GROUP WORK PROJECT # 1
GROUP NUMBER: 6131

MScFE 610: FINANCIAL ECONOMETRICS

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Eduardo Tadeu Santos Gomez Monteiro	United Kingdom	Eduardomonteiro128@gmail.com	
Ibrahim Adeshola Raji	Nigeria	rajiibrahim6@gmail.com	
Krishna Vamsi Regulavalasa	India	krishnavamsi8262@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	Eduardo Tadeu Santos Gomez Monteiro
Team member 2	Ibrahim Adeshola Raji
Team member 3	**Krishna Vamsi Regulavalasa**

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.

Skewness

Definition

Skewness measures the deviation of a dataset distribution compared to the normal distribution. A normal distribution would have 0 skewness. The skewness of a distribution is defined as:

Skewness = $\frac{3X-\mu}{\sigma}$

Where:
X is the mean,
μ is the median, and
σ is the standard deviation.

Description

Skewness indicates the asymmetry of the data distribution. Positive skewness means a longer or fatter tail on the right side, and negative skewness means a longer or fatter tail on the left side.

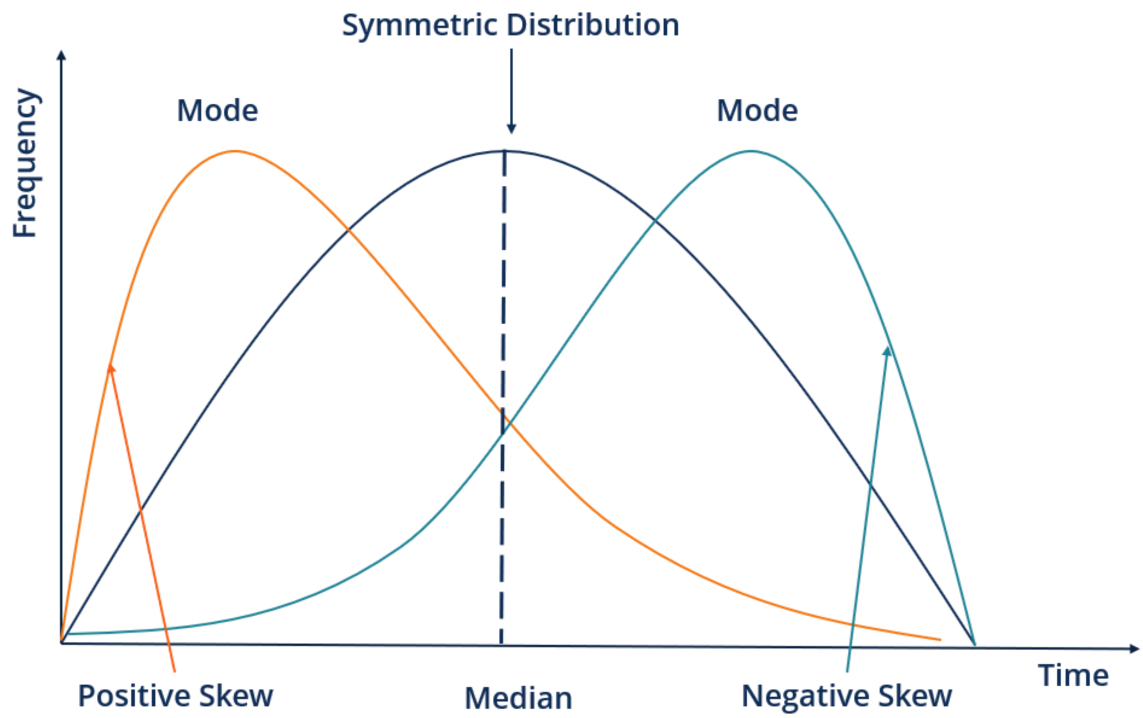


Figure 1 - Positive and negative skewness (Taylor)

Demonstration and Diagram
Using python was generated a positively skewed dataset and then created a histogram.

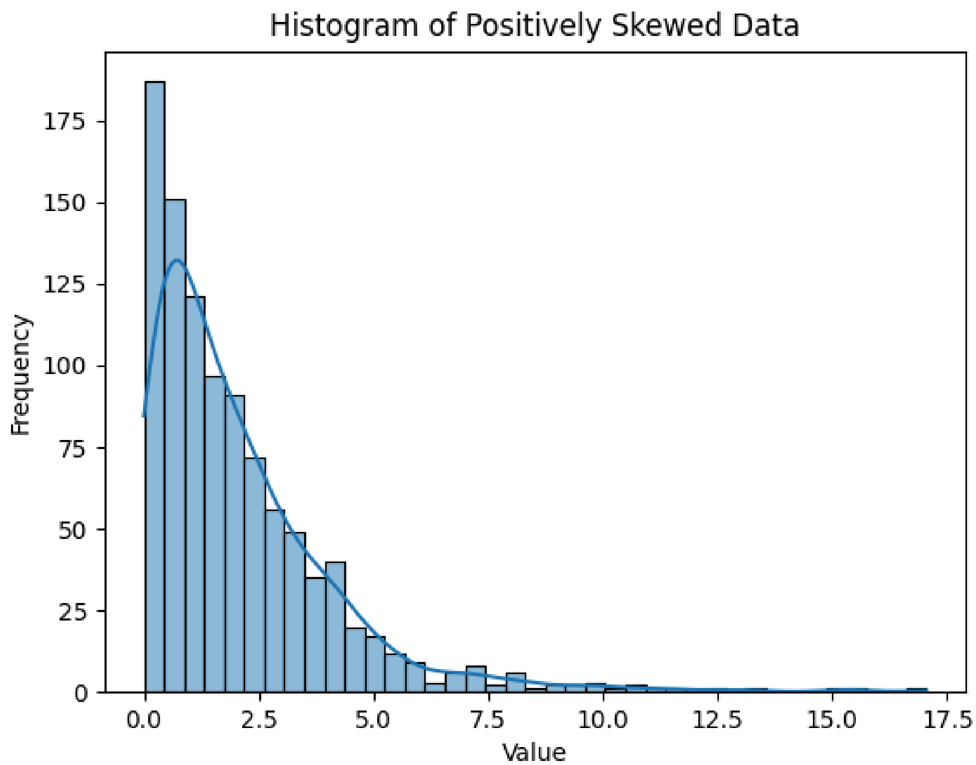


Figure 2 - Generated positively skewed histogram

Diagnosis
There are 2 quick ways to recognize or test for skewness in a dataset, the first would be by checking the data visually by creating a histogram or box plot of the dataset. The second one would be by performing statistical tests by calculating the skewness coefficient using scipy.stats.skew(data) in Python. A skewness coefficient significantly lower or higher than zero indicates skewness.

Damage
Skewed data can cause several issues in statistical modelling and data analysis:

- Biased Parameter Estimates: Many statistical models assume normality, and skewness can lead to biased parameter estimates and predictions.
- Misleading Descriptive Statistics: Measures like the mean may not accurately represent the central tendency in skewed distributions, leading to incorrect conclusions.

Directions
We can perform transformations to reduce skewness like log, square root or box-cox transformations. Or we can use robust statistical measures like using the median instead of the mean as it is less affected by skewness and employing robust regression techniques that are less sensitive to skewed data, such as quantile regression.
By implementing these strategies, we can mitigate the impact of skewness on our model and data analysis, leading to more reliable and interpretable results.

Kurtosis

Definition
Kurtosis is a statistical measure that quantifies the tailedness and peakedness of a probability distribution, relative to the normal distribution. It is defined as the fourth standardized moment, given by:

Kurtosis = $\Sigma(X-\bar{x})/n\sigma$^4

Where:
$\bar{x}$ is the mean,
σ is the standard deviation and
n is the number of observation

Description
Kurtosis measures the concentration of data in the tails and the center of a distribution, indicating the likelihood of extreme values occurring.

Demonstration and Diagram
Visual representation of Apple Inc. stock data using Histogram.

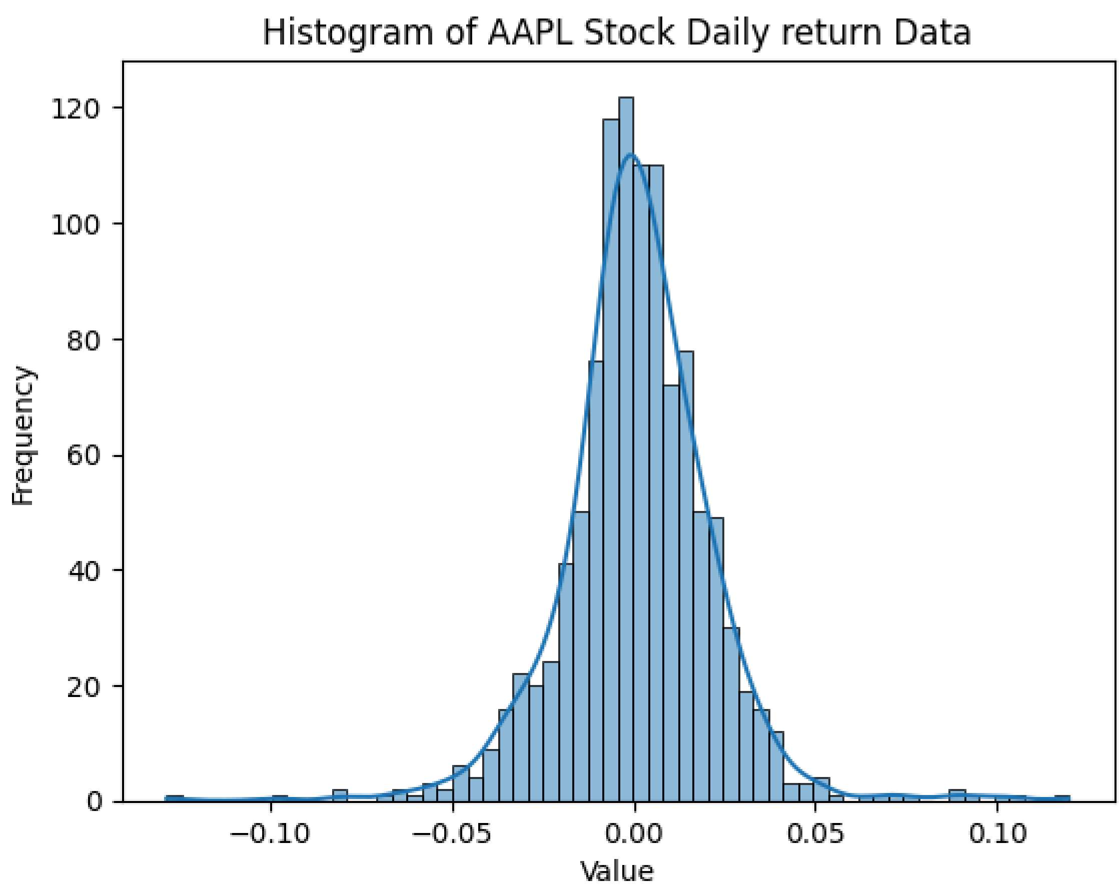


Figure 3 - Generated leptokurtic histogram

Diagnosis
To determine if kurtosis exists in a dataset, calculate the fourth standardized moment using the formula above. A kurtosis value greater than 3 indicates a leptokurtic distribution (heavy tails and a sharp peak), while a value less than 3 indicates a platykurtic distribution (light tails and a flatter peak) (DeCarlo, 1997).

Damage
Ignoring kurtosis can lead to underestimating the probability of extreme events, which can have severe consequences in fields such as finance, where risk assessment is crucial (Xiong & Idzorek, 2011).

Directions
To address kurtosis in data analysis, it's best we use robust statistical methods that are less sensitive to outliers, such as the median and interquartile range instead of the mean and standard deviation. Additionally, using fat-tailed distributions, such as the Student's t-distribution or the generalized extreme value distribution, can better capture the presence of kurtosis in the data (Westfall, 2014).

Over-reliance on the Gaussian distribution

Definition:

A Gaussian distribution, also known as a normal distribution, is defined by the probability density function (PDF):

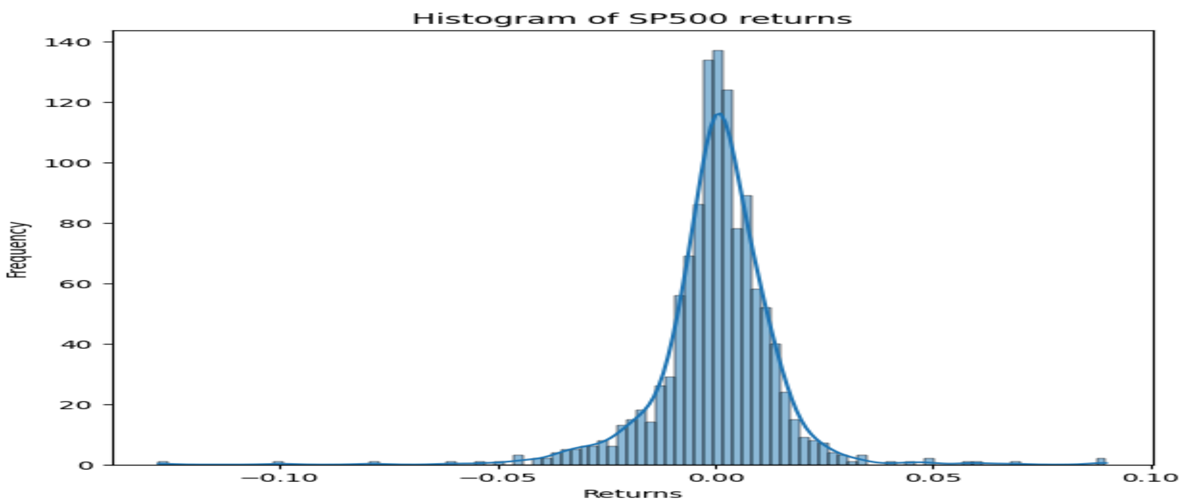
$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

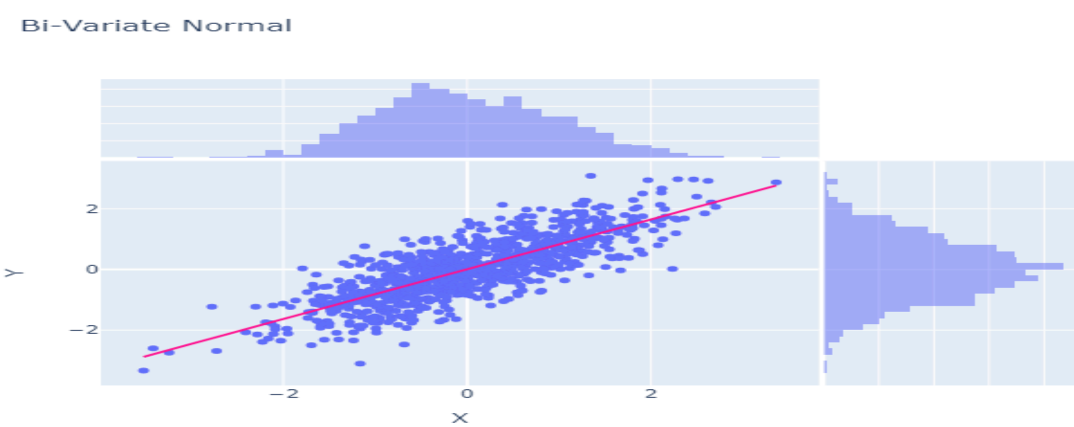
where μ is the mean, σ is the standard deviation, and x represents the variable.

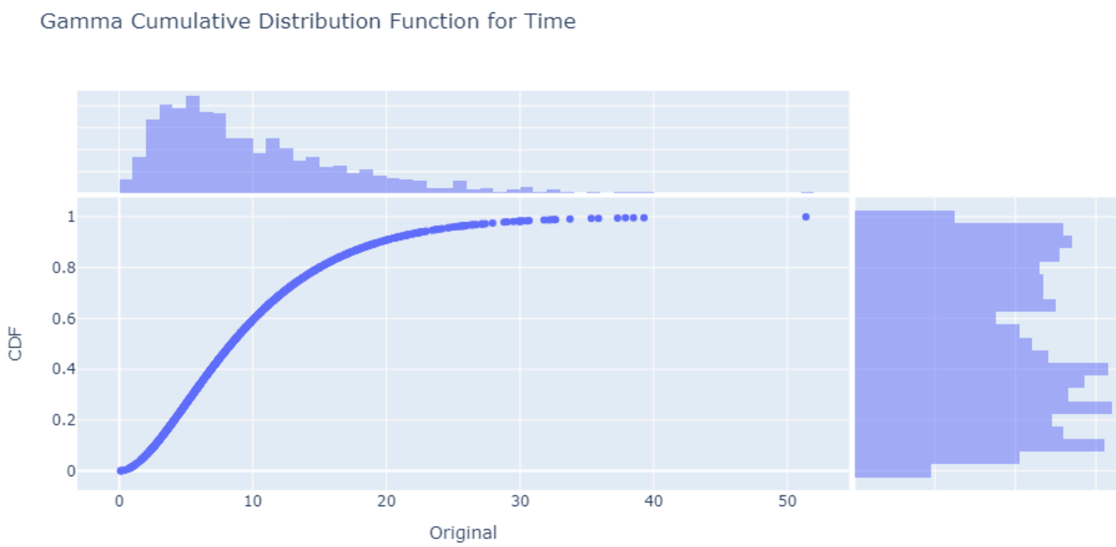
Description :

Over-reliance on the Gaussian distribution in finance means assuming that asset returns follow a normal distribution, which often underestimates the probability of extreme events (fat tails) and leads to mispricing of risk and poor risk management.

Demonstration and Diagram







Diagnosis
To recognize if data deviates from normality:

1)Visual Inspection
2)Statistical Tests

Damage:
Relying too heavily on the Gaussian distribution can lead to:

Underestimation of Risk
Mispricing of Derivatives
Poor Risk Management.

Directions
To address issues arising from non-normal data distributions,we have considered the following methodologies:

Non-Gaussian Models:
Use distributions like t-distribution, GARCH models.
Copulas: without assuming normality , Employ copula functions to model and simulate the dependency structure between multiple financial variables.

Non-stationary

Definition

A time series is said to be non-stationary if its statistical properties, such as mean, variance, and autocorrelation, change over time. Mathematically, a time series

$\{X_t\}$ is non-stationary if:

$E[X_t] \neq$ constant

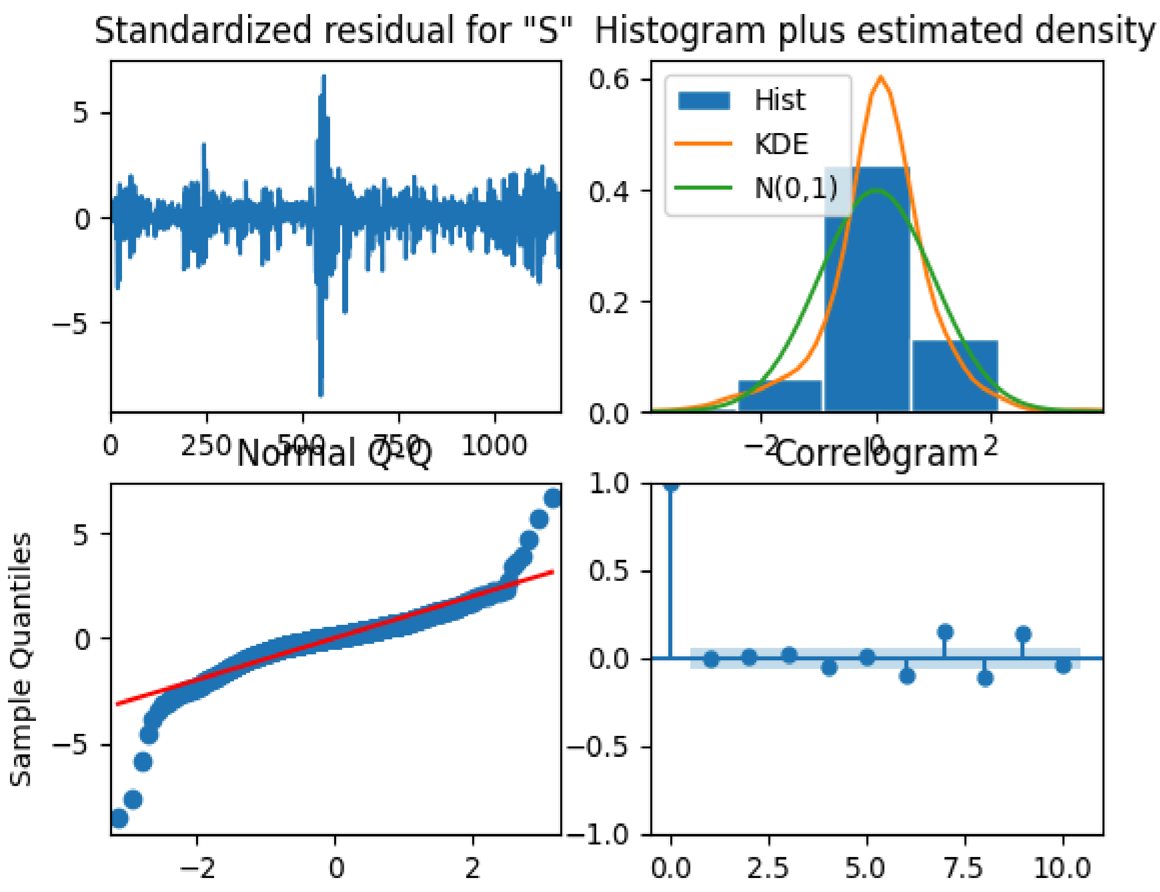
$\text{Var}(X_t) \neq$ constant

$\text{Cov}(X_t, X_{t+k})$ depends on t and k

Description

Non-stationarity implies that the underlying process generating the time series evolves over time, making it challenging to model and predict future values based on past behavior.

Demonstration and Diagram



Diagnosis

By running this code, you will be able to perform the full analysis, including data fetching, differencing, log transformation, ACF and PACF plotting, ARIMA model selection, model diagnostics, and the Ljung-Box test.

Damage
Non-stationarity can lead to misleading statistical inferences, biased estimates, and incorrect forecasting results.

Directions
To address non-stationarity, one can apply techniques such as differencing, detrending, or decomposition to remove the time-dependent components and transform the series into a stationary one (Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M., 2015). Models like ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal ARIMA) are designed to handle non-stationary data by incorporating differencing and seasonal components.(Shumway, R. H., & Stoffer, D. S., 2017)

References

1. Taylor, Sebastian. “Skewness.” Corporate Finance Institute, 8 Mar. 2023, https://corporatefinanceinstitute.com/resources/data-science/skewness/.
2. DeCarlo, L. T. (1997). On the meaning and use of kurtosis. Psychological Methods, 2(3), 292-307.
3. Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. R.I.P. The American Statistician, 68(3), 191-195.
4. Xiong, J., & Idzorek, T. (2011). The impact of skewness and fat tails on the asset allocation decision. Financial Analysts Journal, 67(2), 23-35.
5. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control (5th ed.). John Wiley & Sons.
6. Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: with R examples (4th ed.). Springer.