GROUP WORK PROJECT #1
Group Number: 8564

MScFE 610: FINANCIAL ECONOMETRICS

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Guo Yuxuan	Singapore	y.xuannn03@gmail.com	
Franck Delma Deba Wandji	France	debafranck@gmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	**Guo Yuxuan**
Team member 2	**Franck Delma Deba Wandji**
Team member 3	

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.
N/A

1. **Skewness (Student 1)**

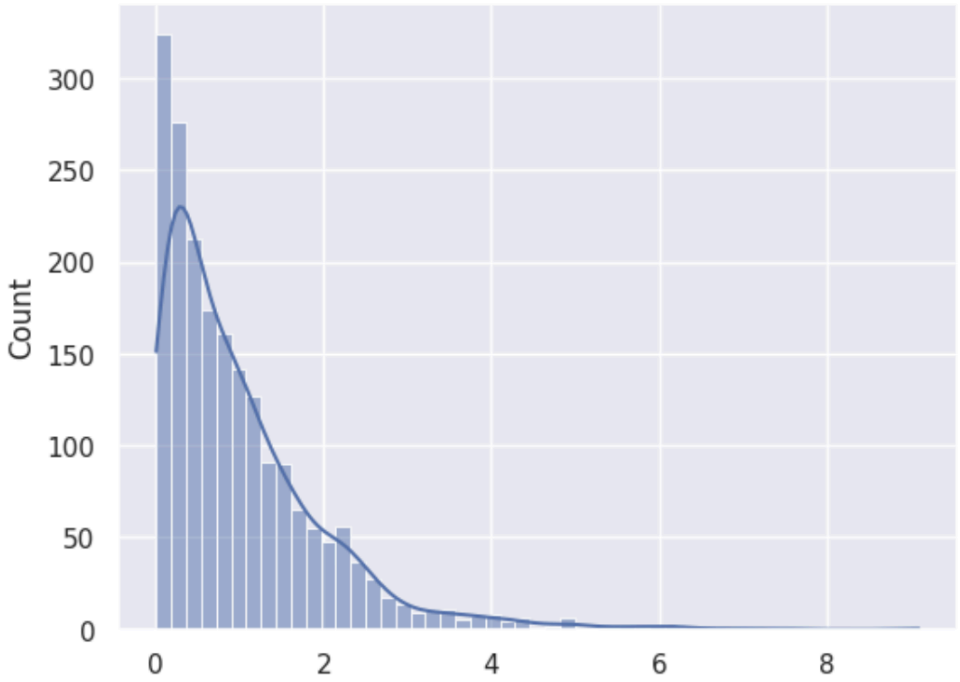
Skewness measures the degree of asymmetry in a probability distribution or the deviation from a normal distribution, which is a symmetrical bell-shaped curve. If skewness is greater than 0, the distribution is right skewed, which suggests that there is a longer tail to the right and hence there are more extreme values at the right side of the mean. If skewness is less than 0, the distribution is left skewed, and there is a longer tail to the left. When skewness is 0, the graph is symmetrical and the data comes from a normal distribution (GeeksforGeeks, 2024).

We demonstrate skewness with data coming from an exponential distribution, which is a known right skewed distribution.

```
[12] right_skewed_data = np.random.exponential(size=2000)
```

```
[15] sns.histplot(right_skewed_data, kde=True)
```

```
<Axes: ylabel='Count'>
```



The easiest way to identify skewness is through plotting the histogram plot of the data and observing the shape of the plot. As we can see from the chart, there is a longer tail to the right, which confirms our hypothesis that this is from a right-skewed distribution. We can also conduct statistical tests like using Pearson's moment of skewness, which compares the mean and the median of the data, to conclude if the data is actually skewed.

Pearson's second coefficient for the simulated data returns 0.8, which is greater than 0, and hence conclude that the data comes from a right skewed distribution.

The damage that skewness might cause is that it might affect the accuracy of some statistical models and tests, which assumes that the data comes from normal distribution. For example, in regression, if the data is highly skewed, it might skew the results of the parameter estimates, or it might not capture the full impact of the extreme values.

There are several ways we can deal with issues of skewness. We can use a box-cox transformation to make the data normally distributed, hence reducing the problem of skewness. In financial studies, it is also common to apply a log transformation on the returns data to transform it into a more symmetrical distribution (Inkiya, 2024). It does this by making the values more equal, reducing the impact of the extreme values, and reducing the variance of the data. However, if transformations are not possible, we can also use non-parametric methods to conduct statistical tests as they do not rely on normality assumptions.

2. Kurtosis / heteroscedasticity (Student 2)

2.1. Definition

Kurtosis measures the "tailedness" of a probability distribution relative to a normal distribution. Given a continuous random variable X under a probability distribution, the excess kurtosis of that distribution is calculated as:

$$Excess\ Kurtosis\ =\ E((X\ -\ E(X))^4)\ /\ E((X\ -\ E(X))^2)\ -\ 3$$

where the -3 adjustment makes normal distribution kurtosis = 0 (DeCarlo, 1997).

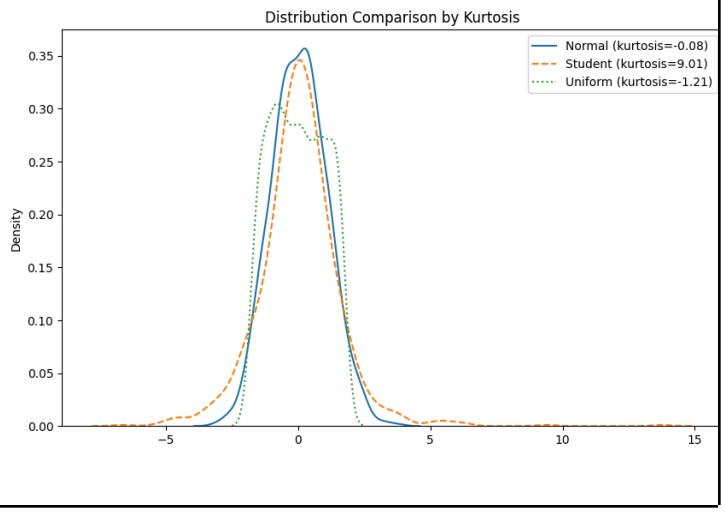
2.2. Description

A positive excess kurtosis indicates heavy tails (more outliers than normal distribution). The distribution is said to be leptokurtic. A negative suggests light tails. The distribution is said to be platykurtic.

2.3. Demonstration

For numerical demonstration, we'll simulate three distributions with different kurtosis properties: Normal kurtosis (normal distribution), Positive excess kurtosis (t-Student distribution), Negative excess kurtosis (uniform distribution).

Simulated distribution	Excess Kurtosis
Normal distribution	-0.08
t-Student distribution (with 3 df)	9.01
uniform distribution	-1.21



2.4. Diagnosis

To formally detect whether the sample data has excess kurtosis, we perform a **hypothesis test**:

2.4.1. Hypothesis Test for Excess Kurtosis

- **Null Hypothesis (H0)**: The sample has no Excess kurtosis (excess kurtosis = 0).
- **Alternative Hypothesis (H1)**: The sample has **non-normal kurtosis** (excess kurtosis different to 0).

Under the null hypothesis, the distribution of the test statistic for excess kurtosis should follow an asymptotic normal distribution for large sample sizes (Westfall, 1993).

- If the p-value is less than the significance level (typically 0.05), we reject the null hypothesis, indicating significant excess kurtosis.
- If the p-value is greater than the significance level, we fail to reject the null hypothesis, suggesting that the sample is consistent with normal kurtosis.

2.4.2. Results from Hypothesis Test

- The test for excess kurtosis on a sample generated from a **normal distribution** resulted in a **p-value of 0.67**. This is greater than 0.05, meaning that we fail to reject the null hypothesis. Therefore, **the sample does not exhibit excess kurtosis**, and it is consistent with a normal distribution.
- On a sample generated from a **Student's t-distribution** wa get a **p-value less than 0.05**. This indicates that we reject the null hypothesis, confirming that **the sample exhibits excess kurtosis**. The distribution has heavier tails compared to the normal distribution, which is typical of the Student's t-distribution.
- On a **sample generated from a Uniform** we get a **p-value less than 0.05**. This indicates that we reject the null hypothesis, confirming that **the sample exhibits excess kurtosis**. In addition the statistic is negative showing that the distribution has thinner tails compared to the normal distribution.

2.5. Damages of ignoring kurtosis

Ignoring kurtosis in while modeling can lead to: Inaccurate predictions and misleading statistical inferences, underestimation of the probability of extreme events which is especially problematic in financial or risk modeling contexts.

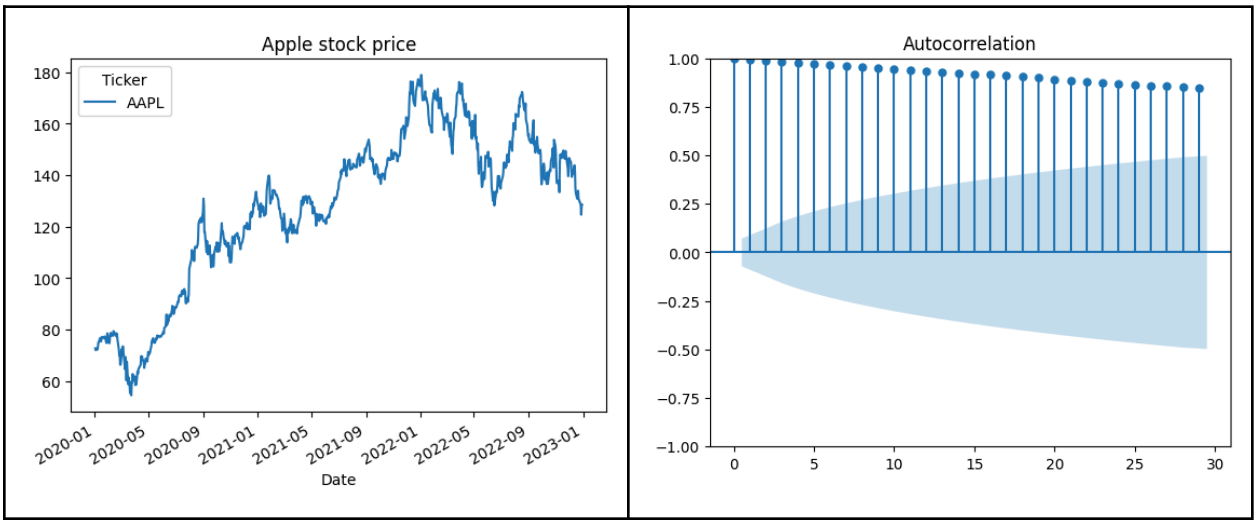
Properly accounting for kurtosis ensures that models are more robust and capable of capturing the true behavior of the data, particularly when extreme events are more likely than assumed under normality.

3. Modeling non-stationarity and finding a unit Root Testing

3.1. Definition: A time series is **stationary** if its finite dimensional multivariate distribution remains the same over time. Formally, a stochastic process X is weakly stationary if its mean and variance are constant over time, and its covariance between to dates depends only on the lag.

3.2. Description: Many real-world time series (e.g., stock prices, GDP, inflation) exhibit non-stationarity, meaning their mean and/or variance depends on time.

3.3. Diagnosing Non-Stationarity: The Autocorrelation Function (ACF) is a tool used to measure the correlation between a time series and its lagged values over time. It is especially useful to visually check for stationarity by identifying patterns such as seasonality or trends in time series data.



The Apple stock price data shown in the figure exhibit non-stationary behavior, as indicated by the presence of a clear trend. This suggests that the time series has a unit root, meaning its statistical properties, such as mean and variance, change over time.

Unlike a stationary data, non-stationary data, on the other hand, displays trend and seasonality. To formally test for non-stationarity, we use **unit root tests**. **Unit root testing** is used to check for non-stationarity and we can use the results of the tests to decide if we would need to apply differencing to make the series stationary.

An example of a time series data that is non-stationary is the random walk (and the Apple stock as illustrated above). The value at time t depends on the value at the previous period and some random error.

The most commonly used unit root test to identify non-stationarity is the Augmented Dickey-Fuller (ADF) test. The null hypothesis of the test is that the time series has a unit root, which indicates that the data is non-stationary. If the p-value from the test is less than 0.05, we

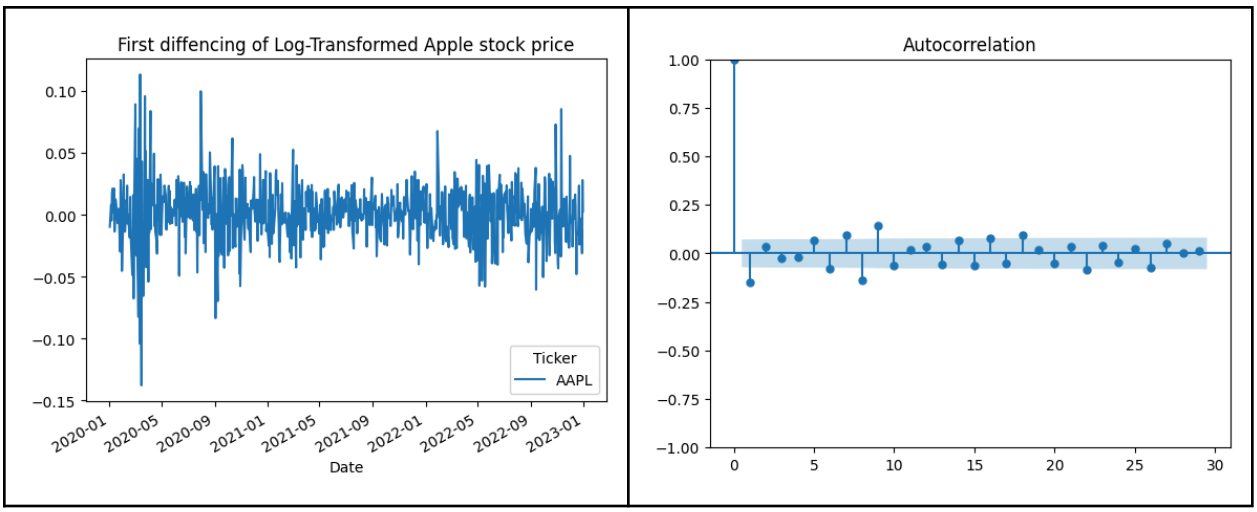
reject the null hypothesis, suggesting that the data does not have a unit root, and is stationary(Dickey & Fuller, 1979).

Applying the ADF test to Apple stock prices yields a p-value of 0.337, indicating that we fail to reject the null hypothesis. This confirms the presence of a unit root, meaning the data is non-stationary.

3. 4. Damage Caused by Non-Stationarity: Ignoring non-stationarity in time series modeling can lead to spurious regression results since all theoretical properties on time series modeling assume stationarity (ARMA).

3.5. Directions

- A common approach to removing a unit root is differencing(Hamilton, 1994). First differencing for example involves subtracting the previous observation from the current observation to remove any trends that might be present.
- If variance increases over time, log transformation stabilizes it. This transformation is common in financial analysis. When analyzing stock prices, it is common to apply the log transformation first before differencing, as this stabilizes the variance across time and can lead to better results when applying statistical models like ARIMA or other time series methods. Where "I" stands for integrated, which means differencing is applied for stationarity.



After applying log transformation and first differencing to the Apple stock price data, the resulting time series appears to be stationary around the zero line. The ACF of the series resembles that of a white noise process, indicating stationarity. Furthermore, the Augmented ADF test yields a p-value of less than 0.05, leading to the rejection of the null hypothesis and confirming that the series is stationary.

To summarize, checking for stationarity is essential in time series modeling. Fortunately, the presence of a unit root that causes non-stationarity can be addressed through successive differencing.

References

1. DeCarlo, L. T. (1997). "On the Meaning and Use of Kurtosis." Psychological Methods, 2(3), 292–307.
2. Dickey, D. A., & Fuller, W. A. (1979). "Distribution of the estimators for autoregressive time series with a unit root." Journal of the American Statistical Association.
3. GeeksforGeeks. “Coefficient Of Skewness,” May 23, 2024. https://www.geeksforgeeks.org/coefficient-of-skewness/.
4. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
5. Inkiya, Vipin Singh. “Mastering Skewness: A Guide to Handling and Transforming Data in Machine Learning.” *Medium* (blog), April 26, 2024. https://medium.com/@vipinnation/mastering-skewness-a-guide-to-handling-and-transforming-data-in-machine-learning-0a42773a026f.
6. Westfall, P. H., & Young, S. S. (1993). "Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment." John Wiley & Sons.