACF plot: the ACF plot for the AR(2) process shows two distinct peaks, one at lag 1 and the other at lag 2. Then, the autocorrelation gradually decreases and tends to zero.

PACF plot: The PACF plot shows distinct peaks at lag 1 and lag 2 and then decreases rapidly to zero. The PACF values of the other lags are close to zero. This indicates that the partial autocorrelation of the AR(2) process exists only at lag 1 and lag 2, which is a second-order AR process.

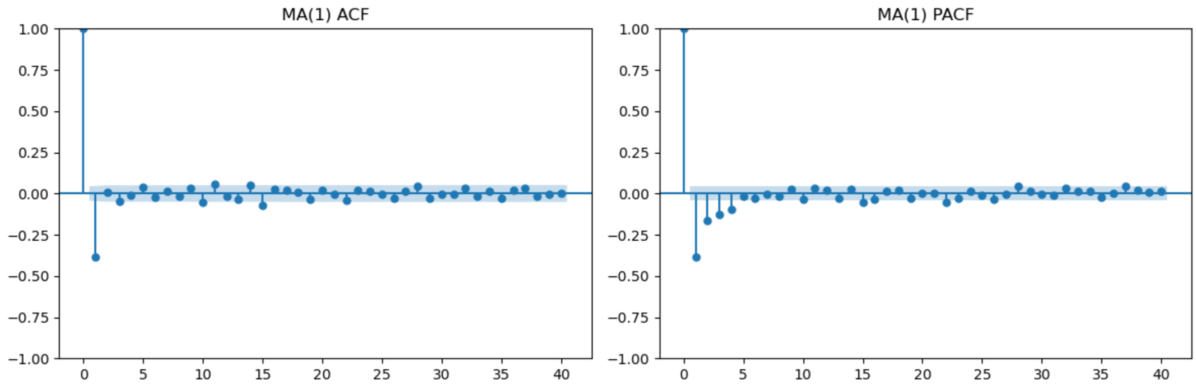
AR(3):



ACF plot: the ACF plot for the AR(3) process shows three distinct peaks, one at lag 1, one at lag 2 and another at lag 3. Then, the autocorrelation gradually decreases and tends to zero.

PACF plot: The PACF plot shows distinct peaks at lag 1, lag 2 and lag 3, and then decreases rapidly to zero. The PACF values for all other lags are close to zero. This indicates that the partial autocorrelation of the AR(3) process exists only at lag 1, lag 2 and lag 3, which is a third-order AR process.

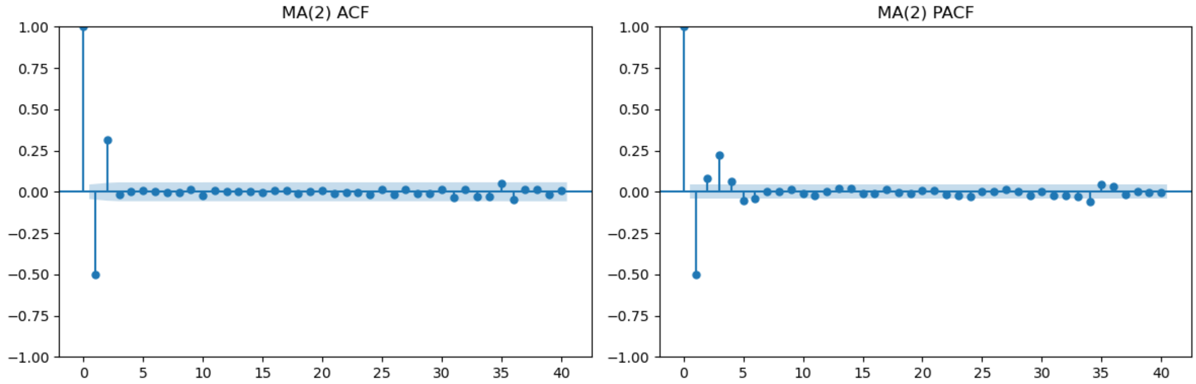
MA(1):



ACF plot: The ACF plot of the MA(1) process has a distinct peak at lag 1 and then drops rapidly to zero. The ACF values of the other lags are close to zero. This indicates that the autocorrelation of the MA(1) process exists only at lag 1.

PACF plot: The PACF plot of the MA(1) process has no significant peak and all lagged PACF values are close to zero. This indicates that the biased autocorrelation of the MA(1) process is close to zero at all lags except lag 1.

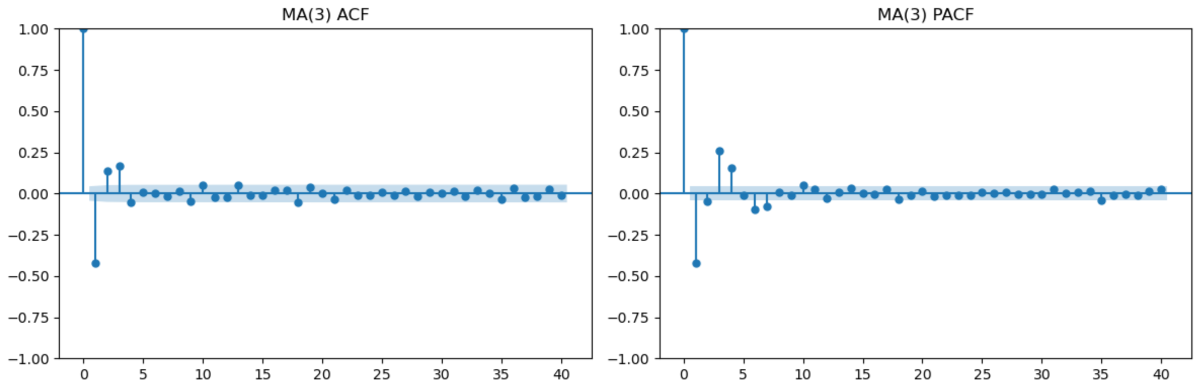
MA(2):



ACF plot: The ACF plot of the MA(2) process has significant peaks at lag 1 and lag 2, and then drops rapidly to zero. The ACF values of the other lags are close to zero. This indicates that the autocorrelation of the MA(2) process exists only at lag 1 and lag 2.

PACF plot: The PACF plot of the MA(2) process has no significant peaks and the PACF values for all lags are close to zero. This indicates that the biased autocorrelation of the MA(2) process is close to zero at all lags except lag 1 and lag 2.

MA(3):



ACF plot: The ACF plot of the MA(3) process has significant peaks at lag 1, lag 2 and lag 3, and then drops rapidly to zero. The ACF values of the other lags are close to zero. This indicates that the autocorrelation of the MA(3) process exists only at lag 1, lag 2 and lag 3.

PACF plot: The PACF plot for the MA(3) process has no significant peaks and the PACF values for all lags are close to zero. This indicates that the partial autocorrelation of the MA(3) process is close to zero at all lags except lag 1, lag 2 and lag 3.